Table of Contents

INTERNATIONAL PRODUCTIVITY MONITOR — NUMBER 42, SPRING 2022

PUBLICATIONS MAIL AGREEMENT NO 40049476
RETURN UNDELIVERABLE CANADIAN ADDRESSES TO:
CSLS
604-170 Laurier Avenue West
Ottawa ON K1P 5V5
email: andrew.sharpe@csls.ca

Editors' Overview

This edition of the *International Productivity Monitor*, No. 42, is a particularly large one (211 pages) as it not only includes four substantive articles as part of our regular line up, but also four articles which represent the first part of our Symposium on Productivity and Well-Being. In this overview we will discuss the other four articles. A separate introduction to the Symposium follows after the first four articles in this volume.

The first two articles make use of input-output data and analysis to better understand the productivity impacts of inter-industry intermediates on productivity. The article by **Tero Kuusi** and **Martti Kulvik** from the Research Institute of the Finnish Economy and **Juha-Matti Junnonen** from Tampere University focuses specifically on the construction industry which has been notorious for weak productivity growth across many countries. By constructing a data set on construction value chains for 12 European countries using data from the World Input–Output Database, the authors find that the upstream industries have been responsible for most of the productivity growth of the value chain in construction. They find an especially large contribution from business services to productivity in the value chain. The authors also find that value chain productivity has much benefited from construction-related patents but have suffered from low efficiency in the use of information technology because of major adjustment frictions.

The second article using input-output data and techniques by **Daniel Lind** of Arenagruppen focuses on the impact of intermediate imports from China on manufacturing productivity growth in high income economies. He finds that productivity in the latter group of countries hugely benefited from the "China shock". This is complementing earlier work by David Autor and others which primarily focused on the effects of intermediate imports from China on labour market. In addition to the effects on value added and employment (and hence on productivity), the author also hints at the role of reduced producer prices and functional specialization in the use of knowledge-intensive intermediate inputs to explain the large productivity effects. The author also points at an even larger productivity effect from intermediate imports from Eastern European economies, whereas intra-trade of intermediate between high income economies has weakened productivity.

The third contribution by **Ulrich Kohli** from the University of Geneva returns to a well-established but still unresolved topic on how to measure trading gains and terms of trade effects on measures of welfare and how to link them in an analytically correct way to productivity. The author shows that most countries, except Canada and the United States, only compute a terms-of-trade effect on income, but fail to take into account a real exchange rate effect which results from the relative-price component when trade is not balanced. This omission also has implications for the relationship of income to productivity as it does affect measure of average and marginal productivity.

The final contribution in our line-up of regular articles is a timely topic looking at how productivity measures have been affected as a result of the COVID-19 pandemic. In his contribution, **Jay Stewart** from the U.S Bureau of Labour Statistics shows how labour productivity and real wages in the United States sharply increased when the pandemic began in the second quarter of 2020. This appears to be the result of a compositional effect on productivity because many low skilled people dropped out of the labour market especially because of the temporary shutdown of firms in leisure and hospitality and other low wage sectors. Indeed an increase in the average levels of skills accounted for the bulk of the extraordinary productivity gain during the quarter. The article shows that as average skills levels have still not returned to their pre-pandemic state, further analysis of how the pandemic has not only created but potentially medium- or long-term effects is needed.

An editorial introduction to the four articles which are published as the first part of our Symposium on productivity and well-being is provided separately.

Productivity Growth in Construction Value Chains

Tero Kuusi and Martti Kulvik
The Research Institute of the Finnish Economy
Juha-Matti Junnonen
Tampere University[1]

Abstract

The construction industry has suffered from low productivity growth in recent decades. Motivated by the economic importance of the industry, we revisit the construction productivity puzzle by analyzing the construction value chains of 12 European countries using data from the World Input–Output, the EU KLEMS databases and complementary datasets. We decompose construction-related value added and productivity contributions to both the construction industry and the rest of the value chain and show that the traditional focus on the construction industry is adversely restrictive for understanding productivity growth in construction activities. There is a substantial contribution of construction-related value added generated in other industries, and the productivity growth in the value chains has, for the most part, been seen outside the construction industry. Furthermore, we show that there is a strong, long-term relationship between construction-related patents and the improvement of total factor productivity in the value chains, but the chains typically do suffer from low efficiency in the use of information technology.

Construction industry is a significant contributor to economic activity in most countries. On average, it accounts for approximately 6 to 9 per cent of economies' gross domestic product (Arditi and Mochtar, 2000). However, productivity growth in the construction industry is commonly and persistently low compared to many manufacturing and service industries (Bankvall *et al.*, 2010; Tran and Tookey, 2011). It is more a rule than an exception that there has been no productivity growth or even a declining productivity in the European construction industries over a period of several decades.

According to O'Mahony and van Ark

1 Tero Kuusi is the Research Director at The Research Institute of the Finnish Economy. Martti Kulvik is the Chief Research Scientist at the same institute. Juha-Matti Junnonen is Project Manager for the Civil Engineering Department at Tampere University. We gratefully acknowledge research funding of the Government's analysis, assessment and research activities for the project "The competitiveness of the Construction Industry and the Quality of Construction in Finland." We are grateful to the project steering group as well as our colleagues at Etla, anonymous referees and editors for their valuable comments. E-mail: tero.kuusi@etla.fi; martti.kulvik@etla.fi; juha-matti.junnonen@tuni.fi.

(2003), annual labour productivity growth in the construction industry was approximately 1 percentage point lower than in the total economy from 1979 to 2001. According to more recent EU KLEMS data[2], the industry has on average shown only negligible labour productivity growth between 2001 and 2015. Such a controversial finding raises questions not only about the origins of the poor performance, but also about the quality of the underlying productivity statistics. This article addresses a key challenge of the latter: the fragmentation of the construction value chains.[3] Before analyzing the fragmentation, we, however, need to share some insights on the complexity of the construction value chains and the operational environment.

Construction is on-site work, while industrialization of construction, with focus on prefabrication, can be seen as a structural action to diminish on-site activities. For example, pre-cast concrete is a manufactured product, while installing pre-cast is construction work. In either case — prefabrication focused or on-site built — flawless communication, precise timing and efficient logistics pose an arduous and critical triad to any construction project. Much of the technological progress in constructed products consists of increasing the amount of work that is done in a plant setting as opposed to on-site, and transporting those components to a construction site for more straightforward installation or erection.

Declining or stagnant labour productivity in the construction industry could be associated with overall gains in the efficiency with which constructed products are installed, but those gains could show up as improvements in productivity in the manufacturing sector instead of the construction industry itself. Plant production enables transparent control of production processes, incurs potential cost savings through coordinated purchases and standardized repetitive work phases, offers natural opportunities for process and material development including easier monitoring thereof, as well as other gains associated typically with economies of scale. An additional benefit particularly valuable in construction is the total control of climate that plant production offers: no rain, no frost, no gusts.

Thus, only a part of the construction value creation is generated by the construction industry on-site, and hence a focus on on-site productivity only hides substantial networks of technologically progressive manufacturing and business services. A comprehensive value chain perspective is important in providing further understanding about the construction industry's or-

2 The EU KLEMS data used in the analysis are discussed in the second section.

3 In what follows, we refer to all construction-related economic activity, including and beyond the construction industry as the construction value chain. The construction industry is narrowly defined according to the ISIC Rev. 4 industry classification (F) and by its productivity growth we refer to the value-added based measurements in the industry. Instead, the construction value chain involves all production activities contributing to the production of the built environment. The value chain constitutes the value added of the construction industry and the value of the intermediate goods and services, both domestic and foreign, used by the industry to produce its gross output. To the extent that other industries provide value added to the construction value chain, they are considered as construction-related activities and according to our methodology, hence a part of the construction value chain. While construction sector has been used both as a synonym to construction industry and as a more general term, we use it only when citing corresponding literature.

ganization and performance, as the value chain approach makes more visible both the substantial role of upstream industries to which construction industry has backward linkages as well as technology and knowledge investments as a source of productivity growth in the entire value chain.

Our work builds on a novel decomposition of the value-added contents of the outputs and contributions of the construction industry and of the other sectors in the upstream value chain. We combine the World Input–Output Database (WIOD) and a method suggested by Los, Timmer, and de Vries, (2016) to measure the value-added content of different economic activities based on the data. Accordingly, we extracted construction activities from the WIOD for 12 European countries.[4] Furthermore, we studied the productivity growth contributions of the industries that participate in the value chains. In particular, we used data on the generated value-add in the value chains to weight the corresponding industry-level productivity growth measurements in the EU KLEMS database and complementary datasets, to thereby account for their contribution of the value-added factor in growth (Wolff, 1994; Timmer, 2017).

While more typical value chain analysis tracks the value creation paths of individual projects and even single products (see Ali-Yrkkö, Seppälä, and Mattila, 2016 and references therein), the approach is not feasible for an entire, highly diversified sector ranging from family house building and land reclamation to oil rig construction. Not only is the construction industry heterogeneous, but also the importance of intermediate goods varies widely — the value of gross output in residential constructions is much greater than in road construction, and installation of underwater pipelines requires huge state-of-the-art machinery. A reciprocal approach using WIOD and KLEMS type data responds to the analysis challenge posed by extensive and complex sectors such as construction.

Our analysis shows that the focus on the construction industry is restrictive from the perspective of understanding productivity growth in construction activities: there is a substantial amount of construction-related value added generated in other industries. We find that roughly half of the total value added in the construction value chains is generated within the construction industry — a proportion common for most observed value chains. The other half of the value added is generated by other industries, involving both manufacturing and business services. Our findings suggest that the role of the business service sector, in particular, is important and has increased in the years from 2001 to 2014. Moreover, the productivity growth in the construction value chains has, for the most part, occurred in the upstream part of the value chain, while the role of the on-site construction industry is weak or even negative. This finding suggests that a focus on the on-site construction industry leads to a suppressing bias in the productivity of construction activities. We also identify a strong long-term relationship between construction-related

4 AUT, BEL, CZE, DEU, DNK, ESP, FIN, FRA, GBR, ITA, NLD, and SWE.

patents and the improvement of total factor productivity (TFP). To this end, we used a panel vector error correction model. Finally, we present how the value chains typically suffer from low efficiency in the use of information technology and due to high administrative costs.

In what follows, we first review the literature. We then introduce our value chain productivity measurement methodology and apply it to study the composition of the construction value chain and the productivity growth in the value chain. We also analyze statistically the effects of patents and information technology in the value chain. Finally, we conclude with a discussion.

Literature Review

This article contributes to several strands of literature. First, it introduces a novel way to analyze the role of different industries by extracting the construction value chains in the global input–output data, more commonly used in an international trade context (Los, Timmer, and de Vries, 2016; Ali-Yrkkö and Kuusi, 2017, 2019). It is one of the few attempts to capture the full economic scope of the construction value chain, beyond the contribution of the construction industry as more narrowly defined in International Standard Classification of All Economic Activities ISIC (ISIC, 2008) and its regional derivatives such as the European Nomenclature of Economic Activities NACE (REGULATION (EC) No 1893/2006).

Previously, Squicciarini and Asikainen (2011) used a discretionary classification of the construction sector to core and supporting (non-core) industries. They extended beyond the core construction sector by adding activities from other sectors that fully or principally depend upon or are functional to core construction activities. Their findings suggest that the indicators for composition, structure, value added, skills, and R&D input and output of the construction sector change substantially when a broader definition of the sector is applied.

Another strand of literature discusses how different features of the construction value chains affect their ability to increase productivity. Construction companies face difficulties in implementing innovation to enhance productivity due to the fragmented characteristic of construction and the high degrees of specialization in its processes, together with production activities carried out within single projects (Winch, 1998; Gann, 2000; Davis *et al.*, 2016). The construction industry also suffers from fragmentation owing to the temporary nature of project execution and the specialism incorporated into a project (Sullivan and Harris, 1986).

The fragmentation brings about well-known problems that may contribute to low productivity growth: capital-heavy approaches to construction bring high fixed costs that are difficult to cut in downturns, and profit margins are slim in the fragmented construction industry. These factors tend to keep investments at low levels (Economist, 2017a). Moreover, due to the complex nature, construction projects are exposed to high risk that is coupled with the problems of imperfect information (Lau and Rowlinson, 2011). The customized nature of most projects, often arising from complex legislation, further limits the usual advantages of size, preventing the generation of bigger, more productive companies. Construction projects are typically tendered out into a cascade of subcontractors, each operating at their industry's low profit margins. The subcontractors have evident incentives to maximize their profit or at least minimize losses — not that uncommon in single subprojects — rather than collaborate to contain overall costs of the entire project.

All in all, the fragmentation often results in numerous structural problems. Projects often lack repeatability and efficiency in performing recurring activities, the resource profiles of value chain members are not strongly shaped by the relationship, and operational decisions about one sub-entity are typically made independently of decisions about other sub-entities (Ketokivi *et al.*, 2017).

Naoum's (2016) study revealed that the rate of labour productivity on-site can be greatly affected by the fragmentation, for example, through ineffective project planning, delays caused by design error and variations, problems in the communications systems, design and buildability related issues such as specifications, and the procurement method. Zhai *et al.* (2009) showed that construction labour productivity is positively related to the use of automation and integration of projects. Ruddock and Ruddock (2011) identified information and communications technology (ICT) capital as the fastest growing input in construction, while it has only a modest share in overall input costs. Productivity growth might be explained by the level of investment in ICT (ICT capital growth); however, problems arise due to the time lag for a new technology to reach its full potential.

These findings correspond with the literature regarding productivity in general. Quite a few studies have shown that the impact of ICT at the industry level plays only a limited role as a source of productivity growth. This finding is reported by, among others, Stiroh (2002a), Draca, Sadun, and Van Reenan (2006), and Inklaar *et al.*, (2008). Complementary innovations in organizations are often needed to foster successful adaptation of ICT (Bresnahan, Brynjolfsson, and Hitt, 2002). Consistently, O'Mahony and Vecchi (2005), Oulton and Srinivasan (2005), and Venturini (2009) report a larger long-term effect. Following a growing body of literature that connects TFP and patents (Romer, 1990; Grossman and Helpman, 1994; Madsen, 2008; Coe, Helpman, and Hoffmaister, 2009), we include into our approach a module on how patents have improved productivity.

The challenges of innovation and ICT investment may become particularly large when their productivity contributions are considered jointly for the whole value chain. A strong interaction emerges in the value chain when the new technology generates positive productivity externalities or there are unmeasured complementary innovations that are made during the adaptation of the technology (Stiroh, 2002b; Basu and Fernald, 2007). The particular importance of a functional value chain strikes us as an intuitive one, as the positive role of technology is likely to accumulate and result in stronger ecosystems (Ketokivi *et al.*, 2017) that foster more efficient formalization of interactions, specialization of firms, and joint real-time decision-making in the value chain. On the other hand, fragmentation of the value chain may make the productivity impact of ICT weaker.

Finally, the literature discusses the measurement techniques of construction industry productivity. The introduction of the EU KLEMS database has made it easier to perform comparative analyses at the industry level. Some papers have analyzed productivity dynamics in construction (Crawford and Vogl, 2006; Abdel-Wahab and Vogl, 2011; Ruddock and Ruddock, 2011), but the measurement is not without problems. Sveikauskas *et al.* (2016) argued using detailed US data that productivity growth in the construction industry may be somewhat greater than previous results suggest. Notably, there have been attempts to include details of project-level dynamics to better understand the increase in the quality of construction outputs (Sezer and Bröchner, 2014). One conclusion is that it should be possible to use the increasing volume of available performance indicator data collected for construction projects, and to thereby improve the quality of the productivity statistics. However, this approach has so far been infeasible due to the limited resources of measurement activities.

Statistical problems and the heterogeneity of data collection practices call for caution in making comparisons of productivity levels across countries (Vogl and Abdel-Wahab, 2015). Acknowledging the difficulties, our approach is to use productivity growth statistics to analyze changes in value chains over time. Moreover, due to the short-term impacts of business cycles (Abbott and Carson, 2012), we focus on the average behavior of the sector over the different phases of the most recent business cycle (2001–2014).

Methodology for measuring productivity in the value chains

In essence, we combined industry-level productivity contributions of different inputs in the EU KLEMS data with the WIOD data to compute the productivity contributions of different industries in the value chain.

Data

In our analyses, we used the 2016 release of the WIOD database.[5] It builds on a set of consistent time series of national supply and use tables that are constructed by harmonizing the corresponding national tables and benchmarking them against the national accounts. The national tables are then used to derive international tables. They build on the disaggregation of imports by country of origin and use category by using bilateral trade data. Finally, the national tables are combined to yield corresponding world tables, which are then transformed into a world input–output table (WIOT) (Dietzenbacher *et al.*, 2013; Timmer *et al.*, 2015, 2016).

The data comprise sector-level World Input–Output Tables (WIOTs) with underlying data for 44 countries and 56 sectors, which serve as a model for the rest of the

5 Timmer *et al.* (2015); http://www.wiod.org/home

world for the period 2000–2014.[6] Together, the countries cover more than 85 per cent of the world GDP (at current exchange rates). WIOTs are built based on National Accounts data, which are extended by means of disaggregating imports by country of origin and using categories to generate international supply and use tables (Timmer *et al.*, 2016).

Our approach combines the WIOD database and the EU KLEMS database (Jäger, 2017, www.euklems.net), the World KLEMS, the WIOD Socio Economic Accounts (SEA) data (available at the WIOD 2016 database), and the Penn World Table (version 10.0, https://www.rug.nl/ggdc/productivity/pwt/). The combination of several datasets is necessary to meet with our method's high requirement of factor input data. The WIOD data show value added contributed by all countries and industries across the world. The contributions can come from any industry and any country within the WIOD database, directly or indirectly needed to produce the final product, in the construction industry, in a given country. However, a further step is necessary to translate the values from each country industry into the implied factor usage.

The EU KLEMS data provides the main factor data for our analysis. EU KLEMS is constructed to provide internationally comparable and consistent time series on outputs, inputs, and productivity by industry (Jäger, 2017). The database includes EU-25 and several other industrialized countries. In general, data for 1970–2005 are available for the old EU-15 nations as well as for the United States, Australia, and Japan. Series from 1995 onward are available for the new EU member states that joined the EU on May 1, 2004. The coverage of the data differs across countries, industries, and variables. In practice, we found that the 12 construction value chains used in our analyses have the best scope of productivity data both within-country and internationally. For the rest of the world economy that is not covered by the EU KLEMS database, we collected complementary data from the world KLEMS, WIOD SEAs and the Penn World Table.

Measurement of the value-added contribution in the value chain

We applied a measurement framework for the decomposition of value added in the construction value chain grounded on hypothetical extraction, a parsimonious mathematical technique based on an input–output representation of the global economy (Los, Timmer, and de Vries, 2016). This approach has clear economic intuition and can easily be applied to the data. It compares the actual, global value-added distribution with a hypothetical distribution in cases where there are no production activities related to construction.

6 The countries have been chosen by considering both the data availability of sufficient quality and the desire to cover a major part of the world economy. They include 27 EU countries and 15 other major countries. Data for the 56 sectors are classified according to the International Standard Industrial Classification Revision 4 (ISIC Rev. 4). The tables adhere to the 2008 version of the System of National Accounts (SNA). The dataset provides World Input–Output Tables (WIOTs) in current prices, denoted in millions of dollars (Timmer *et al.*, 2016). It is notable that we control the monetary inflation component by using VA shares that divide industry VA by the overall value chain VA.

The difference is defined as the value added of construction activities. In the hypothetical world, the construction industry in each observed country seizes the opportunity to generate final goods, as well as intermediate products, to other industry–country pairs.

In our analysis, we first constructed a value-added matrix VA that allocated the total value added into the contributions of different intermediate good producer industries globally across time, countries, and industries. By extracting applicable elements from the input-output tables, we then constructed the counterfactual scenario ($VA*$) and calculated their element-wise difference to provide us with the corresponding value-added contributions $\Delta VA = VA - VA^*$. We describe the methodology rather extensively in the Appendix.

Factor-based productivity contributions to construction output

We next discuss our approach to measuring the productivity contributions of different sectors. The classical KLEMS productivity approach is commonly used to analyze productivity of the construction industry. In an approach that builds on Jorgenson, Gollop, and Fraumeni (1987), gross output production function includes two types of factor inputs, capital (K) and labour (L), and three types of intermediate inputs, energy (E), materials (M), and services (S). This approach offers useful insights into the changes in efficiency with which the inputs are being used in the production process of the industry (or firm), as measured by productivity growth.

Recently, modelling and measuring patterns of substitution and productivity growth at the industry (or firm) level has become both more difficult and less meaningful (Timmer, 2017). With increased outsourcing and offshoring, the share of industry value added in gross output is declining. Consequently, analyses based on industry value added have to rely on strong assumptions of separability. However, as conditions that are jointly necessary and sufficient for the existence of sectoral value-added functions are typically rejected in the data, intermediate inputs should be treated in the same way as factor inputs in the productivity analysis. Thus, the robustness of the KLEMS approach becomes increasingly dependent on proper price measurement of intermediate inputs.[7] These are increasingly hard to measure due to the practice of transfer pricing in multinational enterprises, the difficulty in pricing the flow of intangibles, and an inadequate statistical system to track prices of intermediates when quality is improving (Houseman and Mandel, 2015; Timmer, 2017).

We propose a production function (F) where final output is based on factor inputs only, including both domestic and foreign factors, similar to Wolff (1994) and Timmer (2017). Using the information from

7 Arising methodological complexities concerning the measurement of a value chain function are discussed by OECD (2001). Fundamentally, the deflation of gross output is conceptually straightforward, whereas the volume change for value added combines the volume change for gross output and intermediate inputs, and thereby constitutes a general-form double deflation.

the hypothetical extraction method, discussed above, the flow of intermediate inputs will be netted out so that the production function of a final good can be expressed in terms of factor inputs only. They are located both in the industry where the last stage of production takes place and in other industries (domestic and foreign) contributing to earlier stages of production. The actual contributions are measured individually for each country's construction value chain by using the hypothetical extraction method.

Formally, let F be a translog production function for the construction aggregate product: $f = F(\Lambda, K, T)$, where Λ is the column vector of labour requirements for production, K is similarly a column vector of capital requirements, and T denotes technology. The factor requirements are measured using industry-specific input-to-value-added ratios.

Under the standard assumptions of constant returns to scale and perfect input markets, the productivity decomposition into components of the different industries and the TFP can be derived (see Appendix for further details). The decomposition of the real gross output growth in the construction industry (Y_{t,F^s}) into the contributions of factors and the TFP (π) as residual is:

$$\begin{aligned}\Delta \log(Y_{t,F^s}) = \overline{\alpha^L}(F^s)\Delta \log(\Lambda_t) \\ + \overline{\alpha^K}(F^s)\Delta \log(K_t) \\ + \Delta\pi(F^s)\end{aligned} \tag{1}$$

where the resource-use vectors of all industries (in discrete time) are $\Delta \log(\Lambda_t)$ and $\Delta \log(K_t)$. L_t, K_t, and Y_t are the labour and capital inputs, and the industry's gross outputs, respectively, while $\alpha^L(F^s)$ and $\alpha^K(F^s)$ are constructed Törnqvist shares of the resource costs. They combine the value-added contribution of each industry–country observation to the construction value chain (obtained with hypothetical extraction), and the corresponding measures of the labour and capital cost shares from the productivity data (KLEMS, SEA, Penn).

To add further detail to the analysis, we decomposed labour growth contribution into the components arising from the change in the number of hours and change in the composition of the labour force. Labour is cross-classified in EU KLEMS according to educational attainment, gender, and age, with the aim to proxy for differences in work experience, providing 18 labour categories (3 × 2 × 3 types). It is assumed that service flows are proportional to the hours worked, and wages reflect the relative marginal productivity of labour (Jäger, 2017). This allowed us to decompose the labour input growth into contributions of labour composition LC and number of hours H. However, this approach can be criticized, especially due to the division of labour into gender groups in which wages may not, in fact, reflect productivity differences. Therefore, we reconstructed labour composition indices that only distinguish between educational attainment. In the 2017 EU KLEMS data, we were able to recalculate the composition from 2008 onwards, while we use the original composition for the previous years and in the EU KLEMS-based productivity measurements that are made for comparisons.[8]

Furthermore, we distinguished between ICT capital and non-ICT capital. In the EU KLEMS data, distinctions are made between three ICT assets (office and computing equipment, communication equipment, and software) and four non-ICT assets (transport equipment, other machinery and equipment, residential buildings, and nonresidential structures). ICT assets are deflated using a quality-adjusted investment deflator based on the methodology described in Timmer *et al.* (2007). Capital service flows are derived by weighting the growth of stocks by the share of each asset's compensation in total capital compensation using the Törnqvist index. In this way, the aggregation takes into account the widely different marginal products from the heterogeneous stock of assets by using weights related to the user cost of each asset. The user cost approach is crucial for the analysis of the contribution of capital. This approach is based on the assumption that marginal costs reflect the relative marginal productivity in the corresponding capital type.

A practical caveat of the empirical analysis based on the EU KLEMS data, is that we cannot account for all the involved productivity growth of industries in the value chain. While the share of included value added is large (on average 86 per cent of all value added in the considered construction value chains), it can be argued that merely focusing on the EU KLEMS data might bias our results. To overcome this problem, we thus use alternative datasets to approximate the missing factors.

First, we employed the World KLEMS dataset that includes KLEMS data for Japan, Korea and Russia. This data typically spans from the mid-2000s to the early 2010s. Where data was still missing, we used a combination of WIOD SEA data and the Penn World Table data. The SEAs provide labour input in hours at the industry level for the WIOD countries. To complement this data, we used the Penn World Table data to measure the country-level average labour quality index. We adjust the SEA labour inputs to provide an approximate, yearly labour services index for each industry in each country. For capital services, we use the PENN world table country-level averages for countries where other data are unavailable.

Different datasets are combined by measuring the factor growth components from each dataset, and then replacing observations accordingly, when data are found to be missing. Finally, the WIOD database includes a rest-of-the-world category that aggregates data from small emerging and developing countries. While its contribution to the value chains is negligible, we still provide an approximation of its factor use for the sake of completeness. To this end, we assume that the factor intensities correspond to the Chinese factor intensity at the same time period.

By this, we have completed the description of our methodology. We provide the results in three sections below, each titled with what we regarded as a key finding.

8 We find that the differences are in practice small, and do not affect our main results.

Table 1: Value-add Shares of Different Sectors in Construction Value Chains

	Share of the value chain value VA in 2014 (%)				Change of the share between 2000 and 2014 (percentage points)			
	Primary	Manufacturing	Construction	Services	Primary	Manufacturing	Construction	Services
AUT	3	16	53	28	1.1	0	-6.1	4.9
BEL	4	16	43	37	0.7	-1	-2.3	2.7
CZE	3	14	44	38	-0.7	-5.9	1.1	5.6
DEU	2	15	49	33	0.7	-3.2	1	1.5
DNK	4	16	41	38	0.6	-4.2	-1.7	5.3
ESP	3	11	53	33	0.6	-5.7	-4.4	9.6
FIN	6	19	45	30	1	-4.9	2.2	1.7
FRA	3	15	49	34	0.5	-3.6	3.4	-0.3
GBR	3	12	56	29	0.7	-1.1	-1.1	1.5
ITA	3	13	48	36	-0.1	-5.8	6.6	-0.6
NLD	3	19	44	33	1.8	1.9	-2	-1.7
SWE	4	12	50	34	0.5	-4.1	0.4	3.2
Average	**3**	**15**	**48**	**34**	**0.6**	**-3.1**	**-0.3**	**2.8**

Note: Primary production = industries A and B, manufacturing = industries 10-33, construction = industry F, and Services = all other industries in the international standard industrial classification (ISIC).

Source: WIOD database and authors' calculations

The Role of Services has Increased in Construction Value Chains

We first analyzed the industry composition of value added in our value chains. We decomposed the value added in the chain to components from four sectors: primary, manufacturing, construction, and services. The results are reported in Table 1. As this analysis did not involve productivity measurements, we can analyze the full decomposition of the value added based on the WIOD database.

Our results show that, on average, the construction industry only accounted for roughly 48 per cent of the value-add generated in the value chain in 2014. The second largest contributing sector was services, which generated 34 per cent of the value-add, while manufacturing and primary production generated 15 per cent and 3 per cent, respectively. The results also suggest that there are similarities in the organization of construction activities across countries. For example, in 2014, the share of the construction industry varied within 41 per cent and 56 per cent of total value added. The largest shares of the construction industry were found in the UK (56 per cent), Spain (53 per cent), and Austria (53 per cent), while the smallest shares were measured in Denmark (41 per cent), Belgium (43 per cent), and the Czech Republic (44 per cent).[9]

Over time, there have been some changes in the shares. From 2000 to 2014, the average share of the service sector increased 2.8 percentage points, while the share of the manufacturing sector decreased by roughly the same amount. A closer look at the data showed that this development seems to be associated with a reallocation of tasks in

9 Sector-wise, the largest average value-added shares of non-construction industries in the considered GVCs are for professional, scientific, technical, administrative and support service activities (10 per cent); wholesale trade, except of motor vehicles and motorcycles (5 per cent); rubber and plastics products, and other non-metallic mineral products (4 per cent); basic metals and fabricated metal products, except machinery and equipment (4 per cent); transport and storage (3 per cent); real estate activities (3 per cent); financial and insurance activities (3 per cent); mining and quarrying (3 per cent); retail trade, except of motor vehicles and motorcycles (2 per cent); electrical and optical equipment (2 per cent).

Table 2: Growth and its Components: Comparison of the Construction Value Chain and the Construction Industry

Panel A: Real gross output growth and its components in the construction value chain

	Capital share (%) (a)	Growth components excluding hours: (b) = c+d+g	TFP growth contribution (c)	Capital growth contribution (d)	ICT growth contribution (e)	NIT growth contribution (f)	Labour composition contribution (g)	Hours contribution (h)
AUT	38	0.6	0.2	0.5	0.2	0.2	-0.1	0.2
BEL	41	1.7	-0.4	1.3	0.3	0.8	0.8	0.4
CZE	40	2.1	0.9	1.1	0	1.2	0.1	-0.7
DEU	25	1	-0.3	0.1	0	0.1	1.2	-1.6
DNK	27	0.3	0	0.3	0	0.2	-0.1	-0.5
ESP	39	2.1	2.3	0.6	0	0.7	-0.8	-2.1
FIN	26	0.5	0.5	0.2	0.1	0.2	-0.3	-0.1
FRA	27	0.3	-0.1	0.5	0.1	0.3	-0.1	0.3
GBR	20	0	-0.8	0.4	0.1	0.2	0.4	0
ITA	34	-1	-0.6	-0.1	-0.1	0.1	-0.3	-1
NLD	20	0.9	0.3	0	0	0.1	0.5	-1.6
SWE	38	0.9	-0.3	1.5	0.2	1.2	-0.3	0.8
Average	**31**	**0.8**	**0.2**	**0.5**	**0.1**	**0.4**	**0.1**	**-0.3**

Panel B: Real value-added growth and its components in the construction industry

	Capital share (%) (a)	Growth components excluding hours: (b) = c+d+g	TFP growth contribution (c)	Capital growth contribution (d)	ICT growth contribution (e)	NIT growth contribution (f)	Labour composition contribution (g)	Hours contribution (h)
AUT	36	-1	-1	0.2	0.1	0.1	-0.1	0
BEL	39	2	0.2	1.6	0.2	1.4	0.2	0.3
CZE	30	1.3	-0.4	1.2	0.1	1.1	0.5	-0.4
DEU	5	0	-0.3	0.1	0	0.1	0.2	-1.3
DNK	15	1.1	-0.1	0.2	0.1	0.1	1.1	-0.4
ESP	38	-0.2	-1.8	1.1	0	1.1	0.5	-2.8
FIN	9	-0.2	-0.6	0.4	0	0.4	-0.1	0.5
FRA	18	-1.5	-1.9	0.3	0.1	0.2	0.2	1.1
GBR	10	0.3	-0.1	0.2	0	0.1	0.3	0.6
ITA	27	-1.1	-1.5	0.3	0	0.3	0.1	-0.3
NLD	1	0.2	-0.4	-0.1	-0.1	0	0.7	-1.2
SWE	33	-0.7	-1.8	1.6	0	1.5	-0.4	1.2
Average	**22**	**0**	**-0.8**	**0.6**	**0**	**0.5**	**0.3**	**-0.2**

Note: All columns express annual, average (real, simple mean) percentage point growth contributions 2000-2014. TFP is the total-factor productivity, ICT is information and communications technology capital stock, and NIT is the traditional capital stock. Construction output's growth accounting in the value chain was conducted by using the methodology outlined in Section 3, while the construction industry value-added growth decomposition uses the methodology and data of the real value-added based growth accounting measurement of industry F in the EU KLEMS database.

Source: EU KLEMS, World KLEMS, WIOD SEA, Penn World Table and authors' calculations.

construction activities from the construction industry to various business services. There are three main contributing industries to the increase of the services sector: (1) professional, scientific, technical, administrative, and support service activities (+ 1.6 percentage points); (2) financial and insurance activities (+ 0.5 points); and (3) wholesale trade, except of motor vehicles and motorcycles (+0.4 points).

Productivity Growth in the Construction Value Chain is Higher than in the Construction Industry

We analyze the origins of productivity growth in the value chains using the methodology on factor-based accounting of the industry growth contributions as presented in Section 3.

Our baseline findings show that the construction value chain is rather different from the construction industry in terms of capital intensity and its sources of growth (see Table 2).[10] The average capital intensity (the share of capital income in value added) in the value chain is 9 percentage points greater than that of the construction industry (31 vs. 22 per cent) for the measured part of the value chain. This suggests that the return of production capital per unit of nominal output is higher in construction value chains, indicating that either there is more capital or the production is more profitable.

For a deeper understanding, we analyzed the average real gross output growth rate of the value chain, while excluding the role of merely changing working hours. The average growth rate in the construction value chain has been 0.8 per cent per year, while in the construction industry, the growth (based on the EU KLEMS productivity data on the industry corresponding to ISIC Code F) has been negligible. The rates include the contributions of capital and labour quality deepening and TFP. The difference is explained by many factors: better performance of TFP, capital deepening, and increases in the quality of labour, with the productivity (TFP) growth being the single greatest contributing factor. The finding suggests that the benefits of the organization of construction activities in global value chains may be underestimated when traditional productivity statistics are used.

A few methodological comments should be made. First, it is notable that our choice of correcting the labour composition component by focusing on differences in education has a non-trivial effect on the structure of growth in the value chain. When the EU KLEMS original composition is used, the labour composition effect is estimated to be on average 0.3 percentage points larger, while correspondingly the TFP growth is 0.3 percentage points weaker. While these differences do not affect the overall growth excluding the contribution of working hours, we report an alternative calculation in the Appendix. Furthermore, as we combine data from different sources, not all data includes decomposition of the capital growth into ICT and non-ICT component. When such distinction is not possible, the contribution is merely reported as a part of the overall capital component, while the sub-components are 0 in Table 2. Thus, generally $d \neq e + f$.

Our findings, of course, mask a considerable amount of heterogeneity in country-level construction activities. Particularly interesting is the Belgian construction industry with its marine construction activities — oil platforms, dredging, undersea building, quay construction, etc. — residing at the high end of the productivity distribution. Excluding the contribution of hours, the rate of productivity growth in this industry has been 1.7 pps per year.

10 The first part of the table decomposes real output growth of the entire construction value chain. It constitutes all the construction value chain that includes the value added of the construction industry and the value of the intermediate goods and services, both domestic and foreign, used by the construction industry to produce its gross output. In the latter part, the construction industry is narrowly defined according to the ISIC Rev. 4 industry classification (F) and by its growth decomposition we refer to the value-added based KLEMS measurements for the industry.

Table 3: Growth and its Components: Comparison of the Construction Value Chain and the Construction Industry

Panel A: Contribution of the upstream part of the construction value chain

	VA share (%) (a)	Capital share (%) (b)	Growth components excluding hours: (c) = d+e+h	TFP growth contribution (d)	Capital growth contribution (e)	ICT growth contribution (f)	NIT growth contribution (g)	Labour composition contribution (h)	Hours contribution (i)
AUT	45	41	0.5	0.1	0.4	0.2	0.2	0	0.2
BEL	54	42	0.9	0.1	0.5	0.2	0.1	0.2	0.4
CZE	56	47	0.6	0	0.6	-0.1	0.7	0.1	-0.3
DEU	52	44	0.2	0.1	0	0	0.1	0.1	-0.5
DNK	59	36	0.2	0	0.3	0	0.2	-0.1	-0.2
ESP	46	39	-0.4	0	0	0	0.1	-0.3	-0.6
FIN	56	39	0.4	0.3	0	0.1	0	0.1	-0.4
FRA	51	36	0.5	0	0.3	0.1	0.2	0.1	0
GBR	42	34	0.7	0.2	0.3	0.1	0.1	0.1	0.1
ITA	53	41	-0.5	-0.1	-0.2	-0.1	0	-0.1	-0.6
NLD	51	39	0.2	0.1	0	0	0	0.1	-0.5
SWE	49	43	0.7	0	0.7	0.2	0.4	0	0.3
Average	**51**	**41**	**0.3**	**0.1**	**0.2**	**0.1**	**0.2**	**0**	**-0.2**

Panel B: Contribution of the construction industry part of the value chain

	VA share (%) (a)	Capital share (%) (b)	Growth components excluding hours: (c) = d+e+h	TFP growth contribution (d)	Capital growth contribution (e)	ICT growth contribution (f)	NIT growth contribution (g)	Labour composition contribution (h)	Hours contribution (i)
AUT	55	38	-0.5	-0.6	0.1	0	0.1	-0.1	0
BEL	46	35	1.5	0.1	0.7	0.1	0.7	0.6	0
CZE	44	36	0.4	-0.2	0.5	0.1	0.5	0.1	-0.4
DEU	48	13	1.1	-0.1	0.1	0	0	1.2	-1.1
DNK	41	17	-0.2	-0.3	0.1	0	0	0	-0.3
ESP	54	38	-0.9	-1	0.6	0	0.6	-0.5	-1.5
FIN	44	23	-0.4	-0.3	0.2	0	0.2	-0.3	0.4
FRA	49	26	-1	-1	0.1	0	0.1	-0.2	0.3
GBR	58	17	0.3	-0.1	0.1	0	0.1	0.3	-0.1
ITA	47	31	-0.7	-0.7	0.1	0	0.1	-0.1	-0.5
NLD	49	16	0.2	-0.2	0	-0.1	0	0.4	-1.1
SWE	51	34	-0.4	-0.9	0.8	0	0.8	-0.3	0.5
Average	**49**	**27**	**-0.1**	**-0.4**	**0.3**	**0**	**0.3**	**0.1**	**-0.3**

Note: All columns express annual, average (real, simple mean) percentage point growth contributions 2000-2014. TFP is the total-factor productivity, ICT is information and communications technology capital stock, and NIT is the traditional capital stock. GVC productivity contributions of the different parts' inputs were measured by using the methodology outlined in Section 3, while TFP estimates build on the value-added based measurements of TFP in the EU KLEMS database, as weighted by the Törnqvist shares of individual industries in the GVC value added.

Source: EU KLEMS, World KLEMS, WIOD SEA, Penn World Table and authors' calculations.

Growth has followed consolidation and investments, driven by knowledgeable customers who have demanded extreme precision despite very difficult building environments. These are understandable requirements, since the difference between smooth flow and utter catastrophe lies in the quality of the seam in underwater oil pipes. Increased size and complexity of projects has spurred the development further by forcing companies to use machines instead of labour (Economist, 2017b).

The improvements in growth performance, clearly seen in our data, can be traced to mechanical improvements of tools, increased measurement and use of ICT, and introduction of modular building. However, the cases of Spain and the Czech Republic seem quite different. In those countries, the growth improvements (in both cases 2.1 pps annually when the contribution of hours is excluded) are associated with large reductions of the labour force, which suggests that the initial level of productivity may have been rather low, but more recently, the country has caught up in productivity with respect to other countries.

At the low end, the construction industries in Italy and France show poor growth performance. What is interesting, however, is that in the case of France, our accounting of the value chain TFP growth at least partly offset the poor developments of the industry. This suggests that productivity growth within the construction value chain has shifted more towards upstream industries and away from the construction industry itself: a phenomenon which has not been visible in traditional statistics.

We further decomposed the real gross output growth of the value chain into contributions of the domestic construction industry and those from upstream (all other industries) in the value chain (Table 3). This approach focuses on the different components of the construction value chain measurement in Table 2, but distinguishes between growth components originating in the different parts of the value chain. In the case of the growth contribution to the inputs, it is straightforward.[11]

However, TFP contribution of the total value chain cannot be allocated directly to either part of the value chain. To overcome this shortcoming, we collected value-added growth-based TFP growth estimates from the EU KLEMS dataset, and following Timmer (2017), used them to separate TFP growth contributions of domestic construction industry from the rest of the upstream value chain. Effectively, the GVC-based TFP can be viewed as a weighted average of TFP of the production's last stage and upstream, with the value-added shares of the industries in the value chain as weights. While this approach is not without caveats and can be done only for industries that have KLEMS-based TFP calculations, it may still help to source the GVC-based TFP back to the different parts of the chain.

11 In practice, we allocate real output growth of the entire construction value chain (Panel A of Table 2) to the industry and the rest of the value chain components by first dividing construction value chain's value added to the industry part (ISIC Rev. 4 classification F) and the rest of the value chain part. We then measure the corresponding factor use separately for the different parts.

Table 4: Decomposition of the Real Construction Gross Output Growth to the Foreign Components

	Foreign part of the value chain								
	VA share (%) (a)	Capital share (%) (b)	Growth components excluding hours: b = c+d+g	TFP growth contribution (d)	Capital growth contribution (e)	ICT growth contribution (f)	NIT growth contribution (g)	Labour composition contribution (h)	Hours contribution (i)
AUT	20	41	0.3	0.1	0.2	0.1	0	0	0.1
BEL	29	43	0.6	0	0.4	0.2	0.1	0.1	0.3
CZE	23	43	0.2	0.1	0.1	0	0.1	0	0
DEU	15	45	0.1	0	0.1	0.2	-0.1	0	-0.1
DNK	27	42	0.3	0.1	0.2	0	0.1	0	0
ESP	12	42	-0.1	0	-0.1	0	0	0	-0.2
FIN	21	44	0.2	0.1	0.1	0	0.1	0.1	-0.1
FRA	16	42	0.2	0.1	0.2	0.1	0.1	0	0.1
GBR	12	44	0.2	0	0.2	0	0.1	0	0.1
ITA	12	43	0	0	-0.1	0	0	0	-0.2
NLD	25	42	0.3	0.1	0.1	0	0	0.1	-0.1
SWE	20	42	0.2	0	0.2	0.1	0.1	0	0
Average	**19**	**43**	**0.2**	**0**	**0.1**	**0.1**	**0**	**0**	**0**

Note: All columns express annual, average (real, simple mean) percentage point growth contributions 2000-2014. TFP is the total-factor productivity, ICT is information and communications technology capital stock, and NIT is the traditional capital stock. GVC productivity contributions of the foreign inputs were measured by using the methodology outlined in Section 3, while TFP estimates build on the value-added based measurements of TFP in the EU KLEMS database, as weighted by the Törnqvist shares of individual industries in the GVC value added.

Source: EU KLEMS, World KLEMS, WIOD SEA, Penn World Table and authors' calculations.

When we measure the contribution of the upstream TFP growth in this manner, shown in Table 3, the results suggest that upstream contributed substantially more to the overall productivity growth of the construction value chain. The TFP growth contribution of upstream was roughly 0.1 percentage points per year, while the construction industry's contribution was −0.4 percentage points. However, a significant part of the overall TFP growth in the value chain remains in our analysis unexplained. This is in particular due to the low TFP growth contribution that arises from the EU KLEMS-based measures of the construction industry. As a result, a large portion of the overall GVC-based TFP growth remains unallocated to either parts of the chain.

What might explain these dynamics? One natural explanation for low productivity growth in the construction industry is that there is a shift of the more productive tasks from construction to the upstream part of the value chain. As more productive tasks are shifted to the upstream part of the value chain, the remaining industry tasks are less productive. However, the productivity of the total value chain has nevertheless increased through reallocation of the tasks. This may not appear in the traditional TFP measurements. In particular, if production moves to industries with higher TFP levels in the upstream, it is likely to show up as an increase in the overall productivity through value chain TFP residual beyond the TFP growth measured from the industry-level.

The results may partly reflect measurement problems too. It could be that the growth of the output volume index may be underestimated, as was suggested by previous papers in the literature (Harrison, 2007). Moreover, the validity of the analysis of TFP into the industry location of productivity growth in the GVC depends heavily on the quality of the intermediate input deflator (Timmer, 2017).

We also studied the role of the value chain in output growth by the origin of the supplier (Table 4).[12] In particular, we divided the chain into components that reflect growth components in the domestic and foreign parts of the value chain, and again collected information on the TFP growth from the EU KLEMS measurements.

It turned out that the role of the foreign part was not dominant in productivity dynamics. In terms of the capital deepening and improvements of the labour composition, the foreign part of the chain contributed only roughly 0.2 percentage points per year to the overall productivity growth, whereas the rest can be assigned to the domestic part of the value chain or the overall efficiency gains in it. Note that in Table 4, we only report the growth components in the foreign part, while the domestic part is the residual between it and the overall growth in the chain (Table 2).

Innovations Support Long-term TFP, while Administrative Costs and the Efficiency of ICT Adoption Pose Challenges

TFP of the entire construction value chain reflects the total productivity of all industries that interact within the field of construction. By looking at the complete value chain, we can assess the role of factors that may influence productivity growth. This may help to better understand the determinants of productivity growth of construction activities as well as further validate our approach. As each factor requires separate datasets, our first task was to identify potential factors and justify their relevance. We ended up assessing three factors: innovativeness, administrative costs, and ICT investments. Our reasoning and data collection went as follows.

First, we expected that the degree of construction-related innovativeness is positively related to TFP. While there is no unique way to measure innovativeness, we resorted to one standard measure: the number of construction technology related patents granted in the corresponding country. To this end, we identified all International Patent Classification (IPC) patent classes that we assessed to have a potential link to the construction industry. This yielded a list of 49 patent classes for further analysis.

As the definition of construction is broad, it is most probable that some patent classes are missing, and some might be superfluous. However, we deem the approach to be transparent, straightforward and sufficiently consistent. The list covers a wide variety of different patent classes, including innovations in materials, construction technology, lighting, electricity, and air-conditioning systems (see Appendix 2 for patent classifications). Importantly, these innovations are made not only by the

12 In practice, we allocate real output growth of the entire construction value chain (first part of Table 2) to domestic (omitted in the Table) and foreign (reported in the Table) components based on the nationality of the construction industry. The procedure is similar to the one that we use to construct Table 3.

construction industry but also possibly by other industries in the construction value chain. The inclusion-exclusion boundary was set at patent classes that would most likely be exploited mainly in sectors outside the construction value chain.[13]

Second, we expected administrative costs to lower the efficiency of the value chains, as administrative costs are widely perceived as non-productive additional costs, and then turn into obstacles for optimal allocation of resources. We studied this potential effect using internationally comparable data provided by the World Bank Group's *Doing Business* project from year 2006 to 2014. With a warehouse as the representative example, the project recorded all official costs associated with completing the procedures to legally build a warehouse.[14] The administrative costs are presented as a percentage of the warehouse value.

Third, ICT projects have a virtually universal tendency to exceed original resource allocations, be it in terms of time or costs. The challenges, but also prospects of ICT investments may become particularly large when considered jointly for the entire value chain. Strong positive interactions may emerge in the value chain when new ICT technology generates positive productivity externalities or when there are unmeasured complementary innovations that are made during the adaptation of the technology (Stiroh, 2002b; Basu and Fernald, 2007). Due to these factors, the neoclassical growth assumptions may not apply, and the elasticity of ICT in the production function may not match the measured input share of ICT. As a result, a direct relationship between ICT capital and measured TFP growth may arise. We applied our previously collected growth accounting data to study this question.

For estimating the roles of each factor, we resorted to panel data estimations using yearly data and the value chains of different countries as panel units. We estimated a panel error correction model to analyze the long-term relationship between TFP and the different factors. First we studied the time series properties of our variables of interest. We found that the index of TFP, the cumulative capital and labour contributions — constructed by summing the yearly log-point contribution terms — and the level of patent intensity are trend stationary and cointegrated of order 1.[15]

13 Because TFP growth measures the growth of productivity and is not, per se, related to the size of the sector, we studied the intensity of patent activities by dividing the total number of patents by the number of employees in the construction industry. We total the number of construction-related patent applications to the EPO by applicant country of residence and application year.

14 See, World Bank, *Doing Business* reports 2006-2014. https://elibrary.worldbank.org/. The data include the costs associated with obtaining land use approvals and preconstruction design clearances; receiving inspections before, during and after construction; obtaining utility connections; and registering the warehouse at the property registry. It is calculated as a percentage of the warehouse value. Nonrecurring taxes required for the completion of the warehouse project are also recorded. Sales taxes (such as value added tax) or capital gains taxes are not recorded. Nor are deposits that must be paid up front and are later refunded.

15 By using Im–Pesaran–Shin and Fisher-type tests in Stata (xtunitroot package), we find that the zero hypotheses of all panels having unit roots cannot generally be rejected. However, in the case of patent intensity, it is possible that the variable was (weakly) stationary after controlling for a linear time trend. We also test the cointegration of the variables by using the xtcointtest package in Stata and found that the cointegration relationship cannot generally be rejected, based on Kao, Pedroni, and Westerlund types of cointegration tests.

Cointegration, indeed, indicates that there may be a common growth element showing as a linear relationship between the variables, in the form of a stationary linear combination. Failure to account for it may result in spurious correlations between the variables. Accordingly, we studied separately the short-term dynamics and long-term equilibrium relationships between the different factors and TFP (O'Mahony and Vecchi, 2005).

To establish a long-run relationship between TFP and the different input growth contributions, we first need to make a few, additional methodological remarks concerning the applied statistical model.

We used the so-called mean group estimator developed by Pesaran, Shin, and Smith (1999). Our application, to estimate a common long-term relationship for each construction value chain, was as follows:

Let us denote TFP as π_{it} and the contributions of the different factors as c^{factor} (patents, administrative costs, or ICT) respectively. Then, the relationship for the value chain $i = 1,2,\ldots,12$ and time period $t = 2001,2002,\ldots,2014$ is:

$$\pi_{it} = \theta_{ifactor} c_{it}^{factor} + \mu_i + \epsilon_{it}. \qquad (2)$$

With our variables being I(1)[16] and cointegrated, the error term is I(0) for all i. The corresponding auto-regressive, distributed-lag specification of the relationship between TFP and the contributing variables can be expressed in the error correction form:

$$\begin{aligned}\Delta\pi_t = \phi_i(\pi_{it-1} - \theta_{i0} - \theta_{ifactor} c_{it}^{factor}) \\ + \delta_{ifactor}\Delta c_{it}^{factor} + \delta_{it} t \qquad (3) \\ + \delta_{it}^{SQ} t^2 + \epsilon_{it},\end{aligned}$$

where the first term is the long run cointegration relationship between TFP and input contributions. The θs denote the long-term elasticity of different factor contributions with TFP, the δs are the short-term elasticities, and ϕ_i is the error correction speed of the adjustment parameter. The key parameters of interest are long-term elasticity of patent intensity $\theta_{ifactor}$ and the error correction speed of adjustment.

Table 5 concludes the results of the error correction analysis. We considered three specifications (a-c), separately for the entire construction value chain (GVC TFP), and for the corresponding EU KLEMS-based core construction industry (value-added based TFP). Specification (a) includes patent intensity as the explanatory factor variable. Specification (b) considers administrative costs as the explanatory factor variable. Specification (c) analyzes the relationship between ICT capital growth

16 where I() denotes the order of integration.

17 While in the neoclassical growth model, TFP estimates should be "free" from such factor contributions, the correlation may arise from spillovers, omitted variables, embodied technological progress, measurement errors, or reverse causality (Stiroh, 2002a). In particular, the correlation may turn negative if there are adaptation frictions (Basu and Fernald, 2007). As ICT growth, we used the ICT capital growth component of our previous analysis in case of the whole GVC. For the industry, we used the EU KLEMS ICT capital growth component.

18 We also considered higher order trends, but they do not significantly affect our results. On the other hand, we found that using only a linear trend would be too restrictive an assumption.

component and TFP.[17] All three models include quadratic year trends.[18] Table 5 shows estimates from the different pooled mean group specifications, which allows for heterogeneous short-run dynamics and common long-run relationships. The reported short-term dynamic parameters are the averages of the corresponding value chains.

Our results show that more intensive long-term engagement in patenting activities is systematically linked with value chains that have higher TFP growth. The point estimates of the long-term relationship in Table 5 (row "Patent intensity θ_{PAT}") was 1.028. The coefficient implies the effect of one patent or more per 1,000 employees to the growth rate of TFP. For the whole value chain, the rate by which the current state is corrected towards the long-term relationship is 33.5 per cent per year, as indicated by the speed of the adjustment parameter. In the case of the construction industry only, we found a similar relationship, but this relationship was weaker than in the case of the full value chain (0.726), while the long-term relationship is captured faster.

In the value chain, the analysis suggests that one standard deviation increase in patenting is associated with a long-term increase in productivity through higher TFP by roughly 25 per cent, while the positive effect is one quarter weaker for the industry-only-based TFP.

We then considered the role of administrative costs. Administrative costs show a negative and statistically significant long-term relationship to the GVC-based TFP (−0.019). In this case, however, there was no inertia in reaching the long-term relationship, as indicated by the speed of adjustment close to 100 per cent. This might partially reflect the short dataset that we had for the administrative cost parameter. In case of the industry-based TFP, we found that the long-run relationship is positive. This finding strikes us initially as counter-intuitive. However, it might in fact reflect industry productivity remaining higher where the productive parts of the chain remain in the industry due to administrative costs slowing down development of the larger value chain.

The analysis suggests that one standard deviation increase in administrative costs (1.5 per cent increase in the administrative costs as relative to the building costs) is associated with a long-term decrease in productivity of the value chain by roughly 2.8 per cent through lowered TFP.

Finally, we analyzed the association between an ICT capital growth component and TFP growth. Our results suggest that there are major adjustment frictions in the value chains. The neoclassical ICT capital contributions may have overestimated the effect on productivity, leading to negative correlation with TFP, even in the long run. It may be that productivity growth could indeed be explained by the level of investment in ICT, but problems arise due to the time lag for a new technology to reach its full potential, and such lags may simply extend beyond the length of our data set. The pace of adjustment towards the long-term effect is relatively fast, which may indicate that the data is not sufficient to observe very low-frequency connections.

In the case of the industry-only-based TFP, the low productivity impacts of ICT are pronounced. This might indicate that

Table 5: Results of the Error Correction Model Analysis of the Link Between Underlying Factors and Total Factor Productivity (TFP)

Dependent variable TFP	Construction value chain (a) Patent intensity	Construction value chain (b) Administrative cost	Construction value chain (c) ICT contribution	Construction industry (a) Patent intensity	Construction industry (b) Administrative cost	Construction industry (c) ICT contribution
Pooled mean group normalized cointegrating vector						
Patent intensity	1.028***			0.726***		
(standard error)	(0.253)			(0.166)		
Administrative costs		-0.019***			0.008***	
(standard error)		(0.004)			(0.002)	
ICT comp. contribution			-2.023***			-9.041***
(standard error)			(0.609)			(2.449)
The average short-run dynamic coefficients						
Δ Patent intensity	0.016			0.047		
Δ Administrative cost		0.008			-0.056	
Δ ICT comp. contribution			1.608			-11.166
Linear time-trend comp.	5.445*	0.87	6.841***	-1.511	-9.568	-0.509
Quadratic time-trend comp.	-0.001*	0	-0.002***	0	0.002	0
Constant	-5.5e+03*	-861.192	-6.9e+03***	1529.35	9635.556	516.65
Speed of adjustment	-0.335**	-1.022***	-0.608***	-0.488***	-0.881***	-0.570***
Number of observations	143	88	156	143	88	156
Number of value chains	11	11	11	11	11	11

Note: The confidence levels are *p < .05; **p < .01; ***p < .001

Source: EU KLEMS, World KLEMS, WIOD SEA, Penn World Table and authors' calculations

the benefits, although weak, are better captured by the GVC-based TFP estimates. In the value chain, the analysis suggests that one standard deviation increase in ICT capital growth contribution (1.2 percentage points) is in association with a roughly 2.5 per cent long-term decrease in (TFP) productivity.

To study the robustness of these findings, we considered models where we jointly studied the role of different factors. While this was not possible for administrative costs due to the limited amount of data, a model that included both patents and ICT showed that similar relationships hold true also in a joint model. Moreover, we also tested alternative estimators, namely the Dynamic Common Correlated Effects Estimator. For patents and administrative costs our results were similar, while the ICT effects suggest that the results may not be robust to potential statistical caveats in

19 We used the statistical package xtdcce2 by Ditzen (2018). The package aims at correcting a few potential caveats in the basic model. First, if a lag of the dependent variable is added, endogeneity occurs and adding solely contemporaneous cross-sectional averages is not sufficient any longer to achieve consistent estimates. We also tested for weak cross-sectional dependence in our panel data. Cross-sectional dependence in the error term occurs if dependence between cross-sectional units in a regression is not accounted for. The results are available from the authors upon request.

the setup.[19] Consequently, we acknowledge that the limited size of our dataset allows only to make tentative conclusions from the analysis.

All in all, our analysis suggests that the GVC-based TFP measurements provide intuitive relationships between key underlying factors, and they seem to be stronger than with traditional, industry-only-based TFP estimates. Results suggest that productivity growth in the construction value chains is fostered by innovativeness, while administrative costs and the low efficiency in the use of ICT may hold back their productivity potential. Our analysis also seems to suggest that the investment frictions in ICT are felt more strongly in the industry (larger negative coefficient) and the effects of patents are stronger in the overall value chain. This is indicative for the differentiated effects of the underlying factors in different parts of the value chains.

Conclusions

In this article, we presented our study of the construction value chains and their productivity. We decomposed the value-added contents of the construction outputs of 12 European countries to the contributions of the entire construction value-chain: construction industry and other construction-related sectors in the upstream value chain. We combined the WIOD data and several international productivity datasets. Using the method suggested by Los, Timmer, and de Vries (2016), we measured the value-added content of the value chain through the exclusion of construction activities from the WIOD.

We found that roughly half of the total value added in the construction value chains was generated in the upstream industries of the value chains; a finding that is common in all observed value chains. The rest of the value added was generated by other industries involving both manufacturing and business services. In particular, we found that the role of the business services sector is important and has increased further over the years 2001 to 2014.

We also used information concerning the value chains to measure their overall productivity growth by accounting for the value-added factor contributions of different parts of the value chain (Wolff, 1994; Timmer, 2017). We showed that there has been more productivity growth in construction activities when the productivity improvements in the upstream part of the production chain are considered. There has been a transformation of production toward a larger role for the upstream value chain that had not so far been documented, while the role of the construction industry in total productivity growth was weak. This reallocation of productivity provides at least a partial explanation for the low productivity growth statistics of the construction industry.

We also showed that there is a strong long-term relationship between construction-related patents and the improvement of TFP in the value chain. This strong effect likely reflects positive productivity effects from increased knowledge. On the other hand, our results also suggested that there are major adjustment frictions due to administrative costs and adoption of ICT in the value chains.

All in all, our results show that the focus on the construction industry is a restrictive one when production value chains are more and more fragmented between the construction industry, manufacturing, and business services. A value chain perspective is pivotal in providing further understanding about the organization and performance of construction. A wider perspective makes more visible also the struggles in the adoption of technology striving to make the value chains more efficient. Together, our results suggest that future tools to improve the productivity in construction are likely to be found from more efficient and flexible formalization of interactions in the value chain that are fostered by innovation, more efficient use of ICT, and lowered administrative costs.

References

Abbott, M. and C. Carson (2012) "A Review of Productivity Analysis of the New Zealand Construction Industry," *Australasian Journal of Construction Economics and Building*, Vol. 12, No. 3, pp. 1–15.

Abdel-Wahab, M. and B. Vogl (2011) "Trends of Productivity Growth in the Construction Industry Across Europe, US and Japan," *Construction Management and Economics*, Vol. 29, No. 6, pp. 635–644.

Ali-Yrkkö, J. and T. Kuusi (2017) "Shield the US from Imports! GDP Impacts on Finland and Other European Union Member States," ETLA Reports 76, ETLA Economic Research (The Research Institute of the Finish Economy). https://www.etla.fi/en/publications/shield-the-us-from-imports-gdp-impacts-on-finland-and-other-european-union-member-states/.

Ali-Yrkkö, J. and T. Kuusi (2019) "Brexit and Indirect Impact Routes Through Global Value Chains," ETLA Reports 89, ETLA Economic Research (The Research Institute of the Finish Economy). https://www.etla.fi/en/publications/brexit-and-indirect-impact-routes-through-global-value-chains/.

Ali-Yrkkö, J., T. Seppälä, and J. Mattila (2016) "The Role of the Largest Companies and Their Value Chains in the Economy," ETLA Reports 53, ETLA Economic Research (The Research Institute of the Finish Economy). https://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-53.pdf.

Arditi, D. and K. Mochtar (2000) "Trends in Productivity Improvement in the US Construction Industry," *Construction Management and Economics*, Vol. 18, No. 1, pp. 15–27.

Bankvall, L., L. E. Bygballe, A. Dubois, and M. Jahre (2010) "Interdependence in Supply Chains and Projects in Construction," *Supply Chain Management*, Vol. 15, No. 5, pp. 385–393.

Basu, S. and J. Fernald (2007) "Information and Communications Technology as a General-purpose Technology: Evidence from US Industry Data," *German Economic Review*, Vol. 8, No. 2, pp. 146–173.

Bresnahan, T.F., E. Brynjolfsson, and L.M. Hitt (2002) "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-level Evidence," *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 339–376.

Coe, D.T., E. Helpman, and A.W. Hoffmaister (2009) "International R& D Spillovers and Institutions,"*European Economic Review*, Vol. 53, No. 7, pp. 723–741.

Crawford, P. and B. Vogl (2006) "Measuring Productivity in the Construction Industry," *Building Research & Information*, Vol. 34, No. 3, pp. 208–219.

Davis, P., T. Gajendran, J. Vaughan, and T. Owi (2016) "Assessing Construction Innovation: Theoretical and Practical Perspectives," *Construction Economics and Building*, Vol. 16, No. 3, pp. 104–115.

Dietzenbacher, E., B. Los, R. Stehrer, M. Timmer, and G.J. de Vries (2013) "The Construction of World Input-Output Tables in the WIOD Project," *Economic Systems Research*, Vol. 25, No. 1, 71-98.

Ditzen, J. (2018) "xtdcce2: Estimating Dynamic Common Correlated Effects in Stata," *The Stata Journal*, Vol. 18, No. 3, pp. 585 - 617.

Draca, M., R. Sadun, and J. Van Reenen (2006) "Productivity and ICT: A Review of the Evidence," CEP Discussion Papers 749, Centre for Economic Performance, London School of Economics.

Economist (2017a) "The Construction Industry's Productivity Problem," August 17. https://www.economist.com/leaders/2017/08/17/the-construction-industrys-productivity-problem.

Economist (2017b) "Building Under Water: Marine Contractors Have Made Huge Leaps in Productivity," August 17. https://www.economist.com/business/2017/08/17/marine-contractors-have-made-huge-leaps-in-productivity.

Gann, D. (2000) *Building Innovation: Complex Constructs in a Changing World*, Thomas Telford Publishing.

Grossman, G.M. and E. Helpman (1994) "Endogenous Innovation in the Theory of Growth," *Journal of Economic Perspectives*, Vol. 8, No. 1, pp. 23–44.

Harrison, P. (2007) "Can Measurement Error Explain the Weakness of Productivity Growth in the Canadian Construction Industry?" CSLS Research Report Number 2007-01, Centre for the Study of Living Standards, April. http://www.csls.ca/reports/csls2007-01.PDF.

Houseman, S.N. and M. Mandel (eds.) (2015) *Measuring Globalization: Better Trade Statistics for Better Policy*, Volumes 1 and 2, W.E. Upjohn Institute.

Inklaar, R., M.P. Timmer, and B. van Ark (2008) "Market Services Productivity Across Europe and the U.S. Economic Policy," *Economic Policy*, Vol. 23, No. 1, 139-194.

Jäger, K. (2017) "EU KLEMS Growth and Productivity Accounts: Statistical Module," The Conference Board, European Commission, DG for Economic and Financial Affairs.

Jorgenson, D.W., F.M. Gollop, and B. Fraumeni (1987) *Productivity and U.S. Economic Growth*, Elsevier Science Publishers.

Ketokivi, M., V. Turkulainen, T. Seppälä, P. Rouvinen, and J. Ali-Yrkkö (2017) "Why Locate Manufacturing in a High-cost Country? A Case Study of 35 Production Location Decisions," *Journal of Operations Management*, Vol. 49–51, pp. 20-30.

Lau, E. and S. Rowlinson (2011) "The Implications of Trust in Relationships in Managing Construction Projects," *International Journal of Managing Projects in Business*, Vol. 4, No. 4, pp. 633–659.

Los, B., M.P. Timmer, and G.J. de Vries (2016) "Tracing Value-added and Double Counting in Gross Exports: Comment," *American Economic Review*, Vol. 106, No. 7, pp. 1958–1966.

Madsen, J.B. (2008) "Economic Growth, TFP Convergence and the World Export of Ideas: A Century of Evidence," *Scandinavian Journal of Economics*, Vol. 110, No. 1, pp. 145–167.

Naoum, S.G. (2016) "Factors Influencing Labor Productivity on Construction Sites: A State-of-the-art Literature Review and a Survey," *International Journal of Productivity and Performance Management*, Vol. 65, No. 3, pp. 401–421.

OECD (2001) *Measuring Productivity — OECD Manual: Measurement of Aggregate and Industry-Level Productivity Growth*.

O'Mahony, M. and B. van Ark (eds.) (2003)*EU Productivity and Competitiveness: An Industry Perspective — Can Europe Resume the Catching-up Process?*, European Commission.

O'Mahony, M. and M. Vecchi (2005) "Quantifying the Impact of ICT Capital on Output Growth: A Heterogeneous Dynamic Panel Approach," *Economica*, Vol. 72, No. 288, pp. 615–633.

Oulton, N. and S. Srinivasan (2005) "Productivity Growth and the Role of ICT in the United Kingdom: An Industry View, 1970-2000," CEP Discussion Paper No. 681, March.

Pesaran, M.H., Y. Shin, and R.P. Smith (1999) "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of the American Statistical Association*, Vol. 94, No. 44, pp. 621–634.

Romer, P.M. (1990) "Endogenous Technological Change," *Journal of Political Economy*, Vol. 98, No. 5, pp. S71–S102.

Ruddock, L. and S. Ruddock (2011) "Evaluation of Trends in the UK Construction Industry Using Growth and Productivity Accounts," *Construction Management and Economics*, Vol. 29, No. 12, pp. 1229–1239.

Sezer, A.A. and J. Bröchner (2014) "The Construction Productivity Debate and the Measurement of Service Qualities," *Construction Management and Economics*, Vol. 32, No. 6, pp. 565–574.

Squicciarini, M. and A.-L. Asikainen (2011) "A Value Chain Statistical Definition of Construction and the Performance of the Sector," *Construction Management and Economics*, Vol. 29, No. 7, pp. 671–693.

Stiroh, K.J. (2002a) "Are ICT Spillovers Driving the New Economy?," *The Review of Income and Wealth*, Vol. 48, No. 1, pp. 33–57.

Stiroh, K.J. (2002b) "Reassessing the Impact of Information Technology in the Production Function: A Meta-analysis," Federal Reserve Bank of New York. https://dx.doi.org/10.2139/ssrn.352820.

Sullivan, A. and F.C. Harris (1986) "Delays on Large Construction Projects," *International Journal of Operations & Production Management*, Vol. 6, No. 1, pp. 25–33.

Sveikauskas, L., S. Rowe, J. Mildenberger, J. Price, and A. Young (2016) "Productivity Growth in Construction," *Journal of Construction Engineering and Management*, Vol. 142, No. 10.

Timmer, M.P., M. O'Mahony, and B. van Ark (2007) EU KLEMS Growth and Productivity Accounts: Overview. http://www.euklems.net/.

Timmer, M.P. (2017) "Productivity Measurement in Global Value Chains," *International Productivity Monitor*, Vol. 33, pp. 182–193. http://www.csls.ca/ipm/33/Timmer.pdf.

Timmer, M.P., B. Los, R. Stehrer, and G.J. de Vries (2016) "An Anatomy of the Global Trade Slowdown Based on the WIOD 2016 Release," GGDC Research Memorandum No. GD-162, Groningen Growth and Development Centre, University of Groningen.

Timmer, M.P., E. Dietzenbacher, B. Los, R. Stehrer, and G.J. de Vries (2015) "An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production," *Review of International Economics*, Vol. 23, No. 3, pp. 575–605.

Tran, V.D. and J.E. Tookey (2011) "Labour Productivity in the New Zealand Construction Industry: A Thorough Investigation," *Construction Economics and Building*, Vol. 11, No. 1, pp. 41–60.

Venturini, F. (2009) "The Long-run Impact of ICT," *Empirical Economics*, Vol. 37, No. 3, pp. 497–515.

Vogl, B. and M. Abdel-Wahab (2015) "Measuring the Construction Industry's Productivity Performance: Critique of International Productivity Comparisons at Industry Level," *Journal of Construction Engineering and Management*, Vol. 141, No. 4.

Winch, G. (1998) "Zephyrs of Creative Destruction: Understanding the Management of Innovation in Construction," *Building Research & Information*, Vol. 26, No. 5, pp. 268–279.

Wolff, E.N. (1994) "Productivity Measurement Within an Input–output Framework," *Regional Science and Urban Economics*, Vol. 24, No. 1, pp. 75–92.

Zhai, D., P.M. Goodrum, C.T. Haas, and C.H. Caldas (2009) "Relationship Between Automation and Integration of Construction Information Systems and Labor Productivity," *Journal of Construction Engineering & Management*, Vol. 135, No. 8, pp. 746–753.

Appendix

Measurement of the value-added contributions in the construction value chains

We will next formally represent how we used the exclusion method. First, we partitioned the global input–output table such that we had the construction industry in country s, F^s, and the rest of the world economy r containing all other industries in country s and all industries in other countries c in the world. We can construct Matrix A as follows:

$$A = \begin{bmatrix} A_{F^sF^s} & A_{F^sr} \\ A_{rF^s} & A_{rr} \end{bmatrix}$$

A, that contains the input coefficients a_{ij}, which give the global value units of intermediate goods from industry i that are required to produce one value unit of gross output in industry j. In A, the numbers of the rows and columns are the same and equal the numbers of total national industries (the number of countries, C, times the number of industries, I). For the final demand block, we similarly define a matrix of final demand flows Y, the row elements being different final demand classes (in total: 5 different classes) and columns indicating flows from i to j, with the length $C * I$.

With our decomposition, $A_{F^sF^s}$ represents the purchase requirements of the construction industry from itself in country s, while A_{rF^s} gives the requirements by all other industries for construction products bought from the construction industry of country s. For the final demand block, we can similarly write:

$$Y = \begin{bmatrix} y_{F^sF^s} & y_{F^sr} \\ y_{rF^s} & y_{rr} \end{bmatrix},$$

in which the vectors $y_{F^sF^s}$ and y_{F^sr} represent the values of flows from the construction industry in country s to all final users of its products and to final users in other industries.

We then construct the value chain matrix, VA, that contains industry- and country-specific value-added contributions. The ratios of value added to gross output in industries in countries are contained in a row vector v. The length of this vector equals the numbers of industries, with VA ratios for industries as first elements ($\tilde{v}$) and zeroes elsewhere: $\tilde{v} = [\tilde{v}\ 0]$. Then, we follow Los *et al.* (2018) and collect the actual VA distribution in the global value-chain matrix (VA), that is:

$$VA = v(I - A)^{-1}Y * i$$

in which i is a column vector where all elements are unity, implying that it sums the elements in each of the rows of the matrix Y. The VA matrix has the same dimensions as A, including the contributions of each industry to the overall VA of other industries. The element $(I - A)^{-1}$ is the well-known Leontief inverse, in which I is the identity matrix of appropriate dimensions. When multiplied with final demand, the Leontief inverse calculates the gross output in the industries producing the final products and also the output in industries producing the intermediate inputs required for this (Los, Timmer, and de Vries, 2016). In particular, VA can be interpreted as the limiting value of the infinitely long sum of VA contributions, with the number of stages varying from 1 to ∞.

What amount of value added in industry–country pair j should be attributed to the construction value chain? To measure this, we created a hypothetical world in which the construction industry in country s seizes the opportunity to generate final goods, as well as intermediate products, to other industry–country pairs. Formally, by using our decomposition, we set the intermediate flows $A_{F^sr} = 0$, yielding:

$$A^*(F^s) = \begin{bmatrix} A_{F^sF^s} & 0 \\ A_{rF^s} & A_{rr} \end{bmatrix},$$

and similarly, all the final goods $y_{F^sF^s} = 0$ and $y_{F^sr} = 0$:

$$Y = \begin{bmatrix} 0 & 0 \\ y_{rF^s} & y_{rr} \end{bmatrix}.$$

The hypothetical value added in industry j can be obtained by post-multiplying the hypothetical Leontief inverse with the hypothetical final demand as:

$$VA^*_j(F^s) = v_j(I - A^*(F^s))^{-1}Y^*(F^s) * i.$$

Following the logic of hypothetical extraction, the value added in construction activities for industry–country j can be derived as the difference in VA in the actual and hypothetical situations:

$$\Delta VA_j(F^s) = VA_j - VA^*_j(F^s),$$

and $\Delta VA_j(F^s)$ correctly measures the indirect and direct effects on the value chains that follow from the exclusion of the construction industry in s, F^s.

Importantly, we can study the value-added contribution from any individual sector in the construction value chain by changing vector v_j. In particular, we can focus on the construction industry's contribution to the value chain by instead setting $v_{F_s} = \bar{v}_{F_s}$, while other elements are set to 0. On the other hand, by setting $v = \bar{v}$ for industries other than construction while setting the construction industry elements to 0, we can focus on the rest of the value chain.

Details of the productivity decomposition

Under our assumptions, we can define productivity growth (total factor productivity growth in the production of construction output) in the global value chains of construction by the weighted rate of decline of its total labour and capital requirements:

$$\frac{\delta\pi}{\delta t}(F^s) = -\overline{\alpha^L}(F^s)\frac{\delta\Lambda}{\delta t} - \overline{\alpha^K}(F^s)\frac{\delta K}{\delta t},$$

where $\frac{\delta\Lambda}{\delta t}$ and $\frac{\delta K}{\delta t}$ are vectors of the changes in the labour and capital requirements, respectively, and α^L and α^K are the weights given by a (row) vector of value shares with elements reflecting the costs of labour and capital from all country sectors used in the production of one unit of construction product, respectively. In discrete time, the resource use vectors are $\frac{\delta\Lambda}{\delta t} = \Delta\log(\Lambda_t)$ and $\frac{\delta K}{\delta t} = \Delta\log(K_t)$, where L_t,K_t are the labour and capital inputs.

To measure the value share vectors, we note first that for a single element of the factor share vectors, it holds

$$\alpha_j^L(F^s) = \Delta VA_j(F^s) * \alpha_j^{VA,L}$$

$$\alpha_j^K(F^s) = \Delta VA_j(F^s) * \alpha_j^{VA,K},$$

where $\Delta VA_j(F^s)$ is the value-added contribution of industry–country j to construction value chain s that is obtained after setting $v_j = \bar{v}_j$ and zero otherwise, while the counterfactual without the construction sector is defined by setting $A = A^*(F^s)$ and $Y = Y^*(F^s)$, as defined in the previous subsection. $\alpha_j^{VA,L}$ and $\alpha_j^{VA,K}$ are the KLEMS-based or other productivity data-based measures of the labour and capital shares in industry–country j, respectively. As time is discrete, the value-added content is estimated by using the standard Törnqvist shares of the corresponding yearly factor shares $\overline{\alpha^L} = (\alpha_{-1}^L + \alpha^L)/2$ and $\overline{\alpha^K} = (\alpha_{-1}^K + \alpha^K)/2$. Here, we refer to the year t (α) and year $t-1$ shares (α_{-1}).

Appendix 1: Growth Decomposition of the Value Chain with the EU KLEMS Original Labour Composition

	Real construction gross output growth and its components in the value chain							
	Capital share (%) (a)	Growth components excluding hours: b = c+d+g	TFP growth contribution (c)	Capital growth contribution (d)	ICT growth contribution (e)	NIT growth contribution (f)	Labour composition contribution (g)	Hours contribution (h)
AUT	38	0.6	0.1	0.5	0.2	0.2	0	0.2
BEL	41	1.8	0.1	1.3	0.3	0.8	0.4	0.4
CZE	40	2.1	0.5	1.1	0	1.2	0.5	-0.7
DEU	25	1	0.3	0.1	0	0.1	0.6	-1.6
DNK	27	0.3	-0.4	0.3	0	0.2	0.3	-0.5
ESP	39	2.1	1.3	0.6	0	0.7	0.2	-2.1
FIN	26	0.5	0.3	0.2	0.1	0.2	-0.1	-0.1
FRA	27	0.3	-0.7	0.5	0.1	0.3	0.5	0.3
GBR	20	0	-1.2	0.4	0.1	0.2	0.8	0
ITA	34	-1	-1.1	-0.1	-0.1	0.1	0.2	-1
NLD	20	0.9	-0.1	0	0	0.1	1	-1.6
SWE	38	0.9	-0.7	1.5	0.2	1.2	0.1	0.8
Average	**31**	**0.8**	**-0.1**	**0.5**	**0.1**	**0.4**	**0.4**	**-0.5**

Appendix 2: Patent Classifications

B28B	Shaping clay or other ceramic compositions; shaping slag; shaping mixtures containing cementitious material, e.g. Plaster (foundry moulding b22c;working stone or stone-like material b28d;shaping of substances in a plastic state, in general b29c;making layered products not composed wholly of these substances b32b;shaping in situ, see the relevant classes of section e)
B28C	Preparing clay; producing mixtures containing clay or cementitious material, e.g. Plaster (preparing material for foundry moulds b22c0005000000)
B28D	Working stone or stone-like materials (machinery for, or methods of, mining or quarrying e21c)
B66B	Elevators; escalators or moving walkways (life-saving devices used as an alternative to normal egress means, e.g. Stairs, during rescue to lower persons in cages, bags, or similar supports from buildings or other structuresâ a62b0001020000;â equipment for handling freight or for facilitating passenger embarkation or the like to aircraft b64d0009000000;braking or detent devices characterised by their application to lifting or hoisting gear b66d0005000000)
C04B	Lime; magnesia; slag; cements; compositions thereof, e.g. Mortars, concrete or like building materials; artificial stone; ceramics (devitrified glass-ceramics c03c0010000000); refractories (alloys based on refractory metals c22c); treatment of natural stone
E01B	Permanent way; permanent-way tools; machines for making railways of all kinds (derailing or rerailing blocks on track, track brakes or retarders b61k;removal of foreign matter from the permanent way, vegetation control, applying liquids e01h)
E01C	Construction of, or surfaces for, roads, sports grounds, or the like; machines or auxiliary tools for construction or repair (forming road or like surfaces by compacting or grading snow or ice e01h)
E01D	Bridges (bridges extending between terminal buildings and aircraft for embarking or disembarking passengers b64f0001305000)
E01F	Additional work, such as equipping roads or the construction of platforms, helicopter landing stages, signs, snow fences, or the like
E01H	Street cleaning; cleaning of permanent ways; cleaning beaches; cleaning land; dispersing fog in general (mowers convertible to apparatus for sweeping or cleaning lawns or other surfaces, e.g. To remove snow, or capable of sweeping or cleaning lawns or other surfaces a01d0042060000;cleaning in general b08b)
E02B	Hydraulic engineering (ship-lifting e02c;dredging e02f)
E02C	Ship-lifting devices or mechanisms
E02D	Foundations; excavations; embankments (specially adapted for hydraulic engineering e02b); underground or underwater structures
E02F	Dredging; soil-shifting (winning peat e21c0049000000)
E03B	Installations or methods for obtaining, collecting, or distributing water (drilling wells, obtaining fluids in general from wells e21b;pipe-line systems in general f17d)
E03C	Domestic plumbing installations for fresh water or waste water (not connected to either water-supply main or to waste pipe a47k;devices of the kind used in the ground e03b, e03f); sinks
E03D	Water-closets or urinals with flushing devices; flushing valves therefor
E03F	Sewers; cesspools
E04B	General building constructions; walls, e.g. Partitions; roofs; floors; ceilings; insulation or other protection of buildings (border constructions of openings in walls, floors, or ceilings e06b0001000000)
E04C	Structural elements; building materials (for bridges e01d;specially designed for insulation or other protection e04b;elements used as building aids e04g;for mining e21;for tunnels e21d;structural elements with broader range of application than for building engineering f16, particularly f16s)
E04D	Roof coverings; sky-lights; gutters; roof-working tools (coverings of outer walls by plaster or other porous material e04f0013000000)
E04F	Finishing work on buildings, e.g. Stairs, floors (windows, doors e06b)
E04G	Scaffolding; forms; shuttering; building implements or aids, or their use; handling building materials on the site; repairing, breaking-up or other work on existing buildings
E04H	Buildings or like structures for particular purposes; swimming or splash baths or pools; masts; fencing; tents or canopies, in general (foundations e02d)
E05B	Locks; accessories therefor; handcuffs
E05C	Bolts or fastening devices for wings, specially for doors or windows (latching means for sideboard or tailgate structures for vehicles b62d0033037000;fastening devices for constructional or engineering elements e04, f16b;locks, fastening devices structurally or operatively combined or having significant cooperation with locks e05b;means for operating or controlling wing fasteners in conjunction with mechanisms for moving the wing e05f)
E05D	Hinges or suspension devices for doors, windows or wings (pivotal connections in general f16c0011000000)
E05F	Devices for moving wings into open or closed position; checks for wings; wing fittings not otherwise provided for, concerned with the functioning of the wing
E05G	Safes or strong-rooms for valuables; bank protection devices; safety transaction partitions (alarm arrangements per seg08b)
E06B	Fixed or movable closures for openings in buildings, vehicles, fences, or like enclosures, in general, e.g. Doors, windows, blinds, gates (shades or blinds for greenhouses a01g0009220000;curtains a47h;lids for car boots or bonnets b62d0025100000;sky-lights e04b0007180000;sunshades, awnings e04f0010000000)
E06C	Ladders (e04f0011000000 takes precedence;step-stools a47c0012000000;adaptation of ladders to use on ships b63b, to use on aircraft b64;scaffolding e04g)
E99Z	Subject matter not otherwise provided for in this section
F21H	Incandescent mantles; other incandescent bodies heated by combustion
F21K	Non-electric light sources using luminescence; light sources using electrochemiluminescence; light sources using charges of combustible material; light sources using semiconductor devices as light-generating elements; light sources not otherwise provided for
F21L	Lighting devices or systems thereof, being portable or specially adapted for transportation

Appendix 2: Patent Classifications (cont'd)

F21S	Non-portable lighting devices; systems thereof; vehicle lighting devices specially adapted for vehicle exteriors
F21V	Functional features or details of lighting devices or systems thereof; structural combinations of lighting devices with other articles, not otherwise provided for
F21W	Indexing scheme associated with subclasses f21k, f21l, f21s and f21v, relating to uses or applications of lighting devices or systems
F21Y	Indexing scheme associated with subclasses f21k, f21l, f21s and f21v, relating to the form or the kind of the light sources or of the colour of the light emitted
F24B	Domestic stoves or ranges for solid fuels (for solid fuels in combination with gaseous fuels, liquid fuels or other kinds of energy supply f24c0001020000); implements for use in connection with stoves or ranges
F24C	Domestic stoves or ranges (exclusively for solid fuels f24b); details of domestic stoves or ranges, of general application
F24D	Domestic- or space-heating systems, e.g. Central heating systems; domestic hot-water supply systems; elements or components therefor (using steam or condensate extracted or exhausted from steam engine plants for heating purposes f01k0017020000)
F24F	Air-conditioning; air-humidification; ventilation; use of air currents for screening (removing dirt or fumes from areas where they are produced b08b0015000000; vertical ducts for carrying away waste gases from buildings e04f0017020000; tops for chimneys or ventilating shafts, terminals for flues f23l0017020000)
F24H	Fluid heaters, e.g. Water or air heaters, having heat-generating means, e.g. Heat pumps, in general (steam generation f22)
F25B	Refrigeration machines, plants or systems; combined heating and refrigeration systems; heat pump systems
F25D	Refrigerators; cold rooms; ice-boxes; cooling or freezing apparatus not otherwise provided for (refrigerated showcases a47f0003040000;thermally-insulated vessels for domestic use a47j0041000000;refrigerated vehicles, see the appropriate subclasses of classes b60-b64; containers with thermal insulation in general b65d0081380000;heat-transfer, heat-exchange or heat-storage materials, e.g. Refrigerants, or materials for the production of heat or cold by chemical reactions other than by combustion c09k0005000000;thermally-insulated vessels for liquefied or solidified gases f17c;air-conditioning or air-humidification f24f;refrigeration machines, plants, or systems f25b;cooling of instruments or comparable apparatus without refrigeration g12b;cooling of engines or pumps, see the relevant classes)
F28B	Steam or vapour condensers (condensation of vapours b01d0005000000;condensation during pretreatment of gases prior to electrostatic precipitation of dispersed particles b03c0003014000;steam engine plants having condensers f01k;liquefaction of gases f25j;details of heat-exchange or heat-transfer arrangements of general application f28f)
F28C	Heat-exchange apparatus, not provided for in another subclass, in which the heat-exchange media come into direct contact without chemical interaction (heat-transfer, heat-exchange or heat-storage materials c09k0005000000;â fluid heaters having heat generating means f24h;with an intermediate heat-transfer medium coming into direct contact with heat-exchange media f28d0015000000-f28d0019000000;details of heat-exchange apparatus of general application f28f)
F28D	Heat-exchange apparatus, not provided for in another subclass, in which the heat-exchange media do not come into direct contact (heat-transfer, heat-exchange or heat-storage materials c09k0005000000;â fluid heaters having heat generating means and heat transferring means f24h;furnaces f27;details of heat-exchange apparatus of general application f28f); heat storage plants or apparatus in general
F28F	Details of heat-exchange or heat-transfer apparatus, of general application (heat-transfer, heat-exchange or heat-storage materials c09k0005000000;water or air traps, air venting f16)
F28G	Cleaning of internal or external surfaces of heat-exchange or heat-transfer conduits, e.g. Water tubes of boilers (cleaning pipes or tubes in general b08b0009020000;devices or arrangements for removing water, minerals, or sludge from boilers while the boiler is in operation, or which remain in position while the boiler is in operation, or are specifically adapted to boilers without any other utility f22b0037480000;removal or treatment of combustion products or combustion residues f23j;removing ice from heat-exchange apparatus f28f0017000000)
H05B	Electric heating; electric light sources not otherwise provided for; circuit arrangements for electric light sources, in general

The China Effect on Manufacturing Productivity in the United States and Other High-income Countries

Daniel Lind[1]
Arenagruppen

Abstract

From a macroeconomic perspective and using input-output techniques, this article investigates to what extent, and how, the growing use of Chinese intermediates has contributed to the labour productivity growth within the manufacturing production processes of 22 high-income countries. The main result — based on almost 400 global value chains during the period 2000-2014 — is that this productivity effect is significant and economically relevant. This is also the case for the United States. The effect holds before and after the financial crisis, is robust to different specifications, and is identified in almost all production processes. Three mechanisms behind the identified pattern are — tentatively — identified: reduced employment, reduced prices, and productivity enhancing functional specialization. However, China is not special: the absolute productivity effect of the growing use of Eastern European intermediates seems to be even larger. Finally, China and Eastern Europe are special in relation to the high-income countries: growing intra-trade of intermediates among the high-income countries is associated with weaker productivity growth.

The welfare effects of trade constitute a longstanding issue in economics. With China emerging as the factory of the world, this question has recently attracted renewed attention. A main aspect of this is how China's dominating role within global value chains (GVCs) affects high-income countries (HICs).[2] This article contributes to this literature by addressing the following main question: To what extent, and

1 Daniel Lind is a researcher at Arenagruppen, a think tank associated with the Swedish trade union movement. Lind is responsible for a productivity project financed by the trade unions within the manufacturing sector. He would like to thank seminar participants at the Department of Economics at Örebro University, Business Sweden, and the National Board of Economic Research for fruitful discussions and constructive comments. He is also grateful for statistical advice from Erik Hegelund and Alberto Naranjo. Finally, he is very grateful for helpful and constructive comments provided by two anonymous referees and the editors of the journal. Contact details: daniel.lind@arenagruppen.se. Twitter: @DanielLind_

2 See Autor *et al.* (2013, 2015, 2019, 2020a, 2020b), Dauth, Findeisen, and Suedekum (2014), Acemoglu, Akcigit, and Kerr (2015), Bloom, Draca, and Van Reenen (2016), Feenstra, Ma, and Xu (2019), Pierce and Schott (2020), Jaravel and Sager (2020), Amiti *et al.* (2020), Che *et al.* (2020) and Bloom *et al.* (2021).

how, does the growing use of intermediate inputs (hereafter referred to as intermediates) imported from China affect the productivity growth in the manufacturing sector in the United States and other HICs? The approach is macro-oriented, in the sense that the units of study are countries and sectors.

Recent research on the effects of China on the labour markets in the HICs has shown that a credible answer to the main question requires a vertically integrated perspective.[3] This means that the manufacturing sector should not be treated as an isolated unit, but as a chain of activities that connect sectors and countries through the trade in intermediates. Therefore, the magnitude and character of the labour market effects of the Chinese trade shock are determined by the extent to which this network diffuses the initial effect to all sectors in the HICs.

This type of vertically integrated analysis requires the use of input-output (IO) techniques. Consequently, this article is related to the renewed macroeconomic interest in IO linkages. As argued by Acemoglu and Azar (2020), the association between IO linkages and productivity is an under-researched topic that deserves more attention. Addressing the main question from a vertically integrated perspective, this article represents an attempt to take some small steps in this direction.

The productivity measure used in this article is called vertical labour productivity and is defined as the ratio between the value added and employment generated within the domestic economy in order to produce a manufactured product. Accordingly, this measure includes all upstream/backward activities along the domestic supply chains needed to finalize the product.[4]

The article proceeds as follows. The first section positions the article within the related literatures and presents further questions to be answered, while answering the main question. Next follows a section presenting the main variables, the empirical approach, and some descriptive statistics. After this follows a section that discusses and tests the identification strategy. Next follows a section containing the empirical results, wherein three possible explanations for these results are discussed and empirically tested. The last section concludes and briefly discusses some avenues for future research.

Related Literatures

Global value chains and productivity diffusion

The starting point of this article is the

---

3 Autor, Dorn, and Hansen (2016), Pierce and Schott (2016), Acemoglu *et al.* (2016), Feenstra and Sasahara (2018), and Bloom *et al.* (2019). Autor and Salomons (2018) and Reijnders, Timmer, and Ye (2021) use the same argument for the employment effects caused by technical change.

4 With the argument that the emergence of the GVCs requires a new approach, similar productivity measures have recently been used in Timmer (2017), Gu and Yan (2017), Timmer and Ye (2018, 2020), Pahl and Timmer (2019) and Buckley *et al.* (2020). However, this approach to productivity analysis is not new. Based on the domestic economy, it is found in early IO research on the US economy (Leontief 1953; Carter 1970). It is also a common theme in the evolutionary tradition (Winter and Nelson, 1982; Rosenberg (1982) and in the post-Keynesian tradition (Pasinetti, 1981, 1993). Other examples of this type of productivity research are found in Wolff (1994), Dietzenbacher *et al.* (2000), De Juan and Febrero (2000), and Ten Raa and Wolff (2000, 2001, 2012).

emergence and increased complexity of the GVCs and how China has become a central node within the global production network, dominated by the trade in intermediates.[5] Using a production process perspective, with each production stage adding value to the final product, recent macroeconomic GVC research has focused on how shocks, such as the China shock, are spread around the world through the trade in intermediates.[6] Within this framework, the manufacturing firm decides how much and from where they buy their intermediates. Optimally, the firm should base its decision on the vertically integrated labour productivity adjusted cost-minimization (Antras and de Gortari, 2020). Empirically, the microeconomic GVC research shows, among many other things, that importing firms often have access to more input varieties and use higher quality intermediates.

Along similar lines, the trade in intermediates is at the centre stage in the recent research that endogenizes the IO structure and how it changes over time (Acemoglu and Azar, 2020).[7] When firms cost-minimize their use of intermediates, new input combinations will emerge, due to technical change. If this new combination leads to price reductions, a small change in one sector can cause a major change in the organization of production and affect productivity in many sectors. This diffusion aspect of the trade in intermediates is also considered in the recent research on shock propagation and how it may affect the macroeconomic volatility (Acemoglu *et al.*, 2012, 2015; Acemoglu, Ozdaglar and Tahbaz-Salehi, 2016; Carvalho and Tahbaz-Salehi, 2019). This occurs when some sectors are particularly important as suppliers and when the use of the intermediate is widespread. The empirical analysis in Acemoglu, Akcigit, and Kerr (2015), focusing on US IO tables for 1992, shows that productivity shocks propagate downstream, and their conclusion is that this amplification mechanism is more important than what is typically presumed in the macroeconomic literature. Moreover, the indirect productivity effects, occurring along the supply chains, are quantitatively more important than the productivity effect in the sector that was first hit by the shock.

The China shock

Naturally, another related literature focuses on the question: how do imports from China affect HICs? As mentioned, one insight from this research is that the labour market effects caused by the China shock — driven by domestic reforms, trade liberalizations and new ICT uses — cannot

5 For general overviews of the emergence and consequences of the GVCs, see Baldwin (2016, 2017, 2019), Ponte, Gereffi, and Raj-Reichart (2019), IMF (2019), WTO (2019) and World Bank (2020).

6 Antras and Chor (2021) is a recent survey of this GVC research. On the theoretical side, recent contributions are Caliendo and Parro (2015), Caliendo, Parro, and Tsyvinski (2017), Antras and Chor (2019) and Antras and de Gortari (2020). These frameworks have also been used in counterfactual exercises to quantity the effects of US-China trade tensions (e.g. Caceres, Cerdiero, and Mano, 2019; Ju *et al.*, 2019), productivity shocks in the US economy when IO linkages are present (Caliendo *et al.*, 2019), effects on the US economy from the China trade shock (Caliendo *et al.*, 2019; Rodriguez-Clare, Ulate, and Vasquez, 2020) and the effect of global specialization on the sensitivity for productivity shocks in other countries (Caselli *et al.*, 2020).

7 See also Carvalho and Voigtländer (2015) and Oberfield (2018).

be properly identified by only focusing on the manufacturing sector itself. This approach was used in the seminal analysis of Autor, Dorn, and Hansen (2013), but the whole manufacturing process, including the supply chains, was included in Acemoglu *et al.* (2016). This production process perspective can create both positive and negative employment effects, indicating that the net effect of the Chinese trade shock is ambiguous in sign. The empirical analysis in Acemoglu *et al.* shows that the negative employment effect of this fundamental change in the global economy is more than doubled, as compared to the effect within the manufacturing sector itself. The authors conclude: "Thus, interindustry linkages magnify the employment effects from trade shocks...." (Acemoglu *et al.*, 2016:145). Although the level and sign of this employment effect is still discussed, the standard approach in the macroeconomic literature on the China shock has become to apply a vertically integrated perspective, in the sense that the IO structure is included in the empirical analysis on the effects of China on the labour markets in the HICs.[8]

To the best of my knowledge, no attempt has been made to study the macroeconomic effects of China's intermediate exports on the productivity within the manufacturing production processes among the HICs, i.e. when the productivity among the suppliers are included in the analysis. There are, however, some related research. From a microeconomic perspective, Bloom, Draca, and Van Reenen (2016) investigate the productivity aspect of the China shock. Their main conclusion is that the effect is positive on firm TFP growth in four European countries between 1996-2007. Using instrumental techniques, they find that 30-60 per cent of the TFP growth between 2000-2007 was accounted for by the imports from China. Bloom *et al.* (2021) continue along a similar path, and show that firms in 11 European countries that are more exposed to trade from China increased their productivity-enhancing efforts more than other firms between 1995-2005, while also experiencing a decline in sales. From a propagation perspective, Acemoglu, Akcigit, and Kerr (2015) use the IO structure for the year 1992 to investigate how different types of shocks are spread to almost 400 sectors in the US economy and how they affect value added, employment, and labour productivity. In terms of a trade shock from China, labour productivity is unaffected, since the effects on value added and employment are both negative and of a similar magnitude.

Productivity in an IO setting

From a general IO perspective, Acemoglu and Azar (2020) investigate how changes in individual cells of the Leontief inverse affect TFP. They find that "large" changes — defined as being above the 20th percentile in terms of changes in the number of suppliers — in the composition of intermediates contribute to faster productivity growth in the United States. Over the

8 Autor, Dorn, and Hansen (2016:220) express this in the following way: "A full account of the impact of trade shocks thus requires incorporating input-output linkages between domestic industries." See also Pierce and Schott (2016), Bloom, Draca, and Van Reenen (2016), and Feenstra and Sasahara (2018).

period 1987-2007, between 40-60 per cent of the difference in TFP growth between sectors can be explained by these changes in the intermediate structure.[9] From a GVC perspective, and using vertically integrated productivity measures, one main conclusion in Timmer (2017) and Timmer and Ye (2018, 2020) is that a substantial part of the TFP growth within the manufacturing production process of a group of HICs since the 1990s is generated outside the manufacturing sector itself.

Gu and Yan (2017) follow the same approach. Among a group of HICs during the period 1995-2007, their main result is that there is a substantial difference between the conventional, sectoral-based TFP growth and the TFP measure that includes the supply chains. Moreover, due to imported intermediates produced by industries with high productivity levels, Canada has experienced more rapid productivity growth than EU countries and the United States from participating in the GVCs. Pahl and Timmer (2019) define their vertically integrated labour productivity measure as the ratio between the value added and employment used to produce an exported manufactured product. Based on 58 countries and the period 1970-2008, their main result is that a high level of imported intermediates correlates with a faster vertical labour productivity growth. This result does not, however, seems to hold for the most productive countries.

Questions addressed

Based on the aforementioned literatures, this article investigates the impact of the China shock on the vertical labour productivity within the manufacturing production processes of 22 HICs. The macroeconomic approach has two main advantages. First, it gives overall estimates and establishes the general picture of the productivity effect among close to 400 GVCs of the HICs. These aggregate estimates can, in turn, be broken down into analyses of separate countries and GVCs. Second, it makes it possible to fully exploit the vertical dimension of manufacturing production (Antras and Chor, 2021), i.e. how firms in different sectors and countries interact in order to finalize a product and how this affects productivity outcomes.

While trying to give a credible answer to the main question, and following some of the paths in the China shock literature, this article also addresses the following questions:

1. Is there a China effect on value added or employment – or both?
2. Is there a China effect on prices?
3. Is there a productivity enhancing China effect on the allocation of manufacturing activities performed in different countries?
4. In a comparison with Eastern Europe, are China's intermediates a special case?

9 In a non-competitive (bargaining) framework, see Acemoglu *et al.* (2020) for a further theoretical discussion on how a TFP shock may affect the affected sector's suppliers and customers.

Descriptive Statistics and Empirical Model

Data

The source of data in this article is the World Input-Output Database (WIOD). This type of database has recently emerged through the harmonization of national accounts statistics and trade statistics, and it contains intermediate trade between countries/sectors. Therefore, this dataset has become a necessity in current macro-oriented GVC research (Antras and Chor, 2021).[10] The WIOD covers 43 countries, contains data for the period 2000-2014 and covers 56 sectors, of which 19 are defined as manufacturing sub-sectors.[11] Following the argument and method in Timmer *et al.* (2021), the variables used in this article are expressed in constant prices with base year=2000. Translation into a common currency (USD) is done by market exchange rates.[12] Employment is defined as the number of persons engaged.

Sectors are classified according to the ISIC Rev. 4 and the IO tables follow the 2008 version of the System of National Accounts (SNA).[13] Each manufacturing sub-sector in each country, including its domestic supply chain, will be viewed as a separate GVC (Antras and Chor, 2019; Pahl and Timmer, 2019). With 22 HICs and 19 manufacturing sub-sectors in each country, 418 GVCs, at most, will be included in the empirical analysis. The HICs are EU15 before the 2004 enlargement, Canada, United States, Switzerland, Norway, Australia, Japan, and South Korea.

Variables

The variables used in this article are constructed using IO techniques and the Leontief inverse matrix.[14] By pre- and/or post-multiplication, this matrix is used to create variables that include the upstream/backward activities needed along the supply chains to produce a final product. Therefore, each GVC represents a "composite" sector, as if all production stages were totally vertically integrated.

For each manufacturing production process in each country — i.e. for each GVC — the main variables are the following:[15]

10 For this article, one particularly relevant assumption underlying databases such as the WIOD is that China's intermediates are produced with the same technology regardless to where they are exported.

11 The values of intermediates and final demand in a WIOT can be viewed as endogenous variables, in the sense that they are the result of firm-level decisions on how they optimize the production process (Antras and Chor, 2021).

12 The exact procedure for expressing variables in constant prices in a common currency is as follows. First, the WIOD-researchers use the market exchange rates to convert the national (nominal) values to USD values. Second, they construct time series in t-1 prices (in constant USD). These two datasets (nominal values and t-1 values expressed in USD) are officially released on the WIOD homepage. The third step is to covert these two data sets into time series expressed in constant prices with a base year. This is done in Timmer *et al.* (2021) and this is the procedure followed in this article. To do this, the real growth rate for each particular year (in this case between 2000 and 2001) is equal to ln(value in previous year's prices in 2001/value in current prices in 2000). Starting with the year 2000, these real growth rates are, in the next step, used to calculate the value-added level in constant prices with base year 2000 for each of the years 2001-14.

13 For further details about WIOD, see Dietzenbacher *et al.* (2013) and Timmer *et al.* (2015).

14 See Miller and Blair (2009) for the fundamental ideas behind the IO analysis.

15 See Appendix for further details and the construction of the variables.

Vertical labour productivity: the ratio between the (vertical) value added and (vertical) employment needed to satisfy final demand, including all upstream stages of the domestic production process.

Output multiplier: the gross output needed in the domestic economy in order to produce one unit of final demand, including all upstream stages of the domestic production process.[16]

Import multiplier: the use of imported intermediates per unit of final demand, including all upstream stages of the domestic production process.[17]

Import multiplier from China: the use of imported Chinese intermediates per unit of final demand, including all upstream stages of the domestic production process.

Overall multiplier: the gross output needed to produce one unit of final demand, irrespective of whether the intermediates are domestically or foreign sourced. This variable, constructed by the author, is defined as the sum of the output and import multiplier.

Capital multiplier: the domestic use of the capital stock per unit of final demand, including all upstream stages of the domestic production process.

Vertical gross output: the gross output needed to satisfy final demand, including all upstream stages of the domestic production process.

Descriptive statistics

Table 1 summarizes how the main variables have developed between 2000-2014. Some conclusions emerge. First, with more than a five-fold increase in its absolute level, China's productivity convergence is substantial. Second, the import multiplier has increased over the whole period in both the HICs and the United States, while it has decreased in China after the financial crisis, suggesting a growing self-sufficiency. With only minor changes in the output multiplier, the growing import multiplier in the HICs and the United States implies that the relative use of imported intermediates has grown. Third, the relative use of Chinese intermediates in the HICs and the United States has increased substantially. In absolute terms, the use of Chinese intermediates increased by 420 per cent among the HICs: from 0.0037 in 2000 to 0.0192 in 2014. In the United States, the increase was 490 per cent: from 0.0030 to 0.0178. Following the reduced absolute Chinese import multiplier since the financial crisis, the Chinese use of imported intermediates from the HICs and the United States has been reduced in absolute terms. Finally, China's productivity convergence has occurred alongside a considerable decrease in the capital multiplier, indicating a strong growth in the capital productivity, i.e. less capital is needed to produce one unit of final demand.

Empirical model

Following much of the recent research on the China shock, a linear panel model

16 This variable can be seen as the domestic counterpart to the measure of upstreamness in recent GVC research (Antras and Chor, 2019).

17 This variable is similar to the measures of vertical specialization in the recent GVC research.

Table 1: Descriptive Statistics

Variable	2000	2008	2014
High-income Countries			
Vertical labour productivity level (1000s, USD)	52.8	94.7	104.5
Output multiplier	1.72	1.72	1.68
Import multiplier	0.29	0.32	0.35
Overall multiplier (output + import multiplier)	2.01	2.04	2.03
Import multiplier from China	0.0037	0.0129	0.0192
Import multiplier/Output multiplier	0.17	0.19	0.21
Import multiplier from China/Import multiplier	0.013	0.040	0.055
Capital multiplier	1.30	1.05	1.12
United States			
Vertical labour productivity level (1000s, USD)	70.9	109.3	127.5
Output multiplier	1.80	1.79	1.81
Import multiplier	0.21	0.23	0.24
Overall multiplier (output + import multiplier)	2.02	2.02	2.05
Import multiplier from China	0.0030	0.0109	0.0178
Import multiplier/Output multiplier	0.12	0.13	0.14
Import multiplier from China/Import multiplier	0.014	0.047	0.074
Capital multiplier	1.17	0.96	0.92
China			
Vertical labour productivity level (1000s, USD)	2.7	7.5	15.0
Productivity convergence: HIC (share)	0.05	0.08	0.14
Productivity convergence: US (share)	0.04	0.07	0.12
Output multiplier	2.55	2.62	2.96
Import multiplier	0.15	0.20	0.16
Overall multiplier (output + import multiplier)	2.70	2.82	3.12
Import multiplier from HIC	0.060	0.067	0.056
Import multiplier from US	0.006	0.009	0.006
Import multiplier/Output multiplier	0.059	0.076	0.054
Import multiplier from HIC/Import multiplier	0.40	0.34	0.35
Import multiplier from US/Import multiplier	0.04	0.05	0.04
Capital multiplier	1.54	0.53	0.31

Note: The estimates for the HICs are unweighted averages among the 22 HICs and 19 manufacturing sub-sectors in each country, including all upstream/backward stages of their domestic supply chains (i.e., 418 GVCs). The variables are measured in constant prices with base year=2000. The multiplier variables should be interpreted as: a unit change in final demand generates xx units of the variable in question within the domestic economy, including all upstream/backward stages of the domestic supply chains.

with a fixed effect estimator will be used to empirically analyze the causal effect of the growth in the use of Chinese intermediates on the vertical labour productivity growth within the GVCs of the HICs. Accordingly, the following equation will be the empirical backbone of the article:

$$VLP_{ijt} = \beta_1 X_{1,ijt} + \beta_k X_{k,ijt} + \alpha_i + \varepsilon_{ijt} \quad (1)$$

where VLP_{ijt} is the vertical labour productivity level in GVC i in country j at time t. β_1 is the main coefficient, indicating the average effect of the change in the use of Chinese intermediates on the change in vertical productivity. The β_k vector contains different coefficients depending on which control variables are included in the particular specification, α_i is the GVC-specific intercept and controls for the time-invariant differences between the GVCs that are not included in the regression. ε_{ijt} is the "usual" disturbance, which varies between GVCs and over time.

Identification

General approach

How can it be made likely that any correlation between the growing use of Chinese intermediates and the vertical labour productivity growth can be interpreted in causal terms? In this article, this ever-existing endogeneity problem is addressed using China's vertical labour productivity as the instrumental variable. The support for this choice is found in recent research on the effects of Chinese imports on the HICs. The argument for the IV strategy used in this research is that the effect of China — the China shock — is supply-driven, in the sense that it is mainly caused by political and economic reforms within China.[18] The empirical foundation of this strategy is centred around China's strong productivity growth caused by the reforms (supported by the data in Table 1).[19] This, in turn, has increased HICs' intermediate and final imports from China. More precisely, to solve the endogeneity problem in this setting, the instrumental variable has been the imports from China among HICs not included in the particular study.[20]

Following the same line of reasoning but using China's vertical labour productivity as the instrumental variable has two main advantages. First, China's strong productivity growth represents the core aspect of the supply-argument. Therefore, the chosen strategy opens up for a strong causal interpretation of China's effect on the HICs. Second, the chosen methodology makes it possible to empirically test the exclusion criterion.[21]

Exclusion criterion

Apart from a strong first-stage correlation, the second criterion for being an appropriate IV strategy is that the chosen instrumental variable satisfies the exclusion criterion. This criterion is met if it is reasonable to believe that the use of Chinese intermediates is a main and independent channel through which China's strong vertical labour productivity growth affects the vertical labour productivity growth in the HICs. From a theoretical perspective, this connection is central in a world of GVCs, and particularly so when it comes to China emerging as the factory of the world (Antras and Chor, 2021). To test the exclusion criterion empirically, however, the following question is addressed: is there a positive causal effect from the growth of HICs' import multiplier from China on the vertical labour productivity growth in the HICs, when controlling for China's vertical labour productivity? If this is the case, the import multiplier from China has an independent effect on the productivity growth in the HICs. Table 2 presents the answer to this question.

Using the linear panel data model with the fixed effect estimator presented in equa-

18 After the seminal work by Autor, Dorn, and Hansen (2013), this strategy is often used. Many of the articles already mentioned use some variety of it. See also Antras *et al.* (2017), Acemoglu and Restrepo (2019), Constantinescu, Mattoo, and Ruta (2019), Acemoglu and Azar (2020), and Bloom *et al.* (2021).

19 As an example of this, Antras *et al.* (2017) model the China shock as a productivity increase in the Chinese production of intermediates.

20 For example, when studying the China effect on the US economy, imports from China among a number of — often eight — other HICs are used as the instrument.

21 As will be clear in the main estimations to come, as an instrumental variable China's vertical labour productivity generates highly significant first-stage correlations.

Table 2: Test of the Exclusion Criterion

High-income countries				
Dependent variable: vertical labour productivity	Only Chinese productivity	IV. Including the use of Chinese intermediates	IV. Including the use of Chinese intermediates, lag 1	IV. Including the use of Chinese intermediates, lag 2
China's vertical labour productivity	0.335***	0.02	0.067	0.13
	(0.049)	(0.078)	(0.091)	(0.098)
HICs' import multiplier from China		0.334***	0.394***	0.394***
		(0.065)	(0.072)	(0.076)
Instrument	No	Yes	Yes	Yes
R^2 - within	0.86	0.93	0.90	0.85
N	268	268	250	232

Note: Linear fixed effect IV estimations (2SLS). Other included regressors are: output multiplier, import multiplier, and vertical gross output. Vertical gross output is used to control for the actual level of production, or more precisely: the change in the level of gross output needed to satisfy final demand. All variables except China´s vertical labour productivity is measured by their unweighted average. Years: 2000-2014. Log values. *** = p<0.001. The variables are expressed in constant prices with base year=2000. Since China lacks productivity data for the manufacturing sub-sector Repair and installation of machinery and equipment (C33), robust standard errors in parentheses are adjusted for 18 clusters (manufacturing sub-sectors, including their domestic supply chains). Since China´s vertical labour productivity is a variable on its own in the estimations, and analogously with much of the recent China shock literature, the instrumental variable is the import multiplier from China among the 21 countries in the WIOD not defined as a HIC. In the 2SLS estimation without lag, the F-value of the first-stage regression is 1856 and the elasticity of the instrumental variable is significant at p<0.001 (the F-value should not be lower than 10-15 in order for the IV strategy to be appropriate). The results for the United States are similar to those of the HICs, although China´s vertical labour productivity also seems to have an independent, significant positive elasticity in the 2SLS estimations.

tion (1), and including three regressors — the output and import multiplier controlling for the overall network of intermediate use, and vertical gross output controlling for the actual level of demand — the first column shows that there is a positive and significant correlation between China's productivity growth and HIC productivity growth, indicating a process of diffusion. Then, what happens to this correlation when HICs' import multiplier from China is included in the estimation? Column 2 shows two effects. First, the significant elasticity of China's productivity from the previous estimation disappears. Second, the elasticity of the import multiplier from China turns out to be significant and economically relevant.

This indicates that there is no causal link between China's and HIC productivity growth when the import multiplier from China is unchanged. And the other way around, when China's productivity is held constant, there is still a significant positive effect of HIC use of Chinese intermediates on HIC productivity growth. Using lags, the remaining two columns strengthen this result.[22] Therefore, the growing use of Chinese intermediates seems to be necessary to establish a causal link between China's strong productivity growth and HIC productivity growth. From my viewpoint, these estimations thus give credible empirical support for the argument that the growing use of Chinese intermediates, at least, represents an independent and main channel for the productivity diffusion from China to the HICs. Accordingly, China's vertical labour productivity should qualify as an appropriate instrumental variable within the setting of this article.

22 The fact that the elasticities with lags 1-2 are larger than the elasticity with no lag, indicates the importance of using lags when analysing the productivity effect of Chinese intermediates.

Table 3: Baseline Estimations for Vertical Labour Productivity

Panel A: High-income Countries

Dependent variable: vertical labour productivity	Without Chinese intermediates	With Chinese intermediates	IV: with Chinese intermediates	IV: Chinese intermediates and the overall multiplier	IV: Chinese intermediates and the capital multiplier
Output multiplier	-1.989***	-1.466***	-0.951***		-0.496*
	(0.214)	(0.174)	(0.203)		(0.224)
Import multiplier	0.690***	-0.138	-0.823***		-0.868***
	(0.059)	(0.075)	(0.088)		(0.087)
Import multiplier from China		0.320***	0.587***	0.476***	0.564***
		(0.018)	(0.026)	(0.016)	(0.026)
Overall multiplier				-1.769***	
				(0.231)	
Capital multiplier					-0.350***
					(0.077)
Instrument	No	No	Yes	Yes	Yes
R^2 - within	0.31	0.61	0.40	0.51	0.46
N	5866	5866	5560	5560	5560

Panel B: United States

Dependent variable: vertical labour productivity	Without Chinese intermediate	With Chinese intermediates	IV: with Chinese intermediate	IV: Chinese intermediates and the overall multiplier	IV: Chinese intermediates and the capital multiplier
Output multiplier	-3.459***	–2.004***	-1.089*		-1.041
	(0.353)	(0.341)	(0.553)		(0.614)
Import multiplier	0.949***	0.135	-0.304		-0.303
	(0.121)	(0.165)	(0.209)		(0.204)
Import multiplier from China		0.234***	0.376***	0.334***	0.372***
		(0.031)	(0.061)	(0.025)	(0.061)
Overall multiplier				-1.526***	
				(0.338)	
Capital multiplier					-0.036
					(0.105)
Instrument	No	No	Yes	Yes	Yes
R^2 - within	0.74	0.85	0.83	0.84	0.83
N	283	283	266	266	266

Note: Linear fixed effect IV estimations (2SLS). Robust standard errors in parentheses, in the HIC estimations (US) adjusted for 408 (GVCs, i.e., countries*manufacturing sub-sectors, including their domestic supply chains) (19) clusters (GVCs, i.e., manufacturing sub-sectors, including their domestic supply chains) when no instrumental variable is used, and adjusted for 390 (18) clusters in the IV estimations. The variables are expressed in constant prices with base year=2000. The instrumental variable is China's vertical labour productivity. Vertical gross output is used to control for the actual level of production, or more precisely: the change in the level of gross output needed to satisfy final demand. Years: 2000-2014. Log values. $^{***}p < 0.001,^{**} p < 0.01,^{*} p < 0.05$. In column 3, the F-value of the first-stage regression for the HICs (US) is 321 (175), and the instrumental variable is significant at $p < 0.001$.

Results

Benchmark estimations

To what extent, and how, does the growing use of Chinese intermediates affect the vertical labour productivity growth within the manufacturing production processes of the United States and other HICs? To find a credible answer to this main question, the first steps are taken in Table 3. Using the linear panel data model with a fixed effect estimator, in the first column, the HICs'/United States' productivity is – apart from the control for the business cycle – only regressed against the two variables that describe the overall network of intermediate trade: the output and the import multiplier, respectively. The results for the HICs and the United States are very similar. When keeping the output multiplier (import multiplier) constant, a growing use of imported (domestic) intermediates is positively (negatively) and significantly correlated with a faster productivity growth. This estimation thus suggests that a growing use of imported intermediates per unit of final demand — the core aspect of the new global economy dominated by the GVCs — has been rewarding in terms of productivity growth.[23] On the other hand, an increased domestic specialization in terms of intermediate use per unit of final demand correlates negatively with the productivity growth, when controlling for the import multiplier.[24]

What happens to the import multiplier when the import multiplier from China is included in the estimation? Column 2 shows that the positive elasticity of the import multiplier disappears, both for the HICs and the United States. Instead, the elasticity of the import multiplier from China becomes positively significant. Consequently, when the import multiplier from China is held constant, there is no positive correlation between the growing use of imported intermediates and a faster productivity growth. This is a first indication that the positive productivity contribution from the growing use of foreign intermediates among the HICs seems to be particularly associated with Chinese intermediates.

The next question is: what happens to the elasticities in Column 2 when the IV strategy is applied? First, Column 3 shows that the elasticity of the import multiplier from China is considerably increased in both the HICs and the United States, to 0.587 and 0.376, respectively, indicating that an appropriate IV strategy is important when investigating the size of the causal effect of the China shock on the productivity in the HICs. A 1 per cent increase in the use of Chinese intermediates per unit on final demand thus, on average, leads to a 0.587 (0.376) per cent increase in the growth of the vertical labour productivity in the HICs (United States).[25] Second, the elasticity of the import multiplier turns significantly negative in the HICs, further emphasizing the role of China in the vertical specialisation of the global economy. Third, the explanatory power of this benchmark results in terms of R^2 is particularly large in the United States: more than 80 per cent of the difference in productivity growth between its manufacturing sub-sectors is explained by the estimated model.

However, is the result in Column 3 dependent on how the overall network of intermediates is defined? Substituting the output and import multiplier for the overall multiplier, Column 4 shows that this somewhat reduces the elasticity of the use of Chinese intermediates in both the HICs and the United States, but it is still significant at $p < 0.001$. Another obvious concern is possible omitted variables. Follow-

23 For the HICs, the correlation between the output multiplier and the import multiplier is –0.58. Clearly, to a certain extent the use of domestic intermediates is low when the use of imported intermediates is high – and vice versa. However, between 2000-2014 the mean value of the overall multiplier spans from 1.19 to 3.31, indicating substantial differences between the GVCs in the amounts of intermediates used per unit of final demand.

24 One reason for this somewhat counterintuitive result may be that the yearly changes in the output multiplier are so small that they do not have the strength to counterbalance other forces around. Also, when lags are added this negative elasticity disappears. See Table 4.

25 When applying weighted estimations these results are confirmed. This is also the case when another IV strategy is used, analogously constructed from the recent China shock literature. For further details, see the Appendix.

ing the main question in this article, an important issue is related to the exclusion of the capital stock.[26] Therefore, the estimation in the last column includes the capital multiplier. As can be seen, its elasticity turns out to be significant in the HICs but insignificant in the United States. The negative significance means that in GVCs with a faster reduction in the capital multiplier — i.e. when the use of capital per unit of final demand decreases rapidly — the vertical labour productivity grows even faster. Hence, a more efficient use of the capital stock leads, in line with theory, to faster vertical labour productivity growth. In the United States, the non-significant elasticity of the capital multiplier suggests that its vertical labour productivity growth has been more dependent on TFP growth.[27]

Separate time periods and lag structures

With a vertically integrated perspective and its focus on the process of diffusion within and between countries, it is very reasonable to add lags to the analysis.[28] There are three main reasons for this. First, it reduces the potential problem of reversed causality. Second, it is theoretically reasonable to believe that the productivity effect of the use of Chinese intermediates is not instantaneous: that the (strongest) effect does not necessarily show up in the same year as the transaction is registered. Finally, the results presented in Table 2 indicated the existence of such delayed effects.

Therefore, the question to be answered in this section is: when adding lags to the main estimation in Column 3 in Table 3, how persistent is the productivity effect of the growing use of Chinese intermediates? But first, is there any difference in the China effect before and after the financial crisis? Columns 1-2 in Table 4 give a clear answer: the productivity effect is positively significant in both periods in both the HICs and the United States. For the HICs, the size of the effect is larger before the financial crisis, while the opposite is the case in the United States. A common pattern between the two is, however, that the explanatory power of the estimated model is lower after the financial crisis, indicating that differences in productivity growth rates between GVCs are less dependent on the variables included in the estimation. Moreover, the F-value of the first-stage regression is considerably lower after the financial crisis, supporting the im-

26 Furthermore, the distinction between intermediates and the capital stock can be questioned (Jones, 2013). Both types of "capital" can be viewed as produced factors of production with the same purpose: contributing productively to the finalisation of a product. The only difference is the time dimension, with the more short-lived intermediates defined as current consumption (and not as a capital investment). Corrado *et al.* (2020) argues that the distinction is particularly difficult to uphold when it comes to knowledge-based capital — a type of capital especially important for the HICs.

27 When adding lags to the capital multiplier, the elasticity of the HICs is negatively significant with lags 1-2, while the US elasticity turns weakly negative ($p<0.1$) with lags 1-3. In this sense, the difference between the HICs and the United States is reduced when dynamics, in terms of lags, are included. Moreover, the potential productivity effect of the capital import multiplier from China is not investigated further in this article, neither is the possibility that some parts of the domestic capital stock in the HICs — e.g. its ICT-related parts — have been productivity enhancing.

28 See, for example, Autor, Dorn, and Hansen (2013), Acemoglu, Akcigit, and Kerr (2015), Bloom, Draca, and Van Reenen (2016), Autor, Dorn, and Hansen (2016), Acemoglu and Azar (2020) and Autor *et al.* (2020a).

Table 4: Different Time Periods and Lag Structures for Vertical Labour Productivity

Panel A: High-income Countries

Dependent variable: vertical labour productivity	2000-2008	2009-2014	2000-2014: lag 1	2000-2014: lag 3	2000-2014: lag 4	2000-2014: lag 5
Output multiplier	-1.140***	-1.269***	-0.539**	0.181	0.184	0.388**
	(0.199)	(0.205)	(0.179)	(0.142)	(0.127)	(0.146)
Import multiplier	-0.655***	-0.798***	-0.682***	-0.498***	-0.300***	-0.151**
	(0.093)	(0.105)	(0.085)	(0.069)	(0.054)	(0.057)
Import multiplier from China	0.562***	0.381***	0.542***	0.394***	0.283***	0.213***
	(0.019)	(0.045)	(0.024)	(0.018)	(0.015)	(0.017)
R^2 - within	0.55	0.38	0.38	0.31	0.22	0.15
N	3363	2197	5090	4335	3961	3582

Panel B: United States

Dependent variable: vertical labour productivity	2000-2008	2009-2014	2000-2014: lag 1	2000-2014: lag 3	2000-2014: lag 4	2000-2014: lag 5
Output multiplier	-1.038**	-1.526	-0.447	0.640	0.407	0.617*
	(0.355)	(1.009)	(0.579)	(0.824)	(0.461)	(0.269)
Import multiplier	-0.024	-0.525	-0.584*	-0.674*	-0.461**	-0.290
	(0.136)	(0.505)	(0.251)	(0.298)	(0.165)	(0.154)
Import multiplier from China	0.283***	0.523**	0.436***	0.418***	0.344***	0.296***
	(0.046)	(0.185)	(0.072)	(0.076)	(0.042)	(0.044)
R^2 - within	0.81	0.41	0.77	0.61	0.60	0.61
N	159	107	246	211	193	175

Note: Linear fixed effect IV estimations (2SLS). Robust standard errors in parentheses, adjusted for clusters: 390 clusters for the HICs (GVCs, i.e., countries*manufacturing sub-sectors, including their domestic supply chains) and 18 clusters for the US (GVCs, i.e., manufacturing sub-sectors, including their domestic supply chains) when no lags are used. Vertical gross output is included to control for the actual level of production. The variables are expressed in constant prices with base year=2000. The instrumental variable is China's vertical labour productivity. Log values. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. For 2000-2008, the F-value of the first-stage regression of the HICs (US) is 411 (258), and the elasticity of the instrumental variable is significant at $p < 0.001$. For 2009-2014, the F-value of the first-stage regression of the HICs is 258 (79), and the elasticity of the instrumental variable is significant at $p < 0.001$.

pression that the included variables and the IV strategy performs better for the period considered to be the peak years of the China shock — from China's WTO membership in 2001 until the financial crisis.

The question about the persistence of the Chinese productivity effect is addressed in the last four columns of Table 4. Three conclusions emerge. First and foremost, adding lags indicates a clear persistence: with a five years lag, the effect is still significant at $p < 0.001$ and its size is economically relevant (0.213 and 0.296). A 1 per cent increase in the use of Chinese intermediates five years ago, thus leads to a faster productivity growth with more than 0.2 per cent today.[29] Second, when adding lags, the difference in the size of the effect between the HICs and the United States is turned around, indicating that the lag structure is relatively more important in the latter. Third, when adding lags, the United States' import multiplier also turns significantly negative. This increases the similarity with the aggregate HICs, further emphasizing the role of Chinese intermedi-

29 To further investigate this longer-term effect, the fixed effect estimator is used to estimate the elasticities between two time periods, e.g. between the year 2000 and the year 2014, and between the average of the years 2000/2001 and the average of the years 2013/2014, respectively. This exercise, found in the Appendix, supports the existence of such positive longer-term productivity effects from the growing use of Chinese intermediates.

Table 5: Estimate for Separate Manufacturing Sub-sectors for Vertical Labour Productivity

High-income Countries

Dependent variable: vertical labour productivity	Import multiplier	Import multiplier from China	R^2 - within	F-value of first-stage regression	N
Food	Neg sign	0.893***	0.04	30	328
Textile and clothing	Neg sign	0.600***	0.56	74	305
Wood	Not sign	0.516***	0.57	64	279
Paper	Not sign	0.686***	0.55	58	297
Printing and recorded media	Neg sign	0.491***	0.41	45	308
Coke and refined petroleum	Neg sign	2.232*	..	14	270
Chemicals	Neg sign	0.858**	0.41	98	299
Pharmaceuticals	Neg sign	0.806*	..	82	320
Rubber and plastic	Neg sign	0.673***	0.69	60	303
Other non-metallic products	Not sign	0.518***	0.39	36	294
Basic metals	Neg sign	0.566***	0.30	15	294
Fabricated metals, except machinery and equipment	Neg sign	0.568***	0.64	51	322
Computers and electronics	Neg sign	0.339***	0.78	188	313
Electrical equipment	Neg sign	0.500***	0.78	102	327
Machinery and equipment, n.e.c.	Neg sign	0.481***	0.83	174	327
Motor vehicles	Neg sign	0.416***	0.76	159	323
Other transport equipment	Neg sign	0.422***	0.50	156	309
Furniture	Neg sign	0.645***	0.42	65	322
Repair and installation of machinery and equipment	No obs	No obs	No obs	No obs	No obs

Note: Linear fixed effect IV estimations (2SLS). Other included regressors are: output multiplier and vertical gross output. For each of the manufacturing sub-sector, the domestic supply chain is included in the variables. Robust standard errors, adjusted for 22 clusters (countries). The variables are expressed in constant prices with base year=2000. Elasticities are presented in column 2. Neg sign=negatively significant at least at p<0.05. The instrumental variable is China's vertical labour productivity. Years: 2000-2014. Log values. $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$.

ates in the vertical specialization of the US economy.

Separate manufacturing sub-sectors

In terms of separate manufacturing sub-sectors and their production processes, how widespread is the productivity effect of the growing use of Chinese intermediates? Without enough data for the United States, this question is answered focusing on the HICs. Continuing to use equation (1) and the fixed effect estimator, Table 5 gives a clear answer to this question. But first, the first column shows that no manufacturing sub-sector, including its domestic supply chain, has a significant positive elasticity of the import multiplier, when controlling for the intermediate imports from China. On the contrary, the elasticity of the import multiplier is significantly negative in 15 out of 18 sub-sectors. Once again, this emphasizes the importance of separating out China when analyzing the productivity effects of the general increase in the use of imported intermediates within the manufacturing production processes of the HICs.

When it comes to the import multiplier from China, Column 2 shows that its elasticity is positively significant at least at $p < 0.05$ in all 18 manufacturing sub-sectors, although the results for coke and refined petroleum and pharmaceuticals, respectively, look less robust.[30] At the bottom of the table, there are five sub-sectors in which the fixed effect estimator seems to

30 When using a linear panel model with instruments and applying the fixed effect estimator, an unspecified R^2 value is, however, not necessarily a problem.

generate extra robust results: with large F-values, highly significant elasticities, and a high level of explanatory power.

Mechanisms: Three Possible Explanations

Several theoretical mechanisms can be used to explain the seemingly positive causal effect on vertical labour productivity growth in the HICs of the growing use of Chinese intermediates. Without any claim of being exhaustive, this section tries to shed some preliminary empirical light on three possible mechanisms put forward in recent China shock research.

Value added or employment?

The welfare consequences of the Chinese productivity effect depend on the channels by which the intermediates affect the productivity growth in the HICs. Therefore, and along the lines of Acemoglu, Akcigit, and Kerr (2015) and their argument for more research on the interplay between value added and employment in an IO setting: is the Chinese productivity effect explained by a positive effect on vertical value added or a negative effect on vertical employment — or both?[31] Using the same econometric approach as before, the results in Table 6 indicate that the main channel in the HICs is reduced employment: no matter the length of the lag, a growing use of Chinese intermediates seems to lead to a significant and economically relevant decrease in vertical employment. In terms of value added, there is a weak tendency of a positive, more instantaneous effect; but with longer lags, this effect seems to be reversed, indicating a double effect: reduced employment and reduced value added. For the United States, the result looks less bleak, with a positive longer-run effect on value added and no longer-run negative effect on employment. If these patterns are in accordance with reality, it would be interesting to understand what might explain the difference between the HICs and the United States.

Reduced prices

Within a neoclassical framework, prices are the main channel through which productivity effects are propagated in the production system (Acemoglu *et al.* 2021). Despite that, Jaravel and Sager (2020) argues that there are knowledge gaps about the effect of the China shock on the prices in the HICs. Their starting point is that the strong Chinese productivity growth is likely to lead to reduced Chinese prices, which in turn, through strategic price-setting, will lead to reduced producer prices (and consumer prices) in the HICs. Based on US data, they show that the growing imports from China, between 1991 and 2007, led to reduced domestic prices and therefore to large consumer surpluses.[32] In line with this, and focusing on US manufacturing price indices, Amiti *et al.* (2020) shows that China's entry to WTO in 2001 led to reduced prices; between 2000 and 2006 the

31 The same interplay has also recently been used when studying the effects on labour productivity of a growing use of robots within (US) sectors (Acemoglu and Restrepo, 2020).

32 Jaravel and Sager (2020) estimate that the consumer surplus is large enough to compensate each of the displaced US job caused by the China shock by around dollar $400000.

Table 6: The China Shock on Value Added and Employment

Panel A: High-income Countries

Dependent variable: vertical value added	No lag	Lag 1	Lag 3	Lag 5	Lag 7
Import multiplier from China	0.008	0.117**	0.062	-0.100	-0.138*
	(0.006)	(0.037)	(0.061)	(0.066)	(0.061)
R^2 – within	0.99	0.34	0.05	0.02	0.03
N	5560	5090	4335	3582	2870
Dependent variable: vertical employment					
Import multiplier from China	-0.579***	-0.425***	-0.332***	-0.313***	-0.259***
	(0.026)	(0.036)	(0.058)	(0.057)	(0.055)
R^2 – within	0.83	0.29	0.06	0.06	0.07
N	5560	5090	4335	3582	2870

Panel B: United States

Dependent variable: vertical value added	No lag	Lag 1	Lag 3	Lag 5	Lag 7
Import multiplier from China	-0.004	0.19	0.502**	0.313**	0.164*
	(0.006)	(0.121)	(0.187)	(0.097)	(0.076)
R^2 – within	0.99	0.35	..	0.04	0.09
N	266	246	211	175	140
Dependent variable: vertical employment					
Import multiplier from China	-0.380***	-0.245*	0.084	0.017	-0.002
	(0.065)	(0.111)	(0.158)	(0.083)	(0.074)
R^2 – within	0.92	0.44	0.04	0.16	0.2
N	266	246	211	175	140

Note: Linear fixed effect IV estimations (2SLS). Vertical value added is defined as the value added needed in the domestic economy in order to satisfy final demand, including all upstream/backward stages of the domestic production process. Vertical value employment is defined analogously. See Appendix for further details. Other included regressors are: output multiplier, import multiplier and vertical gross output. Elasticities and standard errors are shown in the table. Robust standard errors, adjusted for clusters: with no lag, the number of clusters is 390 (GVCs, i.e. countries*manufacturing sub-sectors, including their domestic supply chains) for the HICs and 18 (GVCs, i.e. manufacturing sub-sectors, including their domestic supply chains) for the US. The variables are expressed in constant prices with base year=2000. The instrumental variable is China's vertical labour productivity. Years: 2000-2014. Log values. $^{***}p < 0.001,^{**} p < 0.01,^{*} p < 0.05$. In the estimations with no lag, the F-value of the first-stage regression of the HICs (US) is 390 (175), and the instrumental variable is significant at $p < 0.001$.

China shock reduced manufacturing price indices by 7.6 per cent.

Along similar lines, the question addressed in this section is: does the growing use of Chinese intermediates affect the prices within the manufacturing production processes? The answer is found in Table 7. Clearly, all three deflators point in the same direction: a growing use of Chinese intermediates seems to lead to slower price increases. With the importance of including lags (Jaravel and Sager, 2020), this result seems to become more robust when such delayed effects are estimated: the size of the price effect increases with time. Although somewhat less robust, this effect seems to be larger in the United States than in the HICs. Within the setting of this article, the price effect of the China shock means that the level of real vertical value added would have been lower without the growing intermediate trade with China – and, hence, has contributed positively to the Chinese productivity effect.

Functional specialization

The growing importance of intermediates has led to new approaches to measure and understand specialization within and between countries (Pahl and Timmer, 2019). In order to trace where the value added embedded in a manufactured product is generated, one aspect of this is the movement of measurement from gross trade

Table 7: The China Shock Effect on Prices

Panel A: High-income Countries

Dependent variables: vertical price indices	No lag	Lag 1	Lag 3	Lag 5	Lag 7
Gross output price index: imp.multi from China	-0.011**	-0.022***	-0.028***	-0.033***	-0.036***
	(0.003)	(0.004)	(0.025)	(0.005)	(0.004)
Value added price index: imp.multi from China	-0.009**	-0.020***	-0.026***	-0.031***	-0.035***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)
Intermediate price index: imp.multi from China	-0.013***	-0.024***	-0.030***	-0.034***	-0.036***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)

Panel B: United States

Dependent variables: vertical price indices	No lag	Lag 1	Lag 3	Lag 5	Lag 7
Gross output price index: imp.multi from China	-0.053**	-0.082**	-0.073**	-0.031	-0.046*
	(0.018)	(0.025)	(0.024)	(0.025)	(0.018)
Value added price index: imp.multi from China	-0.048**	-0.075**	-0.070**	-0.029	-0.044*
	(0.016)	(0.023)	(0.024)	(0.024)	(0.018)
Intermediate price index: imp.multi from China	-0.060**	-0.090**	-0.077**	-0.034	-0.048**
	(0.021)	(0.029)	(0.026)	(0.025)	(0.018)

Note: Linear fixed effect IV estimations (2SLS). The vertical gross output price index is defined as the sectoral gross output indexes weighted by Leontief´s inverse: diagonal (go index) ∗ Leontief´s inverse. Hence, in this matrix, the column sum of manufacturing sub-sector i in country j is sector i's vertical gross output price index. The vertical value added price index and the vertical intermediate input price index are constructed analogously. Other included regressors are: output multiplier, import multiplier, and vertical gross output. The variables are expressed in constant prices with base year=2000. The table presents the elasticities and robust standard errors, adjusted for clusters: 390 (countries∗manufacturing sub-sectors, including their domestic supply chains) in the HICs and 18 in the US (manufacturing sub-sectors, including their domestic supply chains) when no lags are used. The instrumental variable is China´s vertical labour productivity. Years: 2000-2014. Log values. $^{*} = p < 0.05, ^{**} = p < 0.01, ^{***} = p < 0.001$. For the HICs (US), the F-value of the first-stage regressions is 259 (206) when no lags are used. The R^2 value ranges from 0.70 to 0.84 when no lags are used.

flows to net (value added) trade flows. Another aspect is related to the allocation of production activities between countries needed to finalize a manufactured product. With China emerging as the factory of the world, the HICs seem to have become less specialized in fabrication activities and more specialized in knowledge-intensive parts of the manufacturing production process, such as R&D, marketing, and management (Timmer *et al.* 2019).

This process of functional specialization is supported within the framework of this article. As can be seen from Table 8, a growing use of Chinese intermediates seems to have led to a growing relative use of knowledge-intensive business services (KIBS) intermediates:[33] the overall use of KIBS intermediates per unit of final demand increases when the use of Chinese intermediates increases. The size of this effect is considerably larger in the United States than in the HICs, although the level of significance is similar. The estimations on the domestic and imported use of KIBS intermediates — also presented in the table — respectively shows that this functional specialization is mainly explained by a growing use of imported of KIBS intermediates. In the HICs, the China effect on the domestic use of KIBS intermediates is negative, while it seems to be a more or less instantaneous positive effect in the United States, which disappears when longer lags are used. The overall messages from this exercise should then be that:

33 The KIBS sectors are: M69-70: legal and accounting activities, head offices; M71: architectural and engineering activities; M72: scientific research; M73: advertising and market research; M74-75: other professional, scientific, and technical activities.

Table 8: The Effect of the China Shock on Functional Specialization

Panel A: High-income Countries

Dependent variable: overall multiplier: KIBS	No lag	Lag 1	Lag 3	Lag 5
Import multiplier from China	0.219***	0.177***	0.139**	0.103*
	(0.048)	(0.047)	(0.043)	(0.049)
Dependent variable: domestic multiplier: KIBS				
Import multiplier from China	-0.035	-0.047*	-0.094***	-0.136***
	(0.021)	(0.022)	(0.022)	(0.026)
Dependent variable: import multiplier: KIBS				
Import multiplier from China	0.242***	0.224***	0.233***	0.239***
	(0.035)	(0.035)	(0.033)	(0.034)

Panel B: United States

Dependent variable: overall multiplier: KIBS	No lag	Lag 1	Lag 3	Lag 5
Import multiplier from China	1.412***	1.125***	0.900**	0.745**
	(0.158)	(0.176)	(0.265)	(0.217)
Dependent variable: domestic multiplier: KIBS				
Import multiplier from China	0.275***	0.244**	0.189	0.159
	(0.067)	(0.083)	(0.102)	(0.087)
Dependent variable: import multiplier: KIBS				
Import multiplier from China	1.052***	0.881***	0.712***	0.586***
	(0.109)	(0.113)	(0.169)	(0.139)

Note: Linear fixed effect IV estimations (2SLS). The variable overall multiplier: KIBS is defined as the total use of KIBS intermediates per unit final demand, including all upstream/backward production stages, irrespective if they are domestically or foreign sourced. The other two KIBS multiplier variables are defined analogously, but only measuring the domestic or the foreign use of KIBS intermediates per unit of final demand, respectively. Other included regressors are: output multiplier, import multiplier, and vertical gross output. The variables are expressed in constant prices with base year=2000. The table presents the elasticities and standard errors of the import multiplier from China. The instrumental variable is China´s vertical labour productivity. Robust standard errors, adjusted for clusters: 390 (countries∗manufacturing sub-sectors, including their domestic supply chains) in the HICs and 18 in the US (manufacturing sub-sectors, including their domestic supply chains) when no lags are used. Years: 2000-2014. Log values. $^{*} = p < 0.05, ^{**} = p < 0.01, ^{***} = p < 0.001$. In the six estimations without lags, the lowest F-value of the first-stage regression is 175.

1. the China effect on the functional specialization towards the KIBS part of the manufacturing production process seems to be more pronounced in the United States than in the HICs, and;

2. this functional specialization points in the direction of more KIBS intensive production processes, but not necessarily within the domestic production stages of the HICs (including the United States).

Has this functional specialization had any effect on the vertical labour productivity growth? Table 9 might indicate that this is the case. Both from a cross-sectional perspective and when focusing on changes over time (i.e. the fixed effect estimator), the table indicates that the overall use of KIBS intermediates — both domestically and foreign sourced — is positively correlated to vertical labour productivity. Adding lags to the fixed effect estimator,

34 When the import multiplier from China is included in these estimations, the size and significance of the elasticities presented in this table is, however, reduced. This might indicate that the productivity effect of the overall KIBS use is, at least partly, dependent on the imports of KIBS intermediates from China.

Table 9: Knowledge-intensive Business Services and Productivity and Vertical Labour Productivity

High-income Countries and United States

Dependent variable: vertical labour productivity	OLS	FE – no lag	FE – lag 1	FE – lag 3	FE – lag 5
HIC: Overall multiplier: KIBS	0.041***	0.111***	0.126***	0.159***	0.106**
	(0.006)	(0.024)	(0.027)	(0.024)	(0.034)
US: Overall multiplier: KIBS	0.145***	0.165***	0.190***	0.209***	0.171***
	(0.04)	(0.024)	(0.031)	(0.030)	(0.026)

Note: OLS and linear fixed effect (FE) estimations. Other included regressors are: output multiplier, import multiplier, and vertical gross output. The variables are expressed in constant prices with base year=2000. The table presents elasticities and robust standard errors of the variable Overall multiplier: KIBS, adjusted for clusters: 408 (countries*manufacturing sub-sectors, including their domestic supply chains) in the HICs and 18 in the US (manufacturing sub-sectors, including their domestic supply chains) when no lags are used. Years: 2000-2014. In the fixed effect estimation without lag, the R^2 value amounts to 0.33 for the HICs and to 0.82 for the US. Years: 2000-2014. Log values. $* = p < 0.05, ^{**} = p < 0.01, ^{***} = p < 0.001$.

this effect seems to be rather persistent and economically relevant.[34] In all five estimations, the size of the correlation is larger in the United States than in the HICs. Finally, when the import multiplier from China is included in these estimations (not shown in the table), the size and significance of the elasticities presented Table 9 is reduced. This indicates that the productivity effect of the overall KIBS use is, at least partly, caused by the imports of KIBS intermediates from China.

Is China special?: A comparison with Eastern Europe

If the results presented in this article have anything to say about the operation of real world economies, then one obvious question is: are China's intermediates a special case? Following Bloom, Draca, and Van Reenen (2016) among others, the last empirical analysis of this article therefore tries to give an answer to this question. In so doing, the causal effect on the vertical labour productivity growth of the growing use of Chinese intermediates is compared to the productivity effect of the intermediate imports from two country groups: (1) Eastern Europe and (2) the HICs themselves.[35] With this purpose, another identification strategy has to be used. The reason for this is that China's vertical labour productivity is not the most appropriate instrumental variable for the exports of intermediates from these two country groups. From the reasoning in the section on identification, and analogously following the often-used strategy in the China shock literature, the imports of Chinese intermediates among the 21 non-HIC-countries in the WIOD will be used as the instrumental variable.

Continuing the use of equation (1), Table 10 gives some conclusions. First, with high F-values, the new identification strategy generates satisfying first-stage regressions, and very much so for the effects of China's and Eastern Europe's intermediates on the HICs, respectively. When it comes to the comparison of the productivity effects in the HICs between China and Eastern Europe, the table shows that the pattern is very much the same: a significantly negative import multiplier and a positive and significant import multi-

35 For the HICs (US), the average import multiplier from Eastern Europe amounted to 0.042 (0.0006) in 2000 and to 0.049 (0.0014) in 2014. The corresponding figures for the import multiplier from the HICs are 0.216 (0.071) and 0.221 (0.073).

Table 10: China Compared with Eastern Europe

Panel A: High-income Countries

Dependent variable: vertical labour productivity	No lag	Lag 1
Import multiplier: estimation China	-0.606***	-0.561***
Import multiplier: estimation Eastern Europe	-1.053***	-1.022***
Import multiplier: estimation HIC	3.368***	3.185***
Import multiplier from China	0.500***	0.490***
Import multiplier from Eastern Europe	0.849***	0.838***
Import multiplier from HIC	-6.051***	-5.742***
R^2 – within: China	0.52	0.46
R^2 – within: Eastern Europe	0.18	0.10
R^2 – within: HIC	..	..
F-value of first-stage regression: China	709	614
F-value of first-stage regression: Eastern Europe	449	343
F-value of first-stage regression: HIC	48	51
N	5866	5377

Panel B: United States

Dependent variable: vertical labour productivity	No lag	Lag 1
Import multiplier: estimation China	0.011	-0.096
Import multiplier: estimation Eastern Europe	0.138	0.067
Import multiplier: estimation HIC	1.206***	1.089**
Import multiplier from China	0.270***	0.285***
Import multiplier from Eastern Europe	0.512***	0.527***
Import multiplier from HIC	-1.241***	-1.207***
R^2 – within: China	0.84	0.81
R^2 – within: Eastern Europe	0.72	0.67
R^2 – within: HIC	0.64	0.67
F-value of first-stage regression: China	555	506
F-value of first-stage regression: Eastern Europe	78	77
F-value of first-stage regression: HIC	37	31
N	283	262

Note: Linear fixed effect IV estimations (2SLS). Other included regressors are: output multiplier and vertical gross output. The variables are measured in constant prices with base year=2000. Robust standard errors, adjusted for 412 clusters (GVCs, i.e. countries*manufacturing sub-sectors, including their domestic supply chains) in the HIC estimations and for 19 clusters (manufacturing sub-sectors, including their domestic supply chains) in the estimations on the United States. The instrumental variable is the imports of Chinese intermediates among the 21 non-HIC-countries in the WIOD. Years: 2001-2014. Log values. $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$. The import multiplier from Eastern Europe measures the use of Eastern European intermediates per unit of final demand, including all upstream stages of the domestic production process. The import multiplier from the HICs is constructed analogously.

plier from China/Eastern Europe. However, with larger elasticities for Eastern Europe's intermediates in both the HICs and the United States, the average absolute productivity effect seems to be more pronounced than the China effect. In this respect: China does not seem to be special.

On the other hand, China and Eastern Europe seem to be special in relation to the productivity effect of the use of intermediates imported from the HICs (i.e. their intra-trade of intermediates). The difference is striking. Not the least, after controlling for the import multiplier from the HICs, the import multiplier becomes positively significant and very large in absolute terms. Hence, when the import multiplier from the HICs is unchanged, a growing use of imported intermediates leads to much faster productivity growth. This gives a clear indication that the HICs' imports of intermediates from each other hold back the positive productivity effect of the growing use of imported intermediates, or the trend towards vertical specialization in the global economy. This is also shown by the elasticity of the import multiplier from the HICs. In both the HICs and the United States, this elasticity is negatively significant and large, indicating that an increase in the use of HICs' intermediates – when the overall level of imported intermediates is unchanged – reduces the vertical labour productivity growth within the manufacturing production processes among the HICs in a substantial way.

Final Discussion

Main results

This article contributes to the literature on the China shock by investigating, for the first time, the labour productivity effect of the growing use of Chinese intermediates within almost 400 manufacturing production processes among 22 HICs. Including all stages of the production process — an approach that has lately received renewed attention in several literatures — the main results are the following.

1. Since the Millennium, the growing use of Chinese intermediates has led to a faster vertical labour productivity growth in the HICs and the United States;
2. This is the case both before and after the financial crisis;
3. The effect is identified in all — or almost all — manufacturing subsectors;
4. The effect passes several robustness tests: different definitions of the overall network of intermediate trade, inclusion of the capital stock, inclusion of lags, weighted estimations, and an alternative IV strategy;
5. China is not special: the productivity effect of the growing use of Eastern European intermediates is equally significant and larger in size;
6. A growing intra-trade of intermediates among the HICs have been detrimental to their productivity growth.

Suggested mechanisms

Among several possible mechanisms, the main reasons why a growing intermediate trade with China seems to lead to a faster productivity growth are unclear (Bloom, Draca, and Van Reenen 2016). This article has tentatively tried to shed light on three of the mechanisms discussed in the China shock literature: value added or employment, reduced producer prices, and functional specialization. The results point in the direction that reduced employment — in comparison to value added — is the most important channel behind the positive productivity effect; the growing use of Chinese intermediates reduces producer prices; and the China shock has led to a productivity enhancing functional specialization towards the use of knowledge-intensive business services intermediates. These mechanisms fit into more general arguments based on increased specialization, intensified global competition, new input combinations, and higher quality intermediates.

Future research

This article has only scratched the surface of what is possible within the chosen framework. Continuing with analyses of vertical productivity, some interesting questions are:

1. What would the productivity effect be when investigating the China effect on

all countries in the world?

2. How would the results based on a vertically integrated TFP measure compare to the results based on the vertical labour productivity measure?

3. Are there any intermediates imported from China that contribute more than others to the productivity growth in the HICs?

4. What about the productivity effect from China's exports of the capital goods included in the capital stock?

5. Is there any relation between the HICs' exports of intermediates to China and their vertical productivity growth?

6. In terms of value added, employment, and vertical productivity, does the China effect differ between different parts of the manufacturing production process in the HICs?

7. In terms of competitiveness, does a growing use of Chinese intermediates lead to improved relative productivity among the HICs?

8. In terms of the value added needed to produce the world demand for a manufactured product, does a growing use of Chinese intermediates lead to a larger share of the world market?

Final remark

The emergence of China as the factory of the world represents a rare opportunity to identify causal effects on the level of the global economy (Autor, Dorn, and Hansen, 2016). Considered as a natural experiment, the reforms in China initiated in the late 1970s have, through increased domestic productivity, led to a remarkable increase in the demand for Chinese intermediates among the HICs. This fundamental reorganization of manufacturing production seems, in turn, to have contributed to faster vertical labour productivity growth in the HICs — and, hence, to improved fundamentals for faster real wage growth. Having said that, this result by no means represents the general equilibrium effect on the labour market, but it may make one dimension of the puzzle somewhat more illuminated. And that is good enough.

References

Acemoglu, D., U. Akcigit, and W. Kerr (2015) "Networks and the Macroeconomy: An Empirical Exploration," *NBER Macroeconomics Annual,* Vol. 30, pp. 273-335.

Acemoglu, D., D. Autor, D. Dorn, G.H. Hanson, and B. Price (2016) "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics,* Vol. 34, No. 1, Part 2, pp. 141-98.

Acemoglu, D. and P.D. Azar (2020) "Endogenous Production Networks," *Econometrica,* Vol. 88, No. 2, pp. 33-82.

Acemoglu, D., V.M. Carvalho, A. Ozdaglar and A. Tahbaz-Salehi (2012) "The Network Origins of Aggregate Fluctuations," *Econometrica,* Vol. 80, No. 5, pp. 1977-2016.

Acemoglu, D., A. Ozdaglar and A. Tahbaz-Salehi (2016) "Networks, Shocks, and Systemic Risk," *The Oxford Handbook of the Economics of Networks,* edited by Yann Bramoullé, Andrea Galeotti and Brian Rogers, pp. 1-39 (Oxford: Oxford University Press).

Acemoglu, D. and P. Restrepo (2019) "Automation and New Tasks: How Technology Displaces and Reinstates Labour," *Journal of Economic Perspectives,* Vol. 33, No. 2, pp. 3-30.

Amiti, M., M. Dai, R.C. Feenstra and J. Romalis (2020) "How Did China's WTO Entry Affect U.S. Prices?," *Journal of International Economics,* Vol. 126, September.

Antras, P. and D. Chor (2019) "On the Measurement of Upstreamness and Downstreamness in Global Value Chains," World Trade Evolution: Growth, *Productivity and Employment*, edited by L.Y. Ing and M. Yu, pp. 126-94 (Routledge).

Antras, P. and D. Chor (2021) "Global Value Chains," Manuscript, March. An abridged version of this paper is to be published as a chapter in the 5th edition of the *Handbook of International Economics.*

Antras, P. and A. de Gortari (2020) "On the Geography of Global Value Chains," *Econometrica,* Vol. 88, No. 4, pp. 1553-98.

Antras, P., T.C. Fort, and F. Tintelnot (2017) "The Margins of Global Sourcing: Theory and Evidence from US Firms," *American Economic Review*, Vol. 107, No. 9, pp. 2514-64.

Autor, D., D. Dorn, and G.H. Hanson (2013) "The China Syndrome: Local Labour Market Effects of Import Competition in the United States," *American Economic Review,* Vol. 103, No. 6, pp. 2121-68.

Autor, D., D. Dorn, and G.H. Hanson (2015) "Untangling Trade and Technology: Evidence from Local Labour Markets," *Economic Journal,* Vol. 125, May, pp. 621-46.

Autor, D., D. Dorn, and G.H. Hanson (2016) "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade," *Annual Review of Economics,* Vol. 8, pp. 205-240.

Autor, D., D. Dorn, and G.H. Hanson (2019) "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights,* Vol. 1, No 2., pp. 161-78.

Autor, D., D. Dorn, G.H. Hanson, G. Pisano, and P. Shu (2020a) "Foreign Competition and Domestic Innovation: Evidence from US Patents," *American Economic Review: Insights,* Vol. 2, No. 3, pp. 357-74.

Autor, D., D. Dorn, G.H. Hanson, G. Pisano, and K. Majlesi (2020b) "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," *American Economic Review,* Vol. 110, No. 10, pp. 3139-83.

Autor, D. and A. Salomons (2018) "Is Automation Labor Share-Displacing? Productivity Growth, Employment, and the Labor Share," *Brookings Papers on Economic Activity,* Spring, pp. 1-63.

Baldwin, R. (2016) *The Great Convergence: Information Technology and the New Globalization* (London: Harvard University Press).

Baldwin, R. (2017) "Factory-Free Europe? A Two Unbundlings Perspective on Europe's Twentieth-Century Manufacturing Miracle and Twenty-First-Century Manufacturing Malaise," in *The Factory-free Economy, Outsourcing, Servitization, and the Future of Industry,* edited by Lionel Fontagné and Ann Harrison, pp. 24-66 (Oxford: Oxford University Press).

Baldwin, R. (2019) *The Globotics Upheaval, Globalization, Robotics, and the Future of Work* (Oxford: Oxford University Press).

Bloom, N., M. Draca and J. Van Reenen (2016) "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT, and Productivity," *Review of Economic Studies,* Vol. 83, pp. 87-117.

Bloom, N., K. Handley, A. Kurman, and D. Luck (2019) "The Impact of Chinese Trade on U.S. Employment: The Good, The Bad, and The Debatable," *Manuscript,* July.

Bloom, N., P. Romer, S.J. Terry, and J. Van Reenen (2021) "Trapped Factors and China's Impact on Global Growth," *The Economic Journal*, Vol. 131, No. 633, pp. 156-91.

Buckley, P.J., R. Strange, M.P. Timmer, and G.J. de Vries (2020) "Cathcing-Up in the Global Factory: Analysis and Policy Implications," *Journal of International Business Policy*, Vol. 3, pp. 79-106.

Caceres, C,. D.D. Cerdiero, and R. Mano (2019) "Trade Wars and Trade Deals: Estimated Effects Using a Multi-sectoral Model," IMF.

Caliendo, L. and F. Parro (2015) "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, Vol. 82, No. 1, pp. 1-44.

Caliendo, L., F. Parro, and A. Tsyvinski (2017) "Distortions and the Structure of the World Economy," *NBER Working Paper*, No. 23332.

Caliendo, L., M. Dvorkin, and F. Parro (2019) "Trade and Labor Market Dynamics: General Equilibrium analysis of the China Trade Shock," *Econometrica*, Vol. 87, No. 3, pp. 741-835.

Carter, A.P. (1970) *Structural Change in the American Economy* (Cambridge: Harvard University Press).

Carvalho, V.M. and N. Voigtländer (2015) "Input Diffusion and the Evolution of Production Networks", Barcelona GSE Working Paper Series, No. 759.

Carvalho, V.M. and A. Tahbaz-Salehi (2019) "Production Networks: A Primer," *Annual Review of Economics,* Vol. 11, August, pp. 635-63.

Caselli, F., M. Koren, M. Lisicky, and S. Tenreyro (2020) "Diversification Through Trade," *The Quarterly Journal of Economics,* Vol. 135, No.1, pp. 449-502.

Che, Y., Y. Lu, J.R. Pierce, P.K. Schott, and Z. Tao (2020) "Did Trade Liberalization with China Influence US Elections?" *Manuscript*, December.

Constantinescu, C., A. Mattoo, and M. Ruta (2019) "Does Vertical Specialisation Increase Productivity?," *The World Economy*, Vol. 42, pp. 2385-2402.

Corrado, C., J.E. Haskel, M. Iommi, and C. Jona-Lasinio (2020) "Intangible Capital, Innovation, and Productivity à la Jorgenson: Evidence from Europe and United States," in B. Fraumeni (ed.) *Measuring Economic Growth and Productivity, Foundations, KLEMS Production Models, and Extensions*, (London: Academic Press).

Dauth, W., S. Findeisen, and J. Suedekum (2014) "The Rise of the East and the Far East: German Labor Markets and Trade Integration," *Journal of the European Economic Association*, Vol. 12, No. 6, pp. 1643-1675.

De Juan, O. and E. Febrero (2000) "Measuring Productivity from Vertically Integrated Sectors," *Economic Systems Research,* Vol. 12, No. 1, pp. 65-82.

Dietzenbacher, E., A.R. Hoen, and B. Los (2000) "Labour Productivity in Western European Countries 1975-1985: An Intercountry, Interindustry Analysis," *Journal of Regional Science*, Vol. 40, No. 3, pp. 425-452.

Dietzenbacher, E., B. Los, R. Stehrer, M. Timmer, and G. de Vries (2013) "The Construction of World Input-Output Tables in the WIOD Project," *Economic Systems Research*, Vol. 25, No. 1, pp. 71-98.

Feenstra, R.C. and A. Sasahara (2018) "The 'China Shock', Exports and U.S Employment: A Global Input-Output Analysis," *Review of International Economics*, Vol. 26, No. 5, pp. 1053-1083.

Feenstra, R.C., H. Ma, and Y. Xu (2019) "US Exports and Employment," *Journal of International Economics*, Vol. 120, pp. 46-58.

Gu, W. and B. Yan (2017) "Productivity Growth and International Competitiveness," *Review of Income and Wealth*, Vol. 63, No. 1, pp. 113-133.

IMF (2019) "Global Value Chains: What Are the Benefits and Why Do Countries Participate?," *Working Paper*, No. 18.

Jaravel, X. and E. Sager (2020) "What Are the Price Effects of Trade? Evidence from the U.S Implications for Quantitative Trade Models," Manuscript, August.

Jones, C.I. (2013) "Misallocation, Input–Output Economics, and Economic Growth," in D. Acemoglu, M. Arellano, and E. Dekel (eds.) *Advances in Economics and Econometrics: Tenth World Congress. Volume II: Applied Economics*, (Cambridge: Cambridge University Press).

Leontief, W. (1953) *Studies in the Structure of the American Economy* (New York: Oxford University Press).

Miller, R.E. and P.D. Blair (2009) *Input-Output Analysis: Foundations and Extensions* (Cambridge: Cambridge University Press).

Oberfield, E. (2018) "A Theory of Input-Output Architecture," *Econometrica*, Vol. 86, No.2, pp. 559-89.

Pahl, S. and M. Timmer (2019) "Patterns of Vertical Specialisation in Trade: Long-run Evidence for 91 Countries," *Review of World Economics*, Vol. 155, pp. 459-86.

Pasinetti, L.L. (1981) *Structural Change and Economic Growth: A Theoretical Essay on the Dynamics of the Wealth of Nations* (Cambridge: Cambridge University Press).

Pasinetti, L.L. (1993) *Structural Economic Dynamics: A Theory of the Economic Consequences of Human Learning* (Cambridge: Cambridge University Press).

Pierce, J.R. and P.K. Schott (2016) "The Surprisingly Swift Decline of US Manufacturing Employment," *American Economic Review,* Vol. 106, No. 7, pp. 1632-1662.

Pierce, J.R. and P.K. Schott (2020) "Trade Liberalization and Mortality: Evidence from US Counties," *American Economic Review: Insights*, Vol. 2, No. 1, pp. 47-64.

Ponte, S., G. Gereffi, and G. Raj-Reichart (eds.) (2019) *Handbook on Global Value Chains* (Northhampton: Edward Elgar Publishing).

Reijnders, L.S., M.P. Timmer, and X. Ye (2021) "Labour Demand in Global Value Chains: Is There a Bias Against Unskilled Work?," *The World Economy*, Online version before inclusion in an issue.

Rodriguez-Clare, A., M. Ulate, and J.P. Vasquez (2020) "New-Keynesian Trade: Understanding the Employment and Welfare Effects of Trade Shocks," *NBER Working Paper*, No. 27905.

Rosenberg, N. (1982) *Inside the Black Box: Technology and Economics* (Cambridge: Cambridge University Press).

Ten Raa, T. and E. Wolff (2000) "Engines of Growth in the US Economy," *Structural Change and Economic Dynamics*, Vol. 11, No. 4, pp. 473-89.

Ten Raa, T. and E. Wolff (2001) "Outsourcing of Services and the Productivity Recovery in U.S. Manufacturing in the 1980s and 1990s," *Journal of Productivity Analysis*, Vol. 16, pp. 149-65.

Ten Raa, T. and E. Wolff (2012) *Productivity Growth: Industries, Spillovers and Economic Performance* (Cheltenham: Edward Elgar).

Timmer, M. (2017) "Productivity Measurement in Global Value Chains," *International Productivity Monitor,* Vol. 33, Fall, pp. 182-93. http://www.csls.ca/ipm/33/Timmer.pdf.

Timmer, M.P., E. Dietzenbacher, B. Los, R. Stehrer, and G.J. de Vries (2015) "An Illustrated User Guide to the World Input-Output Database: The Case of Global Automotive Production," *Review of International Economics*, Vol. 23, No. 3, pp. 575–605.

Timmer, M.P., B. Los, R. Stehrer, and G.J. de Vries (2021) "Supply Chain Fragmentation and the Global Trade Elasticity: A New Accounting Framework," *IMF Economic Review*, article not assigned to an issue.

Timmer, M.P., S. Mirodout, and G.J. de Vries (2019) "Functional Specialisation in Trade," *Journal of Economic Geography*, Vol. 19, pp. 1-30.

Timmer, M.P. and X. Ye (2018) "Productivity and Substitution Patterns in Global Value Chains," E. Grifell-Tatjé, C.A. Know Lovell, and R.C. Sickles (eds.) *The Oxford Handbook of Productivity Analysis* (Oxford: Oxford University Press).

Timmer, M.P. and X. Ye (2020) "Accounting for Growth and Productivity in Global Value Chains," in B. Fraumeni (ed.) *Measuring Economic Growth and Productivity: Foundations, KLEMS Production Models, and Extensions*, pp. 413-26 (London: Academic Press).

Winter, S. and R. Nelson (1982) *An Evolutionary Theory of Economic Change* (Cambridge: Harvard University Press).

Wolff, E. N. (1994) "Productivity Measurement Within an Input-Output Framework," *Regional Science and Urban Economics*, Vol. 24, No. 1, pp. 75-92.

WTO (2019) "Technological Innovation, Supply Chain Trade, and Workers in a Globalized World," *Global Value Chain Development Report*.

Appendix: Construction of the Main Variables

Vertical labour productivity

This variable is defined as the ratio between the (vertical) value added and (vertical) employment generated within the domestic production process in order to satisfy final demand.[36] Mathematically and using matrix notation, the vertical value added (VVA) is found by the equation:

$$VVA = VA/GO((I - A)^{-1}FD).$$

$(I - A)^{-1} = L = [l_{ijj}]$ is the Leontief inverse $((I - A)^{-1} = I + A + A^2 + A^3 + ...)$, or the total requirements matrix.[37] l_{ij} is thus a partial derivative and expresses the total effect on domestic production in sector i of a unit change in final demand in sector j ($L_{ij} = \frac{\partial x_i}{\partial f_i}$), including all subsequent rounds of indirect intermediate demand. Therefore, L describes how a change in final demand is transmitted throughout the domestic production system in wider and wider circles. VA/GO is a diagonal matrix with the ratio between domestic sectoral value added and domestic sectoral gross output on the main diagonal and zeros elsewhere. FD is a diagonal matrix with sectoral final demand on the main diagonal and zeros elsewhere. In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in VVA_i is the (vertical) value added needed to satisfy sector j's final demand, including all upstream stages of its domestic production process. Vertical employment is defined in the same way, but with sectoral employment instead of sectoral value added.

Output multiplier

This variable measures the gross output needed in the domestic economy in order to produce one unit of final demand, including all subsequent rounds of indirect intermediate demand generated along the domestic supply chain (Miller and Blair, 2009). In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in L_i is sector j's output multiplier. Although domestically oriented, this variable resembles to the downstream measures frequently used in the GVC literature (Antras and Chor, 2021).

Import multiplier

This variable measures the use of imported intermediates per unit of final demand, including all subsequent rounds of indirect demand for imported intermediates generated along the domestic supply chain. Mathematically, the import multiplier (IM) is found by the equation: $IM = II/GO(I-A)^{-1}$, where II/GO is a matrix with the ratios between sectoral intermediate imports and sectoral gross output. In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in IM_i is sector i's import multiplier, including all subsequent rounds of indirect demand for intermediate imports generated along the domestic supply chain. Although domestically oriented, this variable resembles to the measures of foreign

36 The term "vertical" comes from the description of the column sum dimension often used in IO analysis (Miller and Blair, 2009) and from Carvalho (2014) who argues that a vertical economy is – in contrast to a horizontal economy – an economy where trade in intermediates connect sectors.

37 A is the direct requirement matrix, describing the first round effect on the intermediate demand from a unit change in final demand.

value added in exports in the GVC literature (Antras and Chor, 2021).

Import multiplier from China

This variable measures the use of Chinese intermediates per unit of final demand, including all subsequent rounds of indirect demand for Chinese intermediates generated along the domestic supply chain. Mathematically, the import multiplier from China (IMC) is found by the equation: $IMC = IIC/GO(I - A)^{-1}$, where IIC/GO is a matrix containing the ratios between sectoral intermediate imports from China and sectoral gross output. In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in IMC_i is sector j's import multiplier from China, including all subsequent rounds of indirect demand for Chinese intermediates generated along the domestic supply chain.

Overall multiplier

This variable measures the gross output needed to produce one unit of final demand, irrespective of whether the intermediates are domestically or foreign sourced. It is defined as the sum of the output and import multiplier.[38]

Capital multiplier

This variable measures the use of the capital stock per unit of final demand, including all subsequent rounds of indirect demand for the capital stock generated along the domestic supply chain. Mathematically, the capital multiplier (CM) is found by the equation: $CM = CS/GO(I - A)^{-1}$. where CS/GO is a diagonal matrix containing the ratio between the sectoral capital stock and sectoral gross output on the main diagonal and zeros elsewhere. In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in CM_i is sector j's capital multiplier, including all subsequent rounds of indirect demand for the capital stock along the domestic supply chain.

Vertical gross output

This variable measures the gross output needed to satisfy final demand, including all subsequent rounds of indirect demand generated along the domestic supply chain. Mathematically, vertical gross output (VGO) is found by the equation: $VGO = (I - A)^{-1}FD$, where FD is a diagonal matrix with final demand on the main diagonal and zeros elsewhere. In country i, the column sum for sector j (i.e. $GVC_{i,j}$) in VGO_i is sector j's vertical gross output, including all subsequent rounds of indirect demand along the domestic supply chain.

38 The term "overall multiplier" is my own construct.

Appendix Table A: Weighted Estimations

Panel I: High-income Countries

Dependent variable: vertical labour productivity	Weight=go	Weight=va	Weight=empl.
Import multiplier	-1.006***	-0.994***	-0.864***
	(0.154)	(0147)	(0.131)
Import multiplier from China	0.720***	0.710***	0.683***
	(0.068)	(0.065)	(0.065)

Panel II: United States

Dependent variable: vertical labour productivity	Weight=go	Weight=va	Weight=empl.
Import multiplier	-0.32	-0.367	-0.528*
	(0.255)	(0.265)	(0.237)
Import multiplier from China	0.348***	0.362***	0.401***
	(0.066)	(0.07)	(0.064)

Note: Note. Linear fixed effect IV estimations (2SLS). Other included regressors are: output multiplier and vertical gross output. GO weight = share of real vertical gross output, VA weight = share of real vertical value added, and empl weight = share of vertical employment. Robust standard errors in parentheses, adjusted for clusters. The instrumental variable is China's vertical labour productivity. Years: 2000-2014. Log values. $^* = p < 0.05, ^{**} = p < 0.01, ^{***} = p < 0.001$. The six estimations pass the under-identification test (Kleibergen-Papp rk LM statistic) and the weak identification test (Cragg-Donald Wald F statistic).

Appendix Table B: Another Instrumental Variable

Panel I: High-income Countries

Dependent variable: vertical labour productivity	No lag
Import multiplier	-0.606***
	(0.067)
Import multiplier from China	0.500***
	(0.016)
R^2 – within	0.52
N	5866

Panel II: United States

Dependent variable: vertical labour productivity	No lag
Import multiplier	0.011
	(0.182)
Import multiplier from China	0.270***
	(0.042)
R^2 – within	286
N	0.84

Note: Linear fixed effect IV estimations (2SLS). Other included regressors are: output multiplier and vertical gross output. Robust standard errors in parentheses, adjusted for clusters (GVCs). All variables are expressed in constant prices with base year=2000. The instrumental variable is the import of Chinese intermediates among the 21 countries in the WIOD not defined as an HIC. Years: 2000-14. Log values. $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$. The F-value of the first-stage regression for the HICs (US) is 709 (555), and the instrumental variable is significant at $p < 0.001$.

Appendix Table C: Longer-term Effects of the China Shock

High-income countries

Dependent variable: vertical labour productivity	2000 vs 2014	2000-01 vs 2013-14	2000-02 vs 2012-14	2000-03 vs 2011-14	2000-04 vs 2010-14
Output multiplier	-0.467	-0.874**	-1.014**	-0.939**	-0.713*
	(0.249)	(0.299)	(0.322)	(0.332)	(0.335)
Import multiplier	-0.593***	-0.630***	-0.643***	-0.685***	-0.714***
	(0.101)	(0.107)	(0.122)	(0.132)	(0.142)
Import multiplier from China	0.467***	0.491***	0.539***	0.577***	0.593***
	(0.022)	(0.025)	(0.029)	(0.033)	(0.038)
F-value: first stage	490	418	387	339	298
R^2 - within	0.77	0.71	0.69	0.65	0.58
N	750	753	754	754	756

Note: Linear fixed effect IV estimations (2SLS). Robust standard errors in parentheses, adjusted for 386 clusters (countries*manufacturing sub-sectors, including their domestic supply chains). The instrumental variable is China's vertical labour productivity. Vertical gross output is included to control for the actual level of demand. Years: 2000-2014. Log values. $^{***} = p < 0.001, ^{**} = p < 0.01, ^{*} = p < 0.05$. With two time periods, the fixed effect estimator generates the same result as the first-difference estimator.

Trading Gains and Productivity: A Törnqvist Approach

Ulrich Kohli[1]
University of Geneva

Abstract

This article looks at alternative Törnqvist measures of a country's trading-gain and terms-of-trade effects, as they have been proposed in the literature starting with the seminal work of Diewert and Morrison (1986), and their link to standard measures of productivity. It strongly argues in favour of using the price of domestic final demand as a deflator when computing real Gross Domestic Income (GDI), and, by the same token, the trading gains and labour productivity measures. It shows that the trading gains then generally consist of two parts, a pure terms-of-trade component and an additional relative-price component, the latter of which can be interpreted as a real-exchange-rate effect. National and international statistical agencies, with the notable exceptions of Statistics Canada and the U.S. Bureau of Economic Analysis, tend to report incomplete trading-gain statistics in that they omit the second component. Consequently the real GDI estimates they publish must be viewed as flawed. Taking trading-gains into account has no direct effect on the measurement of total factor productivity, but it does affect the measures of average and marginal labour productivity when related to real GDI and its deflator. Numerical estimates for Switzerland are reported as an illustration.

It is well known that changes in the terms of trade and the real exchange rate of an open economy can have a significant effect on its welfare. Yet, the impact of such changes on a country's real income — as captured by the so-called trading gains — have long been rather neglected by the traditional measures of the national accounts.[2] Admittedly, trading gains tend to be much smaller than productivity advances, but they can nonetheless be significant. Moreover, the two types of gains may be intertwined. The purpose of this article is to document these effects using superlative price and quantity indices.

Among the statistical agencies, the

1 Emeritus Professor of Economics, University of Geneva. Over the years, I have greatly benefited from discussions with numerous colleagues on the issues covered in this article. At the risk of forgetting many names, I would like to especially thank (more or less in chronological order) W. Erwin Diewert, Alan D. Woodland, Elie Appelbaum, Ronald W. Jones, Henryk Kierzkowski, Jean-Christian Lambelet, Jaime de Melo, Kevin J. Fox, and Marshall B. Reinsdorf. I am also very grateful to the editors and to three anonymous referees for helpful comments and suggestions. Email: Ulrich.Kohli@hotmail.com, Ulrich.Kohli@unige.ch.

2 See Geary (1961) for an early and lucid exposition of the need for such a concept. The term "trading gain" seems to have been coined by Burge and Geary (1957); see Neary (1997).

U.S. Bureau of Economic Analysis (BEA) stands out for long having been publishing series of *command-basis* real Gross Domestic Product (GDP), generally interpreted in the literature as real Gross Domestic Income (GDI) (Denison, 1981). Originally, the BEA's approach was to deflate the trade account by the price of imports, as opposed to deflating exports and imports by their own respective prices as it is usually done when computing real GDP. The difference between the two measures was interpreted as the trading gains, or losses.[3]

Before proceeding, it would seem useful to try to define the concept of trading gain more precisely, rather than simply referring to the statistical approach originally used by the BEA. Thus, one might define the *trading gain* (or *loss*) as the extra real domestic income that a country earns (or loses) simply as the result of changes in the relative prices relevant for its international trade. As it will be shown, these relative prices generally involve at least three prices: the prices of imports, exports, and domestic final goods.

It is noteworthy that most statistical agencies do not define real GDI as nominal GDI (equal to nominal GDP by the national accounts identity) deflated by an appropriate price index. Instead, real GDI is still generally computed as real GDP plus the trading gains, however defined.[4] This is all the more surprising given that, for many purposes, real GDI is just as important a macroeconomic concept as real GDP. Real income is essential in explaining aggregate demand and savings, plays a leading role in many fields of economics, like public finance and monetary economics, and it is a better welfare indicator than real GDP. Real income and trading gains also play an important role in many models of international economics, including the modelling of internal and external balance (Salter, 1959; Corden, 1960). Nonetheless, the estimation of real GDI is generally relegated to a side issue and is subjected to the vagaries of the measurement of the trading gains.

Both the *System of National Accounts* (SNA) *2008* and Eurostat's *European System of Accounts* (ESA) *2010* now do recommend that trading gains be treated as an integral part of the SNA.[5] They leave the choice of the price deflator to the individual countries, however, simply suggesting one of the following: the price of imports, the price of exports, an average of the two, a general price index like the consumer price index, or a price index for gross domestic final expenditures. In recent years, many countries have thus begun to publish data on trading gains, mostly using the price of imports as the deflator of the trade account. Some countries opted for a domestic price index instead. Thus, Switzer-

---

3 The *trading gains* are measured relative to a reference period; this is not to be confused with the gains from trade, which traditionally refer to a (hypothetical) closed-economy situation.

4 Thus, the *Export and Import Price Manual* defines real GDI as: "A real income measure defined as the volume of GDP plus the trading gain or loss resulting from changes in the terms of trade," International Monetary Fund (2009b:619).

5 See International Monetary Fund (2009a:317), and European Commission (2013:302). The Stiglitz Commission also recommended that trading gains be taken into account (Stiglitz, Sen, and Fitoussi, 2009:95); Hartwick (2020) also discussed this issue in his very extensive review of national accounting.

land started publishing trading-gain statistics using the gross domestic final expenditure price index as deflator in July 2007.[6] Canada did likewise in December 2008, and the United States followed suit in July 2010 (Statistics Canada, 2016, Chapter 7:28).

While the BEA originally used fixed-weight Laspeyres quantity indices when computing trading gains, it started using chained Fisher price and quantity indices in 1996.[7] Statistics Canada did the same in 2001, but as of today most other countries, including Switzerland, still use the Laspeyres quantity aggregation, albeit in chain form.

The focus in this article is on chained Törnqvist — rather than Fisher — indices. This choice is motivated by their ease of computation and exposition, plus the fact that the Translog functional form, for which Törnqvist indices are exact, can be estimated relatively easily (Christensen, Jorgenson and Lau, 1973; Diewert, 1974). Moreover, there is no known functional form for a GDP function for which Fisher indices are exact, except under some rather restrictive restrictions such as global separability between domestic factor services and output (including import) quantities (Kohli, 1993; Kohli, 2004a; footnote 21). In any case, it is widely acknowledged that the numerical differences between these two superlative indices are typically very small (Diewert, 1976).

We will show that there are compelling arguments in favour of using the price index for gross domestic final expenditures as a deflator when computing real GDI and the trading gains. Moreover, except for the unlikely situation when trade is balanced, the trading gains really consist of two elements, a pure terms-of-trade effect and a further relative-price effect that can be interpreted in some cases as a real-exchange-rate effect. Most statistical agencies only report the first component, which means that their so-called trading-gain estimates are incomplete, and thus misnamed, and, furthermore, that their measures of real GDI are flawed.

The Diewert and Morrison Approach to Terms-of-trade Effects

Our starting point is the seminal *Economic Journal* article by Erwin Diewert and Catherine Morrison (Diewert and Morrison, 1986). They use the GDP-function approach to modelling imports and exports, which treats traded goods as middle products.[8] This approach recognizes the fact that most imports are made up of raw materials and intermediate products, and even most so-called finished products must still go through a number of domestic transformations (such as transportation, insurance, unloading, storage, wholesaling, and retailing), where they re-

6 See Swiss National Bank (2007:page IV); these series were extended back to 1990; thanks are due to Michel Peytrignet, former Head of Economic Affairs, and Christoph Menzel, former Head of Statistics, for their role in having these series published.

7 See Landefeld and Parker (1997); also see Reinsdorf (2010) for a very thorough and detailed analysis of trading gains in the context of chained Fisher indices; Reinsdorf makes a very strong case in favour of the use of the gross domestic final expenditures price index as deflator.

8 See Kohli (1978, 1991) and Woodland (1982); the term middle product was coined by Sanyal and Jones (1982).

ceive domestic value added before eventually reaching final demand. The same holds true for exports that can be viewed as intermediate inputs to the foreign technology. In essence, nearly all international trade takes place during production, rather than after.

Define the country's production possibilities set (Θ_t) as the set of all feasible input and output combinations at time t. Let $p_{i,t}$ be the price of output i at time t and $q_{i,t}$ its quantity, and let $x_{j,t}$ and $w_{j,t}$ be the quantity and the rental price of domestic primary factor j; the corresponding vectors are denoted by $p_t \equiv \lfloor p_{i,t} \rfloor, q_t \equiv \lfloor q_{i,t} \rfloor, x_t \equiv \lfloor x_{j,t} \rfloor, w_t \equiv \lfloor w_{j,t} \rfloor$. For illustrative purposes, and since we are mostly interested in imports and exports, we will assume just three variable quantities: exports (X), imports (M, treated as a negative output), and domestic final expenditures (N, an aggregate of private consumption, government consumption, and investment). Note that the domestic final good is clearly distinct from imports and exports, and it can be therefore interpreted as a non-traded good. Production involves two domestic factors, labour (L) and capital (K), both in fixed supplies at any point in time. Assuming that Θ_t is a convex cone and that production is competitive and profit maximizing, the technology can be represented by a GDP function defined as follows:[9]

$$\pi_t = \pi(p_t, x_t, t) \equiv max_q \left\{ \begin{array}{l} p_{N,t}q_N + p_{X,t}q_X \\ -p_{M,t}q_M : (q, x_t) \in \Theta_t, \end{array} \right\} \quad (1)$$

where π_t is nominal GDP at time t. This GDP function is linearly homogeneous in prices by definition. Moreover, the assumption that Θ_t is a cone implies constant returns to scale, i.e. linear homogeneity in domestic input quantities. It can conveniently be implemented using the Translog functional form; it is well known that this function can provide a second-order approximation to an arbitrary GDP function such as (1).

Let $\Pi_{t,t-1} \equiv \pi_t/\pi_{t-1}$ be the growth factor of nominal GDP. Diewert and Morrison show that it can be expressed as:

$$\Pi_{t,t-1} = P_{Y,t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1} \quad (2)$$

where

$$P_{Y,t,t-1} \equiv \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{\bar{s}_{X,t}} \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{-\bar{s}_{M,t}} \cdot \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{\bar{s}_{N,t}} \quad (3)$$

is a Törnqvist index of the prices of outputs (including imports, treated as a negative output), and

$$X_{t,t-1} \equiv \left(\frac{x_{L,t}}{x_{L,t-1}}\right)^{\bar{\sigma}_{L,t}} \left(\frac{x_{K,t}}{x_{K,t-1}}\right)^{\bar{\sigma}_{K,t}} \quad (4)$$

is a Törnqvist index of the quantities of the fixed domestic factors; $s_{i,t}(i = N, X, M)$ and $\sigma_{j,t}(j = K, L)$ are the nominal GDP shares of output i and input j at time t, respectively, with $s_{X,t} - s_{M,t} + s_{N,t} = 1$ and $\sigma_{L,t} + \sigma_{K,t} = 1$; $\bar{s}_{i,t} \equiv \frac{1}{2}(s_{i,t-1} + s_{i,t})$ and $\bar{\sigma}_{j,t} \equiv \frac{1}{2}(\sigma_{j,t-1} + \sigma_{j,t})$ denote the av-

9 See Diewert (1974), Kohli (1978, 1991), and Woodland (1982) for the properties of GDP functions.

erage share of output i and input j over consecutive periods. Diewert and Morrison demonstrate that both of these indices are exact if the underlying GDP function is indeed Translog. $R_{t,t-1}$, finally, is a measure of total factor productivity (TFP) growth and it is obtained as a residual:[10]

$$R_{t,t-1} \equiv \Pi_{t,t-1}/(P_{Y,t,t-1} \cdot X_{t,t-1}). \quad (5)$$

Considering expression (2), both $X_{t,t-1}$ and $R_{t,t-1}$ are real growth factors and their product yields the real-GDP growth factor, $Y_{t,t-1} \equiv y_t/y_{t-1}$:[11]

$$Y_{t,t-1} \equiv \Pi_{t,t-1}/P_{Y,t,t-1} = X_{t,t-1} \cdot R_{t,t-1}. \quad (6)$$

This expression shows that the two sources of economic growth are the increases in factor endowments, as captured by $X_{t,t-1}$, and increases in productivity, as measured by $R_{t,t-1}$.

Diewert and Morrison convincingly argue, however, that $P_{Y,t,t-1}$ in (2) does also contain a real element, namely the impact of changes in the terms of trade. An improvement in the terms of trade is similar to a technological progress, in that it allows a country to obtain more for less, so to speak. It is as if a country's exports were transformed into its imports by the rest of the world. The fact that this transformation takes place abroad rather than within the country is not relevant from a strictly economic viewpoint, and if this transformation technology becomes more (or less) productive over time, it has very real consequences on the country's income. Diewert and Morrison therefore seek to exclude this real component from $P_{Y,t,t-1}$ and they define the following terms-of-trade effect, $DMA_{t,t-1}$:

$$DMA_{t,t-1} \equiv \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{\bar{s}_{X,t}} \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{-\bar{s}_{M,t}}. \quad (7)$$

This term captures the impact on nominal GDP of changes in import and export prices. The decomposition of nominal GDP growth then becomes:

$$\Pi_{t,t-1} = DMA_{t,t-1} \cdot DMB_{t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1} \quad (8)$$

where

$$DMB_{t,t-1} \equiv \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{1-(\bar{s}_{X,t}-\bar{s}_{M,t})} = \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{\bar{s}_{N,t}} \quad (9)$$

measures the nominal-GDP effect of changes in the price of the domestic final good.

Decomposition (8) has been used in a number of empirical studies to explain the growth of nominal GDP (Fox and Kohli, 1998; Fox, Kohli, and Warren, 2002; Kohli, 1990, 2002). One drawback of this approach, however, is that $DMA_{t,t-1}$, which is supposed to measure a real effect, is not homogeneous of degree zero in prices, unless trade happens to be balanced over consecutive periods (Kohli, 2003, footnote 25; Kohli, 2004a, footnote 19). In

10 It is possible to calculate $R_{t,t-1}$ exactly if the parameters of the Translog GDP function are known (Kohli, 1990).

11 This index of real GDP thus has the implicit Törnqvist form (Kohli, 2004b).

other words, an equiproportionate change in import and export prices would generally lead $DMA_{t,t-1}$ to register a change, even though the terms of trade clearly would not have varied, thus suggesting that $DMA_{t,t-1}$ is not a measure of a pure terms-of-trade effect. By the same token, if trade is not balanced (i.e. if $\bar{s}_{N,t} \neq 1$), the price term $DMB_{t,t-1}$ is not linearly homogeneous in current prices as one would expect it to be: it must therefore still contain a real element. This qualification may not be important for a majority of countries whose trade is close to being balanced, but there are also many countries that do not satisfy this requirement.

Terms-of-trade and Real-exchange-rate Effects

To address the problem of the non-zero price homogeneity of the terms-of-trade effect, a number of different approaches have been proposed in the literature. The idea behind all of them is to recognize, as suggested in the introduction, that, unless trade happens to be balanced, the trading gains do not merely depend on the prices of imports and exports, but also on a third price, the price of the domestic final good. These three prices can be characterized by two price ratios, one of which being the terms of trade, whereas the second ratio can be defined in different ways, each time relative to $p_{N,t}$. It is important, though, that both ratios be taken into account when deriving the trading gains, not only for the estimate of these to be complete, but also to ensure that the decomposition of nominal GDP be linearly homogeneous in prices. Whether trade is balanced or not, $p_{N,t}$, the price of the domestic final good then emerges naturally as the appropriate deflator for real GDI. Some of these approaches are discussed in the on-line Appendix.[12] In what follows, we will use the approach of Kohli (2006a, 2006b, 2007), which is the most appealing from a trade-theoretic viewpoint.[13]

We begin by defining the terms of trade (h_t) as the ratio of export prices to import prices:

$$h_t \equiv \frac{p_{X,t}}{p_{M,t}}. \tag{10}$$

We next define the price of traded goods ($p_{T,t}$) as a weighted geometric mean of the price of exports and imports:

$$p_{T,t} \equiv p_{X,t}^{\lambda} \cdot p_{M,t}^{1-\lambda} \quad 0 \leq \lambda \leq 1. \tag{11}$$

The weight on the price of exports (λ) could be set to ½ in analogy to one of the options mentioned in the introduction. Alternatively, it could be set to the share of exports in total trade in the first period, or the mean share over the sample.

Finally, we define the real exchange rate (e_t) as the price of traded relative to the price of non-traded goods:[14]

12 The on-line Appendix can be found on the Centre for the Study of Living Standards Website: http://www.csls.ca/ipm/42/IPM_42_Kohli_Appendix.pdf

13 Also on this issue, see Kohli and Natal (2014), Macdonald (2010, 2020), Macdonald and Ripsoli (2016), and Reinsdorf (2010).

14 This measure of the real exchange rate is also known in the literature as the Salter (1959) ratio; on this topic, also see Corden (1992).

$$e_t \equiv \frac{p_{T,t}}{p_{N,t}} = \frac{p_{X,t}^{\lambda} \cdot p_{M,t}^{1-\lambda}}{p_{N,t}}. \quad (12)$$

An increase in e_t means, ceteris paribus, a real depreciation of the home currency as internationally traded goods become relative more expensive.

Let $P_{N,t,t-1} \equiv p_{N,t}/p_{N,t-1}$; it can be shown that the following exact decomposition holds if the underlying GDP function has the Translog form:[15]

$$\Pi_{t,t-1} = P_{N,t,t-1} \cdot H_{t,t-1} \cdot E_{t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1}, \quad (13)$$

where

$$\begin{aligned} H_{t,t-1} &\equiv \left(\frac{h_t}{h_{t-1}}\right)^{(1-\lambda)\bar{s}_{X,t}+\lambda\bar{s}_{M,t}} \\ &= \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{(1-\lambda)\bar{s}_{X,t}+\lambda\bar{s}_{M,t}} \\ &\cdot \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{-(1-\lambda)\bar{s}_{X,t}-\lambda\bar{s}_{M,t}} \end{aligned} \quad (14)$$

measures the terms-of-trade effect, and

$$\begin{aligned} E_{t,t-1} &\equiv \left(\frac{e_t}{e_{t-1}}\right)^{(\bar{s}_{X,t}-\bar{s}_{M,t})} \\ &= \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{\lambda(\bar{s}_{X,t}-\bar{s}_{M,t})} \\ &\cdot \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{(1-\lambda)(\bar{s}_{X,t}-\bar{s}_{M,t})} \\ &\cdot \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{-(\bar{s}_{X,t}-\bar{s}_{M,t})} \end{aligned} \quad (15)$$

is the real-exchange-rate effect. Note that the welfare effect of a real depreciation of the home currency (an increase in e_t) depends on the position of the trade account as export revenues and the cost of imports both increase: the net effect is positive if the country is in a surplus position, negative otherwise.

Taken together, these two effects capture the complete trading gains as given by factor $G_{t,t-1}$:[16]

$$\begin{aligned} G_{t,t-1} &\equiv H_{t,t-1} \cdot E_{t,t-1} \\ &= \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{\bar{s}_{X,t}} \\ &\cdot \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{-\bar{s}_{M,t}} \\ &\cdot \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{-(\bar{s}_{X,t}-\bar{s}_{M,t})}. \end{aligned} \quad (16)$$

This approach only differs from the one of Kohli (2003, 2004a) discussed in the online Appendix by the decomposition of the trading gains between a terms-of-trade effect and a relative-price effect. The defining advantage of the approach encapsulated by (16) is that the residual, relative-price effect ($E_{t,t-1}$) has a clear economic interpretation, namely that it is a real-exchange-rate effect.

As recommended by the SNA, we can define real GDI (denoted by z_t with $Z_{t,t-1} \equiv z_t/z_{t-1}$) as real GDP augmented by the trading gains; in terms of growth factors:

15 For a proof, see Kohli (2006a, 2007).

16 Reinsdorf (2010) obtains a similar result in the context of the Fisher aggregation.

$$Z_{t,t-1} \equiv Y_{t,t-1} \cdot G_{t,t-1} = G_{t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1}. \quad (17)$$

Making use of (13), we then find that $p_{N,t}$ can be interpreted as the real GDI price deflator:[17]

$$\Pi_{t,t-1}/Z_{t,t-1} = \Pi_{t,t-1}/(G_{t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1}) = P_{N,t,t-1}. \quad (18)$$

This makes considerable sense since the ultimate objective of domestic income is precisely to purchase domestic goods, at price $p_{N,t}$. Moreover, this shows that it is just as easy to compute the full trading gains as the ratio of two price indices:

$$G_{t,t-1} = Z_{t,t-1}/Y_{t,t-1} = P_{Y,t,t-1}/P_{N,t,t-1}, \quad (19)$$

and real GDI can then be obtained simply by deflating nominal GDP (i.e. nominal GDI by the national accounts identity) by $p_{N,t}$.

The Modified Diewert and Morrison Effect

By comparing (7) with (16), the full meaning of the non-zero homogeneity of $DMA_{t,t-1}$ becomes clear: an equiproportionate change in the prices of imports and exports, other things equal, does have a real effect if trade is unbalanced, not because the terms of trade have changed (they have not), but because, by keeping domestic prices constant, it implies a change in the relative prices of traded and nontraded goods. In that case, $DMA_{t,t-1}$ measures a real-exchange-rate effect, not a terms-of-trade effect. $DMA_{t,t-1}$ can best be described as measuring the contribution of changes in import and export prices to the growth in nominal GDP (and nominal GDI). It is only when trade is balanced that $DMA_{t,t-1}$ gives an accurate measure of the terms-of-trade effect, and indeed of the trading gains, the real exchange-rate effect then being nil.

In later work, Diewert and Lawrence (2006) have rewritten expression (7) in terms of relative prices, in which case the zero homogeneity in prices is achieved. In doing so, they used consumption goods as the numeraire, but in order not to depart unnecessarily from the framework used so far, one can opt for the full set of domestic final purchases instead. The modified Diewert and Morrison term ($DMA'_{t,t-1}$) is therefore as follows:

$$DMA'_{t,t-1} \equiv \left(\frac{p_{X,t}/p_{N,t}}{p_{X,t-1}/p_{N,t-1}}\right)^{\bar{s}_{X,t}} \cdot \left(\frac{p_{M,t}/p_{N,t}}{p_{M,t-1}/p_{N,t-1}}\right)^{-\bar{s}_{M,t}} = \left(\frac{p_{X,t}}{p_{X,t-1}}\right)^{\bar{s}_{X,t}} \left(\frac{p_{M,t}}{p_{M,t-1}}\right)^{-\bar{s}_{M,t}} \cdot \left(\frac{p_{N,t}}{p_{N,t-1}}\right)^{-(\bar{s}_{X,t}-\bar{s}_{M,t})}. \quad (20)$$

Comparing (20) with (16), it appears immediately that $DMA'_{t,t-1} = G_{t,t-1}$, and hence $DMB'_{t,t-1}$, the price term accordingly adjusted, becomes equal to $P_{N,t,t-1}$. That is, the modified Diewert and Morrison term is not a measure of the terms-of-

17 Thus, if $P_{N,t,t-1}$ is computed as a Törnqvist price index, real GDI has the implicit Törnqvist form, just like real GDP; see footnote 11.

trade effect, but of the full trading gains instead. The modified Diewert and Morrison trading-gain measure can then easily be decomposed into a pure terms-of-trade effect and a real exchange-rate effect with the help of (14) and (15).

Note that if one had used the price of consumption as the numeraire as recommended by Diewert and Lawrence (2006), (20) could still be interpreted as a trading-gain index, but (15), the relative-price-effect, could no longer be viewed as a real-exchange-rate effect since changes in the prices of the other nontraded goods (investment and government purchases) would not be caught by (20): they would instead directly affect real income then defined as nominal GDP deflated by the price of consumption goods.

Trading Gains and Productivity

As suggested in the introduction, trading gains and productivity advances are of a similar breed since they both lead to increases in real income for given endowments of primary factors. Moreover, trading gains may affect the measurement of productivity, depending on the definition of productivity that is retained.

One favoured measure of productivity has already been referred to, namely total factor productivity (TFP) as captured by Törnqvist index $R_{t,t-1}$. Identifying the trading gains and adding them to real GDP to obtain real GDI has no impact on the measures of nominal and real GDP. Changes in the prices of exports, imports, and domestic goods are already fully taken into account when computing nominal GDP and its price. Expression (6) remains valid and the measure of TFP is therefore unaffected. For a given change in the endowment of domestic factors as given by $X_{t,t-1}$ if properly measured, $R_{t,t-1}$ is fully determined and thus independent of $H_{t,t-1}$ and $E_{t,t-1}$.[18] The trading gains simply are a benefit in addition to increases in TFP.

More generally, it is noteworthy that if the Törnqvist aggregation is exact for the underlying function, and assuming perfect competition and optimization, a change in any output price, holding technology and factor endowments constant, has no impact on real GDP since it has exactly the same relative effect on nominal GDP and on its price. Put in another way, using a language familiar to trade economists, a change in output (including import) prices will lead to a movement along the production possibilities frontier, but real GDP, adequately measured, is constant along that line.[19] This is not to say that, for given factor endowments and a given technology, a change in the terms of trade or the real exchange rate cannot affect total factor productivity. Quite the contrary: a change in h_t or e_t is likely to have an impact on relative factor rental prices and hence on their income shares, thereby affecting the measure of $X_{t,t-1}$, and, by the same token, the measure of $R_{t,t-1}$ obtained as a resid-

18 Kehoe and Ruhl (2008) reach the same conclusion with a set of different models.

19 Technically speaking, it will be a surface in a three-dimensional space rather than just a line since we are considering three variable quantities.

ual, real GDP remaining unchanged. This, however, is a matter of economic analysis, not an accounting issue. At any point in time, for a given set of output prices, factor endowments and technology, the measure of $R_{t,t-1}$ is independent of whether or not the trading gains have actually been measured and taken into account.

We next consider a second, very common measure of productivity: the average productivity of labour, i.e. the real value added per unit of labour. In fact, we will consider two such measures, one with respect to real GDP and the other with respect to real GDI. We do, however, have a strong preference for the latter given that international trade takes place overwhelmingly in middle products, and thus occurs during the production process rather than afterwards. As such, we view it as problematic to treat trading gains as an afterthought. The singling out of labour might also need a justification. One might think that there is no reason to impute the trading gains to domestic labour since they were obtained from abroad. However, the same could be said when it comes to the production gains resulting from the availability of more advanced equipment, perhaps even imported from abroad. In both cases, though, labour is involved in some way, and it is a convenient shortcut to relate the overall performance of the economy to the work effort: labour is then used as a metric so to speak.[20]

Nonetheless, it might seem a bit far-fetched to include the trading gains in any measure of productivity. One could assert that productivity is a concept intimately linked to the production process and thus to GDP, whereas trading gains are more of an income concept. An improvement in the terms of trade, for example a drop in the price of oil in the case of an oil-importing nation may be the result of pure luck: domestic production factors are without merit in this development and should not be able to claim an improvement in productivity, even though on average their real income will unambiguously increase. This is quite true, but similar situations can occur in a closed economy. Exceptionally poor weather can have a detrimental effect on agricultural production, and hence on measured average labour productivity, without any fault of the farmers who may have been just as hard working as ever. As for the drop in oil prices, it could also result from the completion of a new trans-border pipeline that gives access to a cheaper foreign supplier. As such, it would be difficult to argue that this improvement in the terms of trade is not related to production activities at home and abroad.

Terms-of-trade movements are often viewed as being temporary and likely to self-correct over time, whereas productivity gains due to improving technology are unlikely to be reversed. Admittedly, resource-exporting countries often face volatile terms of trade. Nonetheless, the price cycles may extend over many

20 The measurement of productivity by the real value added per unit of labour is nonetheless often criticized, precisely because it focuses exclusively on one factor of production, namely labour. The wide acceptance of this concept probably has to do in parts because of its early adoption by the Organisation for European Economic Co-operation (OEEC, the ancestor of the OECD) in 1949 under the influence of Jean Fourastié; see Boulat (2006:97).

years and the reversal to mean is not guaranteed. Exporters of industrialized goods are probably less exposed to such volatility. As we shall see below, Switzerland's terms of trade have trended mostly upwards over a 50-year period, while the real exchange rate trended downwards, revealing a steady real appreciation of the currency. Besides, closed economies are not immune either to random, temporary productivity shocks that may be caused by weather, health, social, or political disturbances: standard productivity measures therefore also have to contend with this type of volatility.

More generally, better terms of trade can be the result of a research activity (e.g. market prospection) or of a marketing effort. In a globalized world, firms are constantly searching for new suppliers and additional customers abroad. To the extent that significant quantities of domestic labour and capital are diverted from domestic production to such activities, average labour productivity (and TFP) could be underestimated. Improvement in the terms of trade could also reflect a refinement in the quality of exports that is not fully reflected by the export price and quantity indices. This could also lead to an underestimation of real GDP per unit of labour. Taking the trading gains into account might help to correct for these types of biases.

Better terms of trade can also result from technological advances made abroad. In that sense, the home country may appear to be free-riding on an effort made elsewhere. Note that such a technological advance could also have been made by the foreign subsidiary of a domestic firm and thus have been initiated in the home country. In any case, there is little doubt that globalization and international trade have led to massive transfers of technology and have favoured the international dissemination of productivity gains. For instance, countries throughout the world have greatly benefited from being able to import better and better hi-tech products manufactured in only a handful of countries at ever-lower prices. It is a two-way street, though: while the home country can largely benefit from technological advances made abroad, the rest of the world can also take advantage of the progress made at home.

As already stressed, almost all trade takes place during production, rather than after. In our view the "trade technology," which "transforms" exports into imports, should therefore be treated as an essential element of the country's all-embracing technology. Whether components are transformed into others through a physical process, a chemical reaction, or trade, at home or abroad, should not really matter much to economists. Because it may be difficult in many situations to clearly label what is capital deepening, what is technological progress, what is human capital enhancement, and what are pure trading gains, the line between these concepts tends to be blurred in an integrated world. Given the risk that as a result of measurement errors one development may be wrongly imputed to one or another growth factor speaks in favour of considering all of them jointly. Moreover, the reason why economists are interested in productivity is ultimately that it is income enhancing, and it therefore makes sense to take account of all sources of gains, whether domestic or foreign.

Nonetheless, as mentioned earlier, we will also consider the average labour productivity relative to GDP in what follows; as we shall see, the difference between this "closed-economy" measure and the "open-economy" measure we favour is fully accounted for by the trading gains.

We thus begin by defining $a_{Z,t} \equiv z_t / x_{L,t}$ as real GDI per unit of labour, or, in terms of growth factors:

$$A_{Z,t,t-1} \equiv \frac{Z_{t,t-1}}{X_{L,t,t-1}}, \tag{21}$$

with $A_{Z,t,t-1} \equiv a_{Z,t}/a_{Z,t-1}$ and $X_{L,t,t-1} \equiv x_{L,t}/x_{L,t-1}$. It follows from (17) that this can be expressed as:

$$A_{Z,t,t-1} \equiv (G_{t,t-1} \cdot X_{t,t-1} \cdot R_{t,t-1}) \cdot X_{L,t,t-1}^{-1}. \tag{22}$$

Making use of (4), we find that:

$$\begin{aligned} X_{t,t-1} \cdot X_{L,t,t-1}^{-1} &= \left(\frac{x_{K,t}}{x_{L,t}}\right)^{\bar{\sigma}_{K,t}} \left(\frac{x_{L,t}}{x_{L,t-1}}\right)^{\bar{\sigma}_{L,t}-1} \\ &= \left(\frac{x_{K,t}/x_{L,t}}{x_{K,t-1}/x_{L,t-1}}\right)^{\bar{\sigma}_{K,t}} \\ &= \left(\frac{k_t}{k_{t-1}}\right)^{\bar{\sigma}_{K,t}} \\ &\equiv \; K_{t,t-1}, \end{aligned} \tag{23}$$

with $k_t \equiv x_{K,t}/x_{L,t}$ the capital/labour ratio, and $K_{t,t-1}$ the contribution of capital-intensity changes to economic growth. We thus obtain the following complete Törnqvist decomposition of the growth in this "globalized" version of domestic average labour productivity:

$$\begin{aligned} A_{Z,t,t-1} &= G_{t,t-1} \cdot K_{t,t-1} \cdot R_{t,t-1} \\ &= H_{t,t-1} \cdot E_{t,t-1} \cdot K_{t,t-1} \cdot R_{t,t-1}. \end{aligned} \tag{24}$$

This decomposition is exact if the underlying real GDI function is indeed Translog. Admittedly the last two components are likely to dominate the terms-of-trade and the real-exchange-rate effects, but, in our opinion, the trading gains need nonetheless to be considered to obtain a complete assessment of the change in average labour productivity in the open economy.

Note that it follows from (6) and (23) that the product of the last two components yields the growth in the average labour productivity defined with respect to real GDP, $A_{Y,t,t-1}$, or put another way, the average productivity of labour in a closed-economy setting (Kohli, 2005b):

$$\begin{aligned} A_{Y,t,t-1} &\equiv \frac{Y_{t,t-1}}{X_{L,t,t-1}} \\ &= X_{t,t-1} \cdot R_{t,t-1} \cdot X_{L,t,t-1}^{-1} \\ &= K_{t,t-1} \cdot R_{t,t-1}. \end{aligned} \tag{25}$$

Thus, the only difference between this measure and the one we recommend is the exclusion here of the trading gains.

Yet another important indicator of productivity is the marginal product of labour. As far as workers are concerned, their marginal product is undoubtedly of more interest to them than their average product since the former is directly related to their purchasing power. In the Cobb-

Douglas case, the marginal product of labour is proportional to its average product, but this is generally not true in the case of higher-order functional forms such as the Translog. Under perfect competition and optimization, the marginal product of labour can readily be observed as the real wage rate, $u_{L,t} \equiv w_{L,t}/p_{N,t}$, i.e. the nominal wage deflated by the price of domestic final goods, the GDI price deflator. Note that the nominal wage is an income concept and it therefore would make little sense to use the price of GDP as given by (3) to deflate nominal wages. Domestic residents buy domestic final goods, they do not purchase imports or exports. Thus, in view of (19), the trading gains are automatically taken into account in the definition of the real wage, and the question of whether or not the trading gains should be included in this indicator of productivity is a non-issue.

Recall now that $\sigma_{L,t} \equiv (x_{L,t} w_{L,t})/\pi_t = (x_{L,t} w_{L,t})/(z_t p_{N,t})$; it therefore follows that $u_{L,t} = a_{Z,t} \cdot \sigma_{L,t}$ or, in terms of growth factors:

$$U_{L,t,t-1} = A_{Z,t,t-1} \cdot S_{L,t,t-1}, \qquad (26)$$

where

$$U_{L,t,t-1} \equiv u_{L,t}/u_{L,t-1} \qquad (27)$$

and

$$S_{L,t,t-1} \equiv \sigma_{L,t}/\sigma_{L,t-1}. \qquad (28)$$

Together with (25), this enables us to obtain a complete decomposition of the growth of the marginal product of labour:

$$U_{L,t,t-1} = S_{L,t,t-1} \cdot H_{t,t-1} \cdot E_{t,t-1} \cdot K_{t,t-1} \cdot R_{t,t-1}. \qquad (29)$$

This expression is very handy since each one of its terms can be measured with observed data exclusively. It also shows that, although TFP and capital deepening are almost certainly the main drivers of the growth in the marginal productivity of labour, terms-of-trade and real-exchange-rate effects again cannot be ignored for the decomposition to be complete.

Decomposition (29) is essentially an accounting identity that should hold at any point in time for a given set of output prices, factor endowments, and technology. It is silent, however, as to the economic forces that cause the changes that are being measured. One must recall that all the components of (29) are endogenous to the extent that they all depend on input and output shares. This is of course most obvious for $S_{L,t,t-1}$, unless the underlying technology is Cobb-Douglas, in which case $S_{L,t,t-1} = 1$. The question of how the ratio of the marginal to the average product of labour would change as the result of hypothetical changes in the terms of trade, the real exchange rate, relative factor endowments, and technological progress is an empirical issue, which cannot be answered without a detailed knowledge of the form of the underlying technology. One key parameter is the Hicksian elasticity of complementarity between labour and capital (ψ_{KL}).[21] If ψ_{KL} is greater than one,

21 In the two-input case, the Hicksian elasticity of complementarity is the inverse of the Allen-Uzawa elasticity of substitution.

an increase in the capital-labour ratio will lead to an increase in the labour share, thus meaning that an increase in capital intensity will raise the marginal product of labour by more than its average product. On the other hand, if technological change is mostly Harrod neutral (i.e. labour-augmenting), the passage of time will tend to have an offsetting effect by reducing the labour share for $\psi_{KL} > 1$. Furthermore, although trading gains lead to increases in real domestic income, it is not certain that both factors of production will benefit equally, if at all. It might indeed be the case that one of the two factors becomes worse off — even though the country as a whole is unambiguously better off — if its own income share decreases sufficiently.[22] The sign and the size of the impact of changes in the terms of trade and the real exchange rate on the marginal product of labour depend on the so-called Stolper-Samuelson elasticities, which, in turn, are functions of the parameters of the underlying technology (Kohli, 2010).

Numerical results for Switzerland

Switzerland has, at times, enjoyed very strong improvements in its terms of trade. Given the relatively large size of its foreign trade sector, one would expect this development to have made a significant positive contribution to real GDI. At the same time, Switzerland has experienced a strong real appreciation of its currency and a large trade surplus, which, put together, suggest a negative real-exchange-rate effect. Its total trading gains — or losses — are therefore likely to be nontrivial and it thus seems of interest to have a look at the Swiss data.[23]

Chart 1 shows the path of the Swiss terms of trade (h_t). As it can be seen, they have improved significantly, particularly during the 1980s and 1990s, peaking at nearly 25 per cent by 2003 and falling back somewhat, to 18 per cent by the end of the sample period. The real exchange rate (e_t), on the other hand, fell almost continuously for the first three decades, to reach a level of about 46 per cent below its initial level by 2003, thus revealing a very substantial appreciation, as the relative price of internationally traded goods decreased. The trade balance index, finally, defined here as the ratio of nominal exports to nominal imports, increased steadily starting in the 1980s, thus indicating a growing trade surplus.

We report in Table 1,[24] first column, the Diewert and Morrison terms-of-trade effect, DMA_t, as given by (7), but chained over the entire sample period.[25] Next to

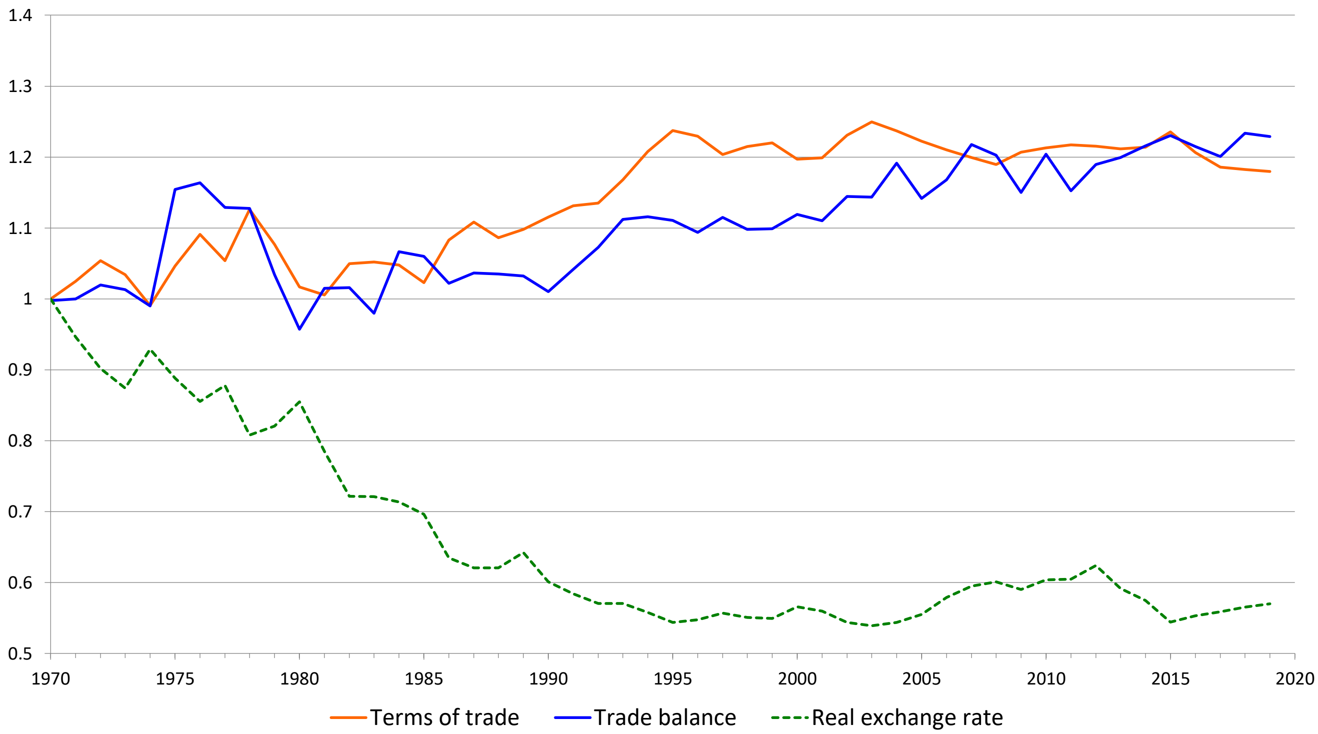
22 This is of course the rule in the well-known two-sector Heckscher-Ohlin-Samuelson model of international trade as the result of the implicit, restrictive non-joint-production hypothesis; see Kohli (1991).

23 These are annual for the period 1970-2019. The output data are taken directly from the OECD data base; the prices and quantities of labour and capital services are derived from the Swiss National Bank and Swiss Federal Statistical Office data bases.

24 See page 16 of this article.

25 Formally, $DMA_t \equiv DMA_{t,t-1} \cdot DMA_{t-1,t-2} \cdot ... \cdot DMA_{1,0} \cdot DMA_0$ with $DMA_0 = 1$, and similarly for the other growth factors.

Chart 1: Terms of Trade, Real Exchange Rate, and Trade Balance, Switzerland, 1970-2019 (1970 = 1.0)



it we report the values of the pure terms-of-trade effect (H_t), and the corresponding real-exchange-rate effect (E_t).[26] In the fourth column, one finds the chained values of the complete trading-gain factor (G_t). For comparison purposes, we also report in columns 5 and 6 estimates of an alternative decomposition of the trading gains briefly discussed in the on-line Appendix, namely the terms-of-trade effect $H_{X,t}$ and the related relative price effect $E_{X,t}$. The corresponding yearly geometric means are reported at the bottom of the table.

Looking first at the values of the terms-of-trade effects, one notes that DMA_t and H_t are fairly well correlated, and they closely reflect the evolution of the terms of trade, weighted by the import and export GDP shares. By 2019, the cumulated effect is somewhat larger for index DMA_t, at nearly 6.5 per cent of real GDP, and just short of 6 per cent for H_t. Chart 2 shows the path of these two measures over the sample period. The deviations between the two indices are largest between 2003 and 2015, a period during which the terms of trade were pretty steady, but with the trade imbalance increasing almost continuously. This is when the non-zero price homogeneity of DMA_t comes into play. H_t and $H_{X,t}$, on the other hand, are highly correlated throughout the sample period.

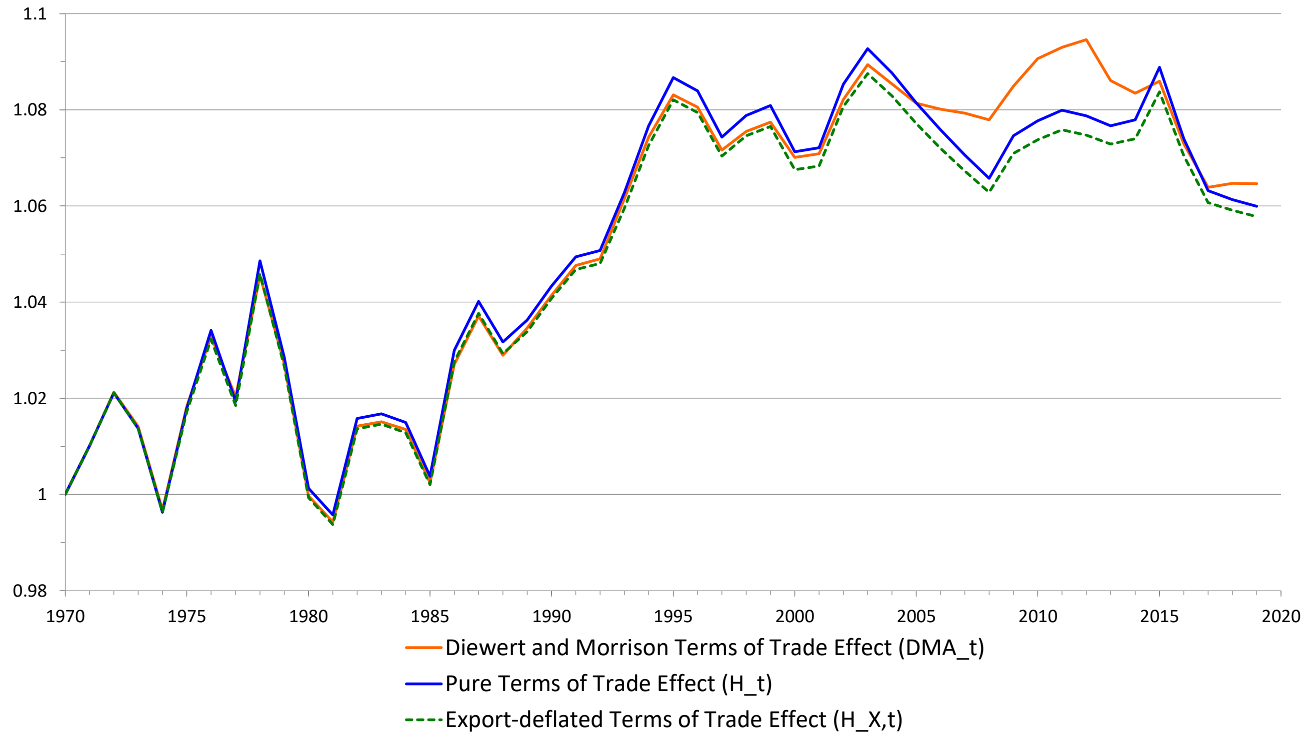
26 For this purpose λ was set to the sample-mean value of s_X, namely 0.5248.

Table 1: Alternative Measures of the Terms-of-Trade, Real-Exchange-Rate, and Trading-Gain Effects, Switzerland, 1970-2019

Year	DMA_t (1)	H_t (2)	E_t (3)	G_t (4)	$H_{X,t}$ (5)	$E_{X,t}$ (6)
1970	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1971	1.0101	1.0101	1.0000	1.0102	1.0101	1.0000
1972	1.0213	1.0212	0.9999	1.0210	1.0211	0.9999
1973	1.0142	1.0137	0.9997	1.0134	1.0137	0.9997
1974	0.9968	0.9963	0.9997	0.9960	0.9963	0.9997
1975	1.0183	1.0179	0.9986	1.0164	1.0172	0.9993
1976	1.0329	1.0342	0.9965	1.0306	1.0323	0.9984
1977	1.0203	1.0193	0.9980	1.0173	1.0185	0.9988
1978	1.0456	1.0486	0.9937	1.0420	1.0459	0.9963
1979	1.0272	1.0285	0.9943	1.0226	1.0267	0.9961
1980	0.9997	1.0013	0.9941	0.9954	0.9993	0.9960
1981	0.9943	0.9957	0.9947	0.9905	0.9938	0.9967
1982	1.0142	1.0158	0.9941	1.0099	1.0137	0.9963
1983	1.0151	1.0168	0.9941	1.0108	1.0146	0.9963
1984	1.0135	1.0150	0.9940	1.0089	1.0128	0.9961
1985	1.0024	1.0038	0.9934	0.9971	1.0020	0.9951
1986	1.0271	1.0300	0.9917	1.0214	1.0276	0.9940
1987	1.0371	1.0402	0.9914	1.0313	1.0377	0.9938
1988	1.0290	1.0317	0.9914	1.0228	1.0293	0.9937
1989	1.0346	1.0363	0.9919	1.0279	1.0339	0.9942
1990	1.0414	1.0433	0.9913	1.0342	1.0408	0.9937
1991	1.0476	1.0494	0.9910	1.0400	1.0468	0.9935
1992	1.0490	1.0507	0.9905	1.0407	1.0481	0.9930
1993	1.0615	1.0628	0.9905	1.0527	1.0595	0.9935
1994	1.0743	1.0767	0.9896	1.0654	1.0726	0.9933
1995	1.0831	1.0867	0.9885	1.0743	1.0821	0.9928
1996	1.0806	1.0840	0.9888	1.0718	1.0795	0.9929
1997	1.0716	1.0743	0.9895	1.0630	1.0704	0.9932
1998	1.0755	1.0789	0.9890	1.0670	1.0746	0.9929
1999	1.0775	1.0809	0.9889	1.0689	1.0765	0.9929
2000	1.0701	1.0713	0.9903	1.0609	1.0675	0.9938
2001	1.0709	1.0721	0.9898	1.0611	1.0683	0.9933
2002	1.0821	1.0853	0.9881	1.0725	1.0807	0.9924
2003	1.0894	1.0927	0.9876	1.0792	1.0875	0.9924
2004	1.0854	1.0877	0.9883	1.0749	1.0829	0.9926
2005	1.0814	1.0815	0.9897	1.0703	1.0772	0.9937
2006	1.0801	1.0759	0.9928	1.0681	1.0720	0.9964
2007	1.0793	1.0706	0.9954	1.0657	1.0672	0.9985
2008	1.0779	1.0657	0.9965	1.0620	1.0628	0.9992
2009	1.0849	1.0746	0.9948	1.0690	1.0710	0.9982
2010	1.0906	1.0777	0.9970	1.0745	1.0738	1.0006
2011	1.0930	1.0799	0.9971	1.0768	1.0758	1.0009
2012	1.0946	1.0788	1.0002	1.0790	1.0748	1.0039
2013	1.0861	1.0767	0.9942	1.0705	1.0729	0.9978
2014	1.0834	1.0779	0.9909	1.0681	1.0740	0.9945
2015	1.0860	1.0889	0.9847	1.0722	1.0838	0.9893
2016	1.0729	1.0740	0.9866	1.0596	1.0705	0.9899
2017	1.0639	1.0632	0.9877	1.0501	1.0607	0.9900
2018	1.0647	1.0614	0.9890	1.0497	1.0591	0.9911
2019	1.0647	1.0599	0.9901	1.0494	1.0578	0.9921
Mean (1970-2019)	1.0013	1.0012	0.9998	1.00010	1.0012	0.9998

Note:
DMA_t: Diewert and Morrison terms-of-trade effect (equation 7)
H_t: Terms-of-trade effect holding $e_t = p_{T,t}/p_{N,t}$ constant (equation 14)
E_t: Real-exchange-rate effect (equation 15)
G_t: Trading gains (equation 16)
$H_{X,t}$: Terms-of-trade effect holding $p_{X,t}/p_{N,t}$ constant (equation A1)
$E_{X,t}$: Relative-price effect (equation A2)
Note that: $G_t = H_t \cdot E_t = H_{X,t} \cdot E_{X,t}$ by (16) and (A3).
Values presented in the bottom row are geometric means.

Chart 2: Alternative Measures of the Terms-of-Trade Effects, Switzerland, 1970-2019 ***(as factors of real GDP)*** **(1970 = 1.0)**

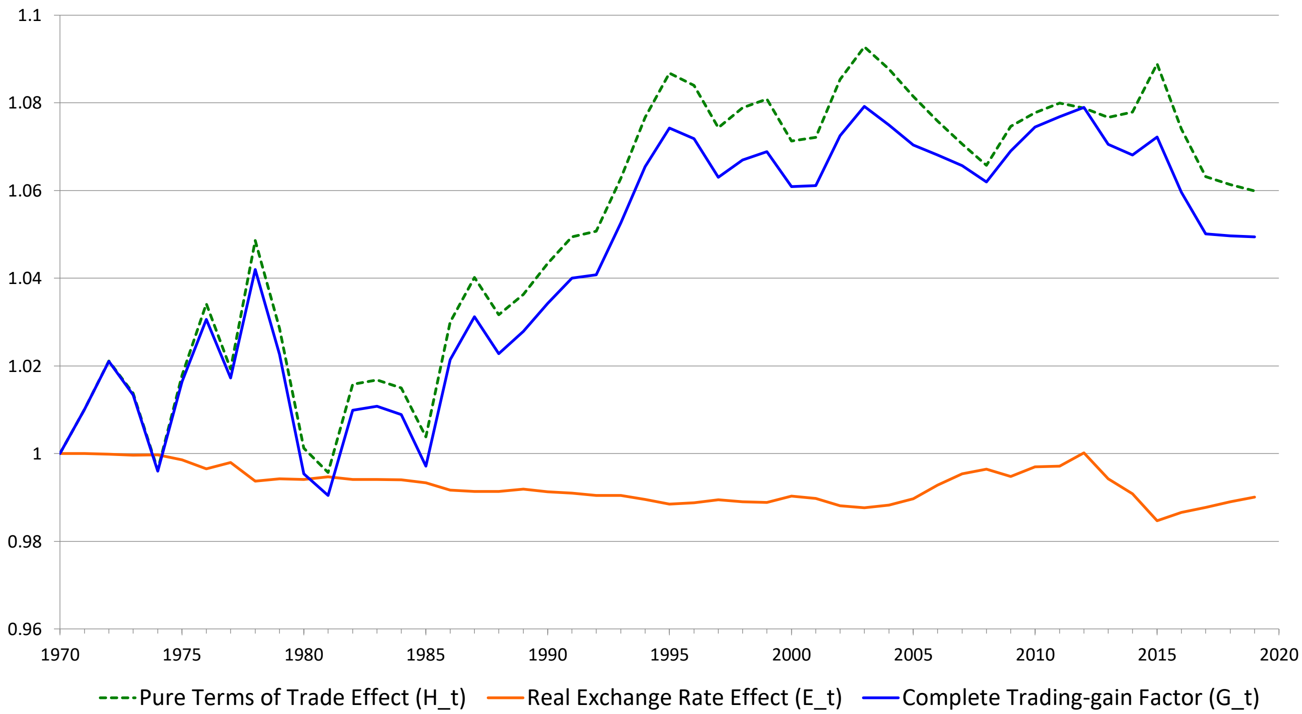


Returning to Table 1, one sees that, by the end of the sample period the cumulated real-exchange-rate effect (E_t) is negative, at about −1.0 per cent. This negative income effect is due to the conjunction of a declining real exchange rate (i.e. a real appreciation) and of a positive trade balance, on average. A similar picture emerges if one considers the relative-price effect $E_{X,t}$, which refers to the price of exports relative to the price of nontraded goods. The total trading gains, finally, show a gain of close to 5 per cent of GDP by 2019. In terms of 2019 prices, this amounts to nearly 36 billion Swiss francs. Admittedly, this is the result of trading gains chained over a 50-year period. Nonetheless, the amount is sizable, particularly if one cumulated the yearly real gains over this period, year after year (using an appropriate real interest rate), and, moreover, considering that for the average country the trading gains must be nil![27]

The trading-gain index, together with our preferred measures of the terms-of-trade and real-exchange-rate components, is depicted in Chart 3. The long-run trends are clearly visible, and so are the shorter-run fluctuations. Looking in more details at the changes through time, the 1985-2005 period stands out. This is when the terms-of-trade effect increased substantially and almost continuously, adding as much as 1.2 percentage points to economic

27 For a sample of 24 OECD member countries covering the period 1970-2012, Switzerland ranked third (behind Australia and Norway) for the relative size of its 2012 trading gains; in terms of the capitalized sum over the entire period, using a 1 per cent real rate of interest, Switzerland came up first with a gain amounting to 168 per cent of its 2012 GDP; see Kohli (2014).

Chart 3: Terms-of-Trade, Real-Exchange-Rate, and Trading-Gain Effects, Switzerland, 1970-2019 ***(as factors of real GDP)*** **(1970 = 1.0)**



growth in 2002, and peaking at over 9.3 per cent of GDP a year later. This effect was somewhat dampened by the simultaneous, negative real-exchange-rate effect, but the trading-gains index still reached a high of nearly 8 per cent in 2003, and remained above the 6 per cent mark until 2015.

It is ironic that it is precisely at the beginning at the new millennium, when the trading gains were reaching new highs, that a sense of growth pessimism became prevalent among Swiss economic actors and observers, coming to a climax at a March 2005 conference held in Zurich and organized by the *Avenir Suisse* think tank.[28] The OECD had just published a report widely interpreted as indicating that Ireland had overtaken Switzerland in terms of real income per capita (Wyplosz, 2005), whereas the data referred in fact to real GDP at purchasing-power-parity exchange rates.[29] Once this confusion exposed, it became apparent that the Swiss economic performance was only half as bad as it looked, and that what could be called the Swiss *growth paradox* — which has Switzerland growing less rapidly than most other countries, and yet always remaining among the front runners in terms of real income per capita — could be explained in parts by the official and public fixation on real GDP, as opposed to real GDI and, even

28 I am grateful to Gerhard Schwarz, the then Director of Avenir Suisse, for having invited me to this conference and for his continuous support.

29 Even though the title of the OECD press release indicated that the comparison pertained to GDP figures, the OECD itself referred to income after just two paragraphs.

more importantly, real Gross National Income (GNI).[30]

We report in Table 2, first column, our estimates of TFP (R_t), as given by (5), cumulated over the entire period. The index, set to unity in 1970, reaches a level of 1.564 by the end of the sample period; this implies an annual average contribution to growth of about 0.92 percentage points. In the second column we report our estimates of the contribution of capital deepening, K_t, as given by (23): its contribution over the sample period is nearly as important as that of TFP, with a 2019 estimate of 1.482 (about 0.81 per cent per year on average). Column 3 shows the estimate of $A_{Y,t}$, the average labour productivity defined with respect to real GDP as given by (25), reaching a level of 2.318 by 2019. The next contributing factor is made up by the trading gains, G_t, as discussed above and reported once again in the fourth column of the table for the sake of completeness. Together, R_t, K_t, and G_t explain the growth in the complete, "globalized," measure of average labour productivity ($A_{Z,t}$), shown in the fifth column. It is found to have well more than doubled over five decades, averaging an annual growth rate of 1.83 per cent. The sixth column of the table shows the value of $S_{L,t}$: it increased by about one tenth over the course of the last half-century. The last column, finally, documents the growth in the real marginal product of labour (U_t). It reached a level of 2.672 by 2019 (just over 2 per cent annually), and it thus exceeded the growth of the average labour productivity $A_{Z,t}$ measure by about 0.2 percentage points per year.

Our results are summarized graphically by Chart 4. Starting at the bottom of the graph, we first show the path of TFP (R_t); this line is next augmented by the path of the capital-deepening contributing factor (K_t) to obtain the path of average labour productivity in terms of real GDP ($A_{Y,t}$); next we have added the contribution of the trading gains to obtain the path of average labour productivity in terms of real GDI ($A_{Z,t}$); finally, multiplying by the labour share index ($S_{L,t}$), we get the path of the real wage rate, interpreted as the marginal product of labour. It is quite clear that the two main engines of growth of the Swiss economy are the increases in TFP and capital deepening. As expected, the contribution of the trading gains is much smaller, although not insignificant. Thus, in the Swiss case, trading gains have contributed close to 0.1 per cent annually to the growth in real wages. In any case, good accounting practices require that this component not be overlooked.

Concluding Comments

As shown above, trading gains are important not just for the measurement of real GDI and the determination of aggregate demand, but also for some measures of productivity when defined in a broad context. We have argued that both the measurement of the average and of the marginal productivity of labour should take trading gains into account since al-

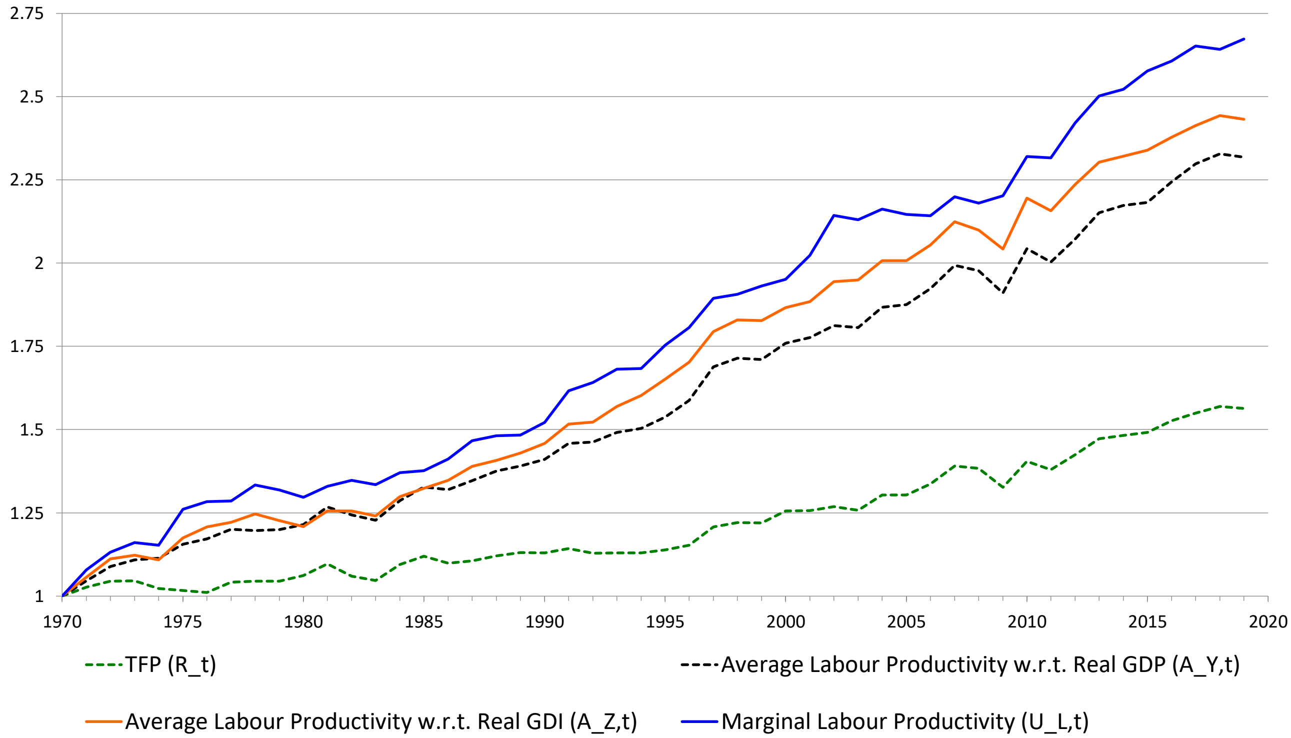
30 See Kohli (2005a); the reactions were virulent, as shown by the following headlines "Krach in der Nationalbank um die Wachstumspolitk" (*Sonntags Zeitung*, March 6, 2005) and "Swiss Pour Scorn on Ireland's Fourth Place in World's Wealthy Elite" (*Irish Times*, March 6, 2005).

Table 2: Alternative Measures of Productivity, Switzerland, 1970-2019

Year	R_t (1)	K_t (2)	$A_{Y,t}$ (3)	G_t (4)	$A_{Z,t}$ (5)	$S_{L,t}$ (6)	$U_{L,t}$ (7)
1970	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1971	1.0266	1.0185	1.0456	1.0102	1.0562	1.0208	1.0782
1972	1.0448	1.0416	1.0883	1.0210	1.1112	1.0187	1.1320
1973	1.0453	1.0602	1.1083	1.0134	1.1231	1.0335	1.1607
1974	1.0228	1.0885	1.1133	0.9960	1.1088	1.0394	1.1525
1975	1.0169	1.1361	1.1553	1.0164	1.1743	1.0731	1.2601
1976	1.0107	1.1594	1.1718	1.0306	1.2077	1.0625	1.2831
1977	1.0418	1.1522	1.2004	1.0173	1.2211	1.0525	1.2852
1978	1.0446	1.1457	1.1968	1.0420	1.2471	1.0695	1.3337
1979	1.0448	1.1482	1.1996	1.0226	1.2267	1.0747	1.3184
1980	1.0620	1.1437	1.2146	0.9954	1.2090	1.0727	1.2968
1981	1.0968	1.1555	1.2673	0.9905	1.2552	1.0594	1.3298
1982	1.0600	1.1731	1.2435	1.0099	1.2558	1.0727	1.3470
1983	1.0464	1.1730	1.2274	1.0108	1.2407	1.0755	1.3344
1984	1.0946	1.1758	1.2870	1.0089	1.2985	1.0552	1.3702
1985	1.1194	1.1860	1.3276	0.9971	1.3238	1.0402	1.3770
1986	1.0989	1.2006	1.3193	1.0214	1.3476	1.0472	1.4111
1987	1.1055	1.2184	1.3469	1.0312	1.3890	1.0557	1.4663
1988	1.1210	1.2274	1.3758	1.0228	1.4072	1.0525	1.4811
1989	1.1302	1.2303	1.3905	1.0279	1.4293	1.0376	1.4830
1990	1.1294	1.2488	1.4104	1.0342	1.4587	1.0428	1.5212
1991	1.1426	1.2763	1.4582	1.0400	1.5165	1.0656	1.6159
1992	1.1291	1.2956	1.4629	1.0407	1.5225	1.0779	1.6411
1993	1.1295	1.3203	1.4913	1.0526	1.5698	1.0712	1.6816
1994	1.1300	1.3308	1.5038	1.0654	1.6021	1.0509	1.6837
1995	1.1382	1.3503	1.5370	1.0743	1.6511	1.0619	1.7533
1996	1.1529	1.3773	1.5879	1.0718	1.7018	1.0612	1.8060
1997	1.2075	1.3980	1.6882	1.0630	1.7946	1.0552	1.8937
1998	1.2201	1.4050	1.7143	1.0670	1.8291	1.0422	1.9062
1999	1.2198	1.4017	1.7098	1.0689	1.8275	1.0568	1.9313
2000	1.2556	1.4011	1.7591	1.0609	1.8663	1.0454	1.9510
2001	1.2567	1.4131	1.7758	1.0611	1.8844	1.0734	2.0227
2002	1.2684	1.4291	1.8126	1.0725	1.9439	1.1024	2.1430
2003	1.2579	1.4358	1.8061	1.0792	1.9492	1.0926	2.1297
2004	1.3036	1.4324	1.8672	1.0749	2.0070	1.0770	2.1615
2005	1.3035	1.4387	1.8753	1.0703	2.0073	1.0689	2.1456
2006	1.3368	1.4384	1.9230	1.0681	2.0539	1.0428	2.1417
2007	1.3903	1.4336	1.9932	1.0656	2.1240	1.0354	2.1993
2008	1.3832	1.4291	1.9768	1.0620	2.0993	1.0387	2.1805
2009	1.3264	1.4400	1.9099	1.0690	2.0418	1.0783	2.2016
2010	1.4045	1.4544	2.0427	1.0745	2.1947	1.0571	2.3201
2011	1.3792	1.4526	2.0035	1.0768	2.1574	1.0734	2.3157
2012	1.4249	1.4543	2.0723	1.0790	2.2360	1.0825	2.4205
2013	1.4725	1.4606	2.1508	1.0705	2.3024	1.0867	2.5020
2014	1.4824	1.4656	2.1726	1.0681	2.3205	1.0865	2.5213
2015	1.4912	1.4631	2.1817	1.0722	2.3392	1.1015	2.5767
2016	1.5262	1.4703	2.2439	1.0596	2.3776	1.0963	2.6066
2017	1.5496	1.4829	2.2980	1.0501	2.4132	1.0987	2.6513
2018	1.5695	1.4830	2.3276	1.0497	2.4432	1.0812	2.6417
2019	1.5637	1.4822	2.3177	1.0494	2.4322	1.0987	2.6722
Mean (1970-2019)	1.0092	1.0081	1.0173	1.0010	1.0183	1.0019	1.0203

Note:
R_t: Total Factor Productivity (equation 5)
K_t: Capital intensity (equation 23)
$A_{Y,t}$: Average labour productivity w.r.t. real GDP (equation 25)
G_t: Trading gains (equation 16)
$A_{Z,t}$: Average labour productivity w.r.t. real GDI (equation 21)
$S_{L,t}$: Labour share factor (equation 28)
$U_{L,t}$: Real wage (equation 27)
Note that: $A_{Y,t} = R_t \cdot K_t, A_{Z,t} = R_t \cdot K_t \cdot G_t$, and $U_{L,t} = R_t \cdot K_t \cdot G_t \cdot S_{L,t} = A_{Z,t} \cdot S_{L,t}$ by (16), (24), (25), (26), and (29).
Values presented in the bottom row are geometric means.

Chart 4: Decomposition of the Marginal Productivity of Labour, Switzerland, 1970-2019 (1970 = 1.0)



most all trade takes place during — rather than after — production. Some domestic labour is involved in almost all transactions with the rest of the world, and international trade is an intimate part of production in a globalized world. The distinction between capital deepening, technological progress, human capital enhancement, and trading gains can be blurred. Some advances could be wrongly attributed to one growth factor rather than to another. This calls for an all-encompassing approach where all income-augmenting forces are considered jointly. In fact, when it comes to the marginal productivity of labour, defining the real wage in terms of anything but the purchasing power of domestic income would make little sense.

It is disappointing that the IMF, the OECD, EuroStat, and the United Nations, among others, do not have the resolution to make explicit recommendations concerning the appropriate trade-balance deflator, basically leaving member countries in the dark as to what the best practices are.[31] Thus, it is up to them whether they want to use $p_{M,t}, p_{X,t}, p_{A,t}, p_{N,t}$, or yet another price index, as a deflator. Moreover, unless trade happens to be balanced, all the so-called measures of the trading gains using a deflator other than $p_{N,t}$ are incomplete since they exclude the relative-price effect resulting from a change in the price of the chosen trade-account deflator relative to the price of domestic final goods. This is why additional components such as

31 Admittedly, the IMF has been advocating the use of the price of gross domestic final expenditures to compute real GDI in some of its own policy work; see Reinsdorf (2020, paragraphs 30, 87).

the real exchange rate effect, $E_{t,t-1}$, are needed. Thus, these official measures are misnamed: they should be viewed at best as measures of the terms-of-trade effects, rather than of the full trading gains. Consequently, the corresponding real GDI estimates must be considered as flawed.

It would appear that most statistical agencies have it backwards. They select a deflator, more or less at random, receiving no strict guidance from the SNA. They then very carefully calculate the (incomplete) trading gain, add it to their estimate of real GDP, and declare it to be real GDI. The implicit GDI deflator is then almost meaningless since it will generally be a function of the prices of imports and/or exports, incorrectly suggesting that a change in the prices of traded goods would change real domestic income for a *given nominal domestic income and a given domestic price level*. Real GDI then becomes some kind of curiosity in the system of national accounts, with no obvious link to the other aggregates. Instead, these agencies and the authors of the SNA should begin by asking themselves what real GDI is supposed to measure. In our view, it should be the real purchasing power that is available domestically, at price $p_{N,t}$.[32] Once that nominal GDP has been deflated in that way to yield real GDI, it is straightforward to compute the trading gain as the ratio of the GDP deflator to the domestic final expenditure price index as shown by (19). The trading gain can then be decomposed into terms-of-trade and real-exchange-rate effects as shown above. This is so simple that it is hard to understand why real GDI and the trading-gain concepts are not standard elements of the macroeconomic toolbox. One can only hope that in its next revision, due in 2025, the SNA will provide definite guidance as to what the best practice is.[33]

Today, Statistics Canada and the Bank of Canada stand out as being the only institutions, to the best of our knowledge, that not only publish real GDI using the price of domestic final expenditures as a deflator, together with the trading-gain estimate, but also the corresponding terms-of-trade and the real-exchange-rate components. As the French saying goes, "nul n'est prophète en son pays!" The Swiss National Bank, that was possibly the very first official institution to publish real GDI statistics using $p_{N,t}$ as the deflator starting in July 2007, inexplicably stopped publishing them in October 2014.

---

32 See Kohli (2004a:97); Reinsdorf (2010, 2020) expresses the same opinion. Oulton (2004), Diewert and Lawrence (2006), and Sefton and Weale (2006) favour the use of the price of private consumption goods as the deflator. However, we see little merit in excluding government expenditures (public consumption) and investment (deferred consumption) since these make up about half of domestic final expenditures, if not more, in many countries. As noted by Reinsdorf, in the absence of evidence to the contrary, it is reasonable to assume that the marginal income arising from trading gains is spent in the same way as average income.

33 There are of course numerous other — and undoubtedly more important — aspects of the System of National Accounts (SNA) that are subject to regular criticism and recommendations, particularly when it comes to the treatment of non-market activities, externalities, the use of natural resources, and many non-economic issues. Our point, though, is that a significant improvement could be made here at basically zero cost.

References

Boulat, R. (2006) "La Productivité et sa Mesure en France (1944-1955)," *Histoire & Mesure*, Vol. XXI, No.1, pp. 79-110.

Burge, R.W. and R.C. Geary (1957) "Balancing of a System of National Accounts in Real Terms," Chapter 2 in United Nations Economic and Social Council, *A System of Price and Quantity Indexes for National Accounts*, E/CN.3/L.46.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1973) "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics*, Vol. 55, pp. 28-45.

Corden, W.M. (1960) "The Geometric Representation of Policies to Attain Internal and External Balances," *Review of Economic Studies*, Vol. 28, pp. 1-22.

Corden, W.M. (1992) "Dependent Economy Model of the Balance of Payments," *New Palgrave Dictionary of Money and Finance* (London: Macmillan).

Denison, E.F. (1981) "International Transactions in Measures of the Nation's Production," *Survey of Current Business*, Vol. 61, pp. 17-28.

Diewert, W.E. (1974) "Applications of Duality Theory," in Michael D. Intriligator and David A. Kendrick (eds.) *Frontiers of Quantitative Economics*, Vol. 2 (Amsterdam: North-Holland).

Diewert, W.E. (1976) "Exact and Superlative Index Numbers," *Journal of Econometrics*, Vol. 4, pp. 115-145.

Diewert, W. E. and D. Lawrence (2006) *Measuring the Contributions of Productivity and Terms of Trade to Australia's Economic Welfare*, report by Meyrick and Associates to the Productivity Commission, Canberra.

Diewert, W. E. and C. J. Morrison (1986) "Adjusting Output and Productivity Indices for Changes in the Terms of Trade," *Economic Journal*, Vol. 96, pp. 659-679.

European Commission (2013) *European System of Accounts, ESA 2010* (Luxembourg: Publications Office of the European Union). https://ec.europa.eu/eurostat/documents/3859598/5925693/KS-02-13-269-EN.PDF/44cd9d01-bc64-40e5-bd40-d17df0c69334

Fox, K.J. and U. Kohli (1998) "GDP Growth, Terms-of-Trade Effects, and Total Factor Productivity," *Journal of International Trade and Economic Development*, Vol. 7, pp. 87-110.

Fox, K.J., U. Kohli, and R.S. Warren Jr. (2002) "Accounting for Growth and Output Gaps: Evidence from New Zealand," *Economic Record*, Vol. 78, pp. 312-326.

Geary, R.C. (1961) "Problems in the Deflation of National Accounts: Introduction," *Review of Income and Wealth*, Vol. 9, pp. 3-8.

Hartwick, J.M. (2020) *Advanced Introduction to National Accounting* (Cheltenham, United Kingdom: Edward Elgar).

International Monetary Fund (2009a) *System of National Accounts 2008* (New York: International Monetary Fund).

International Monetary Fund (2009b) *Export and Import Price Index Manual: Theory and Practice* (Washington, D.C.: International Monetary Fund).

Kehoe, T.J. and K.J. Ruhl (2008) "Are Shocks to the Terms of Trade Shocks to Productivity?," *Review of Economic Dynamics*, Vol. 11, pp. 804-819.

Kohli, U. (1978) "A Gross National Product Function and the Derived Demand for Imports and Supply of Exports," *Canadian Journal of Economics*, Vol. 11, pp. 167-182.

Kohli, U. (1990) "Growth Accounting in the Open Economy: Parametric and Nonparametric Estimates," *Journal of Economic and Social Measurement*, Vol. 16, pp. 125-136.

Kohli, U. (1991) *Technology, Duality, and Foreign Trade: The GNP Function Approach to Modeling Imports and Exports* (Ann Arbor, MI: University of Michigan Press).

Kohli, U. (1993) "A Symmetric Normalized Quadratic GNP Function and the U.S. Demand for Imports and Supply of Exports," *International Economic Review*, Vol. 34, pp. 243-255.

Kohli, U. (2002) "Terms-of-Trade Changes and Real GDP: Evidence from Switzerland," *Quarterly Bulletin of the Swiss National Bank*, pp. 55-63.

Kohli, U. (2003) "Terms of Trade, Real GDP, and Real Value Added: A New Look at New Zealand's Growth Performance," *New Zealand Economic Papers*, Vol. 37, pp. 41-66.

Kohli, U. (2004a) "Real GDP, Real Domestic Income, and Terms-of-Trade Changes," *Journal of International Economics*, Vol. 62, pp. 83-106.

Kohli, U. (2004b) "An Implicit Törnqvist Index of Real GDP," *Journal of Productivity Analysis*, Vol. 21, pp. 337-353.

Kohli, U. (2005a) "Switzerland's Growth Deficit: A Real Problem – But Only Half as Bad as it Looks", in Lukas Steinmann und Hans Rentsch (eds.) *Diagnose: Wachstumsschwäche* (Zurich: Verlag Neue Zürcher Zeitung).

Kohli, U. (2005b) "Labour Productvity vs. Total Factor Productivity," *Irving Fisher Committee Bulletin*, Vol. 20, pp. 106-116.

Kohli, U. (2006a) "Terms of Trade, Real Exchange Rates, and Trading Gains," SSHRC Conference on Price and Productivity Measurement, Vancouver, B.C., Canada, July 2004.

Kohli, U. (2006b) "Real GDP, Real GDI, and Trading Gains: Canada, 1981-2005," *International Productivity Monitor*, Vol. 13, pp. 46-56. http://www.csls.ca/ipm/13/IPM-13-kohli-e.pdf

Kohli, U. (2007) "Terms-of-Trade Changes, Real GDP, and Real Value Added in the Open Economy: Reassessing Hong Kong's Growth Performance," *Asia-Pacific Journal of Accounting and Economics*, Vol. 14, pp. 87-109.

Kohli, U. (2010) "Labour Productivity: Average vs. Marginal," in W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox, and A.O. Nakamura (eds.) *Price and Productivity Measurement*, Vol. 6 (Victoria, B.C.: Trafford Press).

Kohli, U. (2014) "Who Gains Most from Trade?," paper presented at the EMG Workshop, University of New South Wales, Sydney, December 2014.

Kohli, U. and J.-M. Natal (2014) "The Real Exchange Rate and the Structure of Aggregate Production," *Journal of Productivity Analysis*, Vol. 42, pp. 1-13.

Landefeld, J.S. and R.P. Parker (1997) "BEA's Chain Indexes, Time Series, and Measures of Long-Term Economic Growth," *Survey of Current Business*, July.

Macdonald, R. (2010) "Real Gross Domestic Income, Relative Prices and Economic Performance Across the OECD," *Review of Income and Wealth*, Vol. 56, pp. 498-518.

Macdonald, R. (2020) "Measurement of Real Income in the System of National Accounts: An Application to North American Economies," Economic Analysis Research Paper Series (Ottawa, Ontario: Statistics Canada).

Macdonald, R. and L. Rispoli (2016) "How Big was the Effect of Falling Commodity Prices on Canadian Real Incomes Between Mid-2014 and Mid-2016?," *Economic Insights* (Ottawa, Ontario: Statistics Canada).

Neary, J.P. (1997) "R.C. Geary's Contributions to Economic Theory," Working Paper 97/04, University College Dublin, January.

Oulton, N. (2004) "Productivity versus Welfare; or GDP versus Weitzman's NDP," *Review of Income and Wealth*, Vol. 50, pp. 329-355.

Reinsdorf, M.B. (2010) "Terms of Trade Effects: Theory and Measurement," *Review of Income and Wealth*, Vol. 56, S177-S205.

Reinsdorf, M.B. (2020) "Measuring Economic Welfare: What and How?," IMF Policy Paper, March.

Salter, W.E.G. (1959) "Internal and External Balance: The Role of Price and Expenditure Effects," *Economic Record*, Vol. 35, pp. 226-238.

Sanyal, K.K. and R.W. Jones (1982) "The Theory of Trade in Middle Products," *American Economic Review*, Vol. 72, pp. 16-31.

Sefton, J.A. and M.R. Weale (2006) "The Concept of Income in General Equilibrium," *Review of Economic Studies*, Vol. 73, pp. 219-249.

Statistics Canada (2016) *User Guide: Canadian System of Macroeconomic Accounts* (Ottawa, Ontario: Statistics Canada).

Stiglitz, J.E., A. Sen, and J.-P. Fitoussi (2009) *Report* (Paris: Commission on the Measurement of Economic Performance and Social Progress).

Swiss National Bank (2007) "Gross Domestic Product, Trading Gains and Gross National Income," *SNB Monthly Statistical Bulletin*. July.

Woodland, A.D. (1982) *International Trade and Resource Allocation* (Amsterdam: North-Holland).

Wyplosz, C. (2005) "C'est Maintenant Confirmé: le Revenu par Tête (Corrigé pour le Coût de la Vie) Irlandais a Dépassé le Niveau Suisse en 2002," *Le Temps*, January 31. https://www.letemps.ch/economie/recettes-miracle-irlandais-un-exemple-suisse.

Why Was US Labour Productivity Growth So High During the COVID-19 Pandemic? The Role of Labour Composition

Jay Stewart[1]
U.S. Bureau of Labor Statistics

Abstract

In the first few weeks of the COVID-19 recession, around 20 million US workers lost their jobs, with half of those losses occurring in the last two weeks of March 2020. On the tail of these unprecedented job losses, labour productivity grew at an annualized rate of 10.3 per cent in 2020Q2 and the average hourly wage increased sharply. This study examines how these phenomena are related. Because most of the job losses were in low-wage industries or among low-wage workers in high wage industries, the average skill level of the labour force increased substantially. This study finds that this increase in average skill level accounted for 71 per cent (7.3 percentage points) of labour productivity growth in 2020Q2, and that about one-third of the increase in average skill level was due to the change in the distribution of workers across major industries, mainly because of the massive job losses in leisure and hospitality and other low-wage industries. Altogether, changes in the distribution of workers across major industries accounted for 24 per cent (2.5 percentage points) of the 10.3 per cent increase in labour productivity

Much has been written about the impact of the COVID-19 pandemic on the US labour market.[2] The job losses that occurred in late March and early April 2020 were unprecedented. Between mid-March and mid-April of 2020, private sector payroll employment, as measured by the Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) survey, declined by about 20 million jobs. But certain industries and demographic groups were hit harder than others.[3]

---

1 Senior Research Economist at the U.S. Bureau of Labor Statistics. Many thanks to Canyon Bosler, Lucy Eldridge, John Fernald, Matt Russell, Dan Sullivan, the editor, and three anonymous referees for helpful comments. Any views expressed here are mine and do not necessarily reflect those of the BLS.
Email: stewart.jay@bls.gov.

2 See, for example, Bartik *et al.* (2020) and Groshen (2020). Handwerker *et al.* (2020) summarizes a number of these early papers.

3 The next two paragraphs summarize data from BLS Employment Situation News Releases.

The vast majority of these job losses — about 17.4 million of the nearly 20 million jobs lost — were in service-providing industries. This amounted to about a 16.2 per cent decline in employment in just a few weeks. In contrast, goods-producing industries lost 2.4 million jobs, amounting to a smaller, but still large, decline of 11.4 per cent. Within services, employment in the leisure and hospitality industry declined from 16.1 million in March to 8.5 million in April — a decline of 47.1 per cent. Other major industries that saw large decreases in employment between March and April are retail trade (2.3 million), professional and business services (2.2 million), health and education (2.6 million), and other services (1.3 million). Because the declines in employment were concentrated among low-wage workers, there was a sharp increase in the average hourly wage of $1.36, from $28.67 in March 2020 to $30.03 in April. This one-month increase of 4.7 per cent is more than 50 per cent larger than the increase of $0.86 for the one-year period between March 2019 and March 2020.

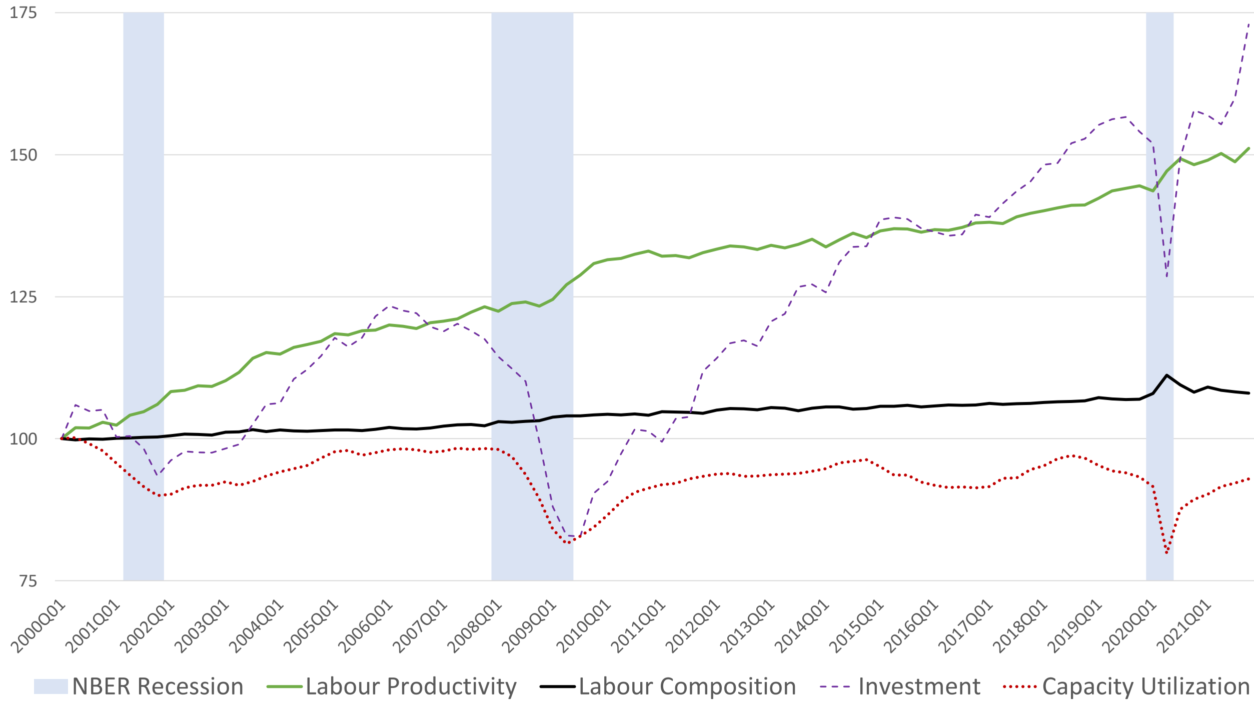
Data from BLS' household survey, the Current Population Survey (CPS), show the impact of the pandemic on different demographic groups and tell a story that is consistent with the establishment-based Current Employment Statistics (CES) data. The CPS data show that between February and April,[4] employment among high school dropouts and high school graduates fell by 26 per cent and 21 per cent, respectively. In contrast, employment of college graduates fell by just 6 per cent. Women lost jobs at a higher rate than men. And younger workers, both men and women, lost jobs at a higher rate than older workers. Around 30 per cent of 20-24-year-olds lost their jobs between February and April, compared with 13 per cent of 45-54-year-olds. Job losses were about the same for 20-24-year-old men and women. But among older workers, job losses were much higher for women than for men.

In 2020Q1, the onset of the pandemic, non-farm business sector labour productivity declined by an annualized rate of 2.5 per cent.[5] But in Q2, when both output and total hours worked labour fell sharply, labour productivity grew by 10.3 per cent. This high growth rate was caused by hours declining at a much faster pace

4 April is compared to February because pandemic-related job losses started showing up in the CPS data for March, whereas the CES data showed a much smaller decline between February and March. The difference is due to the difference in reference periods and how the reference periods interact with the definition of employment. The reference period for the CPS is the week that includes the 12th of the month, which was March 8-14. A person was classified as "employed" if he or she worked at least one hour during the reference week. Therefore, people who lost their jobs in the first week of March would show up as not employed during the CPS reference week (unless they immediately found another job). In contrast, the reference period for the CES is the pay period that includes the 12th of the month. Pay periods can be weekly, bi-weekly, semi-monthly, or monthly, and a worker was included in the payroll employment total if he or she was paid for at least one hour during the pay period. About 2/3 of CES respondents have pay periods longer than one week. Therefore, in cases where the pay period includes the first week of March, any workers who lost their jobs in the first week of March would be included in payroll employment. They would also be included in CES employment total if they did not work during the pay period but were paid.

5 Based on on estimates as of May 5, 2022. The BLS defines the non-farm business sector to include private wage and salary workers (except employees of non-profit organizations), workers in government enterprises (mainly the post office), and self-employed workers (both incorporated and unincorporated). All growth rates are annualized unless otherwise noted.

Chart 1: Indexes of Labour Productivity, Labour Composition, Industrial Capacity Utilization, and Gross Private Domestic Investment, 2000Q1=100



Source: Labour composition index: author's calculations from CPS data. Labour productivity and labour hours: BLS Labor Productivity and Costs program. Investment: Bureau of Economic Analysis. Capacity utilization: Federal Reserve Board of Governors via FRED.

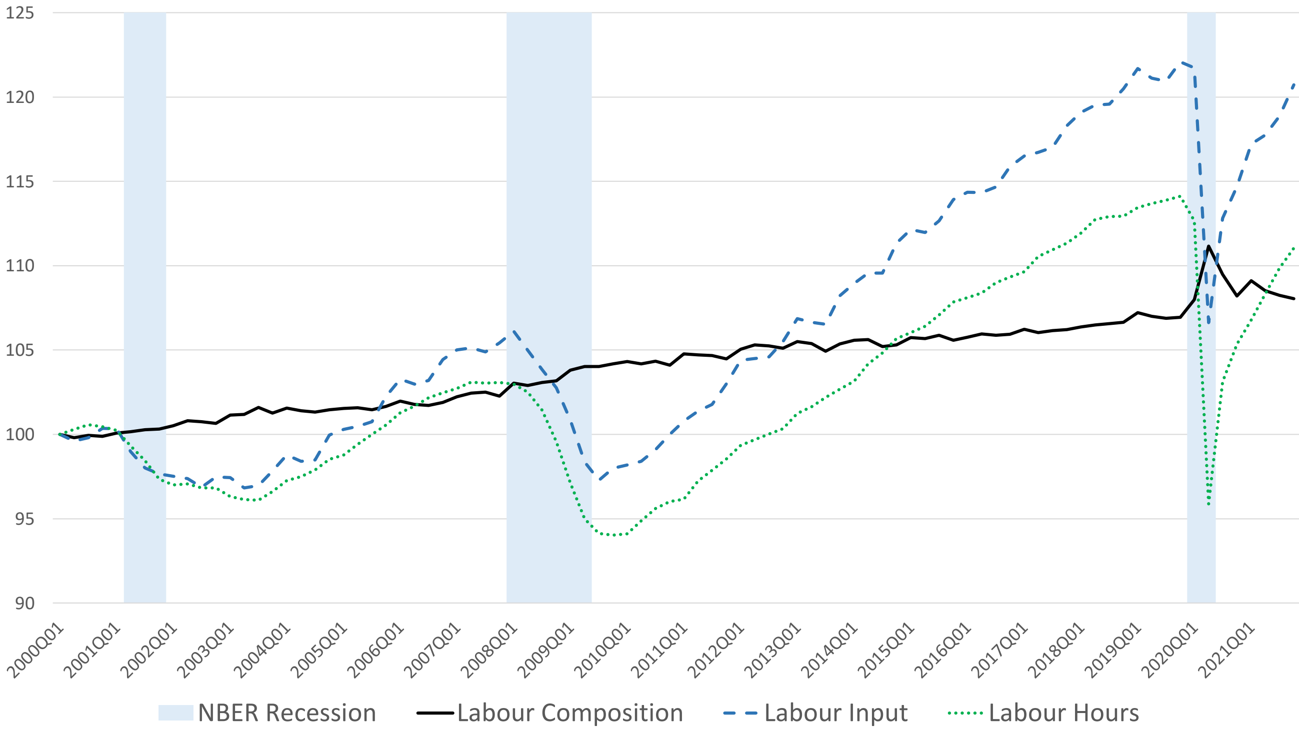
than output. Productivity continued to increase in Q3 by 6.2 per cent, however these gains were a result of a sharp rebound in both output and hours, where gains in output were faster than gains in hours. It may seem puzzling to non-economists that labour productivity growth was so strong in the middle of a pandemic. But we can shed light on these numbers by looking at the three components of labour productivity growth: the growth of total factor productivity (TFP); the change in capital services per hour worked (weighted by capital's cost share); and the change in labour composition (weighted by labour's cost share). Thus, labour productivity can be written as:

$$\dot{LP}_q = \dot{TFP}_q + s_{K,q}(\dot{K}_q - \dot{H}_q) + s_{L,q}\dot{LC}_q \quad (1)$$

where K is capital input, H is total hours worked, LC is labour composition ("quality"), and $s_{K,q}$ and $s_{L,q}$ are the average of q and $q-1$ cost shares of capital and labour. The "dots" indicate percentage growth from the previous quarter.

The main focus of this article is on the contribution of labour composition to labour productivity growth. BLS publishes estimates of labour composition at the aggregate level and by industry in its annual TFP statistics. These annual estimates are usually sufficient because growth of the labour composition index is due to increases in the education and experience of the labour force. Even during periods of rapid changes, labour composition changes slowly and is usually not a significant driver of labour productivity growth. However, the rapid changes experienced in the US economy in 2020 created a need to assess

Chart 2: Indexes of Labour Input, Labour Hours, and Labour Composition, 2000Q1 = 100



Source: Author's calculations from CPS data.

these trends at a higher frequency.

Chart 1 provides some insight as to possible drivers of 2020Q2 labour productivity growth. Ideally, the graph would include each term in equation (1). Unfortunately, there are no quarterly data on capital services. Instead, Chart 1 graphs gross private domestic investment and industrial capacity utilization. Both series exhibited sharp declines, which suggest a decline in capital services, although the sharp decrease in hours worked (Chart 2) suggests that capital intensity may have increased. The sharp increase in the labour composition index suggests that labour composition played a major role in labour productivity growth, which contrasts with previous recessions.

Of the three recessions since 2000, the 2001 recession was the least severe in terms of job losses, although some industries were hit very hard (e.g. travel-related industries). Although it is difficult to see in the indexes, there were several quarters in 2001-2002 with annualized quarterly labour productivity growth rates that exceeded 5 per cent (7.1 per cent in 2001Q2, 5.1 per cent in 2001Q4, and 8.6 per cent in 2002Q1). In those high-productivity-growth quarters, the labour composition index grew only slightly faster than long run trends. In addition, compared to the other two 21st century recessions, there was only a modest decline in investment and industrial capacity utilization.

During the Great Recession, there were several quarters of strong labour productivity growth — in the last three quarters of 2009, labour productivity grew at annualized rates of 8.7 per cent, 5.4 per cent, and 6.3 per cent. At the same time there were significant declines in both capacity utilization and investment, but only slight increases in the labour composition index.

The COVID-19 recession was quite different from the previous recessions. As with the Great Recession, there were sharp declines in investment and capacity utilization; but unlike the Great Recession, both recovered quickly. Investment returned to pre-pandemic levels by 2020Q4 and exceeded pre-pandemic levels in the last two quarters of 2021. The labour composition index also behaved quite differently in the COVID-19 recession, increasing at an unprecedented rate of 2.9 per cent (12.3 per cent annualized) in 2020Q2. Labour composition's contribution to labour productivity growth was 7.3 percentage points (12.3 per cent $\times$ labour share of 0.596) and accounted for 71 per cent of the 2020Q2 growth in labour productivity.

It is worth noting that the large increase in the labour composition index for 2020Q2 was preceded by a larger-than-average increase of 1 per cent (4 per cent annualized) in 2020Q1. Although most of the 2020Q1 job losses occurred in the last two weeks and were not reflected in the establishment survey's employment estimates, employment estimated from the CPS showed a decline of about 2.8 million people between March and April (see footnote 4). As with Q2, the increase indicates that it was primarily low-wage/low-skill workers who lost their jobs in Q1.

In 2020Q3, the story was somewhat different. Labour productivity grew at an annualized rate of 6.2 per cent, while the labour composition index fell by an annualized 5.9 per cent (contributing -3.5 percentage points) to labour productivity growth. This large difference is likely due to greater capital utilization and, to a lesser extent, the rebound in investment. It is also likely that businesses made changes to their production processes to mitigate the impact of social distancing recommendations. About 65 per cent of the decline in the labour composition index was due to within-industry changes in labour composition and about 35 per cent can be attributed to hiring in low-wage industries. It is worth noting that the labour composition index remained above the pre-pandemic level (and above the pre-pandemic trend line) through the end of 2021Q4. Average private-sector employment increased by around 2.4 million jobs in the third quarter, with 95 per cent of the increase due to hiring in services industries with two major industries, retail trade and leisure and hospitality, accounting for 55 per cent of the increase in services.

Additional insights about the COVID-19 recession can be found in the quarterly utilization-adjusted TFP data that are posted on the Federal Reserve Bank of San Francisco (SF Fed) website as a research series.[6] Although those data indicate that 2020Q2 labour productivity growth was "only" 7.5 per cent, which is lower than the official estimate of 10.3 per cent, it is possible to identify the contributions of the three components in equation (1). The SF Fed data show that TFP growth was 17.9 per cent, the contribution of capital deepening (increased cap-

6 These data are based on the methodology outlined in Fernald (2014).

ital per hour worked) was 21.5 per cent and the contribution of the increase in the average skill level of the labour force was 4.0 per cent (a 6.3 per cent growth rate × labour share of 0.62). The large decline in TFP was due mainly to a decrease in capital and labour utilization. After adjusting for utilization, TFP growth was essentially unchanged at 0.1 per cent. The combination of these factors (including the adjustment for utilization) implies that the contribution of capital deepening to labour productivity growth was 3.4 percentage points. The increase in average skill level accounted for 62 per cent of the 2020Q2 increase in labour productivity, which is a much larger portion than in any previous quarter, although somewhat smaller than what was found in this study.

This rest of this article examines the role of labour composition in the sharp increase in labour productivity in 2020Q2. The next section describes this study's data and methodology, which are similar to BLS' official labour composition measure, and compares this study's estimates with the SF Fed labour quality measure and the official BLS measure. The following section shows how labour composition differs across major industries and examines the role of industry composition (the distribution of total hours across industries) on labour productivity growth. The final section summarizes and concludes.

Methods and Data

The quarterly labour composition index presented here is conceptually the same as the official BLS labour composition measure. It is calculated as the growth of "labour input" minus the growth of aggregate labour hours. The growth in labour input is equal to the weighted sum of hours growth across demographic cells, where the weights for each age × education × sex cell are each cell's share of total labour costs. Labour composition growth is given by:

$$Labour\ Comp\ Growth_q = \sum_{c\in C} \bar{s}_{c,q} \cdot \ln\left(\frac{H_{c,q}}{H_{c,q-1}}\right) - \ln\left(\frac{\sum_{c\in C} H_{c,q}}{\sum_{c\in C} H_{c,q-1}}\right)$$

where $H_{c,q}$ is total hours worked by workers in demographic cell c in quarter q, C is the set of all demographic cells, and $\bar{s}_{c,q}$ is the average labour cost share weight, which is defined as:

$$\bar{s}_{c,q} = \frac{1}{2}\left(s_{c,q} + s_{c,q-1}\right)$$

where

$$s_{c,q} = \frac{\sum_{i\in c} \hat{w}_{i,c,q} \cdot H_{i,c,q}}{\sum_{c\in C}\sum_{i\in c} \hat{w}_{i,c,q} \cdot H_{i,c,q}}$$

and the $\hat{w}_{i,c,q}$ are predicted values from a wage equation.[7] Thus, the labour composition index increases when the hours worked by high-wage workers grow faster (or decline more slowly) than hours worked by low-wage workers.

7 This study uses a modified version of the methodology used in BLS' annual TFP growth statistics. The main differences are: (1) wages are estimated using a wage regression rather than using the median wage for each cell; this was necessary because of the small sample size. (2) Coarser demographic definitions were used; in addition, it was necessary to combine a number of cells in small industries to accommodate the smaller sample sizes encountered when generating quarterly statistics. (3) The measure is for the non-farm business sector so that it is consistent with quarterly labour productivity estimates, whereas the official measure used in the TFP statistics is for the private non-farm business sector. A discussion of some of the issues with estimating labour composition can be found in Zoghi (2010).

Data

This study uses data from the monthly Current Population Survey (CPS), which is the only data source that can be used to construct a high-frequency measure of labour composition. This section describes the data and discusses how various issues with the data were addressed.

The CPS collects information on employment, hours worked, usual weekly hours and earnings, industry, and occupation for the week that includes the 12th of the month.[8] It also collects demographic characteristics (age, education, and gender). The data were divided into 50 demographic cells: 5 age categories, 5 education categories, and 2 gender classifications.[9] Although it would have been desirable to use more finely defined demographic cells, the relatively small size of the quarterly CPS samples limit the number of cells that are feasible.[10] Even within this structure, it was necessary to combine very small cells with larger cells.[11]

Information on earnings is collected only in 2 of the 8 CPS rotation groups, known as the Outgoing Rotation Groups (ORGs, which are months in sample 4 and 8).[12] To generate wage rates for the other months in sample, a wage regression was estimated for each reference quarter using the ORG data from that quarter and the previous quarter,[13] and the coefficients were used to generate predicted values. These predicted values were used for all observations, even those with actual data.

The CPS collects hours worked on second jobs every month. But industry on the second job is collected only in the ORGs, and the CPS does not collect wages for second jobs. For analytical purposes, second jobs are treated as separate observations so that each observation represents a job rather than a person. Because industry on second jobs is available only in the ORG data, the ORG weights, which are approximately 4 times as large as the Basic CPS final weights, are used for second jobs.[14]

8 The detailed CPS industries are aggregated into 14 major industries that are the same as the CES "supersectors." Throughout this article, the terms "industry" and "major industry" are used interchangeably.

9 The age categories are: 16-24, 25-34, 35-44, 45-54, and 55+. The education categories are: less than high school, high school, some college, bachelor's degree, advanced degree. The TFP program also stratifies by class of worker (self-employed vs. wage and salary). Self-employed workers were included in the present sample, but class of worker was not used to stratify the sample because the resulting sample sizes of self-employed cells would be too small. In addition, the CPS does not collect wage information for self-employed workers, which means that it must be assumed that self-employed workers earn the same hourly wage as wage and salary workers.

10 Each quarterly sample is composed of three monthly samples

11 For example, the small number of young workers with college degrees were combined with workers in the next age category. And because some of the major industries are small, it was necessary to further combine at least some additional cells in those industries. To make the major industry measures comparable to the aggregate measure, the industry-specific combining of cells was maintained in all calculations.

12 Sampled households are in the CPS for 4 consecutive months, out of the sample for 8 months, and back in the sample for another 4 months. The questions on earnings and the additional questions on second jobs are asked in these "outgoing" rotations because they are more burdensome.

13 Independent variables include age, age squared, education, gender (also interacted with age and age-squared), major industry, and occupation.

14 The Basic CPS final weights were used for the main job of respondents in the ORGs.

This study's sample covers the non-farm business sector. The CPS sample weights were rescaled using data on total hours worked by major industry so that totals match the official estimates used by the BLS Labor Productivity and Costs (LPC) program.[15] It is important to control to major industry totals because the CPS weights account for the distribution of workers by demographic characteristics but not by industry. Finally, the labour composition index was seasonally adjusted and the indexes were rebased such that 2000Q1 = 100.[16]

Chart 2 shows the labour composition index and indexes of labour input and labour hours for the 2000-2021 period. Typically, rapid increases in labour composition are not expected, because shifts in worker experience, skills, and education tend to occur gradually over time. The exceptions occur during periods of rapid change, such as recessions, because job losses tend to be concentrated in lower-wage workers. In the three recessions that occurred during this period, there were large declines in both labour input and labour hours, with labour hours falling by more than labour input. In the 2001 recession, the decline in labour hours was only slightly larger than the decline in labour input, resulting in a slight increase in the labour composition index.

The Great Recession saw a somewhat larger increase in the labour composition index. In contrast, in 2020Q2, hours worked dropped by significantly more than labour input, which caused the sharp increase in the labour composition index. The index declined after 2020Q2 but remained above the pre-pandemic level (and above trend) through the end of 2021.

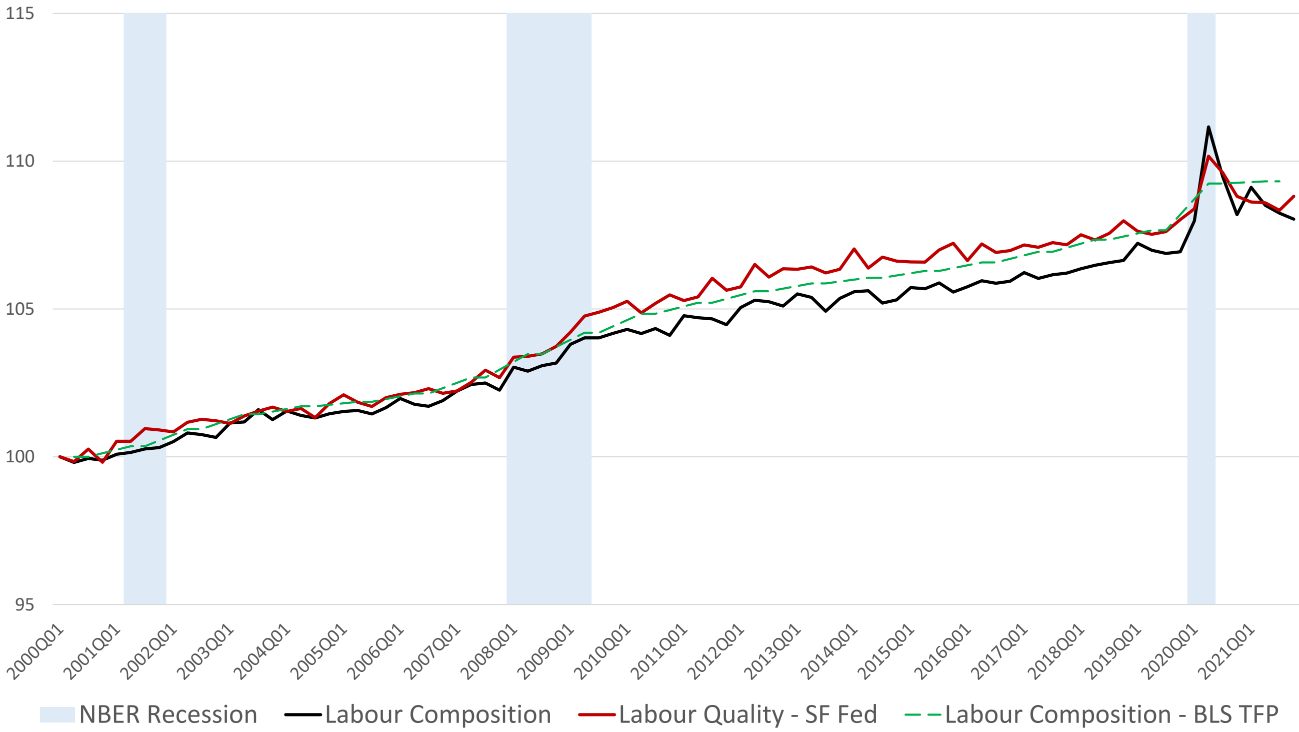
Comparison to other measures of labour composition

As previously noted, the SF Fed posts estimates of quarterly utilization adjusted TFP (and its components) on its website as a research series. The methodology for their labour composition measure, which is referred to as "labour quality," is based on Aaronson and Sullivan (2001), which differs from this study's methodology and the methodology used by the BLS TFP program. The SF Fed estimates of labour quality growth are calculated as the growth in average wages holding the return to demographic characteristics constant. Thus, the Aaronson and Sullivan measure is a quantity-weighted price (predicted wage) index, whereas the BLS measure is a cost-share weighted quantity index, which is consistent with how capital services and labour hours enter into the TFP equation. Because the measures are fundamentally different, the Aaronson and Sullivan measure will be referred to as labour quality and the BLS measure as labour composi-

15 Self-employed workers were included in the rescaling process to ensure that their weights were consistent with those of wage and salary workers.

16 By experimenting with seasonally adjusting the growth rates rather than the indexes, it was found that seasonally adjusting the indexes resulted in a smoother series. More importantly, the LPC program seasonally adjusts levels rather than growth rates.

Chart 3: Comparison of Alternative Measures of Labour Productivity and Labour Composition, 2000Q1=100



Source: Labour composition index: author's calculations from CPS data. Labour Quality – SF Fed: San Francisco Fed. Labour Composition – BLS TFP: BLS Total Factor Productivity program.

tion.[17]

Chart 3 compares the two quarterly labour composition indexes, along with the BLS TFP Program's index.[18] Note that the TFP index, which is annual, was assigned to Q2 to make it easier to see the differences in 2020. All three indexes exhibit similar long-run growth. The average compound growth rates from 2000Q1 through 2019Q4 are 0.34 per cent per year for this study's measure, 0.39 per cent for the SF Fed measure, and 0.37 for the BLS TFP measure. The main differences appear in 2020. Both quarterly measures exhibit a sharp spike in 2020Q2, with this study's modified BLS labour composition measure exhibiting a larger spike (growth of 2.9 per cent vs. 1.6 per cent, which translates to annualized growth rates of 12.3 and 6.3 per cent). Comparing the average values of the indexes for 2019 and 2020, the growth in the labour composition indexes are 2.1 per cent for this study's modified BLS measure, 1.4 per cent for the SF Fed measure, and 1.5 per cent for the BLS TFP measure. However, the higher growth rate for

17 Aaronson and Sullivan calculate the average wage as an hours-weighted mean, where the wage for each observation is the predicted value from a wage regression. Because there is no reason to prefer the coefficients from one quarter over the other, they estimate wage regressions for both the current and prior quarter, calculate growth rates using each set of coefficients, and then take the geometric mean of the two growth rates.

18 Indexes rather than growth rates are compared because growth rates tend to be noisy. The Pearson correlation coefficient between growth rates of this study's labour composition measure and the SF Fed's labour quality measure was only 0.24 between 2000Q1 and 2019Q4. However, when the series was extended through 2021Q4, the correlation coefficient more than doubled to 0.58, mainly due to the spike in 2020Q2.

this study's labour composition measure is entirely due to lower index values in 2019 rather than higher values in 2020.

Taking a closer look at the SF Fed data reveals a few inconsistencies with other data, which suggests that capital deepening played a smaller role in Q2 labour productivity growth and that labour quality/composition played a larger role. The SF Fed data show only a slight decline in the growth rate of capital services in 2020Q2 followed by a sharp slowing of capital growth in Q3, which is inconsistent with the changes in investment for Q2 and Q3 observed in the BEA data.[19] And the SF Fed estimate of labour quality growth, which is an hours-weighted average wage index (holding the returns to demographic characteristics constant), is not consistent with the observed wage changes. The 6.3 per cent annualized growth in labour quality, which translates into a one-quarter increase of 1.6 per cent, is considerably smaller than the observed increase in the average hourly wage of 4.2 per cent between the first and second quarters of 2020.[20] The BLS Employment Situation news release for April 2020 noted that "... the increases in average hourly earnings largely reflect the substantial job losses among lower-paid workers; this change, along with earnings increases, put upward pressure on the average hourly earnings estimates." Given that the CES average hourly wage also is an hours-weighted measure, one would expect the two wage growth measures to be similar.

One possible explanation is that the lower growth in the SF Fed labour quality measure in 2020Q2 is due to the dampening effect of using predicted wages to measure the growth in the wages. The predicted wage for an observation will be closer to the conditional mean than the actual wage. This occurs at both ends of the wage distribution. However, one would expect the impact to be smaller at the lower end of the wage distribution because, given that the wage distribution is right skewed, those values are closer to the conditional mean than wages at the upper end of the distribution. One might wonder why the use of predicted wages does not have the same impact on this study's measure. As noted earlier, the SF Fed labour quality measure is an hours weighted wage index, whereas this study's measure (and the official BLS measure) is a cost-share weighted hours index. For cost-share weights, only the average wages for the demographic cells matter. There is no benefit to having actual wages as long as average predicted wages are close to average actual wages.

-2.9cm

Labour Composition by Major Industry and the Role of Industry Composition

Chart 4 shows the labour composition indexes for 14 major industries. This section will note some general trends and discuss how the labour composition index be-

19 It is worth keeping in mind that short-term declines in investment can have only a limited effect on labour productivity, because investment is a relatively small portion of the capital stock.

20 Based on the average hourly wages for January — March and April — June 2020.

Chart 4: Indexes of Labour Composition by Major Industry

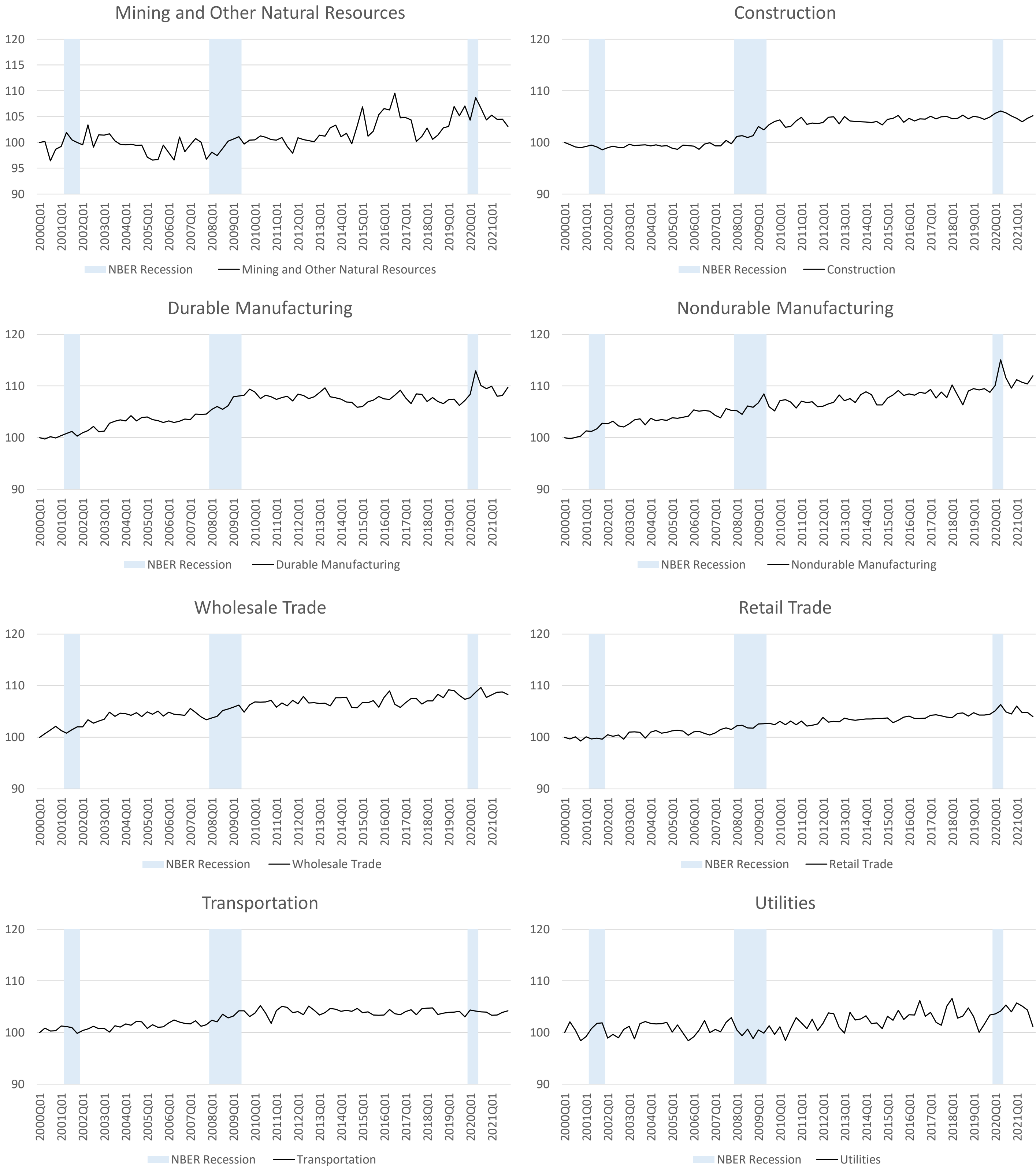
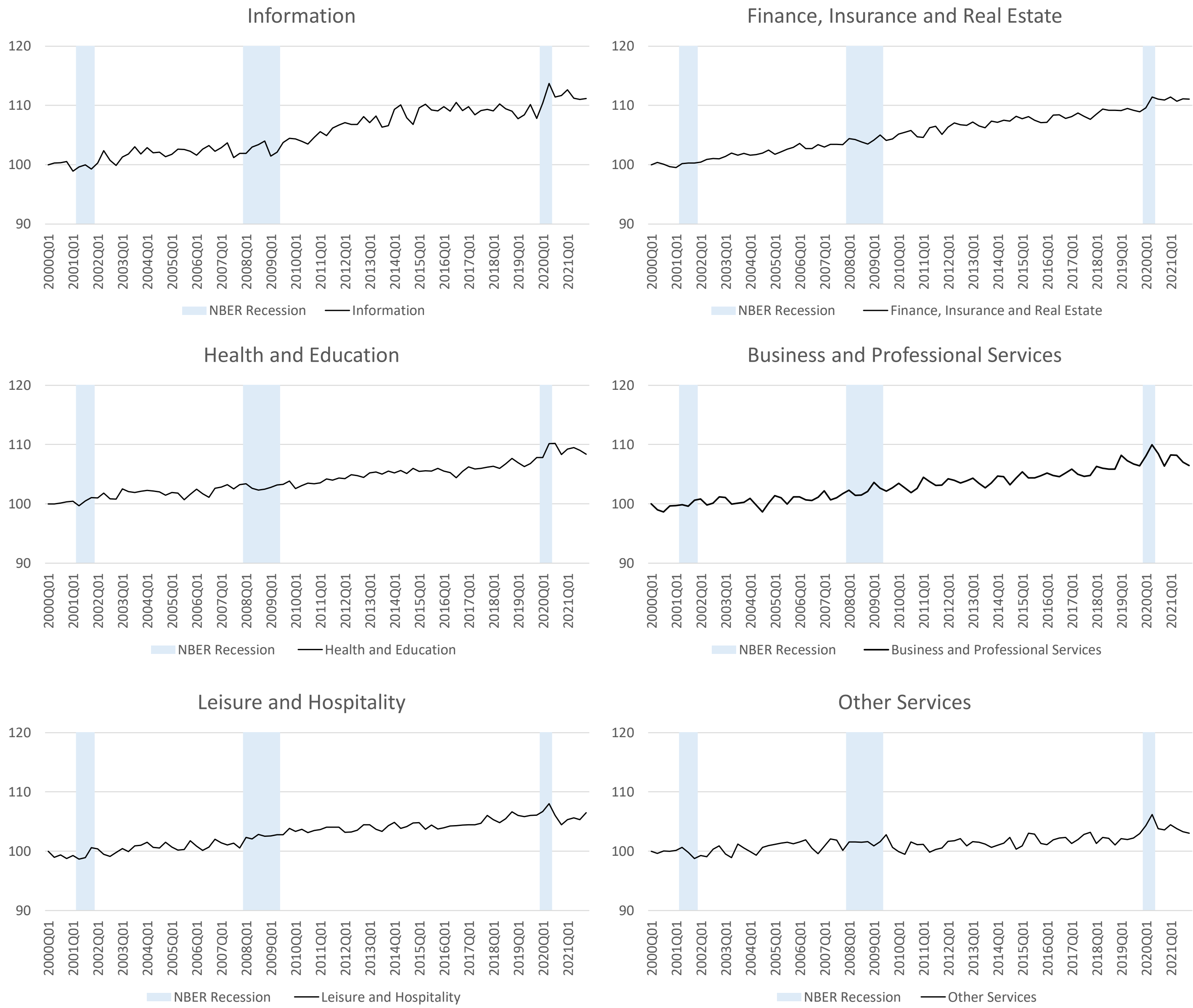


Chart 4 (continued): Indexes of Labour Composition by Major Industry



Source: Author's calculations from CPS data.

haved during the Great Recession. It will then examine how the change in industry composition (the change in the distribution of total hours across major industries) contributed to labour productivity growth in the early months of the COVID-19 pandemic.

The first point to note about Chart 4 is that the trends in labour composition over the 2000-2019 period vary quite a bit by major industry. The labour composition index increased in all major industries, but the utilities, transportation, and other services industries exhibited noticeably slower growth. The largest increases were in non-durable manufacturing, and finance, insurance and real estate (FIRE). Both construction and durable manufacturing exhibited a ratcheting up of the labour composition index around the time of the Great Recession. In construction, the index increased from around 100 to 105 between 2007Q4 and 2010Q1. Looking at occupation data from the Occupational Employment and Wage Statistics (OEWS) survey for approximately the same period, total employment in construction declined by 26 per cent. Employment in high-wage occupations declined by less than the overall decline (14-20 per cent depending on the occupation), while employment in low-wage occupations fell by more (34-50 per cent depending on the occupation).[21]

A similar pattern is seen in durable manufacturing, where the labour composition index increased from around 105 in 2007Q1 to around 110 in 2009Q1. The published OEWS data do not breakdown manufacturing into durable and non-durable, but looking at manufacturing as a whole reveals a similar — though less dramatic — pattern. Between 2007 and 2009, employment in manufacturing declined by about 11 per cent. Employment in high-wage occupations fell by less (1-6 per cent depending on the occupation), while low-wage occupations declined by more (10-18 per cent depending on the occupation).

These patterns are consistent with firms shedding employees during the recession by outsourcing low-wage jobs. But other than these two major industries, there were no sharp changes in labour composition during the Great Recession. Most other major industries exhibited a steadier increase in the labour composition index, which matched the increase in the overall labour composition index.

The behaviour of the labour composition index during the COVID-19 recession of 2020 was quite different in that there are noticeable spikes in some industries in 2020Q2. Some of these spikes (in durable and non-durable manufacturing, retail trade, information, health and education, and other services) are obvious from the graphs in Chart 4. In other major industries, the spikes are less obvious because they do not look that different from the usual quarter-to-quarter variation (construction, wholesale and retail trade, finance, insurance and real estate, leisure and hospitality, and professional and business services).

To examine this further, the two-quarter increases in the labour composition index between 2019Q4 and 2020Q2 for each major industry are compared with the corresponding average two-quarter changes between 2000Q1 and 2019Q4 (see Table 1).[22] The mean of the absolute value of the two-quarter changes was calculated so that positive and negative changes do not offset each other. Looking at column (1), we see that the two industries with the greatest variability are mining and other natural resources and utilities, which is consistent with the figures. These are small industries, so the volatility is not that surprising.

Comparing columns (1) and (3), we can see that two-quarter changes between 2019Q4 and 2020Q2 are significantly larger than the long-run average for most industries. In 8 of the 14 major industries, the

21 High-wage occupations include management, business and financial operations, and architecture and engineering, while low-wage occupation include helpers, security, food services, and cleaning and maintenance—occupations that are more-easily outsourced.

22 Two-quarter averages were calculated because the changes in labour composition started in 2020Q1, but also to mitigate the impact of the quarter-to-quarter variation.

Table 1: **Comparison of Two-Quarter Changes in Labour Composition Index by Major Industry**

	(1)	(2)	(3)	(4)
	Absolute value of two-quarter changes 2000Q1 - 2019Q4		Change between 2019Q4 and 2020Q2	Standard deviations from mean
Major Industry	**Mean**	**Standard Deviation**		
Mining and other natural resources	1.49	1.33	1.48	0.01
Construction	0.48	0.40	1.10	1.55
Durable manufacturing	0.58	0.41	5.35	11.57
Non-durable manufacturing	0.73	0.60	5.82	8.54
Wholesale trade	0.67	0.49	1.29	1.26
Retail trade	0.41	0.33	1.82	4.33
Transportation	0.60	0.48	1.10	1.05
Utilities	1.47	0.90	0.75	0.81
Information	0.88	0.62	5.46	7.44
Finance insurance & real estate	0.44	0.32	2.26	5.62
Business & professional services	0.67	0.47	3.34	5.72
Health and education	0.46	0.35	2.14	4.79
Leisure and hospitality	0.54	0.40	1.81	3.15
Other services	0.72	0.57	3.12	4.20
Non-farm business sector	0.20	0.16	3.95	23.88

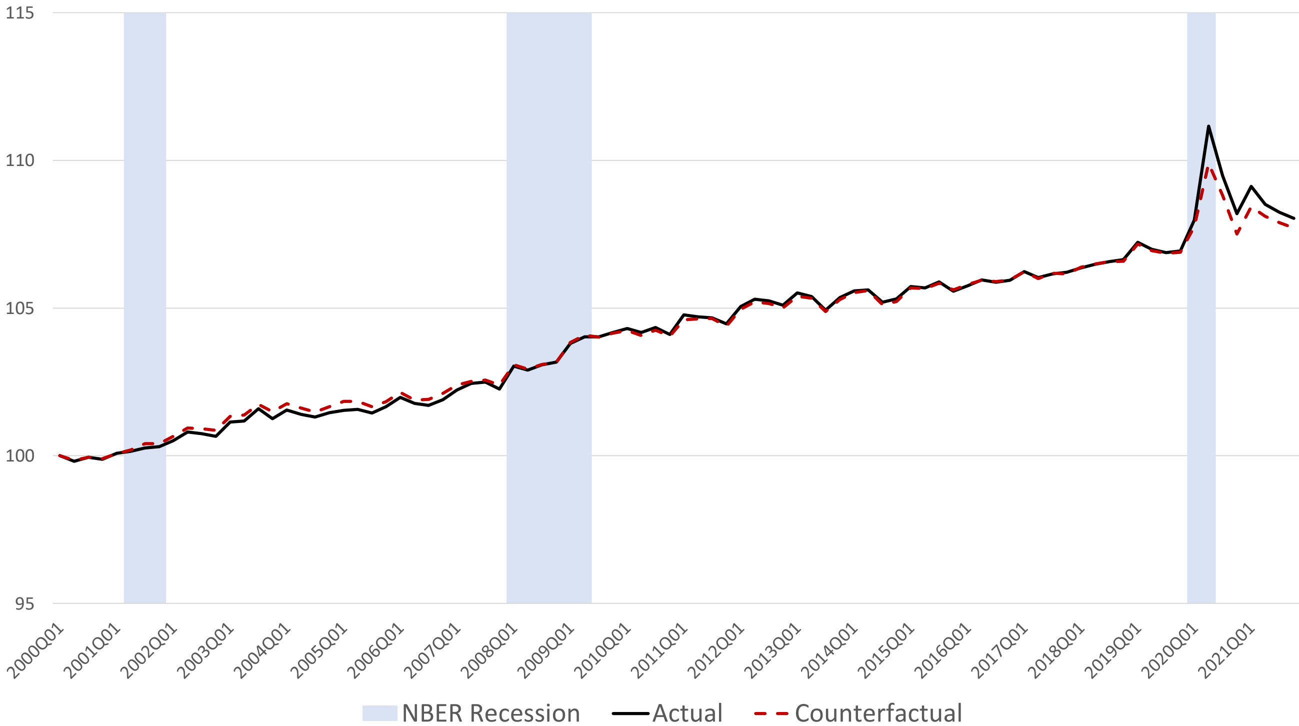
Note: Column (1) shows the average of the absolute value of the 2-quarter percent changes in the labour composition index. Column (2) shows the standard deviation. Column (3) shows the percent change in the labour composition index between 2019Q4 and 202Q2 (all changes are positive). Column (4) shows how far the changes in column (3) are from the long-run mean (in standard deviations).

increases in column (3) are more than three times as large as the average two-quarter changes in column (1). To put it into perspective, column (2) shows the standard deviation of the two-quarter changes in column (1), and column (4) shows how far (in standard deviations) the 2019Q4-2020Q2 changes are from their respective means. As one might have guessed from the figures, the 2019Q4-2020Q2 changes in the labour composition index are well within 2 standard deviations of the mean for mining and other natural resources, construction, wholesale trade, transportation, and utilities. Of the remaining industries, the 2019Q4-2020Q2 was the smallest for leisure and hospitality — just over three standard deviations — compared with more than four standard deviations for the other remaining industries. Interestingly, the increase for the non-farm business sector is nearly 24 standard deviations from the long-run mean. This suggests that changes in the industry composition of employment, specifically the decline in employment in low-wage industries like leisure and hospitality, may have played an important role in the spikes in the labour composition and labour productivity indexes.

The role of industry composition

To examine the contribution of changes in employment across major industries to the growth of labour composition, the labour composition index for the non-farm business sector was compared to a counterfactual index that holds major industry share weights constant between quarters. The counterfactual index accounts for within-industry changes in the labour composition indexes, but not changes in the distribution of workers across industries. The difference between actual and counterfactual indexes is a measure of the

Chart 5: Effect of Changes in Industry Composition



Source: Author's calculations from CPS data.

contribution of the changes that are due to industry composition.

The counterfactual growth rate is calculated as a weighted average of the major industry growth rates, where the weights are the two-quarter average of the industries' shares of the total wage bill:

$$Labour\ Comp\ Growth_q = \sum_{k=1}^{14} \bar{s}_{k,q} \times Labour\ Comp\ Growth_{k,q}$$

where $\bar{s}_{k,q}$ is the average labour cost share weight, which is defined as:

$$\bar{s}_{k,q} = \frac{1}{2}\left(s_{k,q} + s_{k,q-1}\right)$$

where

$$s_{k,q} = \frac{\widehat{W}_{k,q}}{\sum_{k=1}^{14} \widehat{W}_{k,q}}$$

and $\widehat{W}_{k,q}$ denotes total (predicted) wages in industry k in quarter q.

Chart 5 compares the actual and counterfactual labour composition indexes for the non-farm business sector. The two indexes track each other very closely through 2019Q4, except for a slight divergence in the mid-2000s. The largest difference (actual minus counterfactual) was −0.31 in 2005Q1, when the counterfactual index exceeded the actual series. However, the situation changed in 2020Q1 and dramatically so in 2020Q2, when the two series diverged sharply, with the actual index growing faster than the counterfactual.

Table 2 shows how growth rates and index values for the actual and counterfactual indexes changed just before and during the pandemic. The differences in growth rates and index values were small through 2019Q4. There is some divergence in 2020Q1. But in 2020Q2 (in bold), the difference between the actual and counter-

Table 2: Comparison of Actual and Counterfactual Composition Measures

Date	Actual Labour Composition			Counterfactual Labour Composition			Actual minus Counterfactual		
	Index	Growth	Annualized Growth	Index	Growth	Annualized Growth	Index	Growth	Annualized Growth
2019Q1	107.22	0.54	2.19	107.17	0.55	2.20	0.06	0.00	0.00
2019Q2	106.99	−0.22	−0.87	106.94	−0.21	−0.84	0.05	−0.01	−0.03
2019Q3	106.88	−0.10	−0.41	106.86	−0.08	−0.30	0.02	−0.03	−0.11
2019Q4	106.94	0.05	0.22	106.89	0.03	0.11	0.05	0.03	0.10
2020Q1	107.98	0.98	3.98	107.75	0.81	3.27	0.23	0.17	0.71
2020*Q*2	**111.16**	**2.94**	**12.30**	**109.89**	**1.99**	**8.18**	**1.27**	**0.96**	**4.12**
2020Q3	109.47	−1.52	−5.94	108.81	−0.98	−3.87	0.66	−0.54	−2.07
2020Q4	108.19	−1.17	−4.60	107.50	−1.20	−4.73	0.69	0.03	0.13
2021Q1	109.11	0.85	3.45	108.43	0.86	3.47	0.69	−0.01	−0.02
2021Q2	108.51	−0.55	−2.20	108.11	−0.30	−1.18	0.40	−0.26	−1.02
2021Q3	108.24	−0.24	−0.97	107.89	−0.20	−0.81	0.36	−0.04	−0.17
2021Q4	108.04	−0.19	−0.76	107.71	−0.17	−0.67	0.33	−0.02	−0.10

factual indexes was large — an annualized growth rate of 12.3 per cent vs. 8.2 per cent. Thus, had the major industry composition of the labour force remained the same, the growth of the labour composition index would have been 66.5 per cent of the actual growth. Thus, industry composition effects accounted for about 24 per cent (2.5 percentage points) of the increase in labour productivity between 2020Q1 and 2020Q2 (71.2 per cent of labour productivity growth due to labour composition × 33.5 per cent due to industry composition).

Both labour composition indexes fell in the following quarter (2020Q3), but the counterfactual index fell by more, indicating that employment in low-wage industries sectors grew by more than employment in high wage industries. As of 2021Q4, the actual labour composition index was still above the February 2020 level and above trend. The counterfactual labour composition index was still below the actual index, which indicates that employment in low-wage industries has not recovered as much as employment in high-wage industries. Data from the CES confirm this. Employment in most major industries had recovered almost fully. But an important exception is leisure and hospitality, where employment was still nearly 1.9 million (11 per cent) below February 2020 levels.

Given that some major industries are composed of high-wage and low-wage detailed industries, this decomposition underestimates the contributions of changes in employment in detailed industries within major industries. To illustrate, the other services major industry includes a range of more-detailed industries that vary in skill intensity. High-skill industries fall mainly into the repair and maintenance category (automobile, electronic and precision equipment, and commercial and industrial machinery), which were less likely to be affected by shutdowns and consumer hesitancy. The low-skill industries include personal care services, laundry services, and private household services, which were more likely to be impacted. Professional and business services is another major industry that includes both high-wage detailed industries (professional, scientific, and technical services) and low-wage detailed industries (for example, employment services, security, landscaping, and building services). The smaller 2020Q2 spikes in several major industries (mining, construction, wholesale trade, transportation, and utilities) indicate that job losses were distributed more evenly across skill levels and industries within these major industries.

Summary and Conclusions

The COVID-19 recession, which started in March of 2020, saw unprecedented job losses. In a span of just a few weeks, around 20 million jobs were lost. Because most of these jobs were low-wage, changes in the average skill level of the labour force, as measured by the labour composition index, increased sharply. This increase in the labour composition index accounted for about 71 per cent of the 10.3 per cent increase in labour productivity in 2020Q2. Of the 7.3 percentage point growth attributable to labour composition, about 76 per cent (5.5 percentage points) was due to within-industry changes and 24 per cent (1.8 percentage points) due to changes in the distribution of workers across major industries.

As of this writing (May 2022), employment in the US labour market has still not recovered completely. The labour composition index is still above the 2019Q4 level and above the pre-pandemic trend. The counterfactual index, which holds industry composition constant, lies below the actual index. This difference indicates that industry composition was still part of the story and is consistent with the fact that employment in the leisure and hospitality industry was still nearly 1.9 million (11 per cent) below the February 2020 level.

References

Aaronson, D. and D.G. Sullivan (2001) "Growth in Worker Quality," *Economic Perspectives*, Vol. 25, No.4, Federal Reserve Bank of Chicago.

Bartik, A.W., M. Bertrand, F. Lin, J. Rothstein, and M. Unrath (2020) "Measuring the Labor Market at the Onset of the COVID-19 Crisis," *Brookings Papers on Economic Activity*, Summer 2020.

Fernald, J. (2014) "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Federal Reserve Bank of San Francisco Working Paper 2012-19, April.

Groshen, E.L. (2020) "COVID-19's Impact on the U.S. Labor Market as of September 2020," *Business Economics*, November, NABE.

Handwerker, E.W., P.B. Meyer, J. Piacentini, M. Schultz, and L. Sveikauskas (2020) "Employment Recovery in the Wake of the COVID-19 Pandemic," *Monthly Labor Review*, December.

U.S. Bureau of Labour Statistics (2020) "The Employment Situation — April 2020," News Release. https://www.bls.gov/bls/news-release/empsit.htm.

Zoghi, C. (2010) "Measuring Labor Composition: A Comparison of Alternate Methodologies," in K.G. Abraham, J.R. Spletzer, and M. Harper (eds.), *Labor in the New Economy* (University of Chicago Press), pp. 457-485. https://www.nber.org/system/files/chapters/c10834/c10834.pdf.

Introduction to the Symposium on Productivity and Well-being, Part I

Andrew Sharpe
Centre for the Study of Living Standards
Dan Sichel
Wellesley College and NBER
Bart van Ark
The Productivity Institute, University of Manchester[1]

Articles published in the *International Productivity Monitor* have traditionally focused on the production sphere of economic activity and have seldom addressed the relationship between productivity and well-being. Recognizing the increasing attention to well-being issues by economists, government and the general public, this issue of the IPM goes some way to remedy this past lack of attention to well-being by publishing a first symposium of four articles on productivity-well-being linkages.[2] A second symposium of three articles on the same topic will appear in the next issue of the *International Productivity Monitor*. This introduction discusses the background and motivation of the symposium, the organizational process, highlights key issues related to productivity-well-being linkages, and provides a detailed synthesis of the contributions of the four articles.

Background to the Symposium

The Centre for the Study of Living Standards (CSLS), the Ottawa-based not-for-profit economic research organization that founded the *International Productivity Monitor* (IPM) in 2000, has always had a strong interest in well-being issues. In the late 1990s and early 2000s, the CSLS developed the Index of Economic Well-being, a composite index based on consumption flows, stocks of wealth, equality and economic security indicators (Os-

1 Andrew Sharpe is Executive Director of the Ottawa-based Centre for the Study of Living Standards and Founding Editor of the *International Productivity Monitor*. Bart van Ark is Managing Director of the Productivity Institute and Professor of Productivity Studies at the Alliance Business School at the University of Manchester. Dan Sichel is Professor of Economics at Wellesley College, Research Associate at the National Bureau of Economics Research and Chair of the Advisory Committee at the Bureau of Economic Analysis, and Guest Editor of the IPM symposium on productivity-well-being linkages. Emails: andrew.sharpe@csls.ca; bart.vanark@manchester.ac.uk; dsichel@wellesley.edu

2 The authors of the first three of the articles discuss their results in a podcast moderated by Bart van Ark found at https://player.fm/series/productivity-puzzles/productivity-and-well-being

berg and Sharpe, 2002). But there was always a firewall between the CSLS work on well-being and the IPM, with the journal narrowly focused on traditional GDP-based output measures to be used for productivity estimates.[3] The only instances where the IPM went beyond its focus of analysis of traditional productivity issues were articles on the relationship between labour productivity and real wages.[4]

In January 2021, the CSLS entered into a partnership agreement on the IPM with The Productivity Institute (TPI) in the United Kingdom. The mandate of TPI largely focuses on productivity topics, but also encompasses well-being issues. The relationship between productivity and well-being is consequently of great interest to the TPI. The editors Andrew Sharpe and Bart van Ark therefore decided to devote an issue of the IPM to the relationship between productivity and well-being. Dan Sichel, a Professor of Economics at Wellesley College, Chair of the Advisory Committee of the Bureau of Economic Analysis and a member of the IPM's International Advisory Council kindly agreed to join the two editors as a guest editor for the project.

In early 2021, a call for papers was widely distributed to productivity researchers. First drafts of the papers accepted in response to the call were presented at a virtual authors' workshop in November, 2021.[5] After going through the standard refereeing procedures, seven papers from the workshop were accepted for publication in the Symposium. Four of these papers are included in this issue of the IPM as Part I of the Symposium and the remaining three papers will be published as Part II of the Symposium in the next issue in the Fall of 2022.

Context for Understanding Productivity-Well-being Linkages

The productivity and well-being literature appear to exist in two different universes. Productivity is generally understood as relating to efficiency at the firm, industry and aggregate economy level whereas well-being is a much broader concept relating to a wide range of dimensions that contribute to overall life satisfaction and happiness. A full understanding of the complex relationships between these concepts, including the tradeoffs and complementarities, is in its infancy.[6]

The traditional perspective is that productivity is the key to real income growth and that higher incomes are important components of better well-being. It follows then that there is a positive relationship running from productivity to well-

3 The CSLS was of course always well aware of the close two-way relationship between productivity and well-being and addresses this topic in Sharpe (2002).

4 See, for example, the symposium on the decoupling of productivity and pay in the United States, Canada, and the United Kingdom in the Fall 2021 issue of the *International Productivity Monitor* and available at http://www.csls.ca/ipm.asp as well as earlier article on the issue by Sharpe and Uguccione (2017).

5 The program for the workshop is available at https://www.productivity.ac.uk/ipm/workshop-on-productivity-and-well-being-measurement-and-linkages/

6 For a recent survey of the literature on productivity-well-being linkages, see Sharpe and Fard (2022). This research was funded by the International Labour Organization (ILO) in Geneva.

being. This perspective finds support in the annual *World Happiness Report* published by the United Nations, which develops a framework to explain life satisfaction across countries with six explanatory variables. One of these variables is GDP per capita, which is highly correlated with labour productivity. This variable is found to be the most important of the six variables. A doubling of a country's real income would in principle and everything else equal, raise the average level of life satisfaction by 30 per cent. Real-world evidence of this productivity-income effect on well-being is shown by the fact that low-income countries on average (though there are exceptions) have much lower level of life satisfaction than high income countries.

But everything is not equal. Real income growth may be associated with greater work-related stress, longer commuting times, and degradation of the environments, among other negative consequences of economic growth. Since these non-economic factors also are important for well-being, higher productivity does not automatically translate into higher well-being. Indeed, life satisfaction in the United States and other countries has not increased during the postwar period, yet incomes have more than doubled. This is known as the Easterlin paradox and has generated considerable research (Easterlin and O'Connor, 2020). A key conclusion of his research is that well-being is directly related to social factors and is as much of a relative nature as it is of an absolute one. If one's income does not improve relative to one's comparison group, one does not necessarily feel better off even though one's income has increased in absolute terms.

Some of the key aspects in the debate on the relationship between productivity and well-being are highlighted below.

Two-way nature of the relationship between productivity and well-being

The relationship is two-way in nature, running from changes in productivity to changes in well-being and from changes in well-being to changes in productivity. The first direction is the most studied. Productivity growth is widely recognized as the only long-run source of increased per capita income. Increased incomes also generate tax revenues that can be used by governments for transfer payments for the disadvantaged, for public goods, and for the direct provision of government services such as health and education services. The second direction from better well-being to higher productivity is manifested, for example, by happier workers being more productive.

Differences between the level of well-being and well-being efficiency or productivity

Two of the articles in the symposium refer to the efficiency of the generation of well-being defined as the level of well-being attained in relation to the resources available such as capital and natural resources used to generate well-being. This differs from the level of well-being, which is the absolute level of well-being, measured for example, by life satisfaction, abstracting from the resources needed to attain that level.

Definitions of well-being

There is a fundamental distinction made between objective measures of well-being and subjective measures. Historically the focus has been on objective indicators of well-being, such as income, wealth, health, environmental quality. In recent years, subjective well-being, best captured by surveys of life satisfaction, has been receiving much more attention. The papers in this symposium use both measures.

A distinction can be made between overall well-being and economic well-being. The former includes all aspects of well-being, such as political freedom, spirituality, family life as well as standards of living. Economic well-being is obviously more narrow, focusing on material aspects of well-being.

The Symposium Articles

The symposium in this issue of the IPM contains four articles on productivity-well-being linkages. The next issue of the IPM will contain three additional articles. This section of the introduction provides a synthesis of these contributions. Two of the four articles in this symposium are by economists from New Zealand, a very disproportional contribution given the small population of the country. This appears to reflect the high level of importance given to both subjects in the country, as evidenced by the establishment of the New Zealand Productivity Commission in the 2000s and the release of its well known well-being framework for budgets by the New Zealand Treasury in 2019. The country definitely punches above its weight in its contributions to the international debate on productivity-well-being linkages.

A Capital Stocks Approach to Productivity and Well-being (Legge and Smith)

In recent years the capital stocks approach to productivity measurement has been gaining popularity. This approach goes beyond the standard measure of capital as physical investment goods and develops estimates for additional types of capital, namely human capital, natural capital, and social capital. It then uses the conventional total factor productivity growth accounting framework to estimate the contribution to labour productivity or income from the various types of capital.

The lead article in the symposium by **Jaimie Legge**, an independent economic consultant, and **Conal Smith** from the Victoria University of Wellington takes the capital stocks approach to productivity and income and applies it to well-being to estimate the degree to which countries generate well-being, as measured by life satisfaction, from the four types of capital. They find for a given well-being outcome, there is significant cross-country variation in the quantities of the different types of capital used, indicating different degrees of efficiency in the generation of well-being. The Nordic countries, for example, have high levels of well-being, but use considerable amounts of capital to generate these outcomes. On the other hand, certain Eastern European have nearly comparable levels of life satisfaction, but use considerably smaller amounts of capital to attain this level of well-being. From this perspective, these countries are more “efficient” or productive in the uses of the different types of capital.

Of course, it is the absolute level of well-being or life satisfaction that matters most, even if this level requires large quantities of capital inputs. But given resource scarcity, which certainly pertains to the different types of capital, it is very useful to identify which countries are able to transform limited amounts of capital inputs into respectable well-being outcomes. There may be lessons for other lower-resource countries on how to boost well-being.

The authors begin by highlighting two approaches to well-being measurement, that of subjective well-being and that of capabilities. The former is grounded in the utilitarian tradition and sees well-being as something experienced in the mind. The latter, based on the work of Amartya Sen, focuses on well-being as the ability of a person to live the kind of life they have reason to value. While the two approaches are conceptually distinct, the authors argue that in practice this distinction is much less clear and that evaluative measures of subjective well-being, such as overall life satisfaction, capture the most commonly identified capabilities.

The authors then point out that building on the report *The Measurement of Economic Performance and Social Progress* (Stiglitz, Sen, and Fitoussi, 2009), there has emerged a widely used framework for conceptualizing and measuring intergenerational well-being (OECD, 2011). This capital stocks model or framework (Smith, 2018) posits that well-being draws on the stocks of productive resources, namely produced capital, human capital, social capital, and natural capital. The flow of resources from the capital stocks can be used for current consumption, or invested for future consumption. A sustainable level of well-being can be defined as a state where the capital stocks do not decrease over time.

The authors recognize that extensive work has been done on the determinants of subjective well-being, but point out that few contributions to this literature have used the capital stocks approach. Where this approach has been taken, such as OECD (2015), the focus has been on the level of the capital stocks and not how efficiently they are used. The authors' objective is to fill this gap in the literature by developing estimates of total well-being productivity (TWP) based on different types of capital stocks that are methodologically comparable to traditional measures of total factor productivity (TFP). Just as TFP (whether the level of TFP or its growth rate) is a measure of the efficiency with which inputs are used to produce output or income, TWP is a measure of the efficiency by which inputs (the four types of capital) are used to produce well-being, as proxied by life satisfaction.

In the TWP framework, produced capital and human capital are measured and used in the same way as in the TFP framework. On the other hand, social capital and natural capital are generally not included in the TFP framework. In this study, social capital is defined in terms of productive shared norms and values such as trust and the rule of law that allow for constructive engagements between people. Natural capital is more complex and at the broadest level refers to all aspects of the natural environment that support human life and well-being. Many elements of natural capital cannot be monetarized, so natural capital has no single overarching measure.

The authors develop a production function for estimating the relationship between the four types of capital and life satisfaction. Just as TFP is a residual in the standard growth accounting model, TWP is the part of well-being that cannot be accounted for by the four types of capital. They put together a dataset on the four types of capital and subjective well-being for 22 EU countries. The Penn World Tables are used for produced capital, human capital and market TFP. The European Social Survey is used for life satisfaction. The Corruption Perception Index produced by Transparency International is used for trust. Natural capital is proxied by the Biodiversity Indicators Index developed by the UK Natural History Museum. This index captures the impact of human presence on ecosystems and is based on the percentage of original species that remain.

The regression results show that for the full capital stocks model, human capital, produced capital, and social capital are determinants of well-being, but that natural capital has a minimal effect. For market outcomes, the results are generally similar to overall well-being. In contrast, for non-market outcomes, produced capital is less important. An interesting finding is that no correlation is found between market and non-market TFP, suggesting that the production technologies of the two sectors are fundamentally different. The highest levels of TWP are found in Poland and Croatia, even though these countries do not have the highest levels of life satisfaction. It appears that these countries are particularly successful or efficient in transforming their capital stocks into well-being.

The findings from this analysis of the efficiency or productivity of well-being are important for public policy. First, there are large differences in TWP levels across EU countries ranging from 1.6 in Croatia to 0.4 in Bulgaria. This suggests there are ways to increase well-being that do not involve increasing the level of the capital stocks. Second, the different production functions for market and non-market outcomes implies that maximizing market output does not necessarily maximize total well-being as the non-market elements of well-being have very different drivers. In particular, investments in human and social capital, which have positive effects for both market and non-market outcomes, may have a larger effect on overall well-being than produced capital, which appears to have no effect on non-market outcomes. Third, unlike the other three types of capital, natural capital has no relation with overall life satisfaction. When outcomes are decomposed into market and non-market outcomes, natural capital has a negative effect with market outcomes, but a strong positive effect with non-market outcomes. This first finding may reflect the positive impact of resource depletion on market output.

This article represents a novel and highly innovative analysis of a new concept, namely that of well-being efficiency or productivity. But much work remains to be done, especially related to methodology and data, as recognized by the authors. Both TFP and TWP are estimated through a production function as residuals subject to error. The decomposition of TWP for the non-market productivity component is challenging given the different non-market consumption bundles across countries. The data used for the estimation of TWP also needs significant improvement, especially natural capital.

Trust, Well-being and Productivity (Hazledine)

As shown in the first article of the symposium, social capital is increasingly recognized as an important determinant of both productivity and well-being, with social capital often proxied by a measure of trust. But the mechanism by which trust affects productivity and well-being is still poorly understood. In the second article in the symposium, **Tim Hazledine** from the University of Auckland sheds light on this topic.

Hazledine's objective is to measure the role of trust in explaining cross-country differences in the level of labour productivity and self-reported well-being in 136 countries. By trust, Hazledine means socially useful norms and values. He defines two trust variables: the first, called trust, is from a question asked in the World Values Survey; and the second, called deep trust, is estimated as a function of long-standing cultural, historical, geographic, and linguistic factors. He finds that both trust variables have significant bivariate relationships with productivity and well-being. But when these variables are added to standard models for productivity and well-being, they add no explanatory power. Hazledine explains this paradox as follows: while trust affects the determinants of productivity and well-being, it does not operate directly on the two variables but rather indirectly through its effects on the capital stocks. For example, deep trust has positive impacts on human capital, physical capital, and institutional quality, which in turn boosts productivity, while the trust variable similarly affects the determinants of well-being.

Adam Smith was the first to recognize that, for the progress of society, people need to learn to interact effectively with each other, in particular with strangers. More recently, Kenneth Arrow has stressed that every commercial transaction has within itself an element of trust and that much of the backwardness of the world can be explained by the lack of mutual confidence. Trust is essential for economic life and differences in economic success across nations may in part be accounted for by differences in levels of trust.

For his econometric estimation, Hazledine uses a standard neoclassical production function for output and productivity augmented with institutional quality. He also has a "production function for well-being" from the *World Happiness Report* produced by the United Nations. This model or framework explains cross-country differences in self-reported well-being or life satisfaction, in terms of six variables (per capita income, social support, healthy life expectancy, freedom to choose what to do with one's life, generosity with charity, and level of a country's corruption). To both these models, Hazledine adds the two trust variables, but finds they do not improve the fit.

Hazledine persuasively argues that:

> "high trusts levels do not in themselves make people happier or more productive. But high trust demonstrably encourages long-term investment in physical and human capital, and in good institutions, that generate economic prosperity. And

> it somehow contributes to the various cultural and institutional factors that feed well-being."

Hazledine observes that trust is increasing throughout the world, except in the United States. An example of the importance of trust for the economy is the massive and unexpected growth in on-line or digital commerce, which requires that people are willing to buy from strangers. Hazledine points out in the podcast that this success relates in part to the development of rating systems of sellers by purchasers. People generally seem to trust that these ratings are accurate.

Hazledine's article represents an original and important contribution to the literature on the role of trust in the determination of both productivity and well-being. Trust is crucial, but unlike say physical and human capital, it is not a proximate determinant, but an underlying factor or condition for a country's success in terms of both productivity and well-being. Harking back to Adam Smith, this is a common sense finding and we believe few would dissent.

Going forward, tougher questions are why some societies exhibit high trust and others low trust? Even more important, is the level of trust in a society fixed or can public policies, moral suasion or other factors change it? Should trust become one of the variables policy makers consider when developing plans to improve productivity performance or increase well-being, or is trust too bound up with the historical development of a society that it has limited, if any, direct policy relevance?

Time Use, Productivity and Household-Centric Measurement of Welfare in the Digital Age (Coyle and Nakamura)

The digital age has had major ramifications for all aspects of society and the economy, including time use, productivity, and well-being, that we are only just beginning to understand. In the third article in the symposium, **Diane Coyle** from Cambridge University and **Leonard Nakamura** from the Federal Reserve Bank of Philadelphia address the implications of the digital age from the perspective of developing a broader understanding of progress than the standard GDP statistics. Their ambitious objective is to lay the groundwork for a measurement framework for well-being that combines time allocation by activity with monetary measures of well-being and incorporates new ways of measuring productivity in digitalized activities.

Coyle and Nakamura point out that the true budget constraint is the 24-hour day. It cannot be expanded. With the digital revolution, time spent on-line has increased for work, household activities, and leisure. Which of these areas of time use contributes the most to well-being? Do some types of on-line activity actually decrease well-being? Does this overall trend toward increased on-line activity constitute societal progress? These are still open questions. What is needed to shed light on them is a "time lens for progress" that is time use data for the three types of on-line activity by level of satisfaction experienced.

Digital technologies can result in shifts in work tasks between paid and unpaid labour, with implications for productivity.

An example is the use by supermarket customers of self-check-out scanners. Sales are unchanged but paid labour, which is used as the input to calculate labour productivity, is reduced and productivity increases. We have still limited understanding of how important these developments are for both productivity and well-being.

The authors set out an ambitious agenda for measuring broader economic welfare and productivity in terms of a money metric of the well-being afforded by different allocations of time. They highlight the need to take into account the digitally-driven reallocations across the market/home production boundary and the work/leisure boundary. In terms of their perspective on valuation of time, they build on the full-income approach pioneered by Becker. The key requirement for the realization of their agenda is the availability of regular and detailed time use data, including digital activities.

The authors offer advice for statistical offices on how to move forward on this agenda. First, they recommend that statistical offices develop new measures of output that better capture the utility impacts of the changing economy and time use, and produce satellite accounts for these measures. Second, they make the case for statistical offices to broaden their regular collection practices to include data needed for the satellite accounts.

This is a wide-ranging article rich in ideas. It has the potential to stimulate further work in a variety of areas related to the impact of the digital economy.

Links between Productivity and Standard of Living (Oulton)

Unlike the first three articles in the symposium that largely focus on subjective well-being, the fourth article by **Nicholas Oulton** of the London School of Economics and the National Institute of Economic and Social Research relates to an objective measure or metric of well-being or welfare, namely living standards. The author starts out with the premise that it is productivity that accounts for the long-run growth in living standards. He then proceeds to show that this was indeed the case in the UK over the 1977-2019 period when labour productivity accounted for 92 per cent of the increase in living standards.

Oulton's measure of living standards is income-based, but it is not GDP per capita, which he recognizes has weaknesses as a metric of living standards. Rather he uses the concept of median equivalent household disposable income (MEHDI) employed by the UK Office of National Statistics (ONS). He makes the case that this measure is superior to GDP per capita for four reasons. First, the measure is based on median, not average income so it better captures the experience of a typical household in the wake of rising income inequality. If the income gains are concentrated in the top half of the income distribution, average income will rise faster than the median income. Second, given the existence of economies of scale in consumption, Oulton uses an equivalence scale to make adjustments to income for family size. Third, Oulton focuses on the household, not the individual. Household members pool resources, making the household the appropriate unit for decision-making related to labour supply and spending and hence the tracking of living standards. Fourth, income is measured on a post-tax or disposable basis, indicating the purchasing power of the household over goods and services supplied by the market.

Oulton develops a framework to decompose changes in MEHDI into nine factors, with labour productivity measured as GDP per hour worked, being only one of the factors. These eight additional MEHDI determinants are income inequality, the equivalence effect linked to the parameters of the equivalence scale and family size and composition, the share of household income in total income, hours per person employed, the unemployment rate, the labour force participation rate, the relative size of the working age population (16 and over) in the total population, and the relative price of consumer goods compared to the GDP deflator. Over short periods, these factors can be very important for income growth, but over long periods, they are largely offsetting and contribute little to real income growth.

Oulton's results are instructive. He finds the MEHDI advanced at a 1.9 per cent average annual rate in the UK from 1977 to 2019. This was slightly faster than labour productivity growth at 1.7 per cent. Other factors that contributed to income growth were an increase in the household income share of GDP, a greater proportion of the population 16 and over, and a slower rate of increase in consumer prices than in the GDP deflator. On the other hand, median income growth was reduced relative to average income by growing income inequality (growth in average EDHI was 0.24 percentage points higher per year at 2.2 per cent) and by fewer hours worked per person employed.

The importance of productivity growth for living standards is well illustrated in the UK after 2007 when productivity growth plummeted. After averaging 2.3 per cent per year from 1977 to 2007, labour productivity growth collapsed to 0.2 per cent for the 2007-2019 period. MEHDI also fell dramatically, from 2.4 per cent in 1977-2007 to 0.4 per cent in 2007-2019. In other words, all of the 2.0 percentage point fall in living standards of the UK population after 2007 is accounted for by the 2.1 percentage point drop in productivity growth. If the UK wants to increase the living standards of its citizens, productivity growth is the royal road.

One issue that Oulton does not fully address is the implications of his choice of post-tax or disposable income over pre-tax income for the income measure. A comprehensive measure of living standards should extend beyond goods and services produced by the private sector to include the goods and services provided without charge by the public sector, such as health and education. The consumption of these public goods is not currently included as income in MEHDI. One can imagine a scenario where tax rates are increased, reducing disposable income, but the tax revenues are used effectively for the provision of additional and higher-quality public health and education benefits equally shared among the population. The fall in MEHDI would underestimate true developments in the living standards of the population. The effective delivery of public goods and their valuation by the recipients is challenging, but needs to be included in any comprehensive assessment of the link between productivity and living standards. Progress on this front may be a long way off. In the short to medium term, an alternative perspective in living standards measurement is to use pre-tax income instead of disposable income. This is based on the assumption that in democratic societies, through the taxes they pay, citizens receive public services commensurate in value to private goods and services they can purchase with their disposable income.

Take-aways from the Symposium

So far, the symposium has generated some important takeaways (as mentioned, several other articles from the symposium will be published in the next issue of the IPM). The first take-away from the symposium is that it is crucial to differentiate objective or material well-being from subjective well-being or happiness. An income measure, such as the real disposable adjusted household income measure used by Oulton, is a reasonable proxy or metric for material well-being. Subjective well-being, also referred to as happiness, is best measured by long-term satisfaction with one's life. Over time, material well-being increases much more than subjective well-being, which may exhibit no trend. In any discussion on well-being, it must be clear whether material or subjective well-being is the focus of attention.

Second, productivity growth, the foundation of income growth, is much more important for material well-being than subjective well-being. This follows from the fact that many factors other than income affect subjective well-being. In addition, even higher income may not necessarily contribute to greater happiness if any positive long-term effect of income growth on happiness is offset through the effects of comparisons relative to others in the community.

Third, material well-being is, in principle, important to the population. But it is often taken for granted, unless it is declining. The current conjunction where inflation is outpacing wage gains, illustrates the public concern for material well-being when living standards are falling. But steady increases in material well-being generate much less public attention.

Fourth, the digital age has resulted in major changes in time use, with much more time spent on-line for both work, household activities, and leisure. But the implication of these changing patterns of time use for productivity and well-being are unclear. Certainly, the rapid and extensive diffusion of ICT has reduced the labour needed for many routine tasks, boosting labour productivity. Perhaps workers can now devote themselves to more interesting non-routine tasks and obtain greater work satisfaction. Whether this is the case remains unclear.

Fifth, a robust finding is that generalized trust in strangers, which reduces transaction costs, contributes both to higher productivity growth, and to higher material and subjective well-being.

Sixth, through the article by Legge and Smith, this symposium is introducing the concept of total well-being efficiency (TWP) to the literature. In a sense, this is an intuitive concept. How can limited or scarce resources, defined in terms of the four types of capital (produced, human, social, and natural capital) be allocated to produce the highest level of life satisfaction? As shown in the article, countries vary widely in the level of TWP. Understanding the reasons for this variance will occupy researchers for many years to come. For example, differences in TFP across countries are generally linked to differences in institutions and technologies. Does this also apply to TWP?

Seventh, there has been a long and intense debate in the well-being research community on the relative merits of a dashboard of well-being indicators versus a summary indicator or a composite index of well-being. Legge and Smith strongly favour the latter approach. Their preferred summary measure of subjective well-being, life satisfaction, is the dependent variable in their model. Indeed, econometric analysis of the determinants of well-being require summary measures of well-being.

Eighth, going forward, should more attention by both the research community and by public policy makers be given to attempt to quantify well-being in monetary terms so that central agencies of government can more easily incorporate well-being considerations in their budgets? Or should less attention be given to productivity and GDP given that well-being issues are more important and resonate more with the public? Expert opinions differ.

Ninth, none of the articles explicitly examine the role of public policy in improving productivity and well-being. These are big topics well beyond the remit of the articles. Both productivity and well-being are influenced by many factors. From a policy perspective, one must first identify what are the most important factors, then ascertain if these variables are indeed amenable to public policy, and finally determine which public policy levers are most effective. This is a long-term project.

Tenth, in the capital approach to well-being, the different types of capital are inputs in the production of the outcome of life satisfaction. But the inputs themselves may contribute to the well-being of the population and it may be difficult to separate this positive well-being effect from the from output of the "well-being production function." For example, students may experience positive well-being from attending school and the process of accumulating human capital. This is in addition to obtaining higher levels of life satisfaction that result from attaining the qualifications.

Eleventh, in terms of productivity, there is an on-going debate about the importance of digital technologies compared to the major innovations or general purpose technologies of the past, such as the steam engine and electricity. Analysts who see the digital as less fundamental than past major innovations point to the weak aggregate productivity growth in the digital age, except for the second half of the 1990s. Those who take a more positive view of the productivity-augmenting potential of digital technologies argue that the productivity benefits of the digital age are being currently underestimated because of measurement problems, or are forthcoming due to lag effects.

Finally, in recent years, many societies have become more polarized. Some argue that this development, which has implications for well-being, is related to disruptions caused by digital technologies, such as social media. It has been noted that the invention of the printing press in 1440 was followed by two centuries of religious wars. The printing press represented a much less expensive way to communicate ideas, a challenge to conventional religious views. Equally, the digital revolution through social media fosters mass communications. One no longer needs a printing press to express oneself publicly. Gatekeepers are gone. In the long run, this democratization of speech likely represents progress for society. But in the short to medium term, this development can be disruptive and divisive for society, with potential negative implications for well-being.

References

Easterlin, R. and K. O'Connor (2020) "The Easterlin Paradox," *Discussion Paper Series* No, 13923, IZA Institute of Labor Economics. https://player.fm/series/productivity-puzzles/productivity-and-well-being

Osberg, L. and A. Sharpe (2002) "An Index of Economic Well-being for Selected OECD Countries," *Review of Income and Wealth*, September.

OECD (2011) *How's Life?: Measuring Well-being*, OECD Publishing.

OECD (2015) *How's Life? 2015: Measuring Well-being*, OECD Publishing.

Sharpe, A. (2002) "The Contribution of Productivity to Economic Well-being in Canada," in A. Sharpe and S. Rao (eds.) *Productivity Issues in Canada, Industry Canada Research Series* (Calgary: University of Calgary Press).

Sharpe, A. and S.M. Fard (2022) "The Current State of Research on the Two-way Linkages between Productivity and Well-Being," ILO Working Paper 56, March. https://www.ilo.org/global/publications/working-papers/WCMS_839845/lang--en/index.htm

Sharpe, A. and J. Uguccione (2017) "Decomposing the Productivity-Wage Nexus in Selected OECD Countries, 1968-2013," *International Productivity Monitor*, Number 32:, pp. 25–43. http://www.csls.ca/ipm/32/Uguccioni_Sharpe.pdf

Smith, C. (2018) *Treasury Living Standards Dashboard: Monitoring Intergenerational Wellbeing*, Kotata Insight, Wellington.

Stiglitz, J. E., A. Sen, and J.-P. Fitoussi (2009) *Report* (Paris: Commission on the Measurement of Economic Performance and Social Progress).

Well-being and Productivity: A Capital Stocks Approach

Jaimie Legge

Conal Smith
Victoria University of Wellington[1]

Abstract

In the widely used capital stocks approach to conceptualizing intergenerational well-being, the well-being of the current generation is considered a function of produced capital, human capital (labour), social capital, and natural capital. Most discussion of the sustainability of levels of well-being into the future is focused on considering whether the quantity of these capital stocks left for future generations will be the same, larger, or smaller than the quantity available to the current generation. However, the efficiency with which the capital stocks are used to produce well-being also matters. Because the capital stocks approach is grounded in a framework with strong parallels to that underpinning growth accounting, total factor productivity (TFP) provides a potentially useful way of examining this issue.

This article explores the relationship between well-being and TFP. An econometric approach is used to develop methodologically comparable estimates of traditional TFP (where the output in question is national income) and total well-being productivity (where the output is mean national life satisfaction). The differences between the two measures are compared and the impact on this of confounding factors — including the roles of social capital, natural capital, and cultural bias in responses to subjective well-being measures — is explored. We find that there are large differences in total well-being productivity across countries. More generally, interpreting the capital stocks model in terms of an aggregate production function for well-being produces plausible results.

Human well-being is one of the primary goals of public policy. This is reflected in the conceptual framework of standard neoclassical economic analysis which is centred on utility maximization. However, in practice, economic analysis has traditionally focused on income as the primary policy-relevant outcome. This reflects the obvious importance of consumption — and hence income — to human well-being as well as

1 Jaimie Legge is an independent research economist based in Wellington, New Zealand. Conal Smith is a Senior Associate with the Institute for Governance and Policy Studies at the Victoria University of Wellington. Email: conal.smith@vuw.ac.nz.

the conceptual and technical issues associated with measuring well-being in practice. However, in the last 20 years significant progress has been made in the measurement of well-being. The ability to directly measure well-being opens the door to investigating whether the use of well-being, as opposed to income, as the focus for analysis would lead to substantially different policy judgements.

Key developments in the conceptualization and measurement of well-being over the last 20 years have come from two directions. On the one hand, there is a growing body of literature focusing on the measurement of subjective well-being and the use of such measures as a proxy for utility in an economic context (Kahnemn, Diener, and Schwarz, 1999; OECD, 2013a; Frijters *et al.*, 2020). Much of this literature is grounded firmly in the utilitarian tradition and sees well-being as something experienced in the mind. The other main tradition is grounded in the work of Sen and focuses on well-being as the ability of a person to live the kind of life they have reason to value (Sen, 1993). This approach conceptualizes well-being as comprising a vector of distinct capabilities that collectively describe a multi-dimensional frontier within which an individual is able to function.

In principle, these two approaches to well-being are quite distinct. In practice, however, the distinction between the neo-utilitarian and the capabilities approach to well-being is much less clear. The *Report of the Committee on the Measurement of Economic Performance and Social Progress* (Stiglitz, Sen, and Fitoussi, 2009) identifies subjective well-being as an important capability in its own right, suggesting that the distinction between the two approaches is not absolute. Perhaps more importantly, it is clear that some evaluative measures of subjective well-being — such as measures of overall life satisfaction — function empirically as summary measures capturing the impact of the most commonly identified capabilities (Boarini *et al.*, 2013).

Following the release of the Sen/Stiglitz/Fitoussi report, a widely used framework for conceptualising and measuring intergenerational well-being has emerged (OECD, 2011; Arrow *et al.*, 2012; UNECE, 2014). This framework — referred to here as the capital stocks model — draws on the approach to measuring the current well-being of people outlined in Sen, Stiglitz, and Fitoussi (2009). It places this approach in a coherent economic framework where current well-being draws on stocks of productive resources (the capital stocks). Typically, four capital stocks are identified: produced capital, human capital, social capital and natural capital. The flow of resources from the capital stocks can either be used for current consumption (well-being) or re-invested in the capital stocks. An attractive feature of this approach is that a definition of sustainable development that is consistent with the Brundtland declaration on sustainable development (Butlin, 1989) falls directly out of the framework:

> sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

In terms of the capital stocks framework, a sustainable level of well-being is defined as one where capital stocks do not decrease over time (Arrow *et al.*, 2012). This can be considered either in terms of soft sustainability (where the total value of the four capital stocks does not decrease over time) or hard sustainability which requires than none of the four capital stocks is allowed to decrease.

There is an extensive literature on the determinants of current well-being, often focused on the use of an over-arching measure of subjective well-being such as life satisfaction (Boarini *et al.*, 2012; Helliwell, Huang and Wang, 2015; Clark *et al.*, 2018). However, far less attention has been paid to the capital stocks. The most substantive contributions on this front have been from the OECD as part of its Better Life Initiative (in particular, see OECD, 2013b, chapter 6; and OECD, 2015, chapter 3), the World Bank (2006), and Arrow *et al.* (2012). Where capital stocks have been considered the focus has been entirely on the levels of the capital stocks rather than how efficiently they are used (OECD, 2015).

The lack of investigation into the efficiency with which the capital stocks are used to produce well-being represents an important theoretical and empirical gap in the literature. Assuming that the size of the capital stocks and the size of the population whose well-being they need to support are held constant, the Brundtland definition of sustainable development necessarily requires an improvement in the efficiency with which the capital stocks are used if there is to be an increase in sustainable well-being. Put simply, the well-being productivity of the economy and society matters.

This article presents an initial exploration of well-being productivity and its relationship to more conventional productivity measures. Life satisfaction is used as a measure of overall well-being and analysis focuses on the relative importance of the different capital stocks in driving overall well-being. Compared to the extensive literature on the determinants of current well-being (Boarini *et al.*, 2012; Helliwell, Huang, and Wang, 2015, 2017), the focus of this article is less on identifying the causal impact and relative importance of different drivers of subjective well-being and instead centres on developing an estimate of well-being productivity that is methodologically comparable to more traditional measures of total factor productivity.

Section two sets out the conceptual framework and describes the capital stocks model of intergenerational well-being and defines total well-being productivity (TWP) in this context. In the third section, an extended Swan-Solow growth model is used to place the capital stocks model of well-being on a clear conceptual basis and a formal definition of TWP is derived. On the basis of this, an empirical strategy to estimate TWP is proposed and a series of testable hypotheses about the well-being production function and its relation to the four capital stocks are explored.

Section four of the article describes the dataset used to estimate TWP and explore its relationship to more conventional productivity measures. This draws on data from the European Social Survey (ESS) on well-being and cross-country economic statistics from the Penn World Tables (PWT). The Biodiversity Intactness Index (BII) is used to capture variation in natural capital per capita while the Corruption Perceptions Index from Transparency International is used as a measure of social capital. Empirical results are discussed in section five.

The final section considers the implications of the main empirical findings. We find that TWP is only weakly correlated with more traditional measures of total factor productivity (TFP) and that levels of TWP vary widely across countries. The aggregate production functions for market and non-market goods implied by the analysis are quite different, although the importance of the different capital stocks to well-being is affirmed in most model specifications which is consistent with the capital stocks model. An exception is natural capital which is largely non-significant. This may be due to measurement issues or it may reflect that the relationship between natural capital and well-being is negative in the short term due to impacts from current consumption on the natural environment.

Conceptual Framework

In well-being economics, the capital stocks framework is the dominant analytical model used for thinking about intergenerational well-being and sustainability. However, because the measurement of well-being has been the primary focus of well-being economics there has been relatively little development of the capital stocks model beyond the level of a measurement framework. This is reasonable as any empirical analysis of the capital stocks model is dependent on the ability to measure well-being. However, with the emergence of a coherent approach to the measurement of well-being over the last decade, it is now possible to look at the relationship between the capital stocks and well-being.

Before proceeding to outline the model that will be applied to examine TWP, it is useful to review the main approaches to conceptualizing and measuring well-being. The economic literature on well-being identifies two main approaches.[2] The first of these is the so-called capabilities approach (Sen, 1993), while the second is the neo-utilitarian or subjective well-being approach (Frijters *et al.*, 2020).

Sen (1999) defines well-being as peoples' ability to "lead the kinds of lives they value — and have reason to value." In taking this approach Sen grounds well-being in a liberal framework that prioritizes (reasoned) individual choice over other values. Well-being in this sense, Sen argues, can be conceptualized as a set of capabilities that collectively define a multi-dimensional consumption possibility frontier for each person. Within this framework command over market goods and services — measured by

2 In addition to the two approaches that form the focus for the economic literature, a third approach to well-being can be identified in the public health/medical literature. This approach identifies well-being as "wellness" conceived of as positive health states (Roscoe, 2009). Compared to the economic approaches that form the focus of this article, the wellness literature has a narrower focus. Consider that health is commonly identified as a core capability within Sen's approach to well-being and is an major empirical driver of subjective well-being, thus making health a sub-dimension or driver of well-being within the economic approach. In contrast, the "wellness" approach sees well-being as an element of health.

income — is one important dimension of a person's capabilities. However, non-market outcomes such as health status or knowledge and skills also represent important capabilities in that they limit the range of desired functionings that a person can achieve and cannot easily be purchased directly.

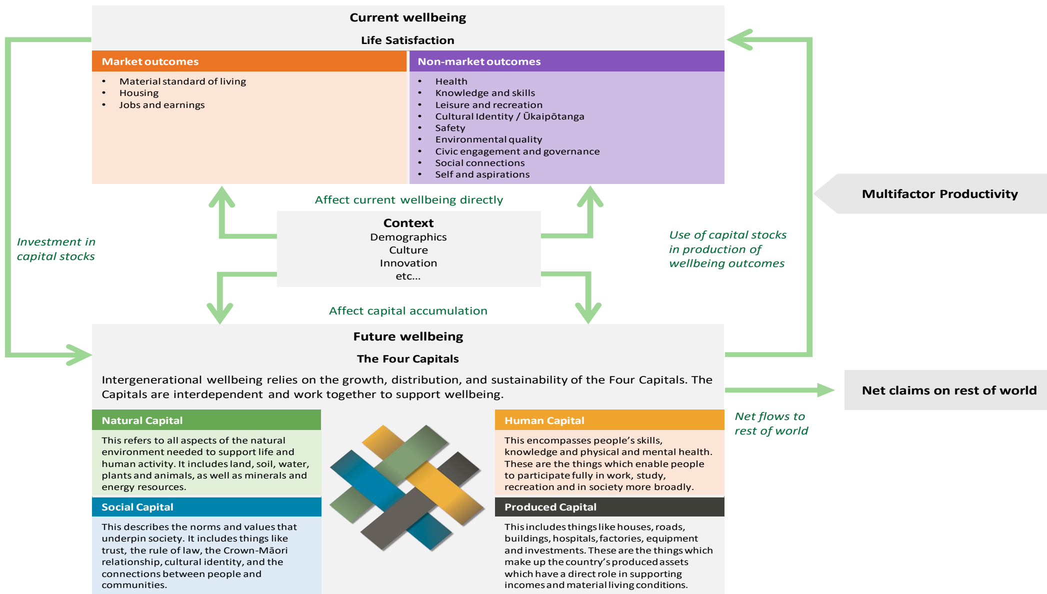
The capabilities approach is widely used in government and related policy contexts (OECD, 2011) for two reasons. First, the capabilities approach is consistent with the standard neo-classical economic framework of ordinal utility and thus integrates easily into conventional policy frameworks. In addition, the multi-dimensional nature of the capabilities framework and the strongly liberal framing of the capabilities approach allows for well-being indicators to be presented in a "dashboard" without the introduction of strong — and potentially contentious — assumptions about the relative importance of different outcomes.

The main alternative to the capabilities approach is the neo-utilitarian conception of well-being. Building on significant evidence that measures of subjective well-being are meaningful and valid (OECD, 2013a) this approach frames well-being in terms of subjective mental states. Fundamentally, a person is deemed to have high well-being if they experience positive mental states. In contrast to the multi-dimensional indicator dashboards used to measure well-being under the capabilities approach, the neo-utilitarian approach tends to focus on the use of a single over-arching measure of subjective well-being. The most commonly used such measure is overall satisfaction with life (OECD, 2013a).

The capital stocks framework builds on the measurement of well-being by placing well-being in an explicitly inter-temporal context and linking well-being as an outcome with the resources required to produce well-being. In effect, the capital stocks model links consumption and the utility function on the one hand (well-being) with resources available for production on the other (the capital stocks). Chart 1 is taken from a report prepared for the New Zealand Treasury (Smith, 2018) and illustrates the capital stock framework. This particular diagram is used because it is relatively simple and it clearly identifies the nature of the resource flows in the model in terms of production and investment, but is fundamentally the same as diagrams of the capital stocks framework from the OECD (2011, 2013b, 2015), Arrow *et al.* (2012), Costanza *et al.* (2016) and others.

It is clear from Chart 1 that the capital stocks model can be thought of in terms of production and consumption. The four capital stocks (natural capital, social capital, human capital, and produced capital) function as factors of production that are combined to produce a range of outputs that either directly contribute to well-being (market and non-market outcomes) or are invested in maintaining the level of the capital stocks. Conceptually, this framework can be seen as an extended version of a Solow-Swan growth model (Solow, 1956; Swan, 1956). This is reflected both in an implicit production function involving the four capital stocks and a decision about the investment rate that determines the maximum sustainable level of market and non-market consumption (and therefore well-being).

Chart 1: The Capital Stocks Framework



Source: Smith, 2018.

Given the focus of this paper on the capital stocks, it is important to be clear about what the capital stocks represent and their role in the model. The scope of produced capital and human capital should be relatively clear as these are used in the same way in the capital stocks model as in growth accounting more generally. Produced capital captures those material assets that contribute to the production process such as roads, buildings, machinery and equipment as well as net financial assets (which represent a claim on the same). Human capital encompasses the productivity ability of human labour including knowledge, skills and the quantity of labour (itself a function of the labour force and participation rates).

Social capital might appear to be a somewhat fuzzy concept, but for the purposes of the capital stocks model, it can be defined in relatively straight-forward terms as productive shared norms and values such as social trust, the rule of law and other intangible assets that allow for constructive engagements between people. Natural capital, on the other hand, is more complex. At the general level, natural capital refers to all aspects of the natural environment that support human life and well-being. This includes not only natural assets used directly in the production process such as minerals, forests, and soil, but also natural assets valued by people for cultural, recreational, or aesthetic reasons and assets valued for the ecosystem services that they provide such as flood control or carbon absorption and sequestration.[3] Unlike produced capi-

3 The issue of climate change provides a useful illustration of the difference between well-being and the capital stocks in the capital stocks model. Current well-being may be enhanced by the use of fossil fuels which allows for higher consumption in the present. However, by exceeding global capacity to absorb atmospheric carbon the use of fossil fuels reduces the natural capital stock. This will impact on the levels of well-being able to be produced for future generations.

tal and (to a lesser degree) human capital – which are traded in the market and can therefore be valued using money as a common metric – natural capital has no single over-arching measure of value and is inherently multi-dimensional.

A second important point regarding the capital stocks model relates to issues of aggregation. While it is possible to analyse the distribution of current well-being across the population (and this is a major focus of the well-being measurement agenda — see, for example, Stiglitz, Sen, Fitoussi, 2009; UNECD, 2014), this is not possible for the well-being of future generations where we do not know the size and make-up of these generations nor their endowments, preferences and constraints. To address this issue, the capital stocks model focuses on the aggregate levels of each capital stock to assess the intergenerational sustainability of well-being. Although it is not possible to know the distribution of the well-being of future generations, it is conceptually possible to assess whether the current generation passes on a greater or lesser total endowment of the resources required to produce well-being (i.e. the capital stocks) to future generations.

While viewing the capital stocks framework through the lens of a Solow-Swan growth model represents a ruthless simplification of a complex issue, such an approach also has significant advantages. In particular, it provides a framework for examining the relationship between the capital stocks and well-being in empirical terms. In contrast to the extensive literature on the measurement of well-being and the determinants of well-being at an individual level, there is comparatively little empirical literature focusing on the relationship between the capital stocks and well-being, and even less that considers this from the perspective of productivity.

The closest study to our approach in terms of scope is Vemuri and Costanza (2006), who model well-being on the basis of capital stocks using data from the UNDP and propose a National Well-being Index based on this analysis. They find natural capital to have a significant impact on life satisfaction along with the joint impact of human and produced capital as reflected in the Human Development Index. Engelbrecht (2015) explores the contribution of both social and natural capital to well-being and finds a significant relationship in both cases. However, neither Vemuri and Costanza nor Engelbrecht directly consider issues of productivity. Another empirical examination of the relationship between well-being and the capital stocks is Qasim and Grimes (2021), who consider how the relationship between genuine savings and well-being varies in the short and long run. Genuine savings is typically defined as aggregate net savings less depreciation in stocks of both natural and produced capital. They find support for the capital stocks model in that genuine savings is negatively related to well-being in the short run but has a positive correlation in the long run. This is consistent with the capital stocks model in that there is a trade-off between savings and consumption in the short run, but in the long run, a higher genuine savings rate implies greater investment in the capital stocks and higher future consumption.

One of the few papers that does investigate the capital stocks model from an empirical perspective, and which also discusses the TFP in this context is Arrow *et al.* (2012). However, the focus of Arrow *et al.* is to define comprehensive wealth (the discounted present value of the capital stocks) rather than to investigate the relationship between the capital stocks and well-being. Consequently, while a conventional measure of TFP is incorporated into their model, Arrow *et al.* do not investigate productivity from the perspective of the efficiency with which the capital stocks contribute to overall well-being. It is, however, precisely this issue that is the focus of this article.

Method

To begin, it is necessary to provide a definition of well-being. Consider the following utility function:

$$U = f(C, Y) \tag{1}$$

where Y is income and C is a vector of non-market outcomes important to a person's well-being. If we are willing to accept a measure of subjective well-being, such as life satisfaction, as a (noisy) proxy for utility then it is possible to empirically estimate a utility function as follows:

$$W_i = \beta_0 + \beta_1 C_i + \beta_2 \ln(Y_i) + \varepsilon \tag{2}$$

In this equation W_i is the life satisfaction (well-being) of person i, C_i is a vector of non-market drivers of life satisfaction (e.g. health status, knowledge and skills, safety) experienced by person i and Y_i is the income of person i. Note that life satisfaction is a bounded measure (typically from 0 to 10) while income is unbounded on the upward side. This imposes the log-linear relationship between life satisfaction and income in equation (2) and is widely supported empirically (Deaton, 2008; Sacks, Stevenson, and Wolfers, 2012). In contrast, C_i is assumed to have a linear relationship with life satisfaction since most of the non-market outcome measures typically included in regressions of this type (Boarini *et al.*, 2013), are bounded themselves.

To incorporate the capital stocks into the model it is necessary to set out an approach to production. The simplest way to approach this is simply to consider well-being as the single output of an aggregate production function. Equation (3) sets out this approach where W_c is mean life satisfaction of country c, $\breve{A}_c$ is TWP for country c, K_c is the per capita (produced) capital stock of country c, and L_c is the per capita human capital stock of country c which is assumed to be a function of the labour utilisation rate and the mean level of education.

$$W_c = \breve{A}_c K_c^{\rho_1} L_c^{\rho_2} \tag{3}$$

While something like equation (3) is implicit in the capital stocks model, this very reduced form approach fails to take the utility function seriously and is difficult to decompose in any useful way to provide an insight into what drives the underlying relationships. An alternative — or possibly complementary approach — is to consider the market and non-market contributions to well-being separately. Equations (4) and (5) below specify respectively an aggregate production function for market goods, which we can assess through income (Y) and a similar production function for non-market goods.

$$Y_c = A_c K_c^{\alpha_1} L_c^{\alpha_2} \tag{4}$$

$$C_c = \ln\left(a_c K_c^{\gamma_1} L_c^{\gamma_2}\right) \qquad (5)$$

Equation (4) is relatively straightforward, with A_c being the TFP of country c, Y_c being per capita income of country c, K_c and L_c capture produced and human capital as in equation (3). Note that this is the standard growth accounting aggregate production function and can be used to estimate TFP. Non-market production — equation (5) — is similar, with a_c being the non-market TFP of country c and C_c being a vector of mean non-market outcomes for country c. For simplicity it is assumed that the production of non-market outcomes and market outcomes is non-rival in terms of K and L.[4]

Given information on Y_c, K_c, and L_c it is possible to estimate α_1, α_2, and A_c, capturing the elasticity of output with respect to produced and human capital respectively and TFP. Taking the log of equation (4) we can estimate the relationship as model (6):

$$\begin{aligned} \ln(Y_c) &= \ln(A_c) + \alpha_1 \ln(K_c) \\ &\quad + \alpha_2 \ln(L_c) + \varepsilon \end{aligned} \qquad (6)$$

Solving equation (6) for A_c is trivial and gives an estimate of TFP as the Solow-Swan residual. While this is not the preferred approach to estimating TFP in most circumstances, it has the appeal here that a similar approach can potentially be applied to equation (5). Estimating A_c and a_c using the same method in turn allows for a comparison between the two measures of productivity without bias introduced due to method effects.

Estimating equation (5) is a little more involved than is the case for equation (4). In particular, we lack a definitive list of non-market outcomes and, even were such a list available, there is no common metric on which we could assess them. Rather than estimating equation (5) directly, it is therefore necessary to approach the issue via measures of overall well-being. In particular, we can estimate the contribution of non-market outcomes to overall well-being by looking at how levels of overall well-being vary after accounting for the impact of market outcomes. Equation (7) presents the country level equivalent of equation (2):

$$W_c = \beta_0 + \theta_c + \beta_1 C_c + \beta_2 \ln(Y_c) + \varepsilon \qquad (7)$$

All variables in equation 7 are country means. The constant θ_c has been introduced to capture cultural response bias that might introduce non-random measurement error across countries. Rearranging (7) we can define $\widehat{W_c}$ as non-market variance in life satisfaction as follows:

$$\widehat{W_c} = W_c - \beta_0 - \beta_2 \ln(Y_c) \qquad (8)$$

If we then substitute in equation (4) this then gives the following identity (9):

$$\begin{aligned} \widehat{W_c} &= \theta_c + \beta_1 C_c \\ &= \theta_c + \beta_1 \ln\left(\alpha_c K_c^{\gamma_1} L_c^{\gamma_2}\right) \end{aligned} \qquad (9)$$

4 In reality, some aspects of the capital stocks will be non-rival and others will be rival. The issue of allocating capital across the non-market and market sectors is left for further work. It should be noted, however, that conceptually the assumption that market and non-market goods are non-rival between equations (3) and (4) is not different to the assumption that the issue of rival uses of capital can be ignored within the equation (3) on its own (i.e. between different market goods).

If a credible control for cultural response bias in life satisfaction can be identified, it is then possible to estimate non-market TFP directly as follows:

$$\widehat{W}_c - \theta_c = \beta_1 a_c + \beta_1\gamma_1 \ln(K_c) + \beta_1\gamma_2 \ln(L_c) + \varepsilon \tag{10}$$

If equation (10) is estimated empirically, we cannot observe γ_1 and γ_2 directly as the coefficients on produced capital per capita and human capital per capita will be $\beta_1\gamma_1$ and $\beta_2\gamma_2$. However, the ratio of the two coefficients $\frac{\beta_1\gamma_1}{\beta_2\gamma_2}$ can be compared directly to the ratio of the two elasticities from equation (4): $\frac{\alpha_1}{\alpha_2}$. Similarly, the residual estimate of non-market TFP from equation (9) will be a linear transformation of actual non-market TFP (i.e. we observe $\beta_1\alpha_c$ rather than α_c). However, since β_1 is a constant and non-market TFP is an index with no natural units, the observed value ($\beta_1 a_c$) is sufficient to identify countries where market TFP and non-market TFP differ.

Empirically estimating the model in equation (9) requires, in addition to the underlying data, good estimates of β_2 (the income coefficient on life satisfaction) and θ_c (cultural response bias in life satisfaction). The former is easy to obtain and can be estimated from a cross-country life satisfaction regression along the lines of that presented in equation (7) or taken directly from the substantial existing academic literature (Sacks, Stevenson, and Wolfers, 2012). Cultural response bias, on the other hand, is more difficult to estimate.

The key challenge in estimating cultural response bias is that it is difficult to distinguish between cultural response bias (a measurement error that should be corrected for) and genuine cultural impacts on well-being or omitted variables affecting life satisfaction (both of which should not be corrected for). A number of approaches have been proposed to identify cultural response bias including the use of anchoring vignettes (Kapteyn, Smith, and van Soest, 2010) and leveraging differences between country of birth and country of residence (Senik, 2014; Exton, Smith, and Vandendreische, 2015). While vignettes require extensive data collection, it is possible to estimate a value for θ_c from any dataset with information on life satisfaction, country of residence and country of birth. The simplest approach[5] to this is as follows:

$$W_{i,r,b} = \beta_0 + \beta_1 D_i + \theta_b + \mu_r + \epsilon \tag{11}$$

In equation (11) $W_{i,r,b}$ is the life satisfaction of individual i residing in country r and born in country b, while D_i is a vector of demographic controls. Finally, θ_b and μ_r are vectors of dummy variables for country of residence and country of birth each having a value of 0 for all countries except for those where the respondent was born and currently resides. From this regression we can interpret the coefficient on θ_b as the impact of having been born in a specific country independently of the impact of current influences on life satisfaction from where one lives (μ_r). Thus θ_b captures the impact of residual social ties to one's country of birth as well as the impact of culture on

5 Adopting a more sophisticated approach to estimating cultural response bias by following Senik (2014) more closely is an obvious extension to this article.

life satisfaction responses.

The approach presented above in equations (4) to (10) breaks TWP down into two elements: market and non-market. This is useful to understand why countries differ in well-being and the relative roles of productivity and the capital stocks in explaining cross-country variation in well-being. Importantly, this provides a framework for empirically assessing aspects of the capital stocks model. In particular, there are three key relationships to be tested:

I. If the capital stocks are not important drivers of non-market outcomes (i.e. $\beta_1\gamma_1 = 0$ or $\beta_1\gamma_2 = 0$) then the capital stocks model is fundamentally broken.

II. We can also compare whether the role of the capital stocks in producing non-market outcomes is similar to that for market outcomes (i.e. test whether $\frac{\beta_1\gamma_1}{\beta_2\gamma_2} = \frac{\alpha_1}{\alpha_2}$).

III. Finally, it is interesting to see whether the relationship between TFP for market outcomes is similar to that for non-market outcomes (i.e. is there a consistent linear relationship between A_c and a_c).

The models discussed above focus on developing an estimate of non-market productivity comparable to traditional estimates of TFP. However, the capital stocks model of well-being usually incorporates four different capital stocks rather than just two: produced capital, human capital, natural capital, and social capital. If measures of natural capital and social capital are available, extending equations (3), (4) and (5) to include the full range of capitals in the capital stocks model is straight forward. If S_c is a measure of country-level social capital, such as generalized trust (Smith, 2020), and N_c is a measure of the overall stock of natural capital then:

$$W_c = \breve{A}_c K_c^{\rho_1} L_c^{\rho_2} N_c^{\rho_3} S_c^{\rho_4} \tag{12}$$

$$Y_c = A_c K_c^{\alpha_1} L_c^{\alpha_2} N_c^{\alpha_3} S_c^{\alpha_4} \tag{13}$$

$$C_c = \ln\left(a_c K_c^{\gamma_1} L_c^{\gamma_2} N_c^{\gamma_3} S_c^{\gamma_4}\right) \tag{14}$$

This extension of the model allows testing the significance of social and natural capital and the impact of their inclusion in the model on the coefficients for produced capital and human capital.

All of the models estimated in the article use a simple cross-sectional regression strategy with robust standard errors to control for clustering of observations at the country level. While a fixed effects regression would be possible with the cross-country panel dataset used here, the residual in such a regression could not be interpreted as a measure of TFP. We are, however, able to test directly for the impact of bias in the cross-sectional model by comparing estimated TFP from the model with methodologically independent estimates of TFP from the Penn World Tables.

Data

Four data sources are used in the empirical section of this article. These are the European Social Survey (ESS), the Corruption Perceptions Index, the Penn World Tables, and the BII (Phillips *et al.*, 2021).[6] Information on life satisfaction and trust is provided by the ESS. The ESS is a biennial survey of attitudes, values, and beliefs run across 38 countries in Europe since 2002. Using the ESS cumulative dataset gives information on 9 waves of the survey covering 2002 to 2018 and 427,656 valid responses. This information is collapsed to produce a cross-country panel dataset containing the mean life satisfaction and mean generalized trust score for each country and survey wave. Individual level data from the ESS is also used to provide an estimate of cultural response bias.

Interpersonal trust is, perhaps, the best single measure of social capital (Smith, 2020) in the sense in which it is used in the capital stocks model (i.e. as a productive resource). However, there is a risk that the correlation between interpersonal trust and life satisfaction at the country level might be biased due to shared method variance (OECD, 2013a). The Corruption Perceptions Index is a composite indicator of public sector corruption produced by Transparency International. It covers 180 countries and is comparable for time series purposes from 2012 onwards. Sources for the Corruption Perceptions Index come from 13 different surveys and expert assessments (Transparency International, 2020). Importantly, these assessments are external to the countries under evaluation meaning that — unlike the ESS trust measure — there is no risk of correlation with life satisfaction due to survey effects or cultural response bias. However, as illustrated in Chart 2, the Corruption Perceptions Index is strongly correlated with generalized trust across countries. On this basis the Corruption Perceptions Index is used as a proxy measure of social capital in the growth regressions that form the core of this article.

Information on GDP, produced capital, human capital, and market TFP[7] was obtained from the Penn World Tables (Feenstra, Inklaar, and Timmer, 2015), covering the same period as for the ESS. Although estimates of TFP in the next section are derived directly from the Solow-Swan residual, the Penn World Table measure of TFP provides a useful validity check to ensure that the cruder approach required here for consistency with the TWP measures is not introducing any systematic bias.

Table 1 presents the variables used in the analysis along with basic descriptive information. Real GDP per capita is output GDP at constant prices (PPP) across countries in 2017 US dollars and divided by population. Following Inklaar, Woltjer, Albarrán and Gallardo (2019), the capital services measure from the PWT divided by population is used for produced capital per capita (K_c). Human capital per capita is an index calculated as persons engaged in the labour market multiplied by average hours worked multiplied by the PWT human capital index divided by population.

6 The dataset constructed herein is available to researchers upon request.

7 The term market here is used to distinguish production that falls within the scope of measured GDP from other wider drivers of life satisfaction such as health status, safety, or social contact rather than in the sense of distinguishing private sector from government activity. Market TFP is therefore used to refer to the Penn World Tables measure of TFP for the total economy.

Chart 2: Comparison of the Corruption Perceptions Index and ESS Interpersonal Trust Scores, 2012 to 2018

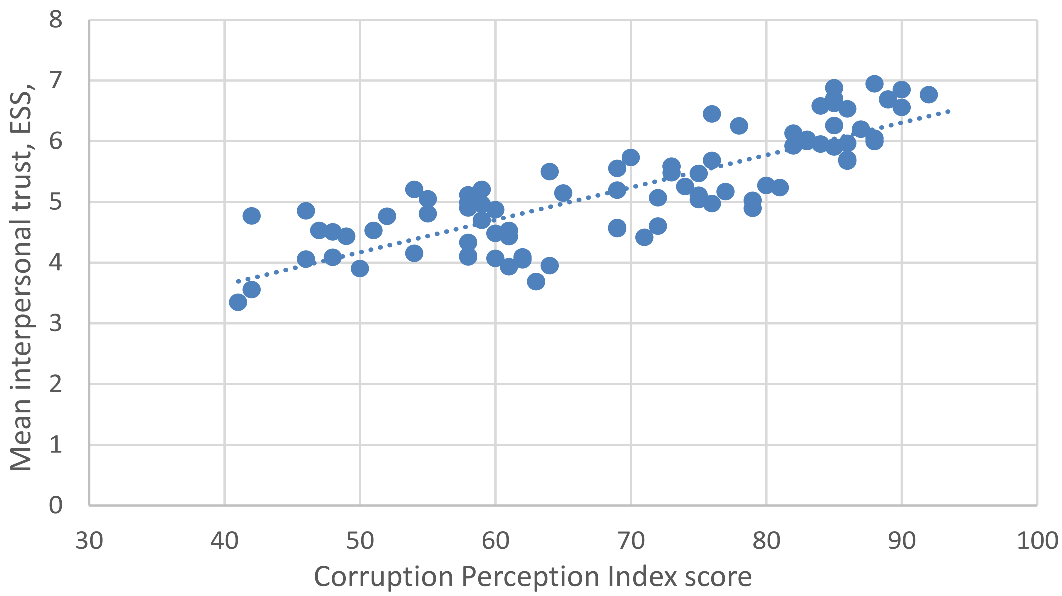


Table 1: Cross-Country Dataset

Variable	Min	Max	Mean	Observations	Country coverage	Years covered	Source
Real GDP per capita in 2017 $US (Y)	13082	92226	35667	206	31	2002-2018	PWT
Capital services level per capita million 2017 $US (K)	0.00037	0.00628	0.00232	206	31	2002-2018	PWT
Human capital per capita (L)	1165	3547	2504	206	31	2002-2018	PWT
TFP at current PPP (cTFP)	0.549	1.511	0.869	206	31	2002-2018	PWT
Mean life satisfaction (W)	4.54	8.54	7.15	206	31	2002-2018	ESS
Mean interpersonal trust	3.35	6.95	5.2	206	31	2002-2018	ESS
Corruption perceptions index (S)	41	92	69	102	34	2012-2020	Transparency International
Biodiversity Intactness Index(N)	0.406	0.96	0.715	223	36	2002-2018	Natural History Museum
Cultural response bias (θ)	-0.321	0.593	0.169	31	31	n/a	ESS - derived

The BII is an index developed by the UK Natural History Museum (Phillips *et al.*, 2021) that summarises the impact of human pressures on ecosystems. It is based on the estimated percentage of the original species that remain and their abundance within a given area. The BII is intended as a proxy measure for total natural capital per capita that is more inclusive than alternative estimates such as that produced by the World Bank (2006) which are built from a "bottom-up" approach with individual components added over time (Engelbrecht, 2015). The land-cover approach taken here avoids the bias due to missing components issues with the World Bank dataset at the expense of greater measurement error. It also helps avoid some of the issues of multicollinearity associated with the World Bank's dollar value estimates of capital stocks.[8]

Previous studies of well-being and the four capital stocks (Vemuri and Costanza, 2006) found multicollinearity between measures of capital caused significant econometric issues in estimating the relationship between different capital stocks and well-being. The datasets used here suffer significantly less from multicollinearity than those used by Vemuri and Costanza. The only statistically significant bivariate correlation between the capital stocks in this study is between produced capital and social capital which are correlated with an r value of 0.67. This difference in capital stock measures is almost certainly due to the fact that the dataset used by Vemuri and Costanza reports the dollar value of the capital stocks – thus ensuring that stock measures are correlated at the country level through price levels[9] — while the measure of human capital used here is a simple index of labour force variables.

Adjusting for cultural response bias is one of the most significant empirical challenges associated with the proposed analysis. The estimates of cultural response bias in Table 1 are derived from an analysis of the ESS based on equation (11). The full results of the model are not reported here[10] as the regression structure is relatively uninteresting and consists largely of two long vectors of dummy variables. Ideally it would be possible to test these estimates against other comparable estimates of cultural response bias, but there are relatively few comparable estimates available in the literature that could form the basis of a direct comparison.

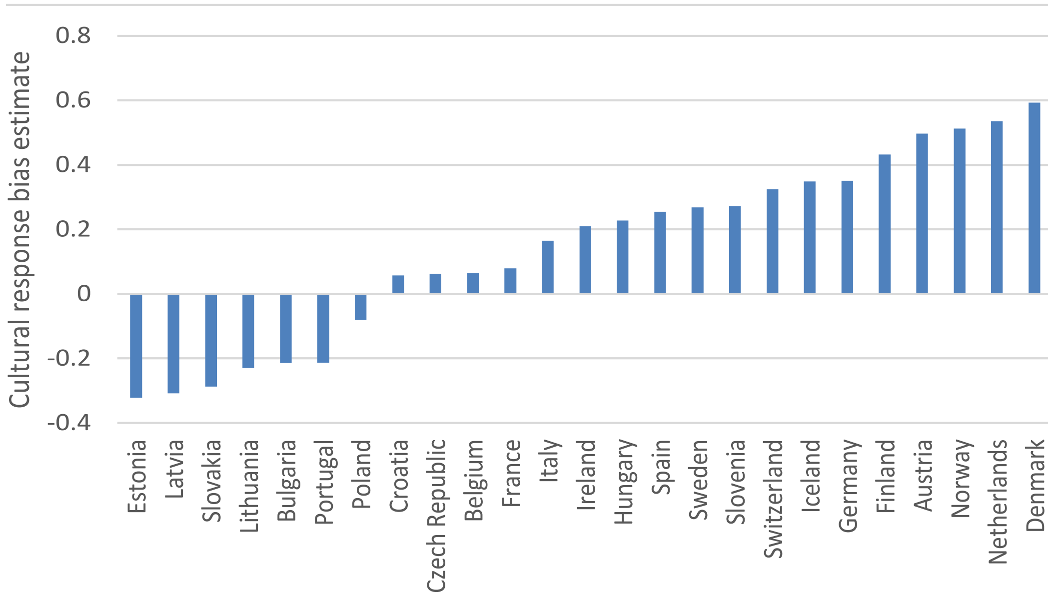
Exton, Smith, and Vandendreissche (2015) use a similar approach to identifying cultural response bias and find that it accounts for a maximum of approximately 20 per cent of cross-country variation in life satisfaction. However, they do not provide country-specific estimates. Senik (2014) uses a slightly more sophisticated version of the same approach and obtains estimates of

8 Additional information on the methodology of the BII can be found at https://www.nhm.ac.uk/our-science/data/biodiversity-indicators/about-the-biodiversity-intactness-index.html.

9 For example, the price of human capital — the wage rate — is a function not only of years of schooling and work experience, but also of the capital to labour ratio in the country and is therefore correlated with measures of produced capital.

10 Full regression results are available on request from the authors.

Chart 3: Estimated Cultural Response Bias



Notes: 1. Y-axis values are mean cultural difference in life satisfaction (0-10)

cultural response bias for a relatively small number of countries. In Senik's analysis the Nordic countries (Norway, Sweden, and Denmark) are characterized by a high positive bias in life satisfaction, while Portugal and France have a small negative bias. The only ex Eastern-bloc country reported by Senik has the largest negative coefficient. This pattern is replicated in Chart 3, which shows the cultural response bias estimates used in this paper.

Results

Table 2 reports the results of a well-being regression based on equations (3) and (12). This captures the combined effect of the capital stocks on well-being from both market and non-market outputs. Columns (A) and (E) correspond to model (3) while columns (D) and (H) correspond to model (12). The intermediate columns (B), (C), (F), and (G) add natural capital and social capital independently to model (3). As a sensitivity test, the same analysis is repeated twice. The first four columns of Table 2 (A) to (D) use mean life satisfaction adjusted for cultural response bias as the dependent variable, while the second four columns (E) to (H) use unadjusted mean life satisfaction. The data underlying these regression models cover the period from 2002 to 2018.

A comparison of the models using adjusted life satisfaction and those using unadjusted life satisfaction shows very little qualitative difference between them with the exception that produced capital (K) has a larger impact on unadjusted life satisfaction under all model specifications than it does on adjusted life satisfaction. Both human capital (H) and produced capital are consistently significant across all model specifications as is social capital (S) when it is included. Natural capital (N) is significant when included alongside human capital and produced capital but loses significance when social capital is added. An examination of the R^2 shows that the nat-

Table 2: Full Capital Stocks Model

Variable	Life Sat (adjusted for cultural response bias)				Life Sat (not adjusted for cultural response bias)			
	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**	**(G)**	**(H)**
$\ln(L)$	1.31**	1.17*	0.99**	0.94*	1.33*	1.20^	0.94^	0.92^
$\ln(K)$	0.79**	0.81**	0.24*	0.27	1.15***	1.17***	0.48^	0.49^
$\ln(N)$		0.53*		0.20		0.50		0.09
$\ln(S)$			1.85***	1.79***			2.25***	2.22***
Adj. R^2	0.457	0.482	0.681	0.682	0.531	0.542	0.733	0.730

Notes: $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05, ^{\wedge}p < 0.1$

Table 3: Market and Non-market Decomposition

Variable	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\widehat{W_c} - \theta_c$	$\widehat{W_c} - \theta_c$	$\widehat{W_c} - \theta_c$	$\widehat{W_c} - \theta_c$
	(J)	**(K)**	**(L)**	**(M)**	**(N)**	**(P)**	**(Q)**	**(R)**
$\ln(L)$	0.32	0.35^	0.23	0.28^	0.93*	0.75*	0.71^	0.61^
$\ln(K)$	0.66***	0.65***	0.51***	0.48***	0.02	0.04	-0.36^	-0.30
$\ln(N)$		-0.12		-0.22		0.67		0.46
$\ln(S)$			0.51**	0.57**			1.25***	1.12**
Adj R^2	0.767	0.769	0.813	0.826	0.085	0.168	0.279	0.310

Note: 1. $^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05, ^{\wedge}p < 0.1$

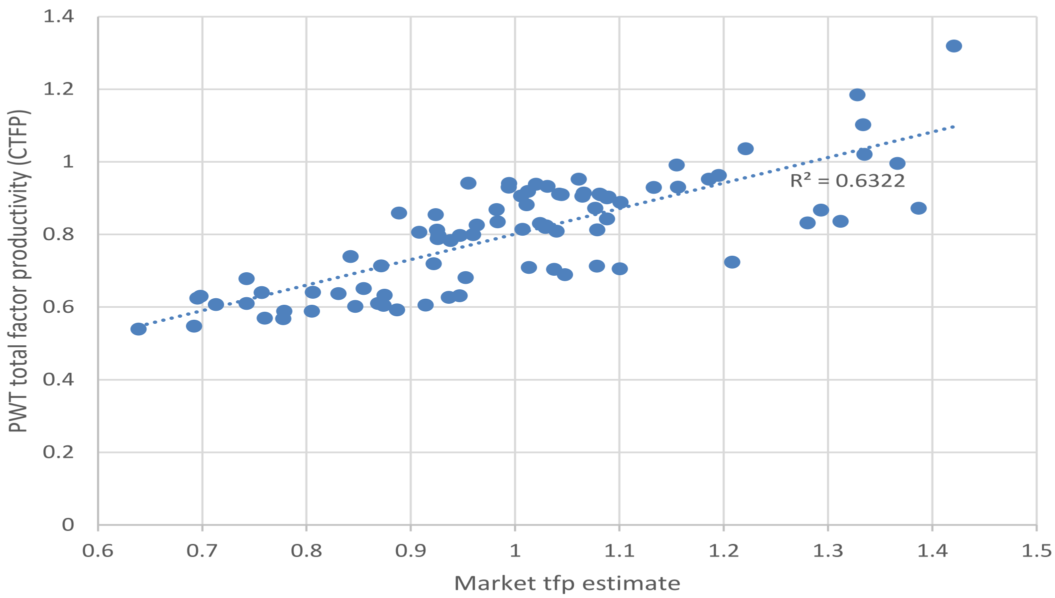
ural capital measure used here adds relatively little to the total variance explained compared to the other three measures.

Broadly speaking the results in Table 1 can be considered supportive of the capital stocks model in that all coefficients have the expected sign and all are significant except natural capital in columns D, F, and H. There is clearly some evidence of an interaction between the social capital measure used here and produced capital, with produced capital having a much lower coefficient once social capital is included in the model. This may reflect the impact of omitted variable bias in the regression where produced capital is correlated with TFP and social capital explains a significant proportion of conventional TFP measures (Smith, 2020; Coyle and Lu, 2020).

Table 3 examines the relative contributions of the capital stocks to market and non-market output. Columns (J) to (M) estimate equation (6) while columns (N) to (R) estimate equation (10). It is apparent that the picture for market outcomes is generally similar to that for overall well-being (Table 2). Human capital, produced capital, and social capital all have positive and significant coefficients. In contrast to Table 2, human capital has a smaller impact than produced capital on market outcomes and is insignificant when social capital is included on its own (L) and is very marginally insignificant in the basic model ($p = 0.104$). The main difference between market outcomes in Table 3 and the results in Table 2 is that the relationship between natural capital and market output is negative and not significant.

The situation for non-market outcomes

Chart 4: Model Estimates of Total Factor Productivity vs Penn World Table Estimates



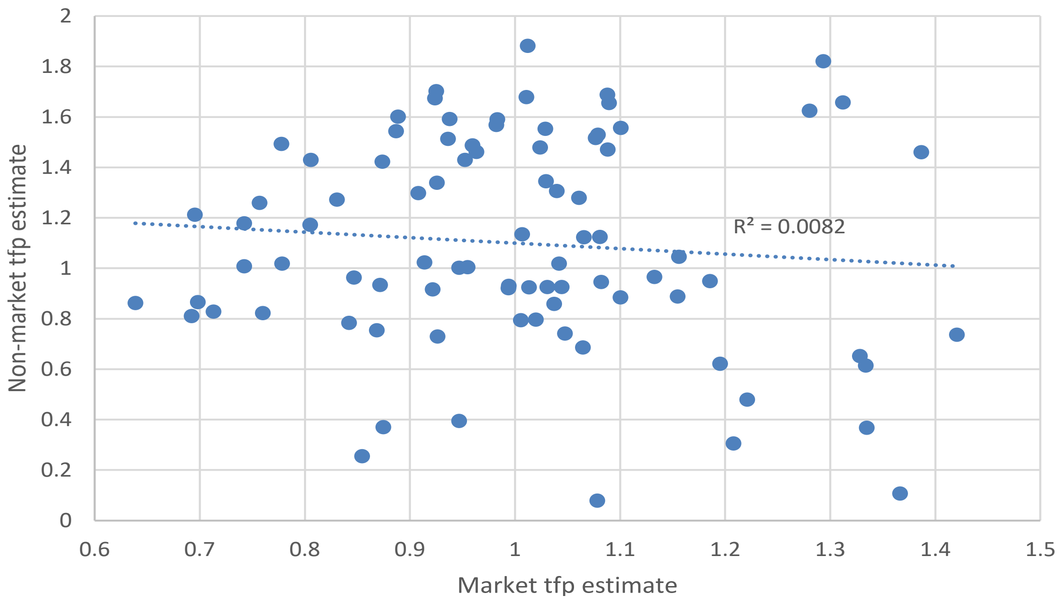
is quite different. Both human capital and social capital are significant in all versions of the model. Produced capital is insignificant in the first two model specifications (N) and (P) but has a marginally significant negative coefficient in model Q, which includes social capital. This result is robust to the choice of adjusted or raw life satisfaction data as the dependent variable and to the choice of mean trust or the corruption perceptions index as the measure of social capital. This counter-intuitive result is likely to be grounded in our approach to estimating non-market well-being (equation 5) in that the empirical estimate of the effect of income on life satisfaction may also capture the positive impact of produced capital on life satisfaction since income and produced capital are correlated with each other.[11]

It should be noted that the coefficients in columns (N) to (R) cannot be directly compared to the coefficients for market goods in columns (J) to (M) as the non-market coefficients represent $\beta_1\gamma_n$ rather than γ_n. Coefficient ratios can be compared between the market and non-market regressions and it is interesting to note that the ratio of the coefficient for human capital to that for social capital is relatively similar across both sets of regressions. However, this is clearly not the case for produced capital.

With the results presented in Tables 2 and 3 it is possible to calculate a range of measures of TFP. These include TWP (TFP with respect to life satisfaction) from columns (A) to (D) of Table 2, market TFP from columns (J) to (M) of Table 3, and non-market TFP from columns (N) to (R) of Table 3. A useful validity test of the models presented in these Tables is to compare market TFP from column (J) of Table 3 to the estimates of TFP from the PWT (cTFP). Chart 4 shows a scat-

11 An example of this is that the non-market benefits provided by a roading system are likely to be highly correlated across countries with the impact of a roading system on market outcomes.

Chart 5: Model estimates of market and non-market productivity



terplot of market TFP against cTFP from the PWT. Although the correlation is only moderate[12], there is a clear linear relationship between the two measures.

Given that the estimate of market productivity is reasonable, it can be compared with an estimate of non-market productivity calculated in a similar way from column (N) of Table 3. This is presented in Chart 5. It is immediately evident from Chart 5 that there is essentially no correlation between market productivity and non-market productivity. This suggests that the production "technologies" of the market and non-market sectors are fundamentally different (i.e the way resources are combined to produce well-being is not similar for market goods and non-market goods).

Moving from non-market productivity, Chart 6 compares TWP to market TFP. Panel A of Chart 6 illustrates the relationship where productivity is calculated on the basis of produced and human capital only (columns A and J). In this instance the impact of social capital is folded into TFP. Panel B of Chart 6 compares productivity estimates based on columns (D) of Table 2 and (M) of Table 3. This gives a narrower measure of TFP with social capital now accounted for in the capital stocks and therefore not reflected in the productivity measure.

Since well-being is considered a function of both market and non-market output in the capital stocks model, it is unsurprising to see that there is a correlation between market TFP and TWP. However, this relationship is weak. It is evident in Panel A, but only barely exists in Panel B. Both panels in Chart 6 show significant differences in TWP across countries. Chart 7 explores this further, presenting the mean TWP over the 2002-2020 period for all the countries covered in Chart 6. Be-

12 Observations with high productivity in PWT but not in the residual are Ireland, Poland, and one observation for Bulgaria.

cause Chart 7 shows country mean values while Chart 6 includes estimates for each country/year observation, Chart 7 contains fewer data points.

One common criticism of TFP as a concept is that measures of it can be hard to interpret. This is doubly the case for the estimates of TWP provided here both because the dataset used is exploratory and because there is little literature to provide the basis for comparison. A few observations, however, can be made. First, accounting explicitly for stocks of social capital changes the picture of the Nordic countries in terms of the production of well-being. With the exception of Denmark — which records a relatively high TWP — most of the Nordic countries perform at around the average level despite relatively high life satisfaction. Norway is actually towards the bottom of the table which is consistent with the country's relatively high level of human, produced, social, and natural capital stocks contrasted against well-being levels not very different to the other Nordic countries.

Similarly, while a cross-country analysis of life satisfaction shows a strong post Eastern-bloc effect associated with lower levels of subjective well-being (Senik, 2014), looking at TWP shows a more diverse picture. While some former Eastern-bloc countries have a very low TWP (Bulgarian, Hungary), others are amongst the best performing (Poland, Croatia). All four countries are associated with similar low levels of social trust, but Poland and Croatia have far better well-being outcomes than would otherwise be expected.

Conclusion

This article investigates the concept of productivity from within the framework of the capital stocks model of well-being. In particular, it estimates TWP — the efficiency with which resources (the capital stocks) are used to produce well-being — as a Solow-Swan residual in a modified cross country growth regression. Although the dataset used here is more exploratory than definitive, it is possible to identify some interesting themes.

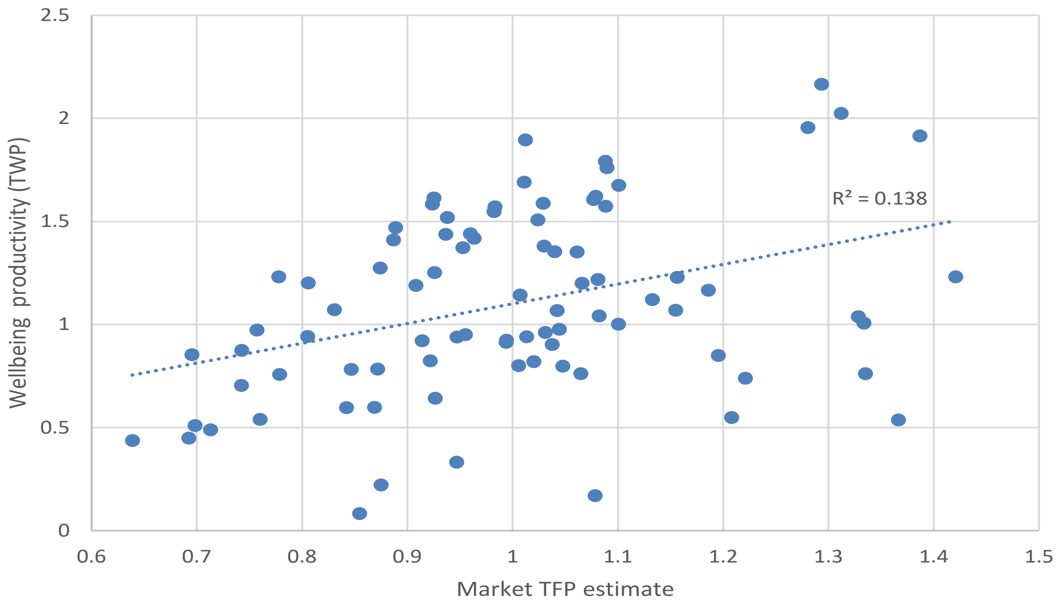
Main findings

There are three key findings from our initial exploratory analysis. First, there is considerable variation in TWP across countries. In other words, once differences in factor endowments are controlled for, there are still important differences in levels of well-being across countries. This is important because it suggests that there are ways to improve well-being that do not involve increasing the levels of the capital stocks. Reconciling the moral imperative to improve the well-being of the population living in less developed countries with the limits of a finite planet is, perhaps, the defining global policy challenge of the present time. Further investigation of TWP is therefore of some potential policy interest if it can offer insights into how some countries are able to achieve higher levels of well-being from a given capital endowment than others.

The second main finding is that the aggregate production functions for market and non-market outcomes appear to be very different. This can be seen both in the different coefficients for the capital stocks in the market production function com-

Chart 6: Well-being productivity compared to TFP with and without including social capital

Panel A: TFP includes social capital



Panel B: TFP does not include social capital

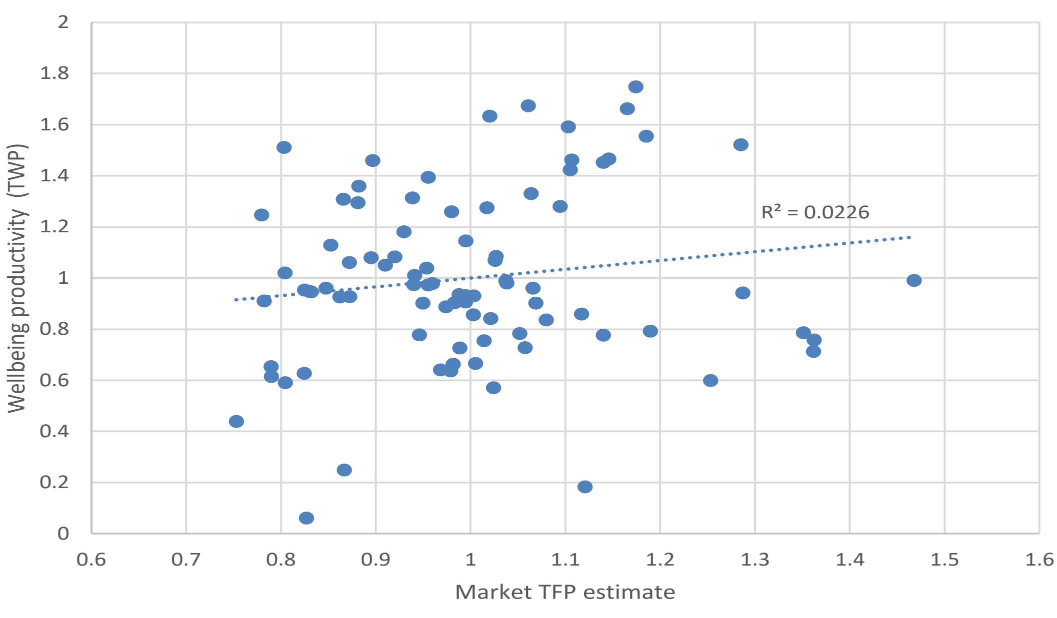
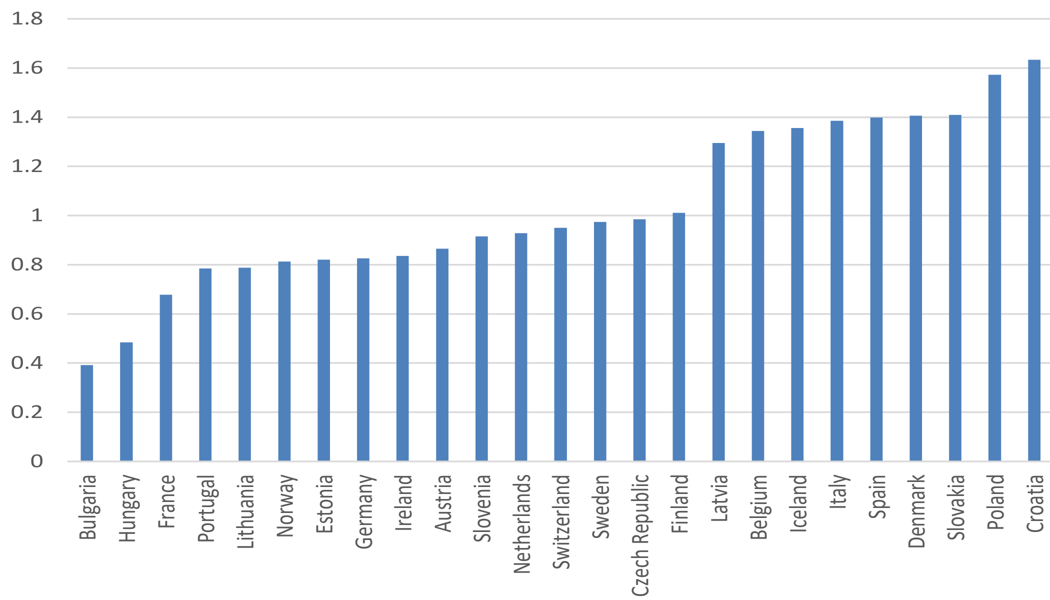


Chart 7: Mean Well-being productivity (TWP), 2002-2020



pared to the non-market production function and also in the lack of correlation between TWP and traditional TFP measures. One empirical implication of this is that policies aimed at maximizing market output will not necessarily maximise total well-being as the non-market elements of well-being have very different drivers. If the relationships estimated in this article hold, it also suggests that investments in human and social capital — which have a clear positive impact on both market and non-market outcomes — might be expected to have a larger impact on overall well-being than investments in produced capital (which has a positive correlation only with market outcomes). This is consistent with the case made elsewhere for the importance of social and human capital (World Bank, 2006; Helliwell, Huang, and Wang, 2017)

Finally, the empirical analysis confirms that the capital stocks are significant in the production function for well-being. The levels of produced, human, and social capital all have the expected relationship with overall well-being which supports the relevance of the capital stocks model as a way of conceptualising intergenerational well-being. Natural capital is an exception here, showing only a weak relationship with life satisfaction which vanishes when social capital is included in the model. The decomposition of well-being into market and non-market outcomes illuminates this issue showing a negative relationship between natural capital and market outcomes but a strong positive relationship between natural capital and non-market outcomes. One hypothesis suggested by this is market output is associated with the depletion of natural capital resources now and in the past (Qasim and Grimes, 2021) which results in a negative relationship between the current level of market output and natural capital. Non-market outcomes, on the other hand, might be associated more closely with nondepleting uses of natural capital.[13]

Limitations

This article is intended to be exploratory, and it is important therefore to acknowledge that it has significant limitations. Three of these are particularly important. First, the residual approach to estimating TFP faces the inherent issue that the residual of any regression analysis will also incorporate the error term. This is compounded in estimating TWP in that it is necessary to adjust life satisfaction to account for potential cultural response bias. This means that the estimate is effectively a residual of a residual, with potential error on both sides of the equation.

The issue of adjusting for cultural response bias, however, goes beyond the issues associated with calculating productivity as a residual. As discussed earlier in the paper, cultural response bias is challenging to estimate. Because it cannot be observed directly and is difficult to distinguish from substantive differences in well-being caused by unobserved omitted variables, cultural response bias is difficult to control for in a robust fashion. Perhaps the best that can be hoped for here is to test the sensitivity of results to estimates of cultural response bias based on different methodologies.

Even if issues in the estimation of TWP are ignored, there are still significant challenges in interpreting the results. The decomposition of TWP into market productivity and non-market productivity illustrates this issue. While market productivity is simply conventional TFP and can be interpreted as such[14], non-market productivity is more complicated to interpret. Because non-market consumption (C_c) is a vector not a quantity (i.e. consists of multiple different outcomes with no obvious common metric such as health status, safety, and social contact), estimated differences in non-market productivity might be due to differences in the relative make-up of C_c across countries rather than differences in the effectiveness with which the capital stocks are used. Different aspects of non-market consumption — such as health status and social contact — might be expected to have different production technologies. With the approach to estimating non-market productivity adopted here it is impossible to distinguish between different non-market consumption bundles and differences in the quality of non-market production technology.

Given the issues identified above, what is the value of attempting to estimate TWP? First, looking at TWP is important simply because the concept is implicit in the most widely adopted approaches to measuring well-being and assessing sustainability. This can be seen in the academic literature on the capital stocks model (Engelbrecht, 2009; Arrow *et al.*, 2012; Qasim and Grimes, 2018), the approach taken by international organizations (World Bank,

13 For example, consider a forest. The use of the forest's wood resources for market outcomes is likely to have a negative impact on the forest ecosystem in a way that the forest's provision of ecosystem services for non-market outcomes (such as air quality or recreational use) does not.

14 Note that the interpretation of conventional TFP is not, itself, uncomplicated. TFP has no natural units and the aggregate production function approach to estimating TFP has been criticized (Felipe and McCombie, 2006).

2206; OECD, 2013, 2015; Hamilton and Liu, 2013), and in the analytical frameworks adopted by governments (OECD, 2016; Ormsby, 2018, National Economic and Social Development Office, 2021). Because the capital stocks model is used to inform and evaluate policy decisions it is important to test it. The limitations identified above exist, regardless of whether the model is used in a quasi-anecdotal fashion to justify indicator dashboards or if it is taken more seriously as a quantitative model. However, it is only by exploring the implications of the capital stocks model in a quantitative fashion that some of these limitations are identified.

It is also important to reflect that the challenges associated with estimating TWP are not unique. Market consumption may have a common metric in terms of market prices, but it is fundamentally just as much a vector of different elements as is non-market consumption. This is of particular relevance in the context of the produced capital stock (K). The so-called Cambridge capital controversy, for example, largely revolved around precisely the issue of whether the capital stock could reasonably be treated as a single quantity when it, in fact, consisted of a wide range of different capital items that were not necessarily good substitutes for each other (Cohen and Harcourt, 2003). What is interesting in this comparison is that, while the criticisms of the notion of a single capital stock are clearly valid, this has not prevented analyses of economic growth based on aggregate production functions contributing useful insights. Modern endogenous growth theory, for example, builds on and extends this framework (Romer, 1994).

Next steps

If the idea of TWP is worth exploring further, what are the next steps in this research agenda? There would appear to be two obvious directions to explore. First, better data would significantly improve the quality of TWP estimates compared to the analysis in this paper. The ESS focuses only on a relatively small number of high-income countries with relatively high levels of well-being and is thus not the ideal dataset from the perspective of examining variation in well-being outcomes. This could be addressed either through extending the analysis to include other similar datasets such as the World Values Survey or various national general social surveys (Fleischer, Smith, and Viac, 2016). Alternatively, the Gallup World Poll would provide a potentially suitable dataset covering a wider range of countries and with better ability to model cultural response bias (Exton, Smith, and Vandendreissche, 2015).

Better measures of the capital stocks are also important. While social capital might seem relatively abstract, the most widely used proxy measures function well (Smith, 2020). Natural capital, on the other hand is extremely difficult to measure. Existing measures tend to be either account for only a small proportion of the total natural capital stock or — as is the case with the proxy measure used in this paper — simply perform poorly.

There is also clearly scope to move beyond the relatively simplistic analytical framework used in this paper. Two obvious extensions would be to explore treating non-market consumption explicitly as a multi-dimensional vector and looking at whether there is evidence of different production "technologies" across the different aspects of non-market production. Introducing non-market consumption also raises the issue as to whether use of the capital stocks is rival across different outputs. Clearly, some elements of the capital stocks are strictly rival in that, if they are used to produce one output, they cannot be used to produce another. However, for other elements this is less the case. An educated worker is more productive in the paid market and is also likely to be more effective in producing non-market outputs.

Finally, if TWP can be measured — even with significant noise — it becomes possible to ask what drives differences between countries. This is a tremendously important policy issue globally, since there is limited scope to increase consumption of some capital stocks globally — particularly natural capital — but low levels of well-being in much of the world suggest that there is likely to be significant pressure to raise well-being. This tension suggests that identifying the drivers of TWP adds a potentially important dimension to growth economics.

References

Arrow, K.J., P. Dasgupta, L.H. Goulder, K.J. Mumford, and K. Oleson (2012) "Sustainability and the Measurement of Wealth," *Environment and Development Economics*, Vol. 17, No. 3, pp. 317-353.

Boarini, R., M. Comola, C. Smith, R. Manchin, and F. de Keulenaer (2012) "What Makes for a Better Life? Determinants of Subjective Well-being in OECD Countries: Evidence from the Gallup World Poll," OECD Statistics Working Paper 2012/3, OECD Publishing.

Butlin, J. (1989) "Our Common Future. By World Commission on Environment and Development. (London, Oxford University Press, 1987, pp. 383£ 5.95.)," *Journal of International Development*, Vol. 1, No. 2, Book Review, pp. 383.

Clark, A., S. Flèche, R. Layard, N. Powdthavee, and G. Ward (2018) *The Origins of Happiness* (Princeton, N.J.: Princeton University Press.)

Cohen, A.J. and G.C. Harcourt (2003) "Retrospectives: Whatever Happened to the Cambridge Capital Theory Controversies," *Journal of Economic Perspectives*, Vol. 17, No. 1, pp. 199-214.

Costanza, R., L. Daly, L. Fioramonti, E. Giovannini, I. Kubiszewski, L. F. Mortensen, . . . and R. Wilkinson (2016) "Modelling and Measuring Sustainable Wellbeing in Connection with the UN Sustainable Development Goals," *Ecological Economics*, Vol. 130, pp. 350-355.

Coyle, D. and S. Lu (2020) "Trust and Productivity Growth — An Empirical Analysis," Bennett Institute Working Paper, University of Cambridge, June.

Deaton, A. (2008) "Income, Health, and Well-being Around the World: Evidence from the Gallup World Poll," *Journal of Economic Perspectives*, Vol. 22, No. 2, Spring, pp. 53-72.

Engelbrecht, H. J. (2009) "Natural Capital, Subjective Well-being, and the New Welfare Economics of Sustainability: Some Evidence from Cross-country Regressions," *Ecological Economics*, Vol. 69, No. 2, pp. 380-388.

Exton, C., C. Smith, and D. Vandendriessche (2015) "Comparing Happiness Across the World: Does Culture Matter?," OECD Statistics Working Paper 2015/4, OECD Publishing.

Feenstra, R.C., R. Inklaar, and M.P. Timmer (2015) "The Next Generation of the Penn World Table," *American Economic Review*, Vol. 105, No. 10, pp. 3150-3182.

Felipe, J. and J. McCombie (2006) "The Tyranny of the Identity: Growth Accounting Revisited," *International Review of Applied Economics*, Vol. 20, No. 3, pp. 283-299.

Fleischer, L., C. Smith, and C. Viac (2016) "A Review of General Social Surveys," OECD Statistics Working Paper 2016/09, OECD Publishing.

Frijters, P., A.E. Clark, C. Krekel, and R. Layard (2020) "A Happy Choice: Wellbeing as the Goal of Government," *Behavioural Public Policy*, Vol. 4, No. 2, pp. 126-165.

Hamilton, K. and G. Liu (2013) "Human Capital, Tangible Wealth, and the Intangible Capital Residual," OECD Statistics Working Papers 2013/02, OECD Publishing. http://dx.doi.org/10.1787/5k4840h633f7-en.

Helliwell, J.F., H. Huang, and S. Wang (2015) "The Geography of World Happiness," World Happiness Report 2015, pp. 12-41.

Helliwell, J.F., H. Huang, and S. Wang (2017) "The Social Foundations of World Happiness," World Happiness Report 2017, pp. 8-47.

Inklaar, R., P. Woltjer, and D.G. Albarrán (2019) "The Composition of Capital and Cross-country Productivity Comparisons," *International Productivity Monitor*, No. 36, Spring, pp. 34-52.

Kahneman, D., E. Diener, and N. Schwarz (eds.) (1999) *Well-being: Foundations of Hedonic Psychology*, Russell Sage Foundation.

Kapteyn, A., J. P. Smith, and A. van Soest (2010) "Life Satisfaction" in Ed Diener, Daniel Kahneman, and John Helliwell, *International Differences in Well-Being*, pp.70-104.

National Economic and Social Development Office (2021) "Ireland's Wellbeing Framework: Consultation Report," Council Report No. 155, National Economic and Social Council.

OECD (2011) *How's Life?: Measuring Well-being*, OECD Publishing.

OECD (2013a) *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing.

OECD (2013b) *How's Life? 2013: Measuring Well-being*, OECD Publishing.

OECD (2015) *How's Life? 2015: Measuring Well-being*, OECD Publishing.

OECD (2016) *Measuring and Assessing Well-being in Israel*, OECD Publishing.

Ormsby, J. (2018) "The Relationship Between the Living Standards Framework and the Sustainable Development Goals," New Zealand Treasury Living Standards Series: Discussion Paper 1806, Office of the Chief Economic Advisor, The Treasury.

Phillips, H., A. De Palma, R.E. Gonzalez, S. Contu, S.L.L. Hill, A. Baselga, L. Borger, and A. Purvis (2021) *Dataset: The Biodiversity Intactness Index — Country, Region and Global-level Summaries for the Year 1970 to 2050 Under Various Scenarios*, Resource: long_data.rds, Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/he1eqmg1.

Qasim, M. and A. Grimes (2021) "Sustainability and Wellbeing: The Dynamic Relationship Between Subjective Wellbeing and Sustainability Indicators," *Environment and Development Economics*, pp. 1-19.

Romer, P.M. (1994) "The Origins of Endogenous Growth," *Journal of Economic perspectives*, Vol. 8, No. 1, pp. 3-22.

Roscoe, L.J. (2009) "Wellness: A Review of Theory and Measurement for Counsellors," *Journal of Counseling & Development*, Vol. 87, No. 2, pp. 216-226.

Sacks, D.W., B. Stevenson, and J. Wolfers (2012) "The New Stylized Facts About Income and Subjective Well-being," *Emotion*, Vol. 12, No. 6, pp. 1181-1187.

Sen, A. (1993) "Capability and Well-being," in Martha Nussbaum, Amartya Sen, *The Quality of Life*, pp. 30-53.

Sen, A. (1999) *Development as Freedom*, (New York: Alfred Knopf).

Senik, C. (2014) "The French Unhappiness Puzzle: The Cultural Dimension of Happiness," *Journal of Economic Behavior & Organization*, Vol. 106, pp. 379-401.

Smith, C. (2018) *Treasury Living Standards Dashboard: Monitoring Intergenerational Wellbeing*, Kotata Insight, Wellington.

Smith, C. (2020) "Trust and Total Factor Productivity: What Do We Know About Effect Size and Causal Pathways?" *International Productivity Monitor*, No. 38, Spring, pp. 3-28.

Solow, R.M. (1956) "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, Vol. 70, No. 1, pp. 65–94.

Stiglitz, J.E., A. Sen, and J.P. Fitoussi (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, The Commission.

Swan, T.W. (1956) "Economic Growth and Capital Accumulation," *Economic Record*, Vol. 32, No.2, pp. 334-361.

Transparency International (2020) "Corruption Perceptions Index Technical Methodological Note."

UNECE (2014) *Conference of European Statisticians Recommendations on Measuring Sustainable Development*, United Nations.

Vemuri, A.W. and R. Costanza (2006) "The Role of Human, Social, Built, and Natural Capital in Explaining Life Satisfaction at the Country Level: Toward a National Well-being Index (NWI)," *Ecological Economics*, Vol. 58, No. 1, pp. 119-133.

World Bank (2006) *Where is the Wealth of Nations? Measuring Capital for the 21st Century*, The World Bank.

Trust, Deep Trust, Productivity and Well-being in 136 Countries

Tim Hazledine[1]
University of Auckland

Abstract

The article explores the role of generalized or social trust (trust between strangers) in explaining cross-country differences in the level of productivity (output per worker) and in self-reported well-being for 136 countries. Trust is measured directly from survey data. In addition, a second trust variable called deep trust is estimated as a function of ancient cultural, historical, geographic, and linguistic factors.

Both trust variables have significant bivariate relationships with each of productivity and well-being, each of which can also be modelled with fairly standard specifications: an augmented production function for productivity, and the multi-variate model of well-being developed in the annual World Happiness Reports. Yet when either trust variable is added to each of the standard models, neither contributes any additional explanatory power.

So where is the bivariate significance of trust coming from? We find that, in every case, one or both of trust and deep trust is significant for the standard determinants of productivity, with deep trust doing better at predicting human capital, physical capital and institution quality, and actual trust being stronger for the well-being determinants. That is, trust in the 21st century appears to not directly contribute to productivity or well-being, but has a substantial effect working through the proximate determinants.

As so often, Adam Smith (1723-1790) said it first and probably best:

> In civilised society [man] stands at all times in need of the co-operation and assistance of great multitudes, while his whole life is scarce sufficient to gain the friendship of a few persons. (Smith, 1776:22 [1998])

Why are the multitudes needed? Because of the division of labour, which in this period just before the Industrial Revolution was the prime source of prosperity (along with still-uncrowded land). Smith carries on to his more famous passage:

> It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our

1 Emeritus Professor of Economics at the University of Auckland. Suggestions and encouragement from Guy Mayraz, Robert Wade, and, especially, John Helliwell, are gratefully acknowledged, as are comments received at the NZAE Annual Conference in Wellington, July 3-5, 2019, and suggestions from my discussant, Sjoerd Beugelsdijk, at the November 16-17, 2021 IPM Productivity Workshop. Very useful comments and queries were provided by two anonymous referees, Andrew Sharpe, and Chris Haun. Email: t.hazledine@auckland.ac.nz.

> dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.
> (Smith, 1776:22 [1998])

But Adam Smith — a moral philosopher by trade — was well aware of the importance of 'humanity' to civilized life, including civilized economic life. All those strangers upon whom each 'man' depends: with no ties of kinship or friendship — well, what is to stop them cheating, robbing, neglecting, even injuring or killing the lonely butcher, etc, in the pursuit of their self-love?

At the least, the investments of time and capital to gain and equip the specialized skills of the chosen trade can make their possessor vulnerable to what we now call 'hold-up problems' — *ex post* revisions of the terms of trade when the specialist's outside options have been run down.

But Smith saw the possibility of something different. He was a prophet in both senses of the word. He foretold a new future, and he helped it come to pass by uncovering the guiding principles of modernity. People needed to learn how to behave properly with each other — in particular, with strangers. And this not — or mostly not — because of what we would now call altruism, and Smith called 'sympathy', which he noted is generally restricted to family and friends. Rather, what we now call bourgeois life requires something quite different from self-regarding altruism:[2] other-regarding adherence to a moral code: the voluntary tempering of self-love in action so as not to harm others, in the overall interest of society as a whole.

In Smith's setting of the division of labour, each new partition of tasks in the cause of specialization and productivity necessarily requires a new 'transaction' — an exchange of a good or service for money, and, in practice, the terms of such exchanges, which may be contingent on uncertain events, can seldom be pinned down unambiguously in advance. Therefore, as Kenneth Arrow put it, in his typically mild way:

> Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence....[3]
> (Arrow, 1972:357)

These brief sentences by a great modern economist can be taken as the first sign of a revival of interest in the 'soft' technologies of specialization and exchange, ne-

2 At least to economists, altruism is seen as someone else's well-being entering an agent's utility function: the agent obtains some of their own utility out of giving utility to others, and does so solely for this reason.

3 These two sentences are actually something of a throwaway digression in a paper subtly reviewing the sociologist Richard Titmuss's famous analysis of gift exchange and its commercial alternative in the matter of obtaining supplies of blood for medical use.

glected for the century and a half through which economics was dominated by the industrial revolution and its 'hard' technologies of substitution of physical capital for labour, as embodied by the late 1960s in the neoclassical production function and the Solow-Swan growth model.

Arrow's words contain the two key propositions which now drive empirical research on what we call generalized trust (trusting and trustworthiness between strangers): the idea that such trust is essential to economic life, and the prediction that follows, to the effect that differences in economic success across nations might therefore, at least partially, be explained by different endowments of generalized trust at the national level. We will in this article seek and identify the role of trust in accounting for the level of economic development, or prosperity, of countries, as manifested or proxied by the level of labour productivity.

A quite recent research program has added a broader concern for the well-being or 'happiness' of citizens, of which economic prosperity turns out to be an important part, but far from the whole. The annual *World Happiness Report* (*WHR*) uses the results of international personal surveys carried out by the Gallup organization. The report develops an empirical model to explain cross-country differences in average self-reported well-being in terms of six core factors, only one of which is per capita incomes — so, a 'production function' for well-being.

This article explores the role of generalized trust in supporting higher levels of economic development as well as levels of well-being, for a panel of recent annual data on a cross section of 136 countries. Two measures of trust are tested: actual trust as reported in Gallup-style personal surveys; and what I will call 'deep trust'; being the value of actual trust predicted by a set of plausibly exogenous cultural and social factors. The results, which are quite striking, can be summarized as follows:

First, simple bivariate regressions of, in turn, the level of productivity (GDP per employed person) and self-reported well-being, on trust deliver strongly significant coefficients for the trust variable, with deep trust performing somewhat better for productivity, and actual trust better for well-being. R^2 values are quite low — around 0.2.

Second, we can with these data easily replicate both the neoclassical GDP production function, usefully augmented by an index of institutional quality, and the *World Happiness Report* core well-being model, with quite high R^2 values.

Third, adding either trust variable to these standard models does not improve the fit.

So, where does the significance of trust in the bivariate models come from? We find that in all (nine) cases, the (three) regressors in the GDP function, and the (six) regressors in the *WHR* well-being model, have a strongly significant bivariate relationship with either trust variable. That is, social trust seems to be an input to the inputs. High trust levels do not in themselves make people happier or more productive. But high trust demonstrably encourages the long-term investments in physical and human capital, and in good institutions, that generate economic prosperity. And it somehow contributes to the various cultural and institutional factors that feed into well-being.

This article contains five main sections. First, we describe and model trust, deploying a perhaps surprising set of deeply exogenous factors found in the literature. The next two sections examine the impact of trust on differences in productivity, and on self-reported well-being, across 136 countries, with a panel of annual data covering various years from 2005 through 2017. These data build on and extend the *World Happiness Report* database. Then there is a brief case study of an apparently anomalous First World country, New Zealand, comparing it with a very similar country — Australia — which has however enjoyed a markedly different productivity performance. A final section concludes.

Modelling Trust and Deep Trust

How to measure and model trusting and trustworthy behaviour? Since the 1980s, the standard data source has been randomized surveys asking versions of the 'trust question' to people in different countries: *Generally speaking, do you believe people can be trusted or not?*, with the answer usually recorded as Yes/No, though sometimes a scale from 1 to 5 is allowed. Although this question literally measures views on the trustworthiness of others, it has been found to predict actual trustworthiness — i.e. the trustworthiness of the respondent — quite well. It seems also understood that respondents are thinking not of their friends or family nor of foreigners, but rather the trustworthiness of strangers in their own country or society.

By combining information from two sources of answers to the trust question — waves of the World Values Survey, and Gallup polls — as described in the Appendix, we are able to present trust data for 136 countries — far more than in any previous study of trust. Table 1 shows the descriptive statistics for trust and all other variables used in the article.

There is a rather large variation in the proportion of people in different countries who believe their fellow citizens to be trustworthy, with the distribution skewed towards the lower end of the range. Nearly three out of four Norwegians are trusting, but the sample average is only 23 per cent, and in the least trusting country — the Philippines — only one in thirty are foolish enough (as it would seem, in this case) to trust others. As some motivation for what follows, all rich countries — Western Europe and the New World — are in the top half of the trust distribution, with the least trusting — France — being at the median.

Can these cross-country variations in social trust be modelled empirically? Algan (2018) provides an up-to-date, insightful, extensive (more than 130 references) but not totally comprehensive review of the literature on the determinants and impacts of generalized trust, and its relationship to the concept of 'social capital'.

There is considerable evidence of systematic inter-regional differences in trust scores within countries, such as between regions within European nations and between the states of the American Union (Algan and Cahuc, 2014). It seems reasonable that such intra-national heterogeneity will generate statistical noise (but not bias) for analyses working with national average data, such that the calculated statistical stability of any coherent results achieved

Table 1: Descriptive Statistics

	TRUST	NOPRODROP	MONARCHY	MUSLIM	CATHOLIC	DIVERSITY
maximum	0.737 *Norway*	1.000	1.000	100	98	0.762 *Singapore*
minimum	0.032 *Philippines*	0.000	0.000	0	0	0.01 *Turkey, etc.*
average	0.231	0.222	0.134	22.50	30.13	0.365
standard deviation	0.141	0.416	0.340	33.92	34.17	0.214

	COLDEST	RGDPO/EMP	RNNA/EMP	HC	INSTITUTIONSAV	log(GDPPOP)
maximum	28 *Panama*	194314 *Ireland*	771062 *Italy*	4.36 *Uzbekistan*	10.85 *Finland*	11.465 *Luxembourg*
minimum	-21.6 *Mongolia*	1728 *Burundi*	3148 *Malawi*	1.13 *Burkina Faso*	-10.36 *Congo (Brazzaville)*	6.377 *Congo (Kinshasa)*
average	11.6	41782	201100	2.66	0.15	9.209
standard deviation	11.0	34108	200285	0.70	5.40	1.155

	LIFE_LADDER	SUPPORT	HLIFEEXP	FREEDOM	GENEROSITY	CORRUPT
maximum	8.02 *Denmark*	0.99 *New Zealand*	76.5 *Hong Kong*	0.985 *Uzbekistan*	0.678 *Myanmar*	0.983 *Hungary*
minimum	2.69 *Syria*	0.29 *Central African Republic*	37.8 *Sierra Leone*	0.258 *Bosnia and Herzegovina*	-0.323 *Greece*	0.035 *Singapore*
average	5.46	0.82	62.5	0.738	-0.001	0.752
standard deviation	1.12	0.12	8.0	0.141	0.167	0.187

with the latter will be underestimates of their true significance.

Bjornskov (2006, 2012 — not referenced in Algan (2018)) finds econometric evidence of some very long run determinants of current trust levels, indicating significant stability of trust over time. As collateral evidence, Bjornskov repeats from Zak and Knack (2001:295) a useful *apercu* from Adam Smith — "the Dutch are most faithful to their word" — and notes that of the "European countries that Smith would likely have had business [with], the Netherlands is to this day the nation with the highest trust score" (2006:3, note 2).

Direct corroboration of intergenerational transmission of trust is provided by Uslaner (2008; cited by Bjornskov), who uncovers a strong tendency for descendants of immigrants to the United States to reveal levels of trust similar to those of the current inhabitants of: the country to which they trace ultimate descent.

Bjornskov (2006) also follows Zak and Knack (2001), and La Porta *et al.* (1997) in finding that variables for predominance of either Muslim or Catholic religions in a country are negative for trust. This is attributed to the hierarchical nature of those religions creating 'vertical bonds of obligation in society that divide rather than

4 Bjornskov (2006:6), following Putnam (1993) through La Porta *et al.* (1997), who also include Eastern (Christian) Orthodox in their list of hierarchical religions – this was not found to be significant here.

unite people socially'.[4] Rather surprisingly, perhaps, constitutional monarchies are also moderately more likely to show higher trust levels, due to a perhaps under-appreciated role of impotent but venerable monarchies as national symbols of stability and cultural unity.[5]

Bjornskov (2006) also includes as regressors per capita GDP and income inequality, but these variables are both likely to be endogenous to trust, and are quietly omitted from the model of Bjornskov (2012). This model adds an even more surprising and authentically long-run exogenous factor: whether or not the predominant language of a country permits the dropping of personal pronouns. Linguists have apparently argued that forbidding the dropping of the pronoun is indicative of greater cultural respect for individual rights, and thence development of stronger trust norms (Lee, 2017).

In his earlier work, Bjornskov (2006) tried a measure of ethnic diversity (from Alesina *et al.*, 2003), which was not empirically very successful. In this article, a measure of religious diversity is constructed and included in the trust model. Diversity could be negative for trust not — or not just — because of prejudice against the presumed trustworthiness of people with different religious beliefs, but if it is harder to *predict* how different people will behave. In other words, diversity could undermine shared norms that facilitate expectations of trusting and trustworthiness.

It is sometimes suggested that climate is an important determinant of cultural traits, and the average daily temperature in the coldest month of the year will be included here — perhaps in harsh climates people have to learn to look after each other more.

Table 1 gives average and extreme values for all the variables used in this article. The variables are defined as follows:

- <u>TRUST:</u> A variable derived from responses to the survey question "*Generally speaking, do you believe people can be trusted or not?*" Values for this variable represent the proportion of individuals in a given country who responded YES to the survey question.
- <u>NOPRODROP:</u> A dummy variable indicating whether or not the main language of a country forbids the dropping of personal pronouns, where 1 = YES, 0 = NO.
- <u>MONARCHY:</u> A dummy variable indicating whether or not a country is a constitutional monarchy. 1 = CONSTITUTIONAL MONARCHY, 0 = NOT A CONSTITUTIONAL MONARCHY.
- <u>MUSLIM:</u> The percentage of a country's population that identifies as Muslim.
- <u>CATHOLIC:</u> The percentage of a country's population that identifies as Catholic.

5 Chaney (2018:647) discusses the view that Western Europe recovered better than Eastern Europe from the devastating 'Black Death' pandemic of 1350, because of more robust peasant organizations in the West. This is possibly attributable to monarchies in Western Europe seeking to strengthen peasant communities in efforts to undermine their real enemies — the nobility.

- DIVERSITY: Represents religious diversity, which is calculated as one minus the sum of the squared 'market shares' of each of five religions (Muslim, Catholic, Other Christian, Hindu, Buddhist) with the shares scaled to add to one. Higher values for the variable represent greater levels of religious diversity.

- COLDEST: The average daily temperature of the coldest month of the year in a given country.

- RGDPO/EMP: The real GDP per employed person in a given country, measured in constant 2011 national prices and expressed in 2011 US dollars. This data is sourced from the Penn World Tables (PWT) database (Feenstra *et al.* 2015) annually in most cases from 2007 to 2017.

- RNNA/EMP: The real capital intensity, or real capital stock per employed person, in a given country, measured in constant 2011 national prices and expressed in 2011 US dollars. This data is sourced from the Penn World Tables (PWT) database (Feenstra *et al.* 2015) annually in most cases from 2007 to 2017.

- HC: An index of a country's human capital based on average years of schooling and returns to education. See footnote for the procedure used to infill values in the event of missing data.[6]

- INSTITUTIONSAV: An index of the average quality of a country's institutions. The variable is an average of six dimensions: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. In the original (Kaufman et al. 2010), the variable is scaled to mean = 0 and ranges from Finland with the best institutions, at least from this perspective, and Congo (Brazzaville) with the worst.

The following variables are taken from the 2018 World Happiness Report:[7]

- LIFE_LADDER: Evaluation, on a scale of 0 to 10, of the respondent's current satisfaction with their life so far, interpreted as subjective well-being (SWB). Values for this variable represent the average response of a given country.

- Log(GDPPOP): Natural log of a country's per capita GDP in 2011 purchasing power parity international dollars, from World Bank, *World Development Indicators*[8]

- SUPPORT: A variable derived from responses to the question "*If you were*

6 Fourteen of our countries have no PWT human capital data, but do have years-of-schooling data as compiled by Barro and Lee (2013). For countries with both data sources, we can estimate a linear relationship between PWT and Barro-Lee numbers which has an $R^2 = 0.933$. We use this to infill HC numbers for the 14 countries.

7 For more information, refer to 'Technical Box 1' in the 2018 *World Happiness Report*.

8 This variable is of course very similar to the PWT variable RGDPO/EMP. The latter can be used in the well-being modelling reported in this section, and it gives very similar (slightly better) econometric results.

in trouble, do you have relatives or friends you can count on...?" where 1 = YES and 0 = NO. Values for this variable represent the proportion of individuals in a given country who responded YES to the survey question.

- HLIFEEXP: Healthy life expectancy at birth, from World Health Organization; country data.
- FREEDOM: A variable derived from responses to the question "Are you satisfied with your freedom to choose what you do with your life?" where 1 = YES and 0 = NO. Values for this variable represent the proportion of individuals in a given country who responded YES to the survey question.
- GENEROSITY: A variable derived from responses to the question "*Have you donated money to a charity in the past month?*" This is calculated as the residual of a regression of GDP per capita on responses to the charity question.
- CORRUPT: An average of the responses to the questions "*Is corruption widespread through the government?*" and "*Is corruption widespread within businesses?*" where 1 = YES and 0 = NO. Values presented for this variable are the average scores for individual countries.

The first row of variables in Table 1 are Bjornskov's deep-rooted determinants of trust. Just 22 per cent of countries have a main language which forbids dropping the personal pronoun (NOPRODROP), with many but not all of these being European languages. Just 13 per cent (eighteen) of the countries are constitutional monarchies, with most of these being European or members of the British Commonwealth. The average proportion of countries' populations that profess the Muslim religion is nearly 23 per cent, and 30 per cent for Catholics. However, I have calculated from the database that the percentage of all the people in the world who are either Catholic or Muslim is just 38 per cent — this total figure being in particular affected by the fact that the most populous country, China, has very few Muslims or Catholics.

Shares of both Muslims and Catholics in the population range from nearly 100 per cent to zero. Diversity, therefore, must have some zero value, but cannot be bounded upwards by 100 per cent, given that there are just five religions categorized. In fact, the most religiously diverse country is Singapore, in which all five religions are represented (even if the ethnic diversity of this country is much less – there are Christians and Buddhists of Chinese ethnicity).[9] The warmest country in the coldest month of the year is Panama and the coldest Mongolia.

Table 2 shows a linear regression model of trust, incorporating all the above-named variables, and estimated — as will be all the econometric models in the article —

9 Wiki reports that, of the 80 per cent of the population of Singapore who are 'citizens or permanent residents,' about 74 per cent are ethnically Chinese, 13 per cent Malay, and 9 per cent Indian. Apparently, the Singapore government does not release or record the ethnicity of the 20 per cent who are migrant or guest workers. I do not know if the latter group are surveyed for the well-being data.

Table 2: Modelling Trust

dependent variable	Constant	NOPRODROP	MONARCHY	MUSLIM	CATHOLIC	DIVERSITY	COLDEST DAY	R^2	n
TRUST	0.304	0.082	0.114	-0.000828	-0.001027	-0.03667	-0.00386	0.437	1376
t-statistic	8.14	3.31	3.63	-2.06	-2.89	-0.72	-4.28		

with the EViews 10 OLS package with cluster-corrected standard errors: countries being the cluster.

All regressors apart from diversity have coefficients of the expected sign and are reaching statistical significance by the usual standards.[10] The overall goodness of fit of the model may or may not seem impressive, given ones priors as to the plausibility of those regressors. No doubt, adding regional dummies (Latin America, Western Europe, etc) would raise the R^2, but it is more satisfying to look for fundamental determinants of social trust.

No doubt, also, there must be other factors — unknown or known — generating trust. Strong candidates are the direct (survey-based) measures of diversity or 'polarization' of views on politics, religion, honesty, and other factors, compiled by Beugelsdijk and Klasing (2016). They find a negative bivariate relationship between polarization and trust with an R^2 of 0.41, for a sample of 75 countries.

From an econometric point of view, the regressors used in the Table 2 model are attractive because, being so deep-seated in history and geography, they can very plausibly be taken as exogenous to measures of trust from contemporary surveys taken in our times. For this reason, I will call the values predicted by the model for each country 'deep' trust. These values will be given their chance to compete with currently surveyed trust in our subsequent analysis of international differences in productivity and well-being.

Trust and Productivity

For Adam Smith, productivity growth came primarily not from the accumulation of capital, both physical and human, but from the reorganization of existing resources through the division of labour. His work and life predated the large-scale application of science and technology to materials and mechanization that would fuel the 19th century Industrial Revolution and thus modern capitalism.

However, Smith did not himself discover or invent the idea of the division of labour. Indeed, his famous example in *The Wealth of Nations* of the productivity gains generated by splitting the manufacture of textile pins into 18 specialized steps was lifted directly and without acknowledgement from the *Encyclopédie* of the French philosopher Denis Diderot, twenty five years earlier.[11]

But what Smith may have been first to do was to examine the division of labour, not as a production engineer, but as a social scientist. He realized the extraordi-

10 A referee has suggested assessing the economic significance of the trust regressors, by multiplying each variable's coefficient from Table 2 by its sample standard deviation from Table 1. The resulting numbers are all in the range 0.3-0.4, or around 15 per cent of the sample average value of TRUST of 0.23.

11 See Katherine Sutherland's 'Explanatory Notes' on Smith, 1776 (1998:467).

nary demands that exploiting the division of labour would put on the coordinating capacity of the economy, by vastly increasing the number and extent of transactions needed in the new system. And he noted that this, inevitably, would take workers and capitalists beyond the safe confines of kith and kin: they would need to deal with strangers, and trust those strangers to behave reasonably honestly and predictably, as noted in the introduction. So can we test the importance of this, empirically, using the new survey data we now have aggregating responses to the standard 'trust question'?

Literature on trust, growth, and productivity

There is now a significant number of studies linking trust to economic growth — that is changes over time in incomes or productivity, rather than cross-sectional differences at a point in time — as surveyed by Bjornskov and Méon (2015) and Smith (2020). These studies seem to strain to achieve statistically significant results, and it is not surprising that this is so. In a cross section of nations there are huge differentials of levels of prosperity, and these differences do not change suddenly. Table 1 reports a more-than hundred-fold difference between real GDP per worker in Ireland compared with Burundi.

Differences in trust may be able to account for these differences in levels of prosperity, but not necessarily differences in year-to-year changes. And those differences are not strongly correlated decade to decade, as Hall and Jones (1999) note. For example, any database on economic growth covering the past thirty years will, or should, include Japan: a high-trust, high-income economy that has hardly grown at all since its great growth spurt in the post-war quarter century.

It *is* surprising, then, that there have been so few studies of trust — or more generally of the 'soft' institutions of social cohesion — linked — as Arrow predicted — to levels of economic development or productivity, not economic growth. Hall and Jones (1999) is pioneering, finding a strong effect on a cross section of output per worker data for 127 countries, of a measure of what they call quality of 'social infrastructure', this being the average of two indexes: one of the quality of protection of private property rights; the other of openness to international trade.[12]

Turning to studies focusing directly on trust as the soft-institution measure, Algan and Cahuc (2010) find a quite large effect of the inherited component of trust on per capita incomes, over time and across 24 countries. Bjornskov and Méon (2015) use total factor productivity (TFP) as their dependent variable, and are able to show a significant bivariate correlation, for 67 countries, between level of TFP and social trust, but this disappears when a measure of countries' legal quality is added to the model. Smith (2020), with a panel database on 32 mainly European countries, also finds a bivariate trust-TFP correlation,

12 Hall and Jones also successfully instrument their social infrastructure measure with two variables that in effect link it to Western Europe: distance from the equator, and prevalence in a country of a European language.

but does not explore the robustness of this to other possible explanatory factors, such as legal quality. Dearmon and Grier (2009), with data on 51 countries, report a linkage between trust and the level of per capita GDP, and also between trust and investment in physical capital — that is, in the change of the capital stock, not its level.

The present study will follow Hall and Jones (1999) in using output per worker, not incomes per capita, because our focus is on the supply side — productivity — rather than incomes, which can have other determinants.

Results

Along with the TRUST and DEEP TRUST variables as defined above, production function data are sourced from the Penn World Tables (PWT) database (Feenstra, Inklaar, and Timmer, 2015) annually in most cases from 2007 through 2017. We see from Table 1 that capital per worker varies internationally even more than output per worker (consistent with diminishing returns), and that the world's best educated citizenry live in Uzbekistan, a former republic of the Soviet Union.

We begin with the simplest 'Smithian' production function, regressing labour productivity (logged RGDPO per person employed) on actual and predicted deep trust (Table 3). Though the overall fit of this model is not high, the trust variables show strong significance, with DEEP TRUST doing the better. That is, a variable created as a linear combination of various ancient cultural, religious and linguistic customs is by itself quite successful at accounting for the vast differences now in countries' material standards of living.

The third regression model shown on Table 3 switches to the standard neoclassical formulation that has output per worker dependent on physical and human capital per worker, with a non-neoclassical gloss in the form of the index of institutional quality.

Not surprisingly, the R^2 of this model is much higher, with both capital measures comfortably significant, and institutions less so. So, what happens if we combine the two models? Both trust variables now have negative coefficients!

So we dig down a level. Does a high level of trust encourage risky long-term investments in physical and human capital? Table 4 gives the answer: trust is indeed a strongly significant determinant of both types of capital, as well of the quality of a country's institutions. And deep trust is a more significant factor than current actual trust.

This last result is particularly interesting. Our model of deep trust is, as noted, subject to error, but the predicted value at least holds its own compared with directly measured trust. This suggests that deep trust is not an instrument for actual trust (at least, in the productivity setting), but, rather, the opposite: actual trust is really an instrument for deep trust. It is deep trust that matters.

That is, deep trust really is important for levels of economic development, but only indirectly, through its encouragement of productive investments. So, if two countries happen to have similar levels of the two types of capital, the more trusting of the two will not obtain an additional productivity boost from this. But in general, workers in high-trust societies are indeed more likely to have more capital to work

Table 3: Modelling Real GDP per Person Employed

Independent Variable	(1)	(2)	(3)	(4)	(5)
Constant	9.437	8.950	3.250	3.298	3.388
t-statistic	*67.36*	*50.44*	*7.82*	*7.85*	*7.89*
TRUST	3.335			-0.274	
t-statistic	*6.95*			*-1.39*	
DEEP TRUST		5.481			-0.789
t-statistic		*9.15*			*-2.66*
log(RNNA/EMP)			0.542	0.541	0.537
t-statistic			*13.17*	*13.22*	*12.88*
HC			0.265	0.278	0.302
t-statistic			*4.22*	*4.44*	*4.51*
INSTITUTIONSAV			0.024	0.028	0.030
t-statistic			*2.57*	*2.61*	*2.95*
R^2	0.203	0.236	0.900	0.901	0.903
Observations	1385	1385	1385	1385	1385

Table 4: Modelling Production Function Inputs

	Dependent Variable					
Independent Variable	log(RNNA/EMP)		HC		INSTITUTIONSAV	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	10.51	9.82	2.13	1.71	-4.78	-7.69
t-statistic	*54.96*	*40.66*	*23.29*	*14.94*	*-7.31*	*-8.97*
TRUST	4.41		2.29		21.32	
t-statistic	*6.86*		*7.97*		*7.36*	
DEEP TRUST		7.42		4.13		34.15
t-statistic		*9.17*		*10.39*		*10.17*
R^2	0.209	0.254	0.213	0.297	0.310	0.341

with, either from their employer's willingness to invest in physical capital or their own willingness to delay entering the paid labour force in order to invest in skills and education.

In particular, there is no systematic role for trust in determining what is called total factor productivity. Indeed, there is little variation in total factor productivity to be determined or explained. The R^2 (=0.9) of the third labour productivity regression in Table 3 — very high for what is basically a cross sectional model — leaves little to be accounted for beyond the contributions of the two capitals and institutional quality.[13]

Adam Smith would probably be surprised by this – that a more trusting division of labour between strangers is not *ipso facto* productivity enhancing. In his pre-industrial revolution world, the accumulation of physical capital cannot get fully under way, because the new technologies in which capital would be embodied were still embryonic. As for human capital — Smith certainly recognized the skills developed by specialization. But he thought apprenticeships were made too long (for monopoly reasons). He was also definitely sceptical — this fuelled by his own fairly useless experience as a student of Balliol College Ox-

13 A full-blown production function model, with log(RGDPO) dependent, and log(RNNA), log(EMP), HC, INSTITUTIONSAV as regressors, has $R^2 = 0.965$

ford — about what now is assumed to be the standard method for increasing human capital: university education.

The productivity-enhancing division of labour on which Smith focused as the main determinant of prosperity in the late 18th century did not necessarily require much physical investment — just the willingness of strangers to cooperate to do the job. But Adam Smith also did not foresee the consequences of truly large scale production from the 19th century onward, with its (apparently) necessary innovation of bureaucratic organization and control systems. This may have reduced the importance of person-to-person trust in the workplace, by internalizing and codifying so many transactions.

How big is the trust effect, overall? From Table 1, the mean value of output per worker is about 42,000, for which the natural log is 10.6. Suppose a country with that value also has a sample-average value of deep trust, which is 0.23 (almost the same as the average of surveyed trust, as reported on Table 1). How much higher, *ceteris paribus*, would be the productivity of this country if it had deep trust at its highest predicted value, which is Norway's 0.50? The difference in deep trust is 0.27, and multiplying this by the deep trust coefficient from Table 3 (=5.5), we obtain a predicted change in the log of output per worker of +1.5, to 12.1, which corresponds to an actual value of output per worker of about US$180,000 — that is, more than four times higher than mean productivity — higher, indeed, than actual productivity in the leading large industrial country — the United States. The estimated deep trust-productivity effect is indeed substantial.

Trust and Well-being

Economists in recent years have looked beyond GDP as an index of economic performance to happiness itself, for which material prosperity may or may not be a significant contributor. Notable research programs include those of Rafael Di Tella and Robert MacCulloch (2008), and the annual (since 2012) *World Happiness Report* (*WHR*) issued by a group led by John Helliwell, Richard Layard and Jeffrey Sachs (2018). The latter uses surveys of samples of national populations whose respondents are asked (by the Gallup polling organization) to evaluate, on a scale of zero to ten, just how satisfied they are with their life so far (the 'Cantril Ladder').

For a panel of 157 countries surveyed (not all in all years) over the 2005 to 2017 period, Helliwell, Layard, and Sachs (2018, Table 2.1) estimate an econometric model with country average Cantril Ladder scores — called 'subjective well-being' or SWB — as the dependent variable. In this model, the natural log of per capita GDP is indeed a reasonably strong predictor, with a coefficient just above 0.3, and so too are several quality-of-life indicators surveyed along with SWB at the individual level: social support (friends in case of need), freedom to make life choices, generosity, and perceptions of corruption, along with healthy life expectancy, measured at the national level.

These are interesting results, and they establish the meaningfulness of individual survey responses to questions about happiness and quality of life, as attested to by DiMaria, Peroni, and Sarracino

(2021). These authors, and Sarracino and O'Connor (2021), take the sensible step of using the analogy with GDP production functions to carry out DEA (Data Envelope Analysis) of the 'productivity' at which a sample of countries convert the *WHR* regressors into well-being, relative to best-practice (frontier) countries.

Smith and Legge (2022) seek to unify the analysis of well-being and GDP production for 38 European countries by using physical and human capital as regressors in both models. Given that GDP is one of the six inputs to well-being in the *WHR* well-being model, we might expect the Smith and Legge procedure to show empirical success, and it does, but with the unexpected, and interesting, finding that human capital is relatively more important for well-being, and physical capital more important for GDP, such that, across these countries, there is no correlation between well-being and productivity.

So, what about trust? The *WHR* team have added the trust score variable to a SWB model, and find that it appears to have a quite strong direct effect on well-being. Comparing the coefficients on trust and on per capita incomes, they deduce, for example, that the increase in trust reported in Poland over the first decade of the new millennium was equivalent in its effect on life satisfaction with a 12 per cent increase in Polish per capita GDP (Helliwell, Huang, and Wang, 2016:11-12).[14]

These results are impressive, but they almost certainly underestimate the total effect of trust on well-being. They pick up the partial direct effect of trust on well-being, holding the other factors constant. But we know from Algan and Cahuc (2010), and the results of this article's second section, that at least one important other factor — per capita incomes or GDP — is in general not held constant when trust changes, and it may well be that some of the other well-being contributors, such as social support and life expectancy, are themselves affected by trust. That is, there may be multicollinearity amongst the regressors which will obscure their true effects on well-being.

The contribution here will be to tease out from the data the channels whereby differences in the long-term component of generalized trust work their way through to a net total impact on well-being.

Data

The basis for our database is the Excel spreadsheet "Table 2.1" supplied with the 2018 World Happiness Report. This has data on subjective well-being for 157 countries, collected by Gallup in its surveys for various years ranging from 2005 through 2017. These data are supplemented with data on the regressors in the *WHR* well-being model, each of these being available for most but not all countries, and for most but not all years. Variable definitions were given above. All variables from individuals' survey responses are averaged to the country level.

Table 1 revealed that there is certainly

14 However, this effect was derived from an econometric model not incorporating all the other well-being factors. When these are added, the coefficient on trust becomes smaller and less stable.

considerable cross-country variability in average subjective well-being to be explained, with Denmark in 2005 at the top of the Cantril Life Ladder with an average score of just over eight out of ten, and war-torn Syria below 3 in 2017.

Availability of support when troubled is generally high, and almost universal in New Zealand. Less happily, there are still countries where citizens cannot expect long healthy lives, though the mean value is much closer to the top than the bottom of the range. Professed freedom to make ones' life choices is apparently almost universal in Uzbekistan, but rather uncommon in Bosnia and Herzegovina — I know not why. However, the mean is well skewed towards the high end of the range. The fairly impoverished people of Myanmar are (relative to incomes) most likely to give to charity; the least generous are the Greeks, whose notorious unwillingness to pay their taxes apparently is not compensated for by a propensity towards charitable giving. It is very sad that almost all Hungarians believe their institutions are corrupt; pleasing but not surprising that citizens of the tightly administered city state of Singapore feel just the opposite.

Overall, it is strikingly clear from these data that income is not the only fact of life that is unevenly distributed around the world (as well as within countries), and the wide range of SWB outcomes may suggest that the differences in the other factors do not cancel out. This in turn implies the existence of some underlying variable or variables which tend to have a similar effect (ie, in the same direction) for most or all of the happiness-determining factors identified in the *WHR*.

Estimation

We use the sample of 136 countries[15] for which we have, or have constructed, data on social trust, and, as before cluster-correct standard errors of estimated coefficients. Table 5 shows the results.

First, we run the simplest trust-only models. As with the production function estimates, trust alone is a successful predictor of well-being, though now actual trust performs better than predicted or deep trust. However, just as with the production model, the direct trust effect does not survive inclusion of other regressors.

Replicating the WHR results, a quite large proportion of the cross-country variation in self-reported well-being is accounted for by incomes and the five non-economic variables, with all six showing statistical significance. The coefficient on log per capita GDP implies an elasticity of SWB with respect to incomes of around 0.3, which apparently is consistent with previous research. Having support in times of trouble is a particularly important factor for well-being.

Now, we repeat the exercise from the third section of the article: here looking for evidence that trust works indirectly through its influence on the direct determinants of well-being. Table 6 shows that such is indeed the case. However, for the

15 Countries that appear in the *WHR* database but not here because trust data are not available include several very small countries (Maldives, Comoros, North Cyprus, Kosovo) and all the 'oil economies' of North Africa and the Middle East, except Iraq.

Table 5: Modelling Self-Reported Well-being

Independent Variable	(1)	(2)	(3)	(4)
Constant	4.580	4.278	-1.720	-1.695
t-statistic	*29.27*	*18.47*	*-3.26*	*-3.23*
TRUST	3.828		-0.192	
t-statistic	*6.82*		*-0.55*	
DEEP TRUST		5.169		-0.388
t-statistic		*5.78*		*-0.61*
LOGGDPPOP			0.352	0.359
t-statistic			*5.43*	*5.25*
SUPPORT			2.438	2.451
t-statistic			*5.74*	*5.98*
HLIFEEXP			0.028	0.027
t-statistic			*3.06*	*3.04*
FREEDOM			0.974	0.960
t-statistic			*3.00*	*2.98*
GENEROSITY			0.709	0.695
t-statistic			*2.42*	*2.32*
CORRUPT			-0.629	-0.635
t-statistic			*-2.02*	*-1.94*
R^2	0.232	0.182	0.730	0.731

five non-economic factors, actual trust is stronger than predicted or deep trust as a determinant. Perhaps this is because well-being is more flexible than the level of a country's economic development.

If we drop deep trust from the *WHR* model, multiply each of the six estimated coefficients by their coefficient in the DEEP TRUST bivariate models shown on Table 6, and then sum these products, we obtain the number 5.43, which is quite close to the coefficient, 5.17, on DEEP TRUST in the bivariate model of Table 5. This demonstrates that all (actually, somewhat more than all) of the overall impact of trust on happiness works indirectly through the regressors identified in the World Happiness Reports.[16]

The Kiwi Conundrum: Comparing Australia and New Zealand

We can apply the productivity and well-being models developed in this article to the case of two countries: Australia and New Zealand. These two countries are very open to each other and quite similar in many cultural respects, but are rather different in their economic policy regimes, and in the outcomes of these in terms of productivity.

The setting here is an enjoyable little scandal playing out about alleged improprieties involved with calculating the World Bank's *Doing Business* 2018 ranking of 190 economies according to an Index aggregating scores on eleven areas of business regulation. The supposed wrong-doing concerns possible attempts by the then Managing Director of the Bank to improve the rating of China, at a time when that country's support was sought for an increase in the Bank's funding. The matter was assigned to an outside law firm to investigate (Machen et al., 2021), and the outcome of

16 Performing a similar exercise for productivity requires re-estimating the Table 3 and 4 bivariate models with DEEP TRUST logged, to match the Cobb-Douglas production function. Again, the sum of products is very close to the overall bivariate effect of (log) trust on productivity.

Table 6: Modelling Trust Determinants of Well-being

	Dependent Variable									
Independent Variable	SUPPORT		HLIFEEXP		FREEDOM		GENEROSITY		CORRUPT	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.741	0.705	56.279	54.053	0.651	0.645	-0.086	-0.068	0.915	0.950
t-statistic	*49.55*	*31.71*	*48.99*	*33.75*	*36.28*	*0.40*	*-4.24*	*-2.05*	*38.85*	*25.10*
TRUST	0.324		27.083		0.377		0.369		-0.707	
t-statistic	*7.20*		*8.05*		*7.11*		*4.03*		*-6.05*	
DEEP TRUST		0.482		36.969		0.403		0.292		-0.859
t-statistic		*6.47*		*6.92*		*4.01*		*1.96*		*-4.84*
R^2	0.145	0.138	0.230	0.184	0.142	0.070	0.096	0.026	0.284	0.182

this was the demise of the published Index.

This furore was in the context of increasing general criticism of the Index itself, on the grounds that it is systematically biased towards favouring a right-wing — even, neoliberal — view of the appropriateness of certain 'business-friendly' policies; in particular, policies conducive to 'flexibility' in hiring and firing workers.[17]

Alarm bells might well have sounded at once on the publication, in 2019, of the 2018 rankings. Top of the list — first, supposedly, amongst 190 economies for the quality of its business regulations — is New Zealand. Yet this country in 2018 was just 22nd in the non-oil First World ranking of per capita GDP, in particular contrast to our nearest neighbour, Australia, which is 12th, with per capita GDP more than 30 percent higher than the smaller country's (Table 7).

Now, these two countries have long been bound together by what may be the most extensive bilateral common market agreement in the developed world, which in particular allows absolutely free mobility of labour, with the result that a rather high proportion of New Zealand (NZ) citizens — perhaps more than 10 per cent — have crossed the Tasman Sea to improve their fortunes in Australia.

This they do with absolutely no difficulty, being quickly employed in Australia at the much higher wages and salaries generated by the sizeable GDP gap. So it does not seem that the quality of human capital is at fault here. Then, could it be a policy/institutional problem? Well, the problem with *this* is that New Zealand has such good, business-friendly policies, as claimed by *Doing Business*, and also noted by Zheng, Duy, and Pacheco (2021) and others – better than Australia. For example, the institutional quality index used in this article, which has much overlap with the World Bank's *Doing Business* methodology, places NZ at third best — just behind Finland and Sweden.

And, it could be noted that the decade in which the income gap widened the most was the 1990s, directly following New Zealand's swingeing Rogernomics neoliberal reforms.[18] So, although the — rather moderate — significance of the INSTITU-

17 On the topic, see Ghosh (2020), Stiglitz (2021), Krueger (2021) and Cárdenas (2021).

18 Named after the very determined Finance Minister, (Sir) Roger Douglas, who pushed through the reforms in the 1984-90 Labour government.

Table 7: Comparing New Zealand and Australia

		PREDICTED	ACTUAL
OUTPUT PER WORKER (RGDPO/emp)	NZ	65,550	70,294
	AU	93,090	100,216
CAPITAL PER WORKER (RNNA/emp)	NZ	441,730	241,543
	AU	349,400	438,047
INSTITUTIONAL QUALITY (INSTITUTIONSAV)	NZ	6.93	10.64
	AU	5.87	9.60
SELF-REPORTED WELLBEING (LIFE LADDER)	NZ	7.32	7.09
	AU	7.31	7.12
TRUST	NZ	0.428	0.548
	AU	0.397	0.503

TIONSAV variable used in Table 3 production function is consistent with very bad institutions and policies being somewhat harmful to prosperity across the world, within the First World group it seems that the case is not at all clear.

Well, can the data and results of this study contribute at all to understanding the productivity gap between New Zealand and Australia?

The first line on Table 7 shows that our production function (the third regression model from Table 3) actually slightly underpredicts actual output per worker in both countries, and by a similar percentage amount. So we can rule out differences in total factor productivity — the two countries do seem to be on the same production function. The measures of human capital are very similar: 3.5 and 3.4, in 2017, consistent with New Zealand emigrants fitting easily into the Australian labour market. The institution quality variable is, of course, higher in New Zealand, which increases the income disparity to be explained by our only other productive input, which is physical capital per worker.

Actual capital intensity is *much* higher in Australia: New Zealanders taking on jobs in Australia are, on average, provided with more productive capital to work with than was supplied by their erstwhile Kiwi employers. The discrepancy is the larger given that our Table 4 model predicts higher capital/labour ratios in New Zealand, due to higher generalised trust levels here.

This is as far as the data and results of this article can take us — not solving the puzzle, but somewhat narrowing it down.[19] I will note casually, however, that an obvious 'smoking gun' is the apparent unwillingness of Kiwi entrepreneurs and managers to build and operate large business corporations. The Australian economy overall is about five times larger than that of New Zealand, but the capitalization of its stock market is about thirteen times greater, including many big firms operating in New Zealand, such as the four major trading banks.

19 The difference in NZ and Australian capital/labour ratios is, of course, well known. Zheng, Duy, and Pacheco (2021) note this as a possible factor to account for their finding from micro data of labour productivity lower in New Zealand than in five small European economies, for firms on their country's productivity frontier, and for firms within the frontier.

So, is there really a problem here? Perhaps New Zealanders just do not have a taste for running and working in the huge bureaucratic structures of the corporate sector. If the real bottom-line of a country's success is well-being, of which material GDP is an important, but not sole determinant, as the World Happiness Report consistently finds, then Table 7 shows little difference between the two countries, with our Table 5 model slightly over-predicting self-reported well-being in both countries. From an accounting perspective, Australia's higher incomes are just about balanced by higher informal support levels, and less corruption in New Zealand.[20] In terms of the Data Envelope Analysis approach applied to European countries by Sarracino and O'Connor (2021), Australia and New Zealand are on the same iso-well-being frontier (just inside the 'efficient' frontier) but at different points of the curve.

Conclusion

When the University of Chicago macroeconomist Robert Lucas turned his attention to income disparities around the world, he was quite shocked. Clearly, capital/labour ratios are smaller in the developing world. But then, according to the neoclassical model, with just one freely available 'blueprint' for converting labour and capital into GDP, the marginal productivity of capital will be higher in poor countries, and therefore international capital flows should be moving from rich to poor countries, to arbitrage the difference in marginal returns. But, Lucas noted, this was not happening, and, if anything, such surplus savings that the elites in poor countries were able to accumulate were siphoned out of the Third World into the capital markets of the West. Lucas was very worried about this:

> The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else (1988:5).

But Kenneth Arrow had already thought about this question, and had suggested the answer to the puzzle, as quoted in Section 1, above. Physical and human capital may be indeed less productive in poor countries because of a shortage of an essential complementary input — 'social capital', as Smith and Legge (2022) put it — provided by social or generalized trust in strangers.

The evidence that has been gathered more recently, including in this article, surely bears out this proposition. In particular, we find that, with respect to trust and its effects, history casts a huge shadow on our lives today. The centuries-old cataclysms — as they usually were — that have shaped our religions, our languages, our governance, are now embodied in the markers of our economic success and our personal well-being. Remarkably, it seems to be at least consistent with the data that

20 It is notable that the 'top ten' countries in each year's *WHR* well-being lists are without exception small or quite small First World economies, with Canada sometimes appearing, as the largest. Finland, Sweden, Norway, Iceland and Denmark are always in the top ten, with Finland at number one in recent years.

all these ancient forces to an impressive extent work through just that one factor: the average level, across a country, of its citizens' trust in each other — in particular, our beliefs in the likelihood of strangers behaving well — trusting, trustworthy — in our country.

What are the policy implications of these findings? On the face of it, it might seem that policy is impossible: trust now is largely dependent on deep historical factors, and we cannot re-write history. The 'middle-income trap' that seems to have prevented any country from attaining first-world economic status over the past quarter century, may be, alas, secure.

But our current state of knowledge on the pathways leading to and from trust is surely not complete enough to justify lack of interest in trust-building policies — such as, for example, the very successful self-reported rating systems that have greatly facilitated the explosion of electronic commerce between strangers. Nor does it warrant lack of concern with what appears, from the analysis of changes in surveyed trust over recent periods by Beugelsdijk and Welzel (2018), to be a significant, though not universal, decline in trust scores across 67 countries.[21]

What about the implications for economics; in particular for the principles of neoclassical economics, with its relentless focus on scarcity and opportunity costs, and exponential growth in incomes — that is, the assumption that, at any point of time, we are constrained by the current production possibility frontier (PPF) such that any choice to have more of something good must be paid for by taking less of something else, whereas over the long term the PPF shifts out without limit?

Helliwell *et al.* (2018:49) write:

> My gold cannot be your gold. But happiness, unlike gold, can be created for all, and can be shared without being scarce for those who give. It even grows as it is shared.

That is, with the (quite important) exceptions of the parts of happiness that are income-dependent, happiness is a public not a private good — perhaps, even, contagious.

However, happiness surely cannot grow exponentially without limit (nor, of course can incomes, in a finite world). If everyone reports a perfect-ten on their Cantril Ladder score — well, that's as good as it can get, is it not? However, there is clearly still enough unhappiness in the world — even within Finland and other high-average happiness countries — to stave off satiation for some time yet. We get back to the belief — or hope — that policy still has — hopefully — its role to play, in particular in building or restoring trust, and the informal support networks that seem entwined with trust.

But, in any case, the range of well-being across tolerably peaceful and competently

21 These authors also identify a plausibly related contemporary trend: an increase amongst young people of "more individualistic, more joyous" attitudes over a surveyed sample of 495,000 individuals across 110 countries. As noted above, Beugelsdijk and Klasing (2016) have documented a quite strong negative relationship between polarization of views within countries and social trust.

managed countries is really not huge. In particular, the range is strikingly limited compared with the range in material prosperity. Even amongst developed countries, per capita incomes in Denmark and Norway are more than twice those in Portugal, and nearly twice Spain and New Zealand, countries are in turn at least twice as materially well off as the leading developing economies, such as Argentina, Brazil and Mexico.

Finally, should we break further from orthodoxy? The *WHR* researchers are themselves quite 'neoclassical' in their assumption of a unique production function for well-being, available on the same terms to all countries, just like a neoclassical GDP production function. For example, Hamilton, Helliwell, and Woolcock (2016) implicitly move around a common happiness production isoquant when they calculate that, for a given level of happiness, the difference between high-trust Sweden and low-trust Italy is worth about a 20 per cent difference in per capita GDP. But what if the Italians in some sense have chosen to live together as a low-trust society, and have developed various behavioural norms and procedures to deal with this in ways not available to Swedes?[22] There is more to be known about the development of widely shared norms, which may not be synonymous with trust.

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wazciarg (2003) "Fractionalization," *Journal of Economic Growth*, Vol. 8, No. 2, pp. 155-194.

Algan, Y. (2018) "Trust and Social Capital," in *For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, OECD, pp. 283-320.

Algan, Y. and P. Cahuc (2010) "Inherited Trust and Growth," *American Economic Review*, Vol. 100, No. 5, pp. 2060-2092.

Algan, Y. and P. Cahuc (2014) "Trust, Growth and Well-being: New Evidence and Policy Implications," in P. Aghion and S. Durlauf (eds.) *Handbook of Economic Growth*, Vol. 2, Elsevier, pp. 49-120.

Arrow, K. J. (1972) "Gifts and Exchanges," *Philosophy and Public Affairs*, Vol. 1, No. 4, Summer, pp. 343-362.

Barro, R. J. and J.-W. Lee (2013) "A New Data Set of Educational Attainment in the World, 1950-2010," *Journal of Development Economics*, Vol. 104, pp. 184-198.

Beugelsdijk, S. and M. J. Klasing (2016) "Diversity and Trust: The Role of Shared Values," *Journal of Comparative Economics*, Vol. 44, No. 3, pp. 522-540.

Beugelsdijk, S. and C. Welzel (2018) "Dimensions and Dynamics of National Culture: Synthesizing Hofstede with Inglehart," *Journal of Cross-Cultural Psychology*, Vol. 49, No. 10, pp. 1469-1505.

Bjornskov, C. (2006) "Determinants of Generalized Trust: A Cross-country Comparison," *Public Choice*, Vol. 130, pp. 1-21.

Bjornskov, C. (2012) "How Does Social Trust Affect Economic Growth?" *Southern Economic Journal*, Vol. 78, No. 40, pp. 1346-1368.

Bjornskov, C. and P.-G. Méon (2015) "The Productivity of Trust," *World Development*, 70, pp. 317-331.

Cárdenas, M. (2021) "After 'Doing Business'," Project Syndicate, October 8, 2021. https://www.project-syndicate.org/commentary/world-bank-after-doing-business-four-criteria-by-mauricio-cardenas-2021-10.

Dearmon, J. and K. Grier (2009) "Trust and Development," *Journal of Economic Behavior & Organization*, Vol. 71, No. 2, pp. 210-220.

22 A well-meaning upper class English settler in Tuscany was advised by her Italian friend: "Do not trust your neighbour. Your neighbour will not expect it."

DiMaria, C.-H., C. Peroni, and F. Sarracino (2021) "From Productivity to Productive Well-Being," presented at Virtual Workshop on Productivity and Well-being: Measurement and Linkages - November 16th and 17th, 2021, International Association for Research in Income and Wealth.

Di Tella, R. and R. MacCulloch (2008) "Gross National Happiness as an Answer to the Easterlin Paradox?" *Journal of Development Economics*, Vol. 86, No. 1, pp. 22-42, April.

Feenstra, R.C., R. Inklaar, and M.P. Timmer (2015) "The Next Generation of the Penn World Table," *American Economic Review*, Vol. 105, No. 10, pp. 3150-3182, available for download at www.ggdc.net/pwt.

Ghosh, J. (2020) "Stop Doing Business," Project Syndicate, September 10, 2020. https://www.project-syndicate.org/commentary/world-bank-should-scrap-doing-business-index-by-jayati-ghosh-2020-09.

Goff, L., J.F. Helliwell, and G. Mayraz (2018) "Inequality of Subjective Well-Being as a Comprehensive Measure of Inequality," *Economic Inquiry*, Vol. 56, No. 4, October, pp. 2177-2194.

Hall, R. E. and C. I. Jones (1999) "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, Vol. 114, No. 1, February, pp. 83-116.

Hamilton, K., J. F. Helliwell, and M. Woolcock (2016) "Social Capital, Trust and Well-being in the Evaluation of Wealth," National Bureau of Economic Research Working Paper No. 22556, August.

Helliwell, J.F., H. Huang and S. Wang (2016) "New Evidence on Trust and Well-being," National Bureau of Economic Research Working Paper No. 22450, July.

Helliwell, J.F., R. Layard, and J. Sachs (2018) *World Happiness Report*, 2018.

Kaufman, D., A. Kraay, and M. Mastruzzi (2010) "The Worldwide Governance Indicators: Methodology and Analytical Issues," Policy Research Paper 5430, The World Bank.

Krueger, A.O. (2021) "The Bretton Woods Credibility Crisis," Project Syndicate, October 7, 2021. https://www.project-syndicate.org/commentary/world-bank-doing-business-scandal-by-anne-o-krueger-2021-10.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1997) "Trust in Large Organizations," *American Economic Review*, Vol. 87, No. 2, pp. 333-338.

Lee, S. (2017) "Rethinking the Relationship Between Pronoun-drop and Individualism with Bayesian Multilevel Models," *Journal of Language Evolution*, Vol. 2, No. 2, July, pp. 188-200.

Lucas, R.E. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, Vol. 22, No. 1, pp. 3-42.

Machen C.R., M.T. Jones, G.P. Varghese, E.L. Stark (2021) "Investigation of Data Irregularities in Doing Business 2018 and Doing Business 2020: Investigation Findings and Report to the Board of Executive Directors," WilmerHale. https://thedocs.worldbank.org/en/doc/84a922cc9273b7b120d49ad3b9e9d3f9-0090012021/original/DB-Investigation-Findings-and-Report-to-the-Board-of-Executive-Directors-September-15-2021.pdf.

Putnam, R. (1993) *Making Democracy Work*, Princeton University Press.

Sarracino, F. and K. O'Connor (2021) "A Measure of Well-being Productivity Based on the World Happiness Report," presented at Virtual Workshop on Productivity and Well-being: Measurement and Linkages - November 16th and 17th, 2021, International Association for Research in Income and Wealth.

Smith, A. (1776) *An Inquiry Into the Nature of and Causes of the Wealth of Nations*, republished as *An Inquiry Into the Nature of and Causes of the Wealth of Nations: A Selected Edition*, edited by Kathryn Sutherland, Oxford University Press, 1998.

Smith, C. (2020) "Trust and Total Factor Productivity: What Do We Know About Effect Size and Causal Pathways?" *International Productivity Monitor*, No. 38, Spring, pp. 3-28. http://www.csls.ca/ipm/38/Smith.pdf

Smith, C. and J. Legge (2022) "Well-being and Productivity: A Capital Stocks Approach," *International Productivity Monitor*, No. 42, Spring. https://www.productivity.ac.uk/wp-content/uploads/2022/07/IPM_42_Smith.pdf

Stiglitz, J. E. (2021) "A Coup Attempt At the IMF," Project Syndicate, September 27. https://www.project-syndicate.org/commentary/coup-attempt-against-imf-managing-director-georgieva-by-joseph-e-stiglitz-2021-09.

Uslaner, E. (2008) "Where You Stand Depends on Where Your Grandparents Sat: The Inheritability of Generalized Trust," *Public Opinion Quarterly*, Vol. 72, pp. 1-14.

Zak, P. and S. Knack (2001) "Trust and Growth," *The Economic Journal*, Vol. 111, No. 470, pp. 295-321.

Zheng, G., H.M. Duy, and G. Pacheco (2021) "Benchmarking the Productivity Performance of New Zealand's Frontier Firms," *International Productivity Monitor*, No. 40, Spring, pp. 27-55. http://www.csls.ca/ipm/40/IPM_40_Zheng.pdf

Appendix Table 1

Dependent Variable: TRUST				
Included observations: 124 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	**0.105171**	0.021908	4.800701	0
TRUSTGALLUP	**0.818429**	0.070269	11.64702	0
CE_EUROPE	**-0.03299**	0.020015	-1.64852	0.1019
CIS	**-0.08241**	0.028324	-2.90968	0.0043
LA	**-0.10454**	0.021342	-4.89827	0
SUB_SAHARA	**-0.18807**	0.028374	-6.62828	0
R^2	0.678252	Mean dependent var		0.238996
Adjusted R^2	0.664619	S.D. dependent var		0.141404

Appendix: Infilling trust data

Ninety eight countries have World Values Survey (WVS) data on proportion of surveyed population agreeing that 'most people can be trusted' for at least one of the 1999-2004, 2005-2009 and 2010-2014 Rounds of the WVS (this means that in, say, one unnamed year of the five years in each Round, the trust question was asked in a country.)

There were 124 country/year rows containing both a WVS trust number and a number from a similar survey asked by the Gallup polling organization. For those country/rows an OLS model was estimated (see Appendix Table 1).

That is, we can predict WVS scores quite well using the Gallup score (which tends to be higher) and a subset of regional dummies.

So, the highlighted estimated coefficients are used to infill WVS score estimates for the 39 countries in our set of 136 countries which have a Gallup trust score but no WVS score.

Time Use, Productivity, and Household-centric Measurement of Welfare in the Digital Economy

Diane Coyle
University of Cambridge
Leonard Nakamura
Federal Reserve Bank of Philadelphia[1]

Abstract

There is substantial interest in developing a broader understanding of economic progress than the standard indicator, real GDP, not least because digital technology is significantly changing both production within the GDP boundary and household activity outside the boundary. Market and household production and leisure now all involve substantial time online. This article describes a measurement framework that would encompass extended utility combining time allocation — over working for pay, producing at home, and leisure — with monetary measures of objective or subjective well-being during each activity and new ways of measuring productivity in digitalized activities. Implementation would require time use statistics in addition to well-being data and direct survey evidence on the shadow price of time. We advocate an experimental set of time and well-being accounts and discuss their data requirements.

Although widely used as shorthand for economic progress, the limitations of Gross Domestic Product (GDP) are well-known (Coyle, 2014). Digital technology is exacerbating GDP's shortcomings on both the production and expenditure sides of the national accounts, as it is significantly changing both production processes and household activity. A change and increase in the use of time is the distinctive feature of

1 Bennett Professor of Public Policy at University of Cambridge and Emeritus Economist at Federal Reserve Bank of Philadelphia, respectively. We are grateful to Lucy Eldridge, Charles Hulten, Penny Mealy, Paul Schreyer, Solomon Tarlin, and participants in the IMF Statistical Forum, a workshop at the Economic Statistics Centre of Excellence, a session at the 2019 Royal Economic Society conference, and a Productivity Institute workshop for their comments. Our thanks also to anonymous reviewers for this journal for their constructive suggestions. Coyle received funding for this work from the Economic Statistics Centre of Excellence and The Productivity Institute. The views expressed are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Emails: dc700@cam.ac.uk, leonard.nakamura@phil.frb.org.

digitalization, pointing to the need to consider the role of time — the only resource in truly fixed supply — in understanding economic progress. In addition to enabling process efficiencies and new business models in production, digitalization greatly enhances private production's reach into the activity of household production and consumption, so to measure its total impact, a fuller measure of output is required.

On the production side, digital technologies are routinizing a growing swathe of service sector activities, such as analyzing legal documents, monitoring financial transactions for fraud, transcription, or writing standard reports, much as automation previously transformed manufacturing. This points toward faster processing time as an important productivity metric for such activities. In the case of other services, however, such as intensive-care nursing or childcare, the need for quality or focused interaction is more important and could point to a slower time to produce as the correct metric of both productivity and consumer welfare.

On the household consumption side, GDP primarily measures monetary transactions, equating inputs to outputs in nominal terms. The household inputs it counts are predominantly paid work hours, while utility is considered as household monetary expenditures on consumption goods. An alternative perspective is Becker's (1965) full income model, where all household hours are considered as inputs, and utility is a function of both time spent and monetary expenditures on goods. In this perspective, household work and leisure time are both inputs into the production of household utility, such that utility is a function of a household consumption technology. Alternatively, one can hew more closely to the notion that GDP is a production concept and include both paid work and household work as production. This perspective dates from the important work of Margaret Reid (1934), who argued for inclusion of household work in overall measures of production. This leads to a full income perspective that has been developed by, among others, Bridgman (2016), who has led its incorporation into a Household Production satellite account of the U.S. and other national income accounts.

The use of digital technology in activities such as online banking and retail is shifting some activities into the household side of the production boundary (Coyle, 2019). Production and consumption are further linked as the automation of routine activities may imply a changing bundle of activities in which slow thinking — as discussed in Kahneman (2011) — is growing in importance. Productivity metrics based on real GDP are often considered to be distinct from questions of measurement of economic welfare or well-being. However, as they require a constant-utility price index to calculate real output terms, they embed an implicit economic welfare framework; but it does not account for either time-savings or quality-time. In this article we consider the scope to consider time spent as an appropriate metric for the digital economy, looking through the lenses of both production and productivity and household activity, consumption and welfare. We propose some additions to statistics that would enable monitoring of productivity and welfare through the lens of time.

This article proceeds in section one with our fundamental framework, in which household well-being is considered to encompass utility throughout the day, Gary Becker's 'full income' approach. In section two, we focus on how digitalization has influenced production and the boundaries between production and consumption. Section three addresses the measurement of well-being and the difficulties of alternative means to assess growth, well-being, the shadow value of time, and well-being while at paid work. In section four we outline the way forward, highlighting the key issues to be addressed as economists and statistical offices grapple with the measurement and meaning of productivity.

Time to Consume

For many digital goods and services, the marginal monetary price of consumption is often zero, but time and attention are required. In both the United States and the UK the average person is estimated to spend the equivalent of about 24 hours, the equivalent of a full day, a week online. This makes a full-income perspective on consumption increasingly relevant as digital activity reaches deeper into our lives. This may be either because consumption is paid for with a barter transaction, as in Nakamura *et al.* (2018), or because consumption products are part of a subscription bundle. When this is true, as Goolsbee and Klenow (2006) point out, then the relevant cost that the consumer faces in choosing what and how much to consume is the shadow value of time.[2] Competition in many digital markets is competition for consumers' attention to advertising (Anderson and Peitz, 2019). Additionally, digitally-produced consumption goods, such as social media and product ratings, are increasingly produced with time contributed by households as well as firms.

A key presupposition of standard measures of inflation and productivity is that the utility of a precisely defined market good remains fixed from period to period. But, as Hulten and Nakamura (2020) point out, the utility of a market good to the consumer is not fixed but is affected by changes in household consumption technology. If the household consumption technology is fixed, then the purchase of a given good today has the same effect on utility as the purchase of that good in the previous period. However, digitalization changes the expected utility of goods. For example, pricing doctor visits or semester hours as if they were constant quality does not take into account improvements in the scientific know-how of the doctors and professors. Similarly, online restaurant ratings and reviews may improve a consumer's ability to better match their tastes to dining options. Enhanced information raises utility without changing the good provided or its supply cost. Indeed, any increase in the precision of a consumer's actionable information raises expected utility.

Furthermore, network externalities change the user value of social media,

2 Rosen (1981) argues that the shadow price of time is a crucial cost in all leisure activities with important implications for the incomes of 'superstar' artists and other entertainment workers.

ecommerce platforms, and so on, over time (Schreyer, 2021). The expected utility of any network good rises as it is more widely adopted, although the good does not change. In these examples, the price change also reflects changes in the quality of the good.

A full-income perspective can help account for such considerations arising from digitalization. One difference between GDP and full income is that the former involves only arm's length, monetary transactions, albeit progressively adding non-market transactions (imputed rent for owner-occupied housing, financial intermediation services indirectly measured, a growing range of intangible investments), whereas full income includes the shadow value of all household time and is thus substantially larger. Another way of describing the difference is that full income takes utility seriously: utility maximization should combine all these choice margins: the individual's choices of market hours, home production hours, leisure, and commodities, subject to the time identity and the usual monetary budget constraint (Steedman, 2001). That is, it measures the full experience of an economic agent during the day, including time spent at paid work, at unpaid household work, and at leisure. Either way, it offers a more complete approach than GDP to economic welfare. One can think of a spectrum from real GDP to full (market plus non-market) income to broad economic welfare as full income plus well-being or quality of life (Heys *et al.*, 2019; Bucknall, Heys, and Taylor, 2021).

When it comes to valuation of full income, there are two main perspectives with very different empirical implications. One, due to Becker, is to view the shadow value of unpaid time as equal to the market wage of the worker, on the grounds that this represents the opportunity cost of leisure or of household work. This is the approach used by the UK's Office for National Statistics in its household production satellite accounts. Another is to view the shadow value of time as equal to the market price of household chores, where the price of household chores is the wage rate of household workers, the route adopted by the US Bureau of Economic Analysis in its satellite accounts. These two perspectives produce very different results, as pointed out in Bridgman (2016), particularly as in recent decades the wage rate of household workers has fallen relative to the average wage.

Yet neither of these two approaches can be seen as bounds on the true value of full income. One reason is that the value of leisure or household work time might exceed the market wage. The monetary wage is only one of the possible gains from paid labour. Paid labour may have intrinsic value of various kinds, including the pleasantness of the task, the meaning of work, or on-the-job learning. Conversely, the wage may also overstate the value of time if the task is unpleasant, the work is viewed as unsavory, or depletes one's human capital. Similarly, hiring a worker to perform household chores may have intrinsic costs or benefits to the employer beyond the wage paid. Households may choose to employ a household worker because they experience social benefits beyond the household chores, such as companionship.

These considerations suggest several possible approaches to estimating economy-wide full income. One is to delve more deeply into self-reports of well-being, to measure the utility economic agents derive from alternative activities on both sides of the production boundary and in leisure. Time use studies with subjective modules are available across a variety of countries and time periods. Ultimately, these might lead to direct monetary evaluations of subjective states. A second approach is to look to self-reports of choices of different possible activities, with economists increasingly looking to surveys to understand time allocation, especially when monetary compensation for behavior changes is included in experiments to ensure incentive compatibility as in Brynjolfsson, Colis, and Eggers (2019). And a third approach is to use parametric models with econometric measurement, which require an estimate of the shadow value of time for households. Ultimately, estimates from these methods need to be combined in a meta-analysis.

The base methodology for the measurement of 'real' GDP is to first create nominal GDP accounting for all monetary transactions (plus some imputations) in the economy, and then to deflate it using period-to-period changes in prices of well-defined products. The theoretical rationale is that the deflation methodology approximates the use of an expenditure function (measuring the cost in today's prices of purchasing last period's utility). Thus deflated GDP is a constant-utility construct.

To consider how to develop alternative measures of nominal and real (constant-utility) full income, we begin with subjective utility (Kahneman, 1999) as the sum of (time-separable) utility over time, $\sum_t U(t)$ (Juster, Courant, and Dow, 1981). In Becker's (1965) simplest full income model, utility is consumption of household commodities, which are created using market goods combined with time needed for preparation and consumption. This time is evaluated by the market wage in his model. Households combine time and market goods to produce basic commodities and combine the inputs via household production functions to maximize utility. Their expenditure function includes expenditure on both market goods and time; these are not independent because time can be converted into more market goods by spending more time at work and less in consumption. There is therefore a single budget constraint and the full price of the goods consists of the sum of the prices of the market goods and time used in production, with an associated allocation of time by the household across the production boundary.[3]

Full income can therefore be considered as the sum of money-metric utility over

3 An additional output of activity is learning (and its inverse, human capital obsolescence). Stigler and Becker (1977) emphasize the intertemporal impact of consumption on the utility of future consumption. Note that learning is an investment activity, whose stream of returns may appear as increases in the productivity of work (both at home and for pay) and in the productivity of leisure time. It has long been recognized that homes and consumer durables are investments that provide a stream of consumption services. There are many additional investment activities, including, but not limited to, home improvement, health care, household innovation, and the raising of children. These investment activities occur over the life cycle and have important impacts on the shadow value of time. It is unclear if these ultimately need to be included in total factor productivity measures, but they have important intergenerational impacts.

time. This stream of utilities subsumes the expenditure of time and of market and household produced goods at each point in time. Indeed, this is the standard form of the utility function used in the economics literature in general and the literature on home production in particular. Measurement of 'real' full income requires inferences about the shadow value of time in all activities measured in the same money-metric. During marketed work and household work, this utility captures the intrinsic enjoyment (or its dislike) associated with the activities of production, including the meaning attached to the activity (such as self-expression). Under this approach the utility or disutility of work (both paid work and household production) naturally comes to the fore. In the simplified Becker analysis, the utility or disutility experienced during market labour is assumed implicitly to be zero, which allows the estimation of the marginal utility of time outside of market labour to be equal to the wage. But if this is not the case, the valuation of leisure – as a marginal choice between paid work and leisure – need not be equal to the wage, but rather the wage plus the utility (or minus the disutility) experienced at work.

The shadow value of time is affected by digitalization. The potential for digitization to influence the utility of consumption, and thus the ultimate productivity of economic activity is modeled directly by Hulten and Nakamura (2017), who take into account the possibility that the household production function is not time invariant, but rather that the Internet and information-generating and aggregating technologies influence utility directly, not just through time and goods. For the additional volume and precision of information leads to better consumption choices, so the ongoing advance of knowledge and its availability to the consumer improve the consumption value of purchased products even when the production processes are unchanged. Moreover, in the consumption of expert services, the advance of knowledge implies that these services are better; yet it is difficult to measure this improvement. As the consumption of services entails the cooperation of the consumer with the producer, the information available to the consumer is often determinative of the value of these services.

In either case — changing utility of work or changing utility of consumption — the relationship between work and leisure come into dynamic flux. And the relationship between money earnings and time changes as well. As De Vries (1994) argues, similar changes (in the opposite direction, increasing the marginal utility of money income) previously helped explain the direction of household activity to paid work and consumption of marketed products in a demand-side structural shift parallel to the supply side technological innovations of the Industrial Revolution. Improvements in household technologies in the 1950s and 1960s likely also led to a similar shift.

A key question is whether the market wage is the correct shadow value of time. Becker's simplifications ignore the portion of utility experienced directly during household and wage labour production. Moreover, by identifying the shadow value of time with the average wage, it ignores the complications due the constraint of a standard workweek causing the marginal value of labour to diverge from its average value. Nevertheless, this framework is more likely to capture a full picture of economic welfare in a digital economy with its zero price goods (Brynjolfsson, Colis, and Eggers, 2019), increased involvement of the household in the economy outside of wage labour (Coyle, 2019), and rapid advances in the application of information (Hulten and Nakamura, 2020). Hulten and Nakamura (2020) and Nakamura, Samuels, and Soloveichik (2018) provide evidence that production measures of output growth may be an order of magnitude smaller than welfare measures for specific innovations. We return below to the question of measurement of full income.

Time to Produce

Time use also offers a distinctive lens on production and productivity, as digital technology is changing production time as well as consumer time. There is no material product in some three quarters or more of economic activity now, yet our productivity intuitions relate to material goods transacted monetarily. From the productivity perspective, digital technologies and the pervasive internet mean there are some significant process innovations under way in terms of the time required to produce, and the production boundary.

For both paid labour and home production, productivity in the sense of minimizing the time required to produce a given outcome is an important variable. In exchange for paid labour we obtain many products we cannot produce ourselves or would take us excessive amounts of time to produce. This is one of the meanings of Adam Smith's pin factory, and it is one of the senses in which Smith ascribes value: "The real price of every thing, what every thing really costs to the man who wants to acquire it, is the toil and trouble of acquiring it. What every thing is really worth to the man who has acquired it, and who wants to dispose of it, or exchange it for something else, is the toil and trouble which it can save to himself, and which it can impose on other people." Moreover, in a primitive society (such as Robinson Crusoe's) Adam Smith says, "Labour was the first price — the original purchase-money that was paid for all things."[4]

Digitalization of more service sectors such as law and accountancy or parts of medicine (tele-health, scrutiny of scans, etc.) is now under way and could in principle be expected to improve productivity through speeding up activities currently done by humans. This is similar to the automation of routine tasks in manufacturing. There is as yet little indication that conventionally measured productivity in many services is improving due to the adoption of digital technologies, and indeed some digitally-intensive services such as computer software have been notably poor productivity performers (Coyle and Chung, 2022). However, the measurement challenges when it comes to service sector productivity are considerable, as there is often no standard unit of volume and adjusting for quality is daunting: the quan-

4 These citations can be found in Ricardo (1819), 12-13 in the section of the Principles where he discusses his differences with Smith over the theory of value.

Table 1: A Time-based Perspective on Production, Consumption and Leisure

	Market production	Home production	Leisure/consumption
	Routine manufacturing	Cleaning, driving; domestic robots, self-driving cars may automate some activities	Daily run, personal care, eating (largely non-automatable because inalienable although some market purchases possible eg nail bars, hairdryers)
Routine	Routine services		
	Examples: payroll processing, checkouts, tax preparation		
	Increasingly: medicine, law, accountancy etc		
	Medicine, legal, consultancy;	Cooking, gardening (may also be purchased in the market)	Cooking, gardening (inherently enjoyable for some people)
	Travel agency, banking;	(Increasingly) Travel agency, banking;	
Non-routine	Non-routine manufacturing;	Creative activities eg vlogs, open source software (some people will seek to monetize these)	Creative activities eg vlogs, open source software (done for enjoyment)
	Car repair, driving, plumbing, decorating;		
		Car repair, driving, plumbing, decorating (may also be purchased in the market)	Theatre, concerts, sport, socializing, eating out

Source: developed by authors.

tity of software produced is not measured by any physical metric such as gigabytes, and its quality is unobservable until much later, if at all.

At the same time, some productivity gains made by companies through automating services have simply transferred time input requirements to households. Examples include the use of call centres which require customers to spend more time navigating menus to access the service they need, or automated checkout machines which have largely substituted unpaid household labour for paid store workers. This has been described as a 'time tax' (Lowry, 2021). On the other hand, some transfers of market activities to the household sector through digitalization, such as the shift to online banking or booking travel, has saved people queueing time or increased the variety and quality of the service.

Table 1 indicates how one might categorize these shifts. The first vertical division is the conventional production boundary between GDP and household production, and the second is the boundary between household productive activities and leisure/consumption time; while the horizontal division distinguishes been routine activities which are progressively being digitally automated and non-routine activities.

In the case of routine activities, welfare gains result from digital enabling the activities to be carried out more quickly. For example, in professional services such as accountancy and law, machine learning means routine tasks such as elements of audit or discovery can be automated and carried out much faster than previously. This is a process innovation enabling the firm to reduce costs; customers should get a better (faster) service, and perhaps pay less for it as well (although this is complicated by information asymmetries and mark-ups). There will be general equilibrium effects too, through accountancy and legal process as an intermediate input to other sectors, and through the shifting tasks, pay and employment of lawyers and accountants (which could decline, like drivers of horse-drawn carriages, or increase, like bank employees (in the aggregate) in the face of ATMs, depending on changes in demand for the sectors' services). The process innovations under way in such sectors are unlikely to be captured directly in GDP and conventional productivity calculations, as this would require a quality adjustment to the sector deflators to turn the time-saving improvements into output metrics. The fact that the process innovations enabled by digital technology manifest as time saved, rather than any other reduced input per unit of output, means they are not captured when the time to produce is omitted from the calculation.

In addition, some routine activities are crossing the production boundary — writing wills is one example, formerly involving lawyers, but now more likely a form downloaded off the internet. Travel agency is an example of a non-routine activity partially crossing from market to home production. Coyle (2019) argues that moves out of marketed activity into home production (such as switching from travel agents to booking trips online from home, or the production of free open source software) have become significant. Shifts between market activity and household activity may change the time required for a given output in subtle ways. For instance, self-service gasoline stations may require some work on the part of the driver, but also less waiting for the gas station attendant. Internet shopping implies time saved in traveling to the store, and not having to wait on a queue at the cash register, but may require more time returning disappointing purchases.

These shifts are still evolving. In retail, for instance, there has been a progression from checkouts that use modest capital equipment (conveyor belts and scanners) and much paid labour time, to self-checkouts using more sophisticated capital equipment and unpaid labour time, to checkout-free stores such as Amazon is pioneering, with highly sophisticated physical and intangible capital and scant labour time. On the whole, it is likely that thanks to digitalization there is a net substitution from market to household time-using production such that the measured productivity of affected sectors is lower than in the counterfactual non-digital world. The failure to consider the time savings in production enabled by digital technologies means the measured productivity figures are at present detaching from 'true' contemporaneous productivity (Coyle, 2019).

In the non-routine cases, economic welfare results from the scope to spend more time to both produce and consume simultaneously a higher quality service (more personalized or tailored to individual need, for example). These are also services where the information gains to either consumer or producer will directly increase consumers' utility, as described in the previous section. Thus productivity and utility are inextricably linked.

There is also an overlap between utility derived from how people spend their time and the productivity of their paid activities, as employee satisfaction can improve productivity. For example, Isham, Mair, and Jackson (2020) conclude from a literature review that positive well-being states increase productivity while negative ones are negatively correlated with productivity. A recent meta-analysis of 339 studies found a strong positive correlation between employee well-being, productivity and firm performance (Krekel, Ward, and De Neve, 2019). A number of different psychological mechanisms have been postulated, such as expectancy theories (the expectation of well-being elicits better performance, e.g. Schwab and Cummings, 1970) or that well-being prompts creativity or more positive attitudes (e.g. Baumeister *et al.* 2007). Oswald, Proto, and Sgroi (2015) found that well-being improvements increased productivity significantly in a lab-based task. Edmans (2011) found a link between reported employee well-being and stock market returns among US companies. Satisfied employees likely gain more utility from their workplaces than dissatisfied ones. This affects the shadow value of their time. An interesting question is whether having higher productivity creates happier workplaces or the reverse: what is the source of these gains?

Challenges in the Measurement of Well-being

We turn now to discussing some potential approaches to the measurement of full income and a time-based perspective on economic welfare, before concluding with the implications for statistical collection. The utility measures based on full income that we propose are, at least potentially, provided with a quantitative metric because of their connection to the consumption and production of goods. How far we can proceed down this road is above all an empirical question. Krueger *et al.* (2009) attempt to integrate aggregate time use figures with well-being results in a "National Time Accounting," calculating a national well-being index that tracks changes over time resulting from changing time use patterns among different population groups. They produce a measure that supplements conventional GDP figures but is not a monetary metric — although they argue a money metric is feasible.[5] How then might it be implemented?

Direct measurement of well-being in time spent

There is a large and growing literature on the measurement of the well-being derived from different activities (Frijters and Krekel, 2021). We argued that how people feel while working for pay, producing at home, or at leisure encompasses all the possibilities for well-being. For a real-terms measure, we can ask, just as we do with dollars, how much time would be required to achieve the same utility as in the previous period.

Indeed, time spent offers a potentially more equitable way of valuing non-market goods. Asking people how much they would be willing to pay for something is always skewed by how much income they have (just as markets overly represent rich

5 "In principle it is possible to estimate the monetary price that people are willing to pay on the margin ... For example, the way workers trade off pay for a more or less pleasant job .... Alternatively, the amount that people are willing to spend on various types of vacations can be related to the flow of utility they receive .... Although it is possible ... to put a dollar value on W in this framework, we shy away from this step ..." (Krueger *et al.*, 2009:15). See also Gershuny (2000)

people's preferences). There is new interest in measuring standard GDP growth more democratically (e.g. Aitken and Weale, 2020). Asking people instead how much time they would be willing to spend could also provide more equitable valuations, as time endowments are equal. A democratic measure of full income might be built around time units. In this metric, productivity is directly expressed as an index which would rise inversely with the decline (or fall with the increase) in the time requirements of a given level of welfare.

In a money metric approach to the shadow value of time, to consider how to implement the well-being in time spent framework, we start as above by thinking about a world in which the shadow value of time is equal to the market wage, and hours are fully adjustable. Suppose that digitalization makes leisure time more valuable in well-being terms. It is possible that market hours worked could fall (if the income effect outweighs the substitution effect) and wages could rise as less labour were offered, raising its marginal productivity. Or the reverse could happen.

One way to capture these effects might be to ask participants directly for evaluations of their well-being during different activities, as is done in some time use surveys. Alternatively, we could ask participants to report their monetary valuations of different activities, in effect their consumer surplus. These types of studies have been used in cost-benefit analysis of government-provided free goods, so there is a well-developed literature (for example, Viscusi, 2018; Small, 2012). There has also been a recent literature on monetary valuation of free digital goods (e.g. Goolsbee and Klenow, 2006; Brynjolfsson, Collis, and Eggers, 2019; Coyle and Nguyen, 2020). Moreover, a series of papers have argued that recent increases in the availability of data on time use provide a robust path forward for the measurement of household production, using parametric modeling (Aguiar, Hurst, and Karabarbounis, 2012; Aguiar *et al.* 2021; and Aguiar and Hurst, 2007, 2016).

We would expect such self-evaluations of either kind to be changing as digitalization is causing relative price changes in terms of time as well as money and could be expected to lead not only to shifts in expenditure and consumption patterns but also to changes at the work/leisure/home-production margins as noted in the previous section.

Absent new time use data, it is impossible to be sure about systematic aggregate changes. But since the launch of the first smartphone in 2007 use of the mobile Internet has become an ever-present activity in many people's lives. This has enabled the rapid growth of new services, from social media to digital apps and platforms, as well as new channels of distribution and access. The available statistics indicate substantial growth in the volumes of data transmitted over mobile and fixed networks during the past decade, with average mobile data usage in OECD countries more than doubling between 2017 and 2020 (http://www.oecd.org/sti/broadband/broadband-statistics/).

Substitutions of this kind may also be hard to pin down through existing time use studies, although these have started to be adapted to the digital age (East *et al.*, 2021). Mobile apps often work in the background, giving us reminders, instructions, messages, and information while we are doing other things. In particular, the availability of many possible actions via a smartphone makes the device particularly useful in periods of downtime, such as waiting or queuing. This may turn periods that would otherwise be ones of boredom into active leisure, or home production, in effect creating newly valuable time out of thin air. Self-reports are one way to explore these dimensions. In principle, time use surveys can capture the primary and alternative activities people are engaged in at a given time, although this is clearly somewhat harder than ascertaining whether somebody is ironing and watching TV at the same time. Time use statistics including the full array of digital activities are essential for understanding the digital economy.

Evaluating well-being

The evaluation of well-being is a core issue for our proposal, and there is a substantial literature on this question. Here we briefly review some of the key open questions for statistical approaches.

The contrast between asking a general question (as in Juster, 1985) and a specific retrospective time period question (as in Kahneman and Krueger, 2006) is related to Kahneman's (1999) distinction between "objective" and "subjective" utility. For objective utility, we want to know how an experience feels in real time. It is evident that our recollection of the past may differ from our moment-to-moment feelings. Gershuny (2013) and Krueger *et al.* (2009) consider self-reports on the enjoyment experienced during different activities, such as at work, driving in traffic, or at leisure out of the home. Gershuny deploys mean activity enjoyment scales, while Krueger *et al.* use unhappiness indexes, measured as the proportion of time during the event when negative feelings are rated as strongest. Both are based on diary self-reports. However, Krueger *et al.* present evidence that, on average, remembered feelings are reflective of moment-to-moment feelings, as detected in surveys conducted with special devices for recording feelings at specific points in time. This is an ongoing area of study, and it is possible that progress could be made since the use of mobile devices for reporting may enable low-cost extensions of these surveys.

Extensive studies by behavioral economists and psychologists on decision-making suggest that we often follow rules of thumb rather than making explicit utility maximization decisions. How does this affect the welfare value of consumption revealed by purchases? Benjamin *et al.* (2012) asked individuals to choose between alternative bundles, such as having a lower rent (20 per cent of income) and a longer commute (45 minutes) or a higher rent (40 per cent of income) and a shorter commute (10 minutes). Moreover, they asked the same individuals whether they believed this choice would lead to higher life satisfaction, greater happiness with life as a whole, or greater felt happiness (subjective well-being). They found that there are systematic differences between the choices people say they would make and what would maximize these various definitions of happiness. They also found that higher life satisfaction is most aligned with choice, while subjective well-being is less so.

Another issue is whether a single dimensional measure, such as happiness is the appropriate way to measure episodic utility. Krueger *et al.* (2009) use five dimensions of feeling and combine them to distill an overall measure of time spent in unpleasantness; a time period is unpleasant when the strongest feeling experienced is negative (stressed, in pain, or sad, as opposed to happy or interested). This allows for the fact that, for example, an episode of work may contain more elements of pain or stress than, say, watching television. Can these multidimensional feelings be placed in a single metric as Krueger *et al.* suggest? For that matter, are scaled self-reports associated with specific activities, whether single dimensional or multidimensional, in turn relatable to scaled self-reports of overall happiness, as in the Cantril scale (that is, how they rate their lives currently on a scale of 0 to 10 with respect to the best possible life they could be leading?) To the extent that the Cantril scale can be related to log measures of income, it may be possible to apply meaningful monetary values to specific activities. In turn, we might be able to associate these feelings with actual expenditures. That is, when someone pays to attend a concert or for a meal, do their feelings line up with their expenditures? Or are the feelings we experience and report partly mediated by the size of our outlays? On the other hand, Kahneman and Deaton (2010) provide evidence that Cantril scale reports and emotional well-being scales are less well correlated with higher incomes, which would limit the value of this strategy. One possibility is to use stated preferences to predict out-of-sample behavioral consequences, as suggested by Bernheim *et al.* (2013). They advocate using econometric techniques to measure the extent to which revealed preferences are predicted by stated preferences.

To the extent we can reconcile the results of different methods, the more confidence we can have in them. But there are several additional challenges in implementing the measurement at an aggregate level of well-being across activities.

For one, subjective reports will differ across individuals. How an individual scores feelings will contain random elements, possibly both person-specific and time- or context-specific. One way to deal with this is to treat these reports as a dependent variable with proxies for true utility on the right hand side, as in Blanchflower and Oswald (2004).

Another caveat is that work can be enjoyable or not, yet even when intrinsic job satisfaction is low, there are benefits from the social attachments and status that come with paid employment. There is evidence that the non-monetary aspects of work are significant, and people seek intrinsic meaning in their paid work (Cassar and Meier, 2018). What's more, the (dis-)utility of work appears to be changing over time as the character of work changes, and there are also substantial variations between groups (Kaplan and Schulhofer-Wohl, 2018; Jahoda, 1981), on the "latent" value of work.

Some home production activities are similarly enjoyable and blend with consumption (including of leisure activities), while others are clearly "chores" (Gershuny and Fisher, 2014). Leisure can also be productive. While we are at leisure, we can come up with good ideas or upload content that others may enjoy and learn from: Sichel and von Hippel (2019) argue that household research and development is substantial relative to private research and development.

Finally, well-being, on the standard Cantril scale, is measured relative to the 'best possible' life. The best possible life changes over time due to economic innovation. That is, novel economic possibilities, such as greater longevity, deeper scientific understanding, tastier food, and more captivating entertainment, may change the definition of the best possible life. This will affect the measurement of well-being over time.[6]

Despite these complexities, to a first approximation we might think that less time spent (holding output constant) in paid labour and home production — that is, in what we call 'work' — are an improvement in welfare. Conversely, increases in time working (either in home production or for pay) given constant output are, in principle, welfare worsening. For leisure, the presumption is the opposite: To a first approximation, the more time allocated to it, the better. It is likely that for many activities there are diminishing returns. How much time one spends at a given activity depends on how rapidly the returns diminish. On the other hand, in general, more time spent at leisure suggests more enjoyment per unit of time for that activity. This is the hypothesis that underlies the Goolsbee and Klenow (2006) analysis of the internet. Of course, unemployment is a bad (forced) "leisure" in that it restricts our ability to obtain the highly productive goods of the marketplace, which may force us back toward the less productive branches of home production. And this overall low level of productivity likely further lowers the enjoyment of leisure time, as we are denied the goods we are accustomed to consuming at leisure.

Boerma and Karabarbounis (2019) draw the distinction between home work and leisure using the substitutability or complementarity of time in production — in home work, time is substitutable with market goods (think washing machines) while in leisure time is complementary with market goods.

Monetary measures of well-being: the shadow price of time

It is important to note that it is when holding income constant that reductions in time spent in either paid labour or household production are leisure- and welfare-enhancing. Holding income constant implies that a monetary measure is required.

The literature often assumes the shadow value of time is given by the wage rate. Are there other ways to assign monetary shadow prices to the feelings of well-being in different uses of time? There are several options. Essentially these correspond to the debate about the relationship between stated preference, stated feelings, and revealed preference measures. Economists place more weight on revealed preference measures, but a good deal of policy-oriented welfare analysis rests upon

6 The best possible life may also be affected by social and political factors, such as the emergence of new rights and freedoms. However, that is outside the scope of this discussion.

stated preferences as providing valuable additional evidence. An excellent example of this can be seen in Small's (2012) discussion of the valuation of travel time as a crucial input into any cost-benefit analysis of transportation policy. He discusses travelers' stated valuation of travel time costs and compares it to their preferences as revealed, for example, by econometric analyses of commuting time-rental trade-offs. He points out that the evidence for the welfare impact of in-vehicle amenities is thin. Amenity questions in this example will become even more salient as we realize the possibility of partially or totally self-driving cars. Reported measures of happiness or other feelings while driving may help bridge this gap.

Self-reports of value of time are one route, asking survey participants directly about their shadow value of time, just as Brynjolfsson *et al.* (2018, 2019) and Coyle and Nguyen (2020) ask about the monetary value of different digital consumption/leisure activities. Such studies introduce monetary scales of utility in the evaluation of goods, asking how much subjects would be willing to pay for a given amenity (such as social media) or how much they would be willing to accept to do without the amenity.

If we were to ask workers how much they would require to work an extra hour at a 'neutral' job — one that, say, requires some concentration but is not stressful — the difference between the pay they would demand for this compared with their current job could be a metric of the utility cost (or benefit) of their work. Pay at the 'neutral' job should reflect the true marginal value of leisure. This would be analogous to the standard use of hedonic wage regressions in order to isolate the marginal benefit or disbenefit of certain job characteristics as compared with average wages. This might help selecting between using the wage rate of the individual or the wage rate of the task performed as the relevant price.

Happiness reports are another approach. For although stated preference studies are widely used in environmental and cultural economics, the more usual approaches to self-reports of utility in the context of the well-being literature are based on arbitrary scales. The best-known of these are the happiness studies, where subjects are asked to report, for example, in terms of the Cantril ladder. While this scale is both arbitrary and context-specific, Deaton (2008) and Stevenson and Wolfers (2008) show that responses across countries are on average well approximated by a linear regression on log income per capita. So self-reports of utility appear to be relatable to a cardinal, monetary measure of utility.[7]

Surveys could ask: What would people be willing to pay for an extra day's vacation, provided their workloads were reduced? What would they have to be paid to work an extra day, assuming their workloads were not reduced? What would they pay for someone else to perform a house-

7 Because the frame for the Cantril ladder is "the best possible life," the definition of the best possible life evolves over time with new discoveries. It is less evident that these happiness measures correlate with measured real GDP over time. Benjamin *et al.* (2012) ask students whether they would choose to have been born about when they were (1990) or in 1950; 87 per cent would choose their actual date, which contrasts with the Cantril ladder results indicating that well-being has remained flat over time.

hold chore (pay rates on some digital platforms would provide alternative evidence of this) or at what pay would individuals work an additional hour at their current jobs or at some benchmark alternative? The answers to such questions could then be related to their wage rates and the measured, experienced utility of labour.

In an alternative approach, not reliant on such methods, Bridgman (2016) uses estimates of the replacement cost of household activities to derive a first version of a household production account. Since the average wage rate for household employees across types of work does not vary very much, we can easily approximate the value of household production if we assume that hired labour is a reasonably good substitute for home production. This approach assumes that the shadow price of time for highly paid workers can be equated to the wage rate of household employees. But if highly paid workers are, say, deeply concerned about their children's education and/or enjoy their interactions with their children, then the shadow price of their time may be substantially higher. The former implies greater household production, but as investment, while the latter adds to consumption (Doepke and Zilibotti, 2019). Diewert, Fox, and Schreyer (2018) show how to estimate the shadow price of household production using the own-wage or the wage rate of potential employees, as well as the case when neither wage rate is applicable.

Alpman, Murtin, and Balestra (2018) take yet another approach, using experienced well-being and time use surveys combined with money measures to estimate directly the monetary value of non-market activities. In essence, they scale money expenditures with estimates of experienced well-being (along the lines of Krueger *et al.*, 2009) within a representative agent framework to estimate total income for a range of countries. Their approach is somewhat ad hoc. Yet they are able to link time use and well-being ratings to different activities to estimate the relative "well-being" valuations of non-market activities and then multiply these by total consumption expenditure to derive a monetary measure of welfare.

Intrinsic well-being at work

The question of how much enjoyment people can derive from work has nagged at economists since the studies that underlie Juster, Courant, and Dow (1981) first revealed how many people value their work. This is consistent with the emphasis in the positive psychology literature on "flow," or satisfying absorption in a meaningful activity (Nakamura and Csikszentmihalyi, 2002). The standard full-income approach assumes that the work itself is neither pleasant nor unpleasant. But some people have jobs they enjoy quite a lot, while others report that they find their work relatively unpleasant. Rothwell and Crabtree (2019) provide survey evidence that job satisfaction beyond the wage is important to workers and correlates with reported well-being. The value of leisure depends then on both the wage the worker receives and how much intrinsic utility they obtain from that job. This may change over time as digitalization changes the character of many jobs, making some (data scientists) more satisfying and others (warehouse workers) far less so; it is possible that the average utility of labour and its distribution as experienced have changed, as argued by Kaplan and Schulhofer-Wohl (2018). Other changes may be occurring if the population is experiencing greater distress, as suggested by Case and Deaton (2017) and Deaton (2018).

Maestas *et al.* (2018) ask workers about their preferences for working conditions, such as flexibility in hours, vacation time, and meaningfulness of the work, and how much they would be willing to accept in pay reductions to change them. This enables them to discuss the extent to which working conditions exacerbate wage inequality. The answers will likely also reflect the shadow value of time. Mas and Pallais (2017) ask similar questions in the context of call centers, where they can also measure the revealed preferences of the workers. The experience of the pandemic has also changed people's preferences over the jobs they hold (the so-called 'great resignation,' Cook, 2021), and the location and hours of work as compared with household production and leisure (Barrero, Bloom, and Davis, 2021).

An additional question raised by Cassar and Meier (2018) is whether the experienced utility measures that we use are adequate for capturing non-monetary incentives that may affect the shadow value of time. In particular, they argue that the meaningfulness of labour, particularly as captured in the mission or purpose of the work (for example, in the non-profit or arts sectors), has an important impact on the pay workers are willing to accept for a given task. A variety of empirical evidence in the human resource management literature bears on this question.

Equally, the utility people receive from different types of non-market production may vary; for example, Lerner and Tirole (2003) suggest that developers of open source software gain three types of utility: enjoyment from the activity, peer esteem, and future rewards in terms of pay and promotion in their career. Juster, Courant, and Dow (1981) and Juster and Stafford (1991) have argued that a more complete welfare accounting might include the underlying utility experience at both paid work and household production.

The Way Forward

The agenda of measuring broader economic welfare and productivity in terms of a money metric of the well-being afforded by different allocations of time, with digitally-driven re-allocations across the production boundary and the work/leisure boundary, must address open questions as discussed above in order to progress.

The key underlying requirement is the need for more detailed and regular time use data, including digital activities. Other open questions concern;

- the concepts and measurement of well-being in different activities;
- measurement of of the shadow value of time in monetary terms;
- the distinction between routinizable and other activities as reflected in changing time allocations;
- the link between well-being in time at work and the quality/productivity of the work.

We have set out a series of questions about the linkages among measures of utility, consumption expenditures, and time allocation to work and leisure, and about the measurement of the shadow value of time. These research questions derive from the earlier seminal work on time use by Becker, Gershuny, Kahneman, Juster, Krueger, and many others. This distinguished tradition is given new urgency not only by the current public debate about the inadequacy of conventional real GDP as a measure of economic welfare or progress but also by the evident significant changes in time use in both consumption and production processes due to digital innovation.

How should statistical agencies move forward on this agenda? There are two parts to our answer. First, statistical agencies need to consider new measures of output that better capture the utility impacts of the changing economy and time use. Many agencies already produce household accounts, such as the BEA and ONS. They could augment these with others, such as the proposed retail satellite account that U.S statistical agencies are establishing under the leadership of the Bureau of labour Statistics. Time saving might be captured in the retail satellite account where we take into account the consumer's time spent shopping (including driving time, as suggested by Mandel, 2017) and spent checking out (as discussed above).

Second, statistical agencies need to broaden their regular collection practices to include the data they will need to support the regular updating of the experimental satellite accounts, so that they can eventually include the results in the main accounts. It may be that aggregate statistics evolve beyond the national accounts largely capturing transactions to move closer to welfare measurement and capturing more of the benefit of innovations. This would correspond to the impetus to go beyond real GDP to official GDP plus (Brynjolfsson *et al.*, 2019) or expanded GDP (Hulten and Nakamura, 2020) or a full spectrum from market activities to broad economic welfare (Heys *et al.*, 2019)

In addition, we want to move from measurement of shifting time allocations to money metrics. There is therefore a rich research agenda concerning the meaning of self-reports on different methodologies (with unbiased self-reports difficult to obtain and so challenging for official statistical production), the utility derived from different activities at leisure and at work, the best approach to applying a money metric, and the potential need for more than one dimension to measure economic welfare.

In addition to the digital transformation that has been our focus, these questions arise in the context of pandemic, ecological crisis and geopolitical conflict. Citizens are unsurprisingly asking questions about how official measures capture well-being changes. Health outcomes, in a world in which some countries' health care expenditures can exceed ten per cent of GDP, are an increasingly important part of measured real growth. At the same time, the relationship between such real expenditures and either direct health outcomes or well-being are not closely connected. Health outcomes, as the pandemic has shown, are the outcome of health shocks and of prior health expenditures and accumulated human capital. As a consequence, well-being may be worsened by health shocks, regardless of the efficacy of the health care system. Such shocks would, in principle, be registered as a decline in measures of full income. These costs include excess deaths, millions of COVID and long COVID patients, and the psychological and educational tolls of isolation, fear, and disruption, as well as direct economic costs.

In addition, the unprecedented speed with which the pharmaceutical firms and governments were able to develop, trial, approve, and manufacture vaccines that are highly protective against the new coronavirus and its variants is a credit to advances in the world economy and evidence of its puissance. This appears to be allowing much of the world to return to more normal levels of activity, a feat that could be valued as worth trillions of dollars.

How should we incorporate these events into both the time series of GDP and full income? In current SNA practice, we generally do not see either the full, extraordinary costs of the pandemic or the extraordinary economic benefits from the innovative ideas, development and distribution of vaccines. The metrics include only the expenditures on the development of the vaccine made by pharmaceutical companies and governments, and then the costs of producing and distributing the vaccine, largely borne by governments.

The rise in global temperatures and other ecological impacts associated with climate change raise similar questions about how to develop a measure of well-being that incorporates all relevant considerations. For example, in the formulation developed by Partha Dasgupta (2021), nature impacts the economy as a set of assets or resources and also as a direct influence on the environment in which consumption and other economic activities take place.

Both types of challenge illustrate the growing wedge between standard national accounts measures, productivity and well-being, and further underline the case for new approaches to measurement. In GDP economists have constructed a measure based on expenditure and output, imperfectly adjusted through deflation to link to underlying utility. The effort to develop an improved measure of economic well-being, reflected in the growing attention paid to measurement issues, is unlikely to have as well-defined or uncontroversial a quantification as current measures of GDP until this research agenda is much further advanced. Agreement on measurement is more likely to come about if we examine economic well-being through multiple lenses and work toward an understanding about the most convincing ways to measure it.

How might this quantification be established as a long-term means of evaluating a national economy's contribution to the welfare of its residents? Macroeconomists and policymakers currently rely upon GDP and its components to answer this question. If there is an increasing gap between the answer supplied by measures of GDP and measures based on welfare, then it may be that a measure of welfare should become part of the system of official statistics. Establishing this additional accounting may be crucial if economists are to be able to discuss economic policy issues meaningfully, in a context in which there is growing public questioning of whether real GDP growth is an adequate measure of broad economic progress. However, this task will require a sustained dialogue between government statisticians and the economics profession at large.

References

Aguiar, M., M. Bils, K.K. Charles, and E. Hurst (2021) "Leisure Luxuries and the Labor Supply of Young Men," *Journal of Political Economy*, Vol. 129. No. 2, February, pp. 337-382.

Aguiar, M. and E. Hurst (2007) "Measuring Trends in Leisure: The Allocation of Time Over Five Decades," *Quarterly Journal of Economics*, Vol. 122, No. 3, August, pp. 969-1006.

Aguiar, M. and E. Hurst (2016) "The Macroeconomics of Time Allocation," in John Taylor and Harald Uhlig (eds.), *Handbook of Macroeconomics*, Elsevier, Vol. 2.

Aguiar, M., E. Hurst, and L. Karabarbounis (2012) "Recent Developments in the Economics of Time Use," *Annual Review of Economics*, Vol. 4, pp. 373–397.

Aitken, A. and M. Weale (2020) "A Democratic Measure of Household Income Growth: Theory and Application to the United Kingdom," *Economica*, Vol. 87, June, pp. 589-610.

Alpman, A., F. Murtin, and C. Balestra (2018) "Unveiling the Monetary Value of Non-Market Activities Using Experienced Well-Being and Time Use Surveys," paper presented at the 35th IARIW General conference. http://www.iariw.org/copenhagen/alpman.pdf.

Anderson, S.P. and M. Peitz (2019) "Ad Clutter, Time Use and Media Diversity," Discussion Paper Series CRC TR 224, No. 140, University of Bonn and the University of Mannheim, December.

Barrero, J.M., N. Bloom, and S. Davis (2021) "Why Working from Home Will Stick," NBER Working Paper 28731, National Bureau of Economic Research, April.

Baumeister, R.F., K.D. Vohs, C.N. DeWall, and L. Zhang (2007) "How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather than Direct Causation," *Personality and Social Psychology Review*, Vol. 11, No. 2, pp. 167-203.

Becker, G.S. (1965) "A Theory of the Allocation of Time," *Economic Journal*, Vol. 75, pp. 493–517.

Benjamin, D.J., O. Heffetz, M.S. Kimball, and A. Ree-Jones (2012) "What Do You Think Would Make You Happier? What Do You Think You Would Choose?," *American Economic Review*, Vol. 102, No. 5, pp. 2083–2110.

Bernheim, B.D., D. Bjorkegren, J. Naecker, and A. Rangel (2013) "Non-Choice Evaluations Predict Behavioral Responses to Changes in Economic Conditions," NBER Working Paper 19269, National Bureau of Economic Research.

Blanchflower, D. G. and A. J. Oswald (2004) "Well-Being over Time in Britain and the USA," *Journal of Public Economics*, Vol. 88, No. 7-8, pp. 1359–1386.

Boerma, J. and L. Karabarbounis (2021) "Inferring Inequality with Home Production," *Econometrica*, Vol. 89, No. 5, pp. 2517-2556.

Bridgman, B. (2016) "Accounting for Household Production in the National Accounts: An Update, 1965–2014," *Survey of Current Business*, Vol. 96, No. 2, pp. 1–5.

Brynjolfsson, E., F. Eggers, and A. Gannamaneni (2018) "Measuring Welfare with Massive Online Choice Experiments: A Brief Introduction," *AEA Papers and Proceedings*, Vol. 108, pp. 473-476.

Brynjolfsson, E., A. Collis, W.E. Diewert, F. Eggers, and K.J. Fox (2020) "Measuring the Impact of Free Goods on Real Household Consumption," *AEA Papers and Proceedings*, Vol. 110, pp. 25-30.

Brynjolfsson, E., A. Collis, and F. Eggers (2019) "Using Massive Online Choice Experiments to Measure Changes in Well-Being," *Proceedings of the National Academy of Sciences*, Vol. 116, No. 15, pp. 7250-7255. https://doi.org/10.1073/pnas.1815663116.

Bucknall, R., S. Christie, R. Heys, and C. Taylor (2021) "GDP and Welfare: Empirical Estimates of a Spectrum of Opportunity," ESCoE Discussion Paper 2021-08, Economic Statistics Centre of Excellence, July. https://www.escoe.ac.uk/publications/gdp-and-welfare-empirical-estimates-of-a-spectrum-of-opportunity/.

Case, A. and A. Deaton (2017) "Mortality and Morbidity in the 20th Century," *Brookings Papers on Economic Activity*, Spring, pp. 397–443.

Cassar, L. and S. Meier (2018) "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, Vol. 32, No. 3, pp. 215–238.

Cook, I. (2021) "Who Is Driving the Great Resignation?," *Harvard Business Review*, September 15. https://hbr.org/2021/09/who-is-driving-the-great-resignation.

Coyle, D. (2014) *GDP: A Brief but Affectionate History*, Princeton University Press.

Coyle, D. (2019) "Do-it-yourself Digital: the Production Boundary, the Productivity Puzzle and Economic Welfare," *Economica*, Vol. 86, pp. 750-774. https://doi.org/10.1111/ecca.12289.

Coyle, D. and D. Nguyen (2020) "Free Goods and Economic Welfare," ESCoE Discussion Paper 2020-18, Economic Statistics Centre of Excellence, December. https://www.escoe.ac.uk/publications/free-goods-and-economic-welfare-escoe-dp-2020-18/.

Coyle D. and J.-M. Chung (2022) "Diagnosing the Productivity Slowdown: Which Sector Matter," The Productivity Institute Working Paper 018, March. lhttps://www.productivity.ac.uk/wp-content/uploads/2022/03/WP018-Diagnosing-the-UK-productivity-slowdown-FINAL-150322.pdf.

Dasgupta, P. (2021) *The Economics of Biodiversity: The Dasgupta Review*, February. https://www.gov.uk/government/collections/the-economics-of-biodiversity-the-dasgupta-review.

Deaton, A. (2008) "Income, Health, and Well-Being Around the World: Evidence from the Gallup World Poll," *Journal of Economic Perspectives*, Vol. 22, No. 2, pp. 53-72.

Deaton, A. (2018) "What Do Self-Reports of Wellbeing Say About Life-Cycle Theory and Policy," NBER Working Paper 24369, National Bureau of Economic Research, March.

De Vries, J. (1994) "The Industrial Revolution and the Industrious Revolution," *The Journal of Economic History*, Vol. 54, No.2, pp. 249-270.

Diewert, W.E., K.J. Fox, and P. Schreyer (2018) "The Allocation and Valuation of Time," paper presented at the 35th IARIW General Conference. http://www.iariw.org/copenhagen/fox.pdf.

Doepke, M. and F. Zilibotti (2019) *Love, Money, and Parenting, Princeton University Press.*

East, S., W. King, C. Payne, G. Vassilev, and S. Wallace (2021) "The ONS Online Time Use Survey: Creating a Feasible Platform for 21st Century Time Use Data Collection," paper prepared for the 36th IARIW Virtual General Conference, August 23. https://iariw.org/wp-content/uploads/2021/08/TheONSOnlineTimeUseSurvey_paper.pdf.

Edmans, A. (2011) "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices," *Journal of Financial Economics*, Vol. 101, No. 3, pp. 621-640.

Frijters, P. and C. Krekel (2021) Chapter 2, *A Handbook for Wellbeing Policy-Making: History, Theory, Measurement, Implementation, and Examples*, Oxford University Press.

Gershuny, J. (2000) Chapter 8, *Changing Times: Work and Leisure in the Postindustrial Society*, Oxford University Press.

Gershuny, J. (2013) "National Utility: Measuring the Enjoyment of Activities," *European Sociological Review*, Vol. 29, No. 5, pp. 996–1009.

Gershuny, J. (2018) "Gender Symmetry, Gender Convergence and Historical Work-Time Invariance in 24 Countries," University of Oxford Centre for Time Use Research Working Paper, February. https://www.timeuse.org/sites/default/files/2018-02/CTUR%20WP%202%202018_1.pdf.

Gershuny, J. and K. Fisher (2014) "Post-industrious Society: Why Worktime Will Not Disappear for Our Grandchildren," Oxford University Sociology Working Paper 2014-03, April. https://www.sociology.ox.ac.uk/materials/papers/wp20143.pdf.

Goolsbee, A. and P.J. Klenow (2006) "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet," *American Economic Review*, Vol. 96, No. 2, pp. 108–113.

Heys, R., J. Martin, and W. Mkandawire (2019) "GDP and Welfare: A Spectrum of Opportunity," ESCoE Discussion Paper 2019-16, Economic Statistics Centre of Excellence.

Hulten, C. and L. Nakamura (2017) "Accounting for Growth in the Age of the Internet: The Importance of Output-Saving Technological Change," NBER Working Paper 23315, National Bureau of Economic Research, April.

Hulten, C. and L. Nakamura (2020) "Expanded GDP for Welfare Measurement in the 21st Century," in C. Corrado, J. Miranda, J. Haskel, and D. Sichel (eds.), *Measuring and Accounting for Innovation in the 21st Century*, NBER Studies in Income and Wealth, University of Chicago Press.

Isham, A., S. Mair, and T. Jackson (2020) "Wellbeing and Productivity: A Review of the Literature," CUSP Working Paper No. 22, Centre for the Understanding of Sustainable Prosperity, University of Surrey.

Jahoda, M. (1981) "Work, Employment, and Unemployment: Values, Theories, and Approaches in Social Research," *American Psychologist*, Vol. 36, No. 2, pp. 184–191.

Juster, F.T. (1985) "Preferences for Work and Leisure," in F.T. Juster and F.P. Stafford (eds.), *Time, Goods, and Well-being*, Survey Research Center, University of Michigan, pp. 333–351.

Juster, F.T., P.N. Courant, and G.K. Dow (1981) "A Theoretical Framework for the Measurement of Well-Being," *Review of Income and Wealth*, Vol. 27, No. 1, pp. 1–31.

Juster, F.T. and F.P. Stafford (1991) "The Allocation of Time: Empirical Findings, Behavioral Models, and Problems of Measurement," *Journal of Economic Literature*, Vol. 29, No. 2, pp. 471–522.

Kahneman, D. (1999) “Objective Happiness,” in D. Kahneman, E. Diener, and N. Schwartz (eds.), *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage Foundation, pp. 3-25.

Kahneman, D. (2011) *Thinking Fast and Slow*, Farrar, Strauss, and Giroux.

Kahneman, D. and A. Deaton (2010) “High Income Improves Evaluation of Life but Not Emotional Well-being,” *Proceedings of the National Academy of Sciences*, Vol. 107, No. 38, September 10, pp. 16489-93.

Kahneman, D., E. Diener, and N. Schwartz (eds.) (1999) *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage Foundation, pp. 3-25.

Kahneman, D. and A.B. Krueger (2006) “Developments in the Measurement of Subjective Well-Being,” *Journal of Economic Perspectives*, Vol. 20, No. 1, pp. 3-24.

Kaplan, G. and S. Schulhofer-Wohl (2018) “The Changing (Dis-)utility of Work,” *Journal of Economic Perspectives*, Vol. 32, No. 3, pp. 239-258.

Krekel, C., G. Ward, and J.-E. De Neve (2019) “Employee Wellbeing, Productivity and Firm Performance,” CEP Discussion Paper 1605, Centre for Economic Performance, London School of Economics.

Krueger, A.B., D. Kahneman, D. Schkade, N. Schwartz, and A.A. Stone (2009) “National Time Accounting: The Currency of Life,” in A.B. Krueger (ed.), *Measuring the Subjective Well-Being of Nations*, University of Chicago Press, pp. 9–86.

Lerner, J. and J. Tirole (2003) “Some Simple Economics of Open Source,” *Journal of Industrial Economics*, Vol. 50, No. 2, pp. 197–234.

Lowry, A. (2021) “The Time Tax,” *The Atlantic*, 27 July. https://www.theatlantic.com/politics/archive/2021/07/how-government-learned-waste-your-time-tax/619568/.

Maestas, N., K.J. Mullen, D. Powell, T. von Wachter, and J.B. Wenger (2018) “The Value of Working Conditions in the United States and Implications for the Structure of Wages,” NBER Working Paper 25204, National Bureau of Economic Research, October.

Mandel, M. (2017) “How Ecommerce Creates Jobs and Reduces Income Inequality,” *Progressive Policy Institute*. https://www.progressivepolicy.org/wp-content/uploads/2017/09/PPI_ECommerceInequality-final.pdf.

Mas, A. and A. Pallais (2017) “Valuing Alternative Work Arrangements,” *American Economic Review*, Vol. 107, No. 12, pp. 3722–3759.

Nakamura, J. and M. Csikszentmihalyi (2002) “The Concept of Flow,” in C.R. Snyder and S.J. Lopez (eds.), *Handbook of Positive Psychology*, Oxford University Press, pp. 89–105.

Nakamura, L., J. Samuels, and R. Soloveichik (2018) “‘Free’ Internet Content: Web 1.0, Web 2.0, and the Sources of Economic Growth,” Federal Reserve Bank of Philadelphia Working Paper 18-17, May.

Oswald, A.J., E. Proto, and D. Sgroi (2015) “Happiness and Productivity,” *Journal of Labor Economics*, Vol. 33, No. 4, pp. 789-822.

Reid, M. (1934) *The Economics of Household Production*, Wiley.

Ricardo, D. (1819) *On the Principles of Political Economy and Taxation*, 1st American Edition, Georgetown: Joseph Milligan.

Rosen, S. (1981) “The Economics of Superstars,” *American Economic Review*, Vol. 71, December, pp. 845-858.

Rothwell, J. and S. Crabtree (2019) “Not Just a Job: New Evidence on the Quality of Work in the United States,” *Gallup* https://www.gallup.com/education/267590/great-jobs-lumina-gates-omidyar-gallup-quality-report-2019.aspx.

Schreyer, P. (2021) “Accounting for Free Digital Services and Household Production: An Application to Facebook,” paper prepared for the 36th IARIW Virtual General Conference, August 24. https://iariw.org/wp-content/uploads/2021/08/Schreyer_paper.pdf.

Schwab, D.P. and L.L. Cummings (1970) “Theories of Performance and Satisfaction: A Review,” *Industrial Relations*, Vol. 9, pp. 408-430.

Sichel, D. and E. von Hippel (2021) “Household Innovation, R&D, and New Measures of Intangible Capital,” *Review of Income and Wealth*, Vol. 67, September, pp. 639-658.

Small, K.A. (2012) “Valuation of Travel Time,” *Economics of Transportation*, Vol. 1, No. 1–2, pp. 2–14.

Steedman, I. (2001) *Consumption Takes Time: Implications for Economic Theory*, Routledge.

Stevenson, B. and J. Wolfers (2008) “Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox,” *Brookings Papers on Economic Activity*, Spring, pp. 1–87.

Stigler, G.J. and G.S. Becker (1977) “De Gustibus Non Est Disputandum,” *American Economic Review* Vol. 67, March, pp. 76-90.

Viscusi, W.K. (2018) *Pricing Lives*, Princeton University Press.

The Link Between the Standard of Living and Labour Productivity in the UK: A Decomposition

Nicholas Oulton[1]

Centre for Macroeconomics, London School of Economics, and National Institute of Economic and Social Research

Abstract

According to Paul Krugman, "Productivity isn't everything, but in the long run it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker." But productivity and the standard of living are different concepts and are measured in different ways, so the question is, what is the linkage between them? Productivity is typically measured by GDP per hour. The standard of living has potentially many aspects such as health, longevity, personal security, and relationships. But here I take a narrower view and stick to the national accounts. So the standard of living is measured by the household disposable income of the median individual. I use the median rather than the mean so that inequality is taken into account. I develop a decomposition of the growth of median household income which relates it to the growth of productivity via eight additional factors, one of which is inequality; four other factors are measures of labour market performance. I apply this decomposition to the UK over the period 1977 to 2019. I find that productivity growth was by far the most important factor in accounting for the growth of living standards which was substantial up to 2007; rising inequality prior to 2007 retarded the growth of living standards but not by much. Since 2007 productivity growth has collapsed as has also the growth of living standards. The fall in the latter has been mitigated somewhat by a fall in inequality.

1 The author is a member of the Centre for Macroeconomics at the London School of Economics, and a Fellow of the National Institute of Economic and Social Research, London. This article was presented to the "Productivity and Well-being: Measurement and Linkages" Workshop, November 16 and 17, 2021, hosted by The Productivity Institute at Manchester University Business School. I am grateful to participants, particularly my discussant John Fernald, for helpful comments. I also thank Rachel Soloveichik for an enlightening discussion about equivalence scales. For helpful comments in revising the article, I am grateful to Josh Martin, Andrew Sharpe, and two anonymous referees. Email: n.oulton@lse.ac.uk.

According to Paul Krugman (1994, Chapter 1), "Productivity isn't everything, but in the long run it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker." Though this seems intuitively likely, the link between the two concepts is not straightforward. A standard measure of productivity (understood here to mean labour productivity) is GDP per hour worked (an improvement on Krugman's GDP per worker). A reasonable measure of what might be termed economic welfare or the standard of living is what the UK's Office for National Statistics calls "median equivalized households disposable income." Households disposable income (HDI) is the income that people actually receive from all sources, including cash benefits, and after taxes on income. The median not the arithmetic mean is studied, since we are interested in the experience of the typical person and the mean may be distorted by the gains accruing to the rich (the top 1 per cent or top 0.1 per cent). We may also be interested in the welfare of other groups, say the bottom 20 per cent or 5 per cent. And in a welfare context household income per equivalent adult is better than just household income per person since households differ in size and by whether or not they contain children; use of an equivalence scale allows for the different needs of different groups.

The aim of this article is to present a decomposition of the growth of economic welfare which links it through a series of factors to the growth of productivity. These factors include inequality, the relative prices of consumption and output, the share of households in national income, household composition, and a set of labour market factors including unemployment and labour force participation. The decomposition is then quantified on UK data for the period 1977-2019. So at the end we will be able to say how much of the growth of living standards is due to the growth of productivity and how much to the growth of these other factors, at least in a statistical sense.

Of course, welfare (or well-being) in the broad sense is multi-dimensional and amongst the aspects excluded from purely economic welfare as defined here are leisure, personal freedom and autonomy, a fulfilling emotional life, and economic security (which includes the value provided by the social safety net even to those who do not currently need to make any use of it). Good health and a long life expectancy are also clearly an important part of welfare in the broad sense. Though all these aspects of welfare may be linked to productivity in some way I do not pursue these issues here and focus solely on economic welfare.

The article has four main sections. The first section discusses the relationship between GDP and welfare. The second section focuses on a particular concept of welfare, namely "median equivalized households disposable income" (median EHDI). This concept allows us to respond to two criticisms of GDP as a welfare measure, namely that it is remote from the incomes that ordinary people receive and it ignores inequality: GDP per head is the arithmetic mean of GDP but this can be rising even if it is only the rich who are getting richer; indeed rising GDP per head is compatible with the poor getting poorer. The median EHDI concept has the advantage that it can be readily measured in practice, in the UK from 1977 onward. Household disposable income is part of the System of National Accounts (SNA) and the median part comes from household surveys. Later in this section, I present the decomposition which links our productivity measure, GDP per hour worked, with our welfare measure. Section 3 then presents and discusses the results of implementing the decomposition on UK data. Section 4 concludes.

GDP and Welfare

GDP is and always was intended to be a measure of output and income, not of welfare. In current prices, it measures the value of goods and services produced for final consumption, private and public, present and future; future consumption is covered since GDP includes output of investment goods. Converting to constant prices allows one to calculate growth of real GDP over time (or, using PPPs, differences between countries across space).[2]

Though not a *measure* of welfare, GDP can be considered a component of welfare. The volume of goods and services available to the average person clearly contributes to welfare in the wider sense, though of course it is far from being the only component. So one can imagine a social welfare function that has GDP as one of its components along with health, inequality, human rights, etc.

GDP is also an *indicator* of welfare. In practice, in cross-country data, GDP per capita is highly correlated with other factors that are important for human welfare. In particular, it is positively correlated with life expectancy, negatively correlated with infant mortality, and negatively correlated with inequality. In other words, richer countries tend to have longer life expectancy, lower infant mortality, and lower inequality, although this last relationship is not a linear one: some middle-income countries have high inequality, but nonetheless the richest countries, if we exclude a few petrostates, are also the most equal ones (Oulton, 2012a, Chart 3). Correlation is not necessarily causation, though one might certainly make the case that higher GDP per capita causes improved health (Fogel, 2004; Deaton, 2013).

I have argued that there is nothing wrong with the concept of GDP as long as it is correctly understood as a measure of output, though there is room for disagreement about where the production boundary should be set (Oulton, 2021). But equally there is no need to stick with GDP as just an indicator of welfare if we can do better and measure welfare directly. A first step is to use the rest of the SNA, augmented by data on distribution, the approach taken in this article. But some would go much further. According to the influential Commission on the Measurement of Economic Performance and Social Progress (the Stiglitz-Sen-Fitoussi Commission), policy should be concerned

2 The distinction between output and welfare can be seen very clearly when the effects of an exogenous, favourable change in the terms of trade are analysed in a small open economy. Under competitive assumptions this raises real consumption and welfare but leaves GDP unchanged. This is the conclusion of economic theory and also of the SNA when correctly applied (Oulton, 2021).

with well-being, and well-being is multi-dimensional (Stiglitz, Sen, and Fitoussi, 2009:15):[3]

> "To define what well-being means a multidimensional definition has to be used. Based on academic research and a number of concrete initiatives developed around the world, the Commission has identified the following key dimension that should be taken into account. At least in principle, these dimensions should be considered simultaneously:
>
> i. Material living standards (income, consumption and wealth);
> ii. Health;
> iii. Education;
> iv. Personal activities including work;
> v. Political voice and governance;
> vi. Social connections and relationships;
> vii. Environment (present and future conditions);
> viii. Insecurity, of an economic as well as a physical nature."

Few will disagree that these dimensions of life are important for human welfare and no one can object to improved measurement. There is clearly a role for government in measuring and tracking these dimensions. To what extent, however, a dimension like "social connections and relationships" should be objects of government policy is open to question. It is doubtful that effective policy levers exist. And, even if they did, the scope for a vast extension of the reach of government will not suit every taste.

A promising area that could lead to a wider concept of welfare is health. Life expectancy rose steadily throughout the 20th century if we ignore the world wars and the 1918 flu pandemic. Pre-Covid at least it was still rising on average in the 21st century. This means that people have more years in which to enjoy the higher consumption they now receive, an aspect of welfare which is not captured just by the GDP statistics. But recently the United States has seen a rise in mortality among less-educated, middle-aged whites due it seems to self-harming behaviour – drug and alcohol dependency, accidents and suicide (Case and Deaton, 2017), so-called "deaths from despair." Whether this is a specifically American phenomenon, related perhaps to deficiencies in the US social safety net (Edin and Shaefer, 2015), or whether the same phenomenon is appearing in other developed countries is not yet clear.[4]

If one sticks to measurement and is

3 The Stiglitz-Sen-Fitoussi Report led to the OECD's "Beyond GDP" programme, subsequently rebranded as "GDP and beyond".

4 Life expectancy has increased in the UK over the last 40 years, albeit at a slower pace in the last decade. This is of course compatible with considerable divergence in life expectancy across income groups and geographies. The years 2018 to 2020 saw a small decrease (7 weeks) in male life expectancy attributable to the Covid-19 pandemic; female life expectancy is so far unaffected. It is too early to say whether the reduction in male life expectancy will prove permanent or temporary (Office for National Statistics, 2021).

somewhat less ambitious than the Stiglitz-Sen-Fitoussi report, then further progress is possible. Jones and Klenow (2016) use an expected utility framework to combine measures of life expectancy, inequality and consumption to construct what they call a consumption-equivalent welfare measure for a large sample of countries. Their measure turns out to be highly correlated with GDP per capita.

Should the welfare measure be explicitly adjusted for inequality?

A more ambitious path than the one followed in this article is to construct a welfare measure which explicitly incorporates value judgements about inequality. One of the best-known of these measures is based on the Atkinson index of inequality (Atkinson, 1970):

$$Z = \left(\frac{1}{N} \sum_{i=1}^{N} y_i^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}, \quad \varepsilon \geq 0$$

where Z is social welfare, y_i is the income of the i-th person, N is the number of people, and ε is a parameter measuring "inequality aversion." If $\varepsilon = 0$ then society cares nothing for inequality, in which case the Atkinson measure reduces to GDP per head.[5]

In the standard treatment, of which the Atkinson index is an example, inequality is bad *per se*, though people may differ in the extent to which they are inequality averse. I would argue that our moral intuitions about inequality are too complex to be wholly captured by this formulation. In particular, the crucial issue of merit or desert is omitted. If the Atkinson approach were the whole story, then social welfare would be raised by abolishing two institutions (among others): the national lotteries run in many countries and the Nobel prizes. Both increase inequality unambiguously. Indeed, Nobel prizes must be the most unequally distributed of all forms of income: only a dozen or so individuals receive one each year out of a world population of some 8 billion. Nobel prizes could be justified on Rawlsian grounds: monetary incentives are needed to induce the effort required to make discoveries that benefit everyone, including the worst off. But suppose that it could be conclusively shown that the monetary rewards are not necessary, and that the prize winners (and their less-successful colleagues) would have expended the same effort in exchange for just the honour and glory alone? I suspect that most people would still be quite happy to see the winners receive a monetary reward, even if it was not economically required. This is because they are perceived to deserve it. With national lotteries, a different form of desert comes into play. In the UK version, some winners receive £20 million or more, and, in one sense, no one is worth this amount. But anyone can buy a lottery ticket and, as long as the lottery process is perceived as fair (not rigged), most people are quite happy with the outcome.

Merit or desert is a complex issue and it may be that people's views are not entirely consistent. Who receives the money and for what purpose may well make a difference. The large rewards paid to professional foot-

5 See Jorgenson (2018, section 4) for a recent discussion of welfare analysis in the Atkinson tradition.

ballers are seen by most people as justified (as long as they are playing well), but not the similar-sized rewards paid to bankers, especially after the global financial crisis.

In summary, it is not clear that the Atkinson index would meet with universal approval, even setting aside the issue of greater or lesser "taste" for inequality (the parameter ε).

Two arguments against using GDP and the SNA to measure welfare

Two arguments are often used to disparage GDP and related measures. The first is that raising GDP is irresponsible because of the environmental damage this would cause. A striking example of this argument is in a report from the UK's premier scientific association, the Royal Society (Royal Society, 2012). There it was claimed that in order to allow for a modest increase in the material standard of living of the world's poorest, consumption in richer countries must be reduced, according to my calculations by about 37 per cent in the UK case (Oulton, 2012b). However, this type of argument should not be taken as a criticism of the validity of GDP (and the related national accounts concept of consumption), concepts which the argument itself deploys. Rather, it is really about the feasibility of future growth of GDP, however desirable this would otherwise be.

A second argument for the irrelevance of GDP to realistic policy debates relates particularly to the United States. It is often claimed that in the United States there has been a virtual disconnect between productivity and living standards since the 1970s: productivity has grown massively but living standards have stagnated. This claim is then often extended to other rich countries including Britain, without much evidence. It is non-controversial that income inequality has been rising for decades in the United States, but does this mean that the typical household has received no benefit from growth? A comprehensive examination of these issues appears in Wolff, Zacharias, and Masterson (2012); and Jorgenson (2018).

The results of Wolff, Zacharias, and Masterson, as interpreted by Oulton (2012b), reveal quite a different picture. They define a number of income concepts that are superior to GDP as a measure of household welfare. Their preferred measure is what they call the Levy Institute Measure of Economic Well-Being (LIMEW). This is household income after tax and benefits, including non-cash benefits, and also with an allowance for household production. They put their preferred measure on an equivalent basis; "equivalent" means that corrections are made for changing household size and composition. For LIMEW they report both the median and the arithmetic mean of equivalent income. Their measure is similar to but broader than the one in the present article. They estimated the LIMEW over the period 1959–2007 and for various sub-periods. Since measuring economic welfare over time is the objective, they convert each measure to real terms using the CPI (actually the CPI-U).

Over their whole period median equivalent LIMEW rose at 1.01 per cent per year while GDP per capita rose at 2.18 per cent per year So the typical American certainly benefited from a growth in living standards over this period. But on the other hand there is a big gap between these two growth rates. Some, but only some, of this gap is due to rising inequality. Mean equivalent LIMEW rose at 1.31 per cent per year So if inequality had remained constant the standard of living of the typical American would have risen faster by 0.30 percentage points per year. Much of the remainder of the gap between their measure of welfare and GDP per capita is due to the choice of deflator. They mostly use the CPI-U but if they had used the deflator for personal consumption expenditure instead this would have knocked about another 0.45 per cent percentage points per year off the gap, while raising the growth of living standards by the same amount. Arguably, the deflator for personal consumption expenditure is a better measure in this context since it has a wider coverage of what people are actually consuming.

The conclusion is that the median US individual has gained significantly from economic growth since 1959. This remains the case even though the median individual would have gained more (to the extent of 0.30 per cent per year) if inequality had not widened.[6] Furthermore, and contrary to a common view, the largest gains were in the 1980s. These gains continued, albeit at a slower rate, in the 1990s and even into the 2000s (Oulton, 2012b, Table 2).[7]

I now turn to an examination of welfare and productivity in the UK case.

Measuring Economic Welfare in the UK

Median equivalized households disposable income

The starting point of the analysis for the UK is the Office for National Statistics' (ONS) concept of "median equivalized households disposable income" (Meidan EHDI). Disposable income here is defined as income from private sources (wages, pensions, dividends, interest), plus cash benefits minus taxes on income (principally income tax and employees' contributions to National Insurance) and council tax.[8] See ONS (2017) and ONS (2016) for a guide to data sources. There is a more ambitious concept of disposable income, "Net Household Adjusted Disposable Income," where capital consumption (depreciation) attributable to households is subtracted and "social transfers in kind," namely the provision of state services which accrue to households such as free education and health care (this last is the "adjusted" part), are added (ONS, 2014). The problem with the more ambitious concept is that capital consumption and social transfers in kind, though available in the national accounts, are not available at the level of individual households, and so can-

6 Jorgenson (2018, section 4) reaches a similar conclusion using two different versions of an Atkinson-type social welfare function, egalitarian and utilitarian. With either one, he finds that since 1973, United States efficiency increases have more than offset the rise in inequality, and have consistently raised standards of living. Thus, Jorgenson's conclusion for the United States is, with a very different methodology, in line with the results of Wolff, Zacharias, and Masterson (2012) just discussed.

7 A parallel literature examines the relationship between wages and productivity and asks whether the two have become decoupled. See Mishel and Bivens (2021) for the US case. For the UK Teichgräber and Van Reenen (2021) answer this question in the negative.

8 Council tax is the name of the residential property tax levied on households in the UK.

not be used to analyse inequality.[9]

The equivalence scale which the ONS uses is "the modified OECD scale" (Anyaegbu, 2010). Under this a couple has a weight of 1, a single adult a weight of 0.67. A second or subsequent adult has a weight of 0.33, as do dependents aged over 12. Children of 12 or under have a weight of 0.2. This is the way that the equivalization process is described in the official documentation though it can lead to misunderstanding. In fact the incomes of all persons in a household are added together, then this total is assigned to each member of the household. Finally, each income is *divided* (not multiplied) by the total of the weights assigned to each person in the household.[10] The mean and median equivalized income (and other quantiles) for the whole sample are then calculated across individuals, not households. The median equivalized income is therefore that of the typical individual, not that of the typical household.[11]

Households disposable income (HDI) as a proportion of GDP at market prices averaged 64 per cent over our period. So, 36 per cent of GDP is not assigned to households, suggesting that it has no impact on household welfare. This is clearly far too extreme a conclusion. To the contrary, it might be argued that all this 36 per cent accrues to households one way or another. The strongest case here is with the undistributed profits of companies which on average raise share prices and therefore accrue to the owners, ultimately households, though often ownership is mediated via pension funds or insurance companies. But the bulk of the 36 per cent is "social transfers in kind," mainly free primary and secondary education and free health care; social transfers in kind constituted 23 per cent of gross HDI in 2019.[12]

Clearly, if one is sick, it is better to receive free health care than not. But it is better not be sick in the first place. So rising government expenditure on health care should not be translated automatically into higher welfare. Even so, some allowance for health and educational expenditures would be desirable. But this would require us to allocate these expenditures across house-

9 Note that the data used here, whether from the national accounts or from surveys, are strictly for households and do not include income accruing to Non-profit Institutions Serving Households (NPISH).

10 An example may help. Suppose a household has just one adult member whose income is £9,000. The weight for a single adult is 2/3, so this person is assigned an equivalized income of £13,500 (= £9,000 ÷ 2/3). Now consider a second household with two adults; the first adult has an income of £30,000 and the second one of £9,000. The total household income is £39,000. The sum of the weights is 2/3 + 1/3 = 1. So each of the two adults is assigned an income of £39,000 which is 4.33 times the income of the person in the first household. Without equivalization and assuming income sharing in the second household, each person in the latter would receive £19,500 or 2.17 times the single person in the first household. So with equivalization each person in the second household is calculated to be much better off than the single person, relative to the position without equivalization. I am grateful to the ONS for helping me to understand this issue better.

11 Though it is widely accepted that one should make some adjustment for household size and composition, the equivalence scale used by the ONS and by the author is rather crude. It would be preferable to have the scale vary with income and prices (Deaton and Muellbauer, 1980, Chapter 8). I am grateful to Rachel Soloveichik for helpful discussion on this.

12 See Table 6.2.5 of the 2020 Blue Book, downloadable as bb20chapter06hnsectorfinal.xlsx from https://www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2020. As mentioned earlier depreciation (capital consumption) is included in Median EHDI though ideally it should be excluded. However it only constituted 3-5 per cent of HDI in our period; see Table 6.2.1 of the 2020 Blue Book at the same URL.

holds, which would be far beyond the scope of the present article. Hence a pragmatic case can be made for focusing on income which accrues directly to households and which they are free to spend as they wish.

For comparisons over time we need to deflate nominal Median EHDI by a price index. The ONS uses a consumer price index or in practice the CPIH (i.e. the CPI with both owner-occupied and rented accommodation included) but with council tax excluded. This is in line with theoretical work suggesting that for a welfare measure the whole of income, the part saved as well as the part consumed, should be deflated by the price index of consumption since the purpose of saving is to change the time pattern of consumption (Weitzman, 1976, and Sefton and Weale, 2006).[13]

A more basic criticism of Median EHDI as a welfare measure is that it makes no allowance for the utility of leisure: £10 (net of tax) earned through an extra hour of labour is counted as £10 of additional welfare, even though the additional consumption (current and/or deferred) is bought at the price of one hour less of leisure. This is the intuition behind the suggestion of Basu and Fernald (2002), and Basu *et al.* (2012) that the growth of TFP is an appropriate measure of welfare change. That is to say, even if Median EHDI is the appropriate measure of income, we should subtract from its growth the growth of labour input weighted by labour's share. This amounts to valuing an hour of leisure at the hourly wage. I have not chosen to go down this route due to doubts that the labour market is in equilibrium. One piece of evidence supporting this doubt is that a substantial fraction of part-time workers say that they would like to work longer hours than they do. Also, it turns out that in the UK case the leisure correction would have only a small impact (see below).

The decomposition

To measure welfare we are focusing on Median Equivalized Households Disposable Income ($EHDI^{median}$) as defined above. Denote the corresponding arithmetic mean by $EHDI^{mean}$. Let us take our aggregate productivity measure to be GDP per head (below we extend the decomposition to GDP per *hour*).

The transformation between productivity on the right-hand side and welfare on the left-hand side can be written as the product of a set of factors as follows:

$$EHDI^{median} \equiv \left(\frac{EHDI^{median}}{EHDI^{mean}}\right) * \left(\frac{EHDI^{mean}}{HDI/N}\right) * \left(\frac{HDI}{GDP}\right)\left(\frac{GDP}{N}\right) \quad (1)$$

This relates (nominal) median EHDI through a series of factors to (nominal) GDP per capita (GDP/N) where N is the population. This is just an identity but the factors can be given an economic interpretation and also can be tracked over time. The factors are:

13 Oulton (2004) quantifies the Weitzman measure for the United States and develops a growth-accounting-style decomposition of it into the contributions of labour, capital, and an analogue of TFP.

1. $\frac{EHDI^{median}}{EHDI^{mean}}$: a measure of inequality. If this rises, inequality is falling. If the distribution of $EHDI$ is (approximately) lognormal, then this ratio equals $\exp(-\sigma^2/2)$ where σ^2 is the variance of log income. In the lognormal case all measures of inequality such as the Gini are monotonically related to the parameter σ^2.

2. $\frac{EHDI^{mean}}{HDI/N}$: a measure of household composition or of the effect of equivalization. The numerator is the arithmetic mean of equivalized HDI while the denominator is the mean of HDI without equivalization, where N is population. If household size is rising due to more single persons partnering up, with other things the same, then the mean of the equivalized HDI rises in relation to the mean of the unequivalized distribution. This is because the equivalence scale embodies the idea that two can live more cheaply than one.

3. $\frac{HDI}{GDP}$: the share of total income accruing to households. This is in part a measure of the size of the welfare state. The ratio rises if the government spends proportionately more on transfers or reduces tax. It also rises if net foreign income accruing to households rises as a proportion of GDP. It falls if corporations distribute less of their profits back to households in the form of dividends.

So far the decomposition is similar to the one employed by Nolan, Rosser, and Thewissen (2018). The main difference is that they use Gross National Income (GNI) rather than HDI for most of their analysis. GNI includes income accruing to sectors other than households (NPISH, corporations and the government). But we are not quite ready yet to analyse Krugman's contention since the right-hand side of equation (1) features GDP per capita not GDP per hour. These two concepts can be linked through a second identity:

$$
\begin{aligned}
GDP/N &\equiv \left(\frac{H}{N}\right)\left(\frac{GDP}{H}\right) \\
&\equiv \left(\frac{H}{(1-u)L}\right)(1-u) \\
&\quad * \left(\frac{L}{N^{wa}}\right)\left(\frac{N^{wa}}{N}\right) \\
&\quad * \left(\frac{GDP}{H}\right)
\end{aligned}
\tag{2}
$$

where H is aggregate hours worked, L is the number of people in the labour force (employed plus unemployed), u is the unemployment rate, and N^{wa} is the number of people of working age (defined here as those aged 16 and over). The factors on the right-hand side can be given the following interpretation:

4. $\frac{H}{(1-u)L}$: hours per worker, or labour intensity.

5. $1-u$: 1 minus the unemployment rate (u).

6. $\frac{L}{N^{wa}}$: the labour force participation rate.

7. $\frac{N^{wa}}{N}$: the proportion of the population which is of working age, defined here as those aged 16+ as a proportion of the total population. Traditionally the working age population

has been defined as 16-59 for women and 16-64 for men but this seems unrealistic given that pension age is being aligned for both sexes, compulsory retirement has been abolished, and more people are working into their 70s.

These four additional factors all have a natural economic or demographic interpretation.

Putting equations (1) and (2) together the full decomposition is:

$$
\begin{aligned}
EHDI^{median} \equiv & \left(\frac{EHDI^{median}}{EHDI^{mean}}\right) \\
& * \left(\frac{EHDI^{mean}}{HDI/N}\right) \\
& * \left(\frac{HDI}{GDP}\right) \\
& * \left(\frac{H}{(1-u)L}\right) \\
& * \ (1-u)\left(\frac{L}{N^{wa}}\right) \\
& * \left(\frac{N^{wa}}{N}\right)\left(\frac{GDP}{H}\right).
\end{aligned} \tag{3}
$$

This decomposition is for median household income but could be adapted for any other quantile, such as equivalized HDI at the lowest quintile, the poorest fifth, or $EDHI^{quin1}$. With $EDHI^{quin1}$ on the left-hand side the first ratio on the right-hand side must then be changed to

$$\frac{EHDI^{quin1}}{EHDI^{mean}}$$

which can also be interpreted as a measure of inequality.

This decomposition applies at a point in time or in other words income and output are in current prices. But the main interest is in tracking changes over time, i.e. we want to relate real HDI to real GDP. Real GDP is related to nominal GDP by the implicit GDP deflator, P^{GDP}, while real HDI is related to nominal HDI by an index of consumer prices, P^{CE}. The ONS employs a version of the CPIH which includes owner-occupied and rented housing but excludes council tax (since the latter is subtracted from HDI). The decomposition now becomes one between real median EHDI and real productivity:

$$
\begin{aligned}
\frac{EHDI^{median}}{P^{CE}} \equiv & \left(\frac{EHDI^{median}}{EHDI^{mean}}\right) \\
& * \left(\frac{EHDI^{mean}}{HDI/N}\right) \\
& * \left(\frac{HDI}{GDP}\right) \\
& * \left(\frac{H}{(1-u)L}\right) \\
& * \ (1-u)\left(\frac{L}{N^{wa}}\right) \\
& * \left(\frac{N^{wa}}{N}\right)\left(\frac{P^{GDP}}{P^{CE}}\right) \\
& * \left(\frac{GDP}{P^{GDP}H}\right).
\end{aligned} \tag{4}
$$

Now we have introduced an eighth factor:

8. $\frac{P^{GDP}}{P^{CE}}$: the price of GDP as a whole relative to the price of consumption. This can be thought of as reflecting technological trends, i.e. relative rates of productivity growth in different industries. For example, the price

of investment goods may be falling relative to that of consumption goods but the opposite may be the case for government services such as health and education (at least as conventionally measured). It may also reflect changes in the terms of trade: the price of consumption is influenced by the price of imports while the GDP deflator is influenced by the price of exports.

The (logarithmic) growth rate of living standards can now be thought of as the sum of the growth rates of the eight factors on the right-hand side of equation (4) plus the growth of productivity. So equation (4) yields an additive decomposition relating the growth of living standards to the growth of productivity. An increase in any of the eight factors raises living standards in relation to productivity.

With the possible exception of the relative price factor, all the other factors in equation (4)have natural limits, whether logical or economic (e.g. the unemployment rate must lie between zero and one). Hence productivity is the only long run driver of living standards though the same may not be true in the short run.

There are a number of ways in which the decomposition could be expanded if that were thought likely to yield further insights. For example, the income share of households, HDI/GDP, could be broken down further to show the separate contributions of taxes and benefits to changes in this ratio. And productivity growth itself could be broken down into the contributions of TFP and capital deepening.

Of course the proposed decomposition is not unique. A silly alternative to equation (2) is the following:

$$\begin{aligned}\frac{GDP}{N} &\equiv \left(\frac{H}{N}\right)\left(\frac{GDP}{H}\right)\\ &\equiv \left(\frac{H}{G}\right)\left(\frac{G}{N}\right)\left(\frac{GDP}{H}\right)\end{aligned}$$

where G is goals scored in the English Premiership. So $\frac{H}{G}$ is the number of (whole economy) hours required to score a goal in the Premiership and $\frac{G}{N}$ is the number of goals per head of population. These two factors clearly yield no insights into productivity or welfare.

A second objection is that the decomposition of equation (4) is by definition just an identity. So a theory would clearly be preferable. But a theory covering all the factors in (4) would have to be very broad. Here is a sketch of one possibility to illustrate the difficulties. Suppose that technical progress has been biased towards skills which require more education. The wages of the less-skilled, particularly males, have therefore declined. These individuals find themselves at a disadvantage in the marriage (or partnership) market. Since partnership is assortative by education levels, less educated women are less likely to find satisfactory partners and so have fewer children. So fertility declines, accelerating the ageing process (which is due to greater longevity in the first instance). The less educated turn off from conventional politics because it seems less and less attuned to their interests, more to the interests of the better educated. This in turn sparks a reaction in the form of populism, which leads to policy changes affecting the size of the state, etc. All of this is of course highly controversial. And it would take a great deal of work to check whether the evolution of the factors is consistent with the theory just sketched. Even if it is, one would also have to see whether some alternative theory could explain the same facts at least as well. In the meantime the decomposition can serve as a guide to further research.

Results for the UK

The UK data

Equivalized and non-equivalized household disposable income, mean and median, and by quintile, come from a ONS spreadsheet entitled hdiireferencetables201920update.xlsx downloaded on 1st June 2021. These data underlie the regular *Statistical Bulletin on Household Income Inequality.*[14] The data are for calendar years up to 1993, thereafter for fiscal years (April to March); I have ignored this break. The source is the Living Costs Survey supplemented from Fiscal Year Ending (FYE) 2017 by the Household Finances Survey. Currently 17,000 private households are surveyed. Estimates of income from 2001/2002 onwards have been adjusted by the ONS for the under coverage of top earners, using data from Her Majesty's Revenue and Customs (HMRC). Both mean and median EHDI are deflated in the source by a special version of the CPIH (the CPI including both owner-occupied and rented accommodation) which excludes council tax; the council tax element of the CPIH is excluded since HDI excludes council tax payments. The data on EHDI go back only to 1977 so this is the starting point of the analysis. These data currently stop in 2019/2020 so this marks the endpoint (which is conveniently also the start of the Covid-19 pandemic).

GDP (in current prices and in chained volume form), HDI, population and the labour market (hours worked, employment, and unemployment) can be obtained from the UK's national accounts, all downloadable from the ONS website. The GDP deflator is calculated as the ratio of the current price measure of GDP to the chained volume measure, both at market prices. A full description is in the Appendix.

Results

Table 1 shows the average growth rates of the standard of living and of productivity over this period. It also shows growth over the sub-periods 1977-1990, 1990-2007, and 2007-2019. These sub-periods are so defined since 1990, 2007 and 2019 are all cyclical peaks. Over the whole 43-year period the standard of living actually grew somewhat faster than productivity (1.88 per cent per year compared to 1.73 per cent per year) but this was not true in the central sub-period (1990-2007). Across the three sub-periods the average growth

14 The latest version is at https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householdincomeinequalityfinancial/financialyearending2020.

15 The years 1977-1990, which correspond closely to the premiership of Margaret Thatcher (1979-1990), saw faster growth of living standards than in either of the two subsequent sub-periods. Although the poorest quintile did worse than the richest one in this period, it still did better than the same quintile did in the latest sub-period, 2007-2019.

Table 1: Trends in the Standard of Living and of Productivity in the United Kingdom, 1977-2019 (average annual per cent rate of change)

	1977-1990	1990-2007	2007-2019	1977-2019
Standard of living (median)	3.07	1.96	0.47	1.88
Productivity	2.35	2.34	0.21	1.73
Standard of living by quintile				
Lowest (poorest) quintile	1.21	2.41	0.14	1.39
Highest (richest) quintile	4.43	2.19	0.03	2.26
Memo item				
Real mean EHDI	3.69	2.44	0.47	2.12

Note: 1. EHDI: Equivalised Household Disposable Income, deflated by the CPIH excluding council tax;
2. Labour force: persons employed (employees plus self-employed) plus unemployed.
Source: Office for National Statistics; see text for details.

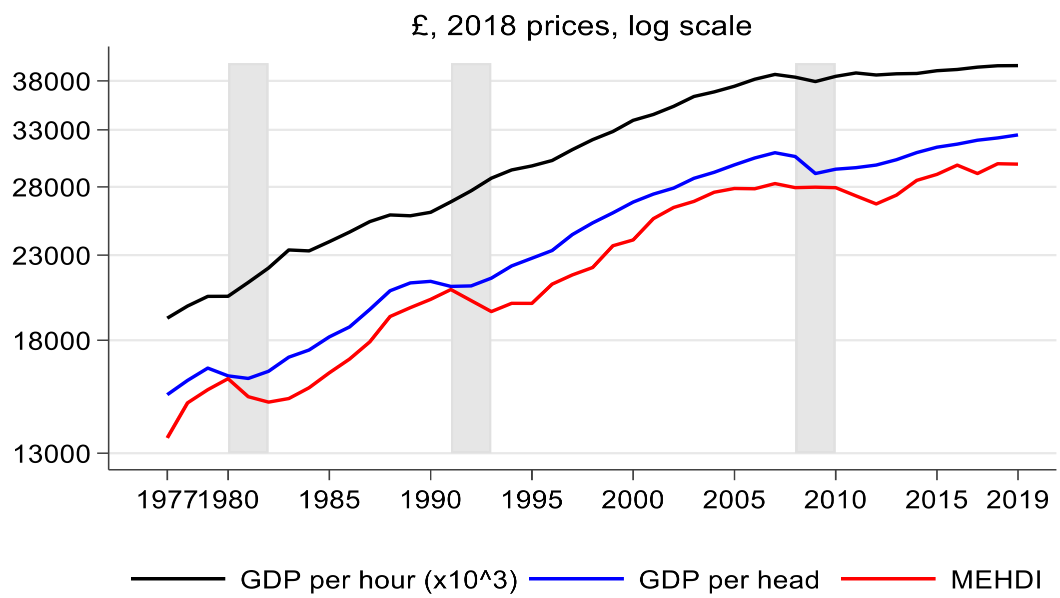
rate of living standards (the overall median EHDI) has been falling.[15] But productivity grew at almost the same rate 1990-2007 as it did over 1977-1990. The years since the global financial crisis have seen a collapse in the growth of both measures.

To check whether there is anything unusual about the median, Table 1 also shows the growth of living standards in the lowest quintile (the poorest fifth) and in the highest (the richest fifth) of the distribution of EHDI. In the first sub-period, 1977-1990, the highest quintile did much better than the lowest one. In the subsequent two sub-periods the lowest quintile did a bit better than the highest. Nonetheless over the whole 43 years the highest quintile did best.

Chart 1 shows the growth of our two main measures and also of an intermediate measure, GDP per head, over the whole period. (In this and subsequent charts grey bars mark recessions, defined as years in which on a quarterly basis GDP was mostly falling: 1980-1981, 1991-1992, and 2008-2009.) Broadly speaking all three measures move in line with each other. When productivity growth is high so too is the growth of living standards. And when productivity growth crashes in the most recent period, so too does the growth of living standards. However, despite the strength of the long-term relationship, at an annual frequency they are not closely related at all. The correlation between the annual growth rates of living standards and of productivity in the whole 43 year period is only 0.045 (which is not significant at conventional levels). Regressing the growth of living standards on its own lag and contemporaneous and lagged productivity growth does not improve things: neither contemporaneous nor lagged productivity growth is significant and the fit is poor.

Table 2 shows the decomposition for the median. The biggest single factor, and the only one apart from the demographic effect (the growth of the proportion of the population aged 16 and over) which is consistently positive, i.e. favourable to living standards, is productivity. Over the whole period it accounts for 92 per cent of the growth of living standards. But since 2007 its contribution has been much smaller, only 45 per cent. This figure is rather misleading though. It is 45 per cent of a much smaller number than in the other sub-periods. The exceptional nature of the years since 2007 is also apparent from this

Chart 1: Productivity and Welfare in the United Kingdom, 1977-2019



Notes: 1. Shaded bars mark recessions;
2. MEHDI: Real Median Equivalized HDI.
Source: Office for National Statistics; see text for details

table. Many of the other factors change sign and become larger in absolute value.

It is also interesting to compare the first sub-period (1977-1990) with the second (1990-2007) since living standards rose most rapidly in the first while productivity growth was virtually the same. Looking at the first two columns of Table 2, we can see that the largest changes in the factors between these two sub-periods were firstly in the relative price effect and secondly in the equivalization effect. Rising inequality had a negative effect up to 2007; the size of this effect varied little between the two sub-periods.

Up to 2007, the four labour market variables (labour intensity, unemployment rate, labour force participation rate and the working age proportion) are not collectively very important. They accounted for +5.4 per cent of growth in living standards in 1977-1990 and −7.8 per cent in 1990-2007. But after 2007 with the collapse of productivity growth the picture changes. Collectively the labour market variables now account for nearly half (+46.3 per cent) of the meagre growth in living standards that actually occurred.

Each of the eight factors in Table 2 (apart from productivity) will now be discussed in turn with the help of charts 2-8.

1. Inequality: median EHDI relative to mean EHDI (Chart 2)

Median EHDI fell relative to mean EHDI from 1977 to 2007, in other words, inequality was rising, but since then the opposite has occurred, i.e. inequality has fallen. So rising inequality reduced living standards in relation to productivity from 1977 to 2007 but the opposite occurred during and after the Great Recession. If

Table 2: Contributions to growth in the standard of living in the United Kingdom, 1977-2019 (percentage points of total change)

	Factor	Measure	1977-1990	1990-2007	2007-2019	1977-2019
	Standard of living *Contributions (%)*	Growth rate of Median EHDI (% p.a.)	3.07	1.96	0.47	1.88
1.	Inequality	Median EHDI/Mean EHDI	-20.2	-24.4	109.1	-12.8
2.	Equivalisation	Mean EHDI/Mean HDI	-8.1	22.7	-115.7	-2.7
3.	Share of households in total income	HDI/GDP (both in current prices)	5.5	-0.3	69.3	7.6
(4.-7.)	Labour market	Sum of factors 4-7	5.4	-7.8	46.3	2.7
4.	*Labour intensity*	*Hours per person employed*	*-10.0*	*-14.1*	*2.4*	*-10.8*
5.	*1 minus the unemployment rate*	*1 minus the unemployment rate* $(1-u)$	*-4.0*	*5.8*	*28.1*	*2.4*
6.	*Labour force participation rate*	*Labour force/population aged 16+*	*6.2*	*-4.7*	*14.4*	*2.1*
7.	*Demographic effect*	*Proportion of population aged 16+*	*13.2*	*5.2*	*1.4*	*9.0*
8.	Relative price	GDP deflator/CPIH	41.0	-9.0	-54.3	13.0
9.	Productivity	GDP per hour worked	76.4	119.1	45.3	92.2
		TOTAL	**100.0**	**100.0**	**100.0**	**100.0**

Note: 1. EHDI: Equivalised Household Disposable Income, deflated by the CPIH excluding council tax;
2. Labour force: persons employed (employees plus self-employed) plus unemployed.
Source: Office for National Statistics; see text for details.

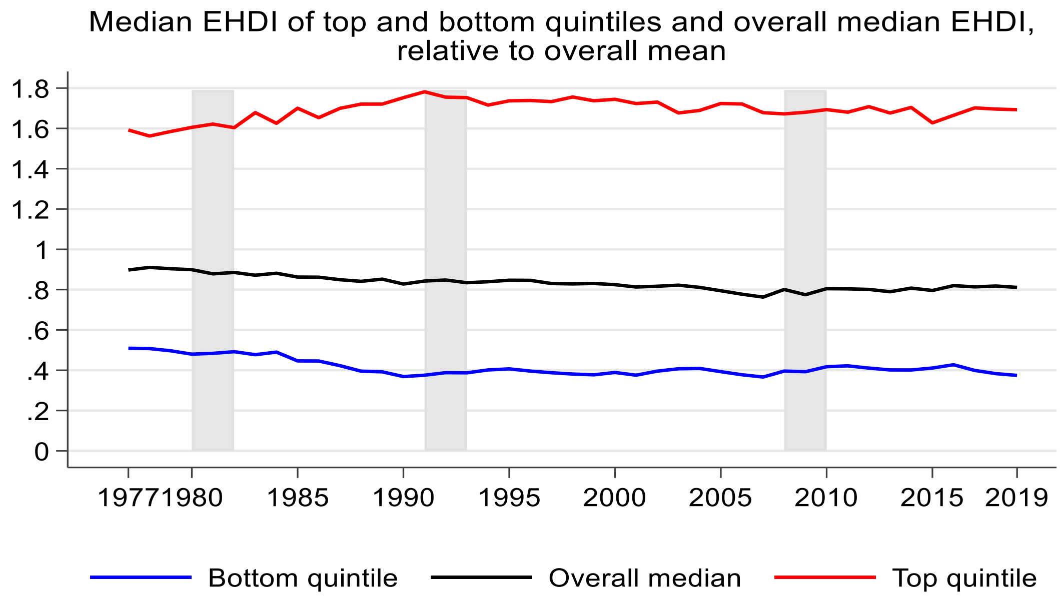
we were focusing on the experience of the poorest quintile, Chart 2 shows that inequality stopped rising earlier, in the early 1990s, and thereafter has remained fairly constant.

2. Equivalization: equivalized mean HDI relative to actual mean HDI (Chart 3)

This factor compares two arithmetic means: the mean of equivalized HDI and the mean of actual (non-equivalized) HDI. In principle any changes in the ratio of the two means should be due to variations in household composition. For example, if household size is rising then more people are partnering up. So with the same incomes individuals are getting better off since two can live more cheaply than one: this is what equivalization is designed to measure. Taken literally, the chart suggests that household size hit a low point around 1995 and thereafter rose till the onset of the Great Recession.

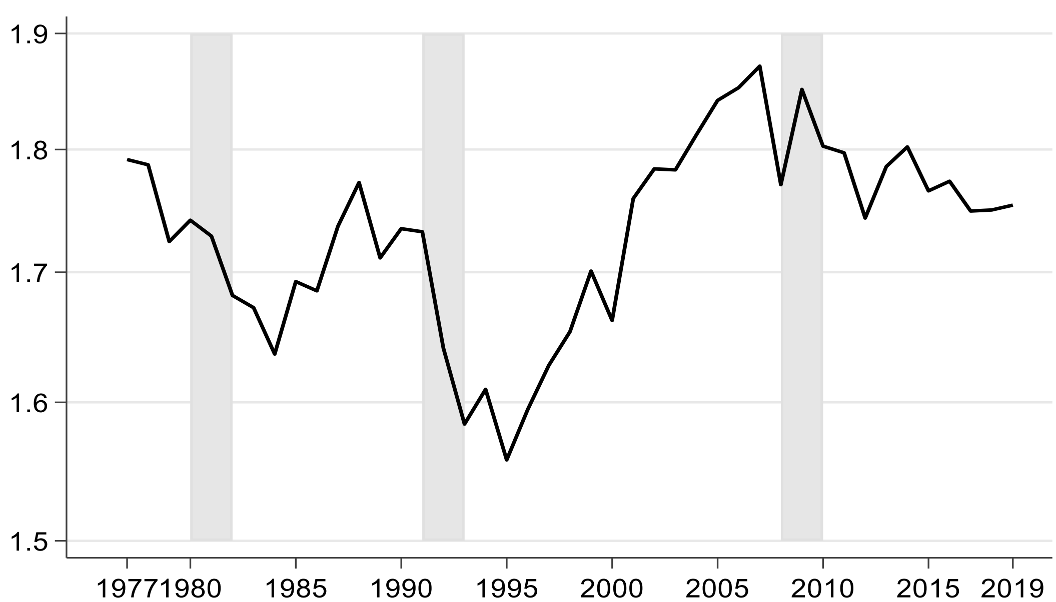
But there may be another factor at work. "Equivalized mean HDI" is the arithmetic mean of equivalized incomes across individuals and comes from a sample survey (currently, the Living Costs and Food Survey supplemented by the Survey on Living Conditions). "Actual mean HDI" is HDI from the national accounts divided by total population. Since the two series come from different sources they may not be fully consistent, despite the fact that all these series

Chart 2: Inequality in UK EHDI, 1977-2019



Note: 1. Shaded bars mark recessions;
2. See text for definitions of variables.
Source: Office for National Statistics; see text for details

Chart 3: Ratio of Equivalized Mean HDI to Actual Mean HDI



Note: 1. Shaded bars mark recessions;
Source: Office for National Statistics; see text for details.

have the status of "National statistics," i.e. meet various quality standards as defined by the regulator, the Office for Statistics Regulation.[16]

There is another reason for doubting whether this factor is actually measuring the effect of equivalization. Mean equivalized HDI from the surveys is always substantially larger than actual mean HDI from the national accounts (the ratio of the two means over the whole period is 1.73); this by itself should not affect the analysis of growth rates. But mean HDI from the surveys, *without* equivalization, is also substantially larger than actual mean HDI from the national accounts and follows a very similar path to the *equivalized* mean from the surveys.[17] In fact, the average size of households did not change between 1990 and 2019: in both years it was 2.6 persons for the non-retired and 1.5 for the retired. So it appears that this factor is not in practice capturing the effect of equivalization. The issue requires further investigation.

3. Household share of total income: HDI/GDP (Chart 4)

This ratio rose sharply by some 7 percentage points between the late 1980s and early 1990s, before falling again until the Great Recession began; thereafter it has been rising again. Both HDI and GDP are in current prices and come from the national accounts. These large swings must therefore mainly reflect changes in taxes and cash benefits. After falling inequality, the rise in the household share was the largest single factor supporting living standards following the Great Recession. This presumably reflects the welfare state doing its job. The remaining factors cover different aspects of the labour market.

4. Labour intensity: weekly hours per worker (Chart 5)

Labour intensity has fallen steadily over this period though a bit faster during the three recessions. British workers now work three hours per week less than they did in 1977. This no doubt reflects in part the growth of part-time working. Fewer hours per week reduces living standards in relation to productivity, though recall that there is no attempt here to put a value on additional leisure (assuming it to be voluntary).

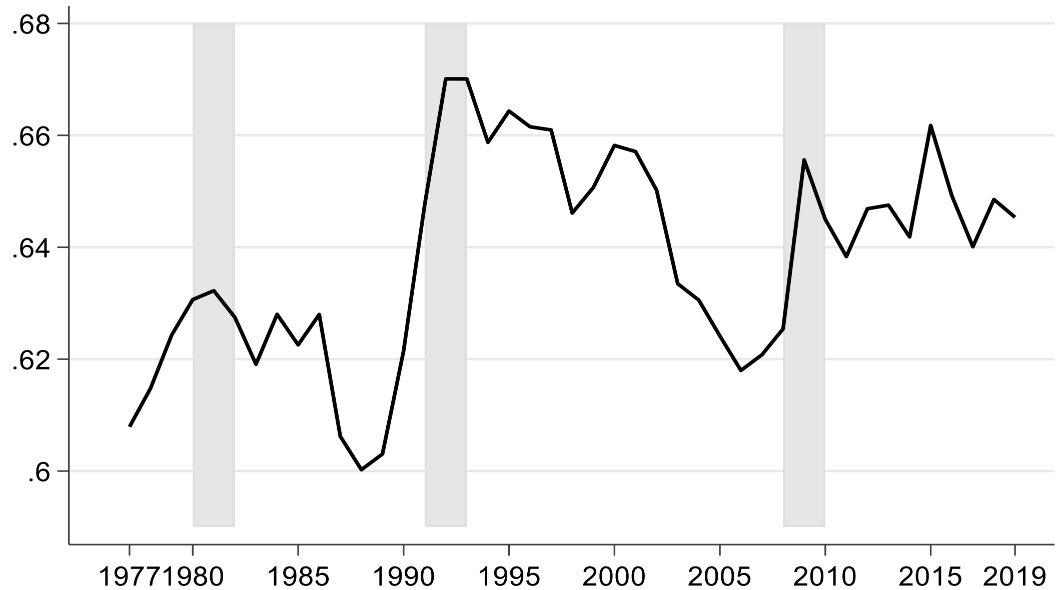
5. Unemployment rate (Chart 6)

Apart from hours worked this is the only variable which is markedly influenced by recessions. The rate rose sharply in all three recessions; during and after the 1980-81 recession it nearly doubled. But apart from these spikes it has been on a downward trend, and by 2019, it was lower than it had been in 1977. But given the amount of commentary and political attention devoted to unemployment it is surprising at

16 The population covered by the national accounts is wider than that of the surveys which exclude the institutional population (residents of care homes, students in student accommodation (halls of residence), prisoners, NHS workers in NHS accommodation, members of the Armed Forces living in barracks, and people living in hotels/BBs/homeless shelters as well as the homeless and travelling communities). But it is hard to believe that changes in the size of the institutional population could account for such wide swings as seen in Chart 3.

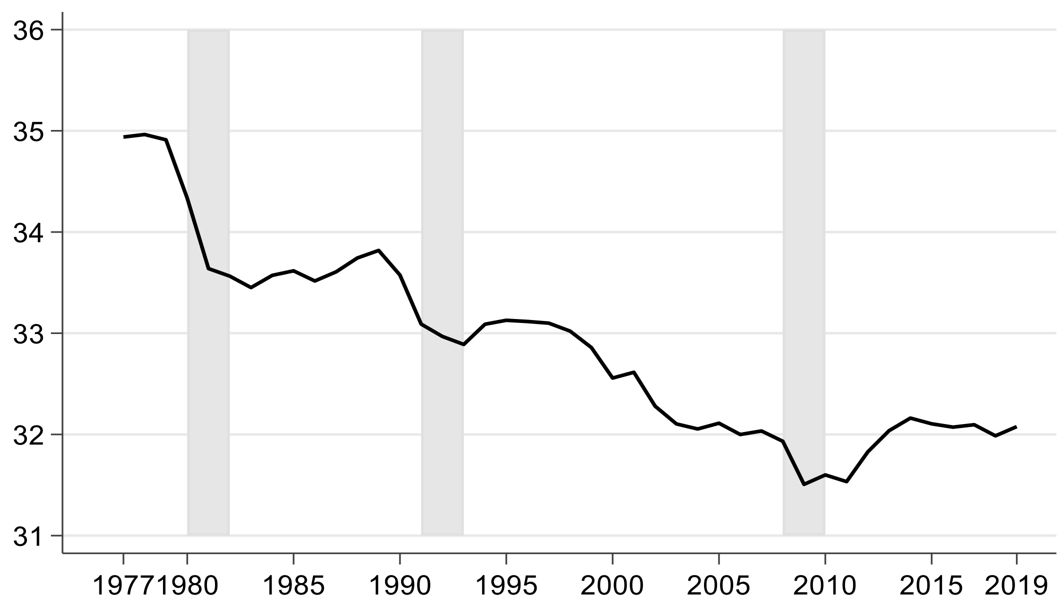
17 I am grateful to the ONS for providing me with this series. See https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/adhocs/13864timeseriesofnonequivalizedhouseholddisposableincomeandhouseholdcharacteristicsuk.

Chart 4: Household Share of Total Income: HDI/GDP



Note: 1. Shaded bars mark recessions;
2. HDI (household disposable income) and GDP in current prices.
Source: Office for National Statistics; see text for details.

Chart 5: Labour Intensity: Weekly Hours per Worker



Note: 1. Shaded bars mark recessions;
2. Total weekly hours divided by total in employment (inc. self-employed).
Source: Office for National Statistics; see text for details.

first sight how small an effect it has had on economic welfare as measured here. The explanation is presumably that the welfare measure does not allow for the insecurity and loss of self-esteem that many people undoubtedly feel on becoming unemployed.

6. Labour force participation (Chart 7)

Labour force participation (measured as the employed plus the unemployed as a proportion of the population aged 16+) peaked in 1988, then fell till 1995. After that it rose steadily till the end of our period. But in 2019 it was still lower than it had been in 1988. So changes in participation were broadly favourable to living standards from 1995 onward, even though participation in 2019 was lower than it had been in 1988. This may seem a surprisingly downbeat assessment of the role of labour force participation given the amount of attention given to the so-called "jobs miracle" in the UK: from 1995 to 2019 employment rose by almost 7 million or 27 per cent. But most of these new jobs went to foreign-born workers (Oulton, 2019), whose numbers also swelled the population. So there was comparatively little effect on economic welfare as measured here which is on a per capita basis.

7. Adult population (aged 16+) as proportion of total (Chart 7)

The proportion of the population aged 16+ rose up until the Great Recession. This was favourable to living standards but the effect was reversed after that.

8. Relative price: the price of output relative to that of consumption (Chart 8)

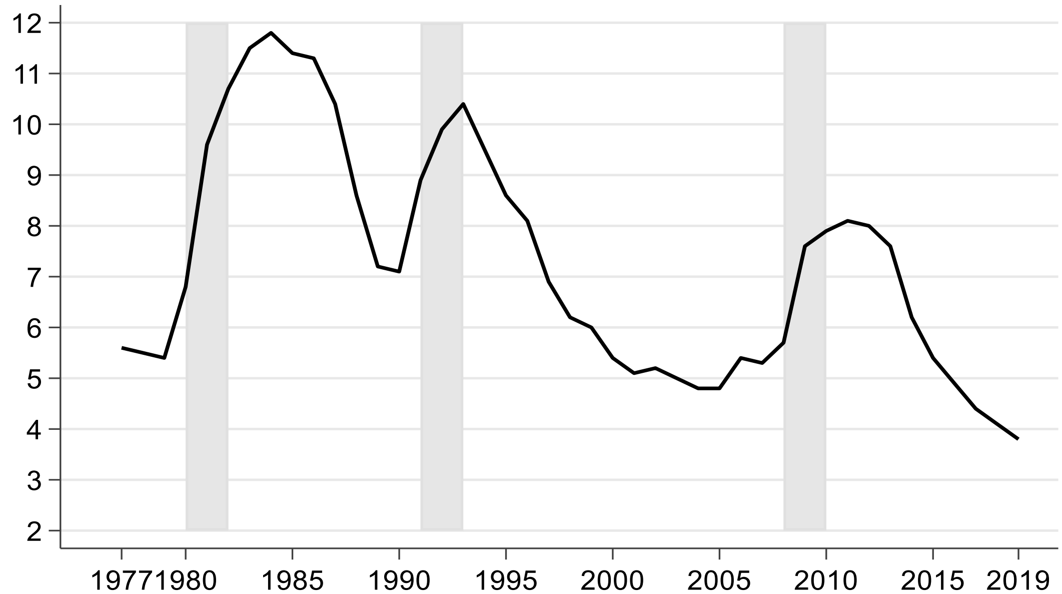
The relative price of output (GDP), the GDP deflator relative to the CPIH, increased sharply from 1977 to 1989, falling slightly until 2000 and thereafter showing no clear trend. The sharp rise in the earlier years may be due to the strengthening of the real exchange rate. This occurred partly because of a rise in the price of petroleum products leading to a boom in exports of oil and gas from the North Sea (which was just then coming on stream) and partly because of high interest rates resulting from tight monetary policy. Either way, the price of exportables rose relative to that of importables. The CPI is influenced by the latter but not the former so it fell relative to the GDP deflator.

Finally, what effect would allowing for leisure (as suggested by Basu and Fernald (2002)) have on these results? If we consider the growth of annual hours worked per person aged 16+, then this has drifted down over most of the years since 1977, i.e. leisure has increased, though it rose a bit after 2007. From 1977 to 2007, hours worked per person aged 16+ were on average falling at 0.25 per cent per year so multiplying by labour's share would raise the growth rate of the standard of living by only about 0.15 percentage points per year.

Conclusion

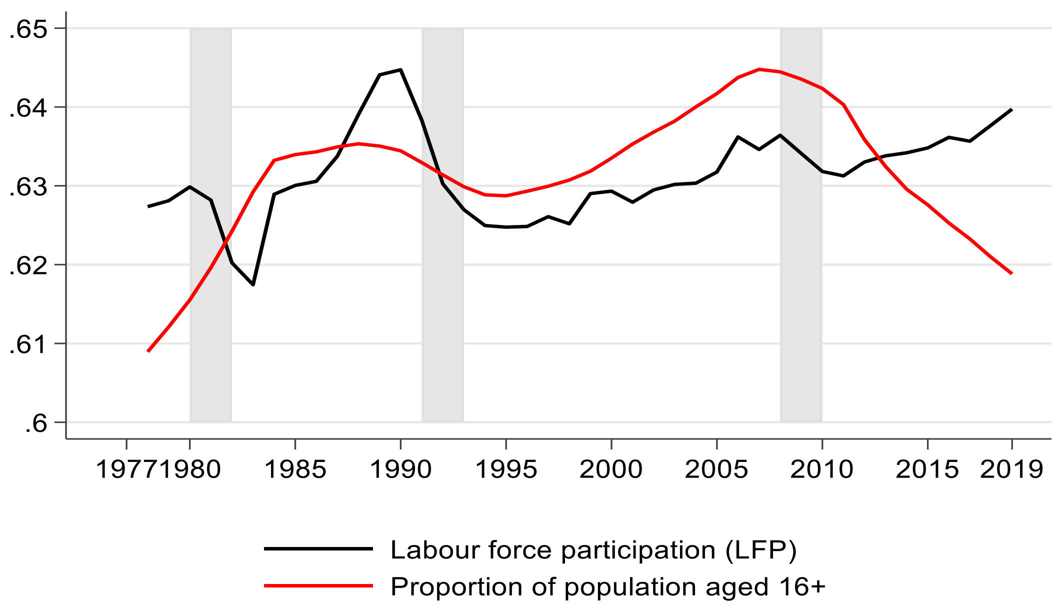
A decomposition has been developed to relate the growth of living standards (or economic welfare), measured by median household disposable income per equivalized adult, to the growth of productivity, measured by GDP per hour worked. The decomposition involves eight factors, each of which can be given an economic or demographic interpretation. The decomposition

Chart 6: Unemployment Rate, per cent



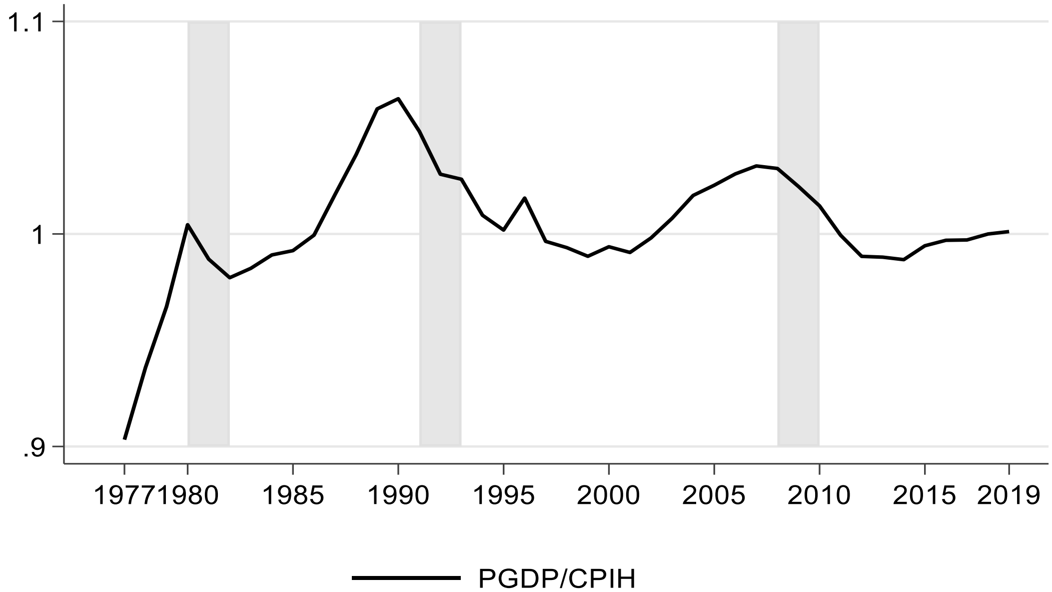
Note: 1. Shaded bars mark recessions;
2. Unemployment rate: Unemployed/(employed plus unemployed)
Source: Office for National Statistics; see text for details.

Chart 7: Labour Force Participation and Proportion Aged 16+



Note: 1. Shaded bars mark recessions;
2. LFP: Employed plus unemployed as proportion of all aged 16+)
Source: Office for National Statistics; see text for details.

Chart 8: The GDP Deflator Relative to the CPIH



Note: 1. Shaded bars mark recessions;
2. CPIH: CPI including housing but excluding council tax. 2018 = 1.0.
Source: Office for National Statistics; see text for details.

has been quantified over the period 1977-2019.

The main findings are as follows:

1. At an annual frequency there is essentially no relation between growth of productivity and growth of living standards.

2. Over a longer time horizon Krugman's intuition is verified. Productivity and living standards move together over the whole 43-year period 1977-2019 and also within the three sub-periods corresponding roughly to business cycles which span these 43 years.

3. Applying the statistical decomposition developed here, productivity growth was much the most important factor accounting for living standards up till 2007. Over 2007-2019 it accounted for only 45 per cent of the growth of living standards. But this was a period in which productivity grew very slowly (0.21 per cent per year).

4. Until 2007 inequality was increasing but had only a relatively minor effect on retarding the growth of living standards. After 2007, inequality declined and this had a modest effect in mitigating the effect of low productivity growth on living standards: living standards grew at 0.47 per cent per year compared to 0.21 per cent per year for productivity.

5. The labour market and demographic factors played only a minor role up to 2007. After 2007, they helped to support living standards.

6. The relative price effect (GDP deflator relative to the CPIH) was favourable to living standards in 1977-1990, unfavourable after 2007.

A decomposition by itself cannot explain anything. But it can be used as a diagnostic tool. If all the factors except productivity stayed constant, then welfare and productivity would grow at the same rate. Or we might find that the factors are all changing, but in an offsetting fashion. Or the growth of the factors taken together may impart an upward or downward movement to welfare relative to productivity. But the message of Table 2 seems unequivocal: if you want to raise living standards you have to raise productivity.

References

Anyaegbu, G. (2010) "Using the OECD Equivalence Scale in Taxes and Benefits Analysis," *Economic & Labour Market Review*, Vol. 4, No. 1, January. https://doi.org/10.1057/elmr.2010.9.

Atkinson, A.B. (1970) "On the Measurement of Inequality," *Journal of Economic Theory*, Vol. 2, No. 3, pp. 244–263.

Basu, S. and J. G. Fernald (2002) "Aggregate Productivity and Aggregate Technology," *European Economic Review*, Vol. 46, No. 6, pp. 963-991.

Basu, S., L. Pascali, F. Schiantarelli, and L. Serven (2012) "Productivity and the Welfare of Nations," NBER Working Paper 17971, National Bureau of Economic Research. http://www.nber.org/papers/w17971.

Case, A. and A. Deaton (2017) "Mortality and Morbidity in the 21st Century," *Brookings Papers on Economic Activity*, Spring, pp. 397-443.

Deaton, A. and J. Muellbauer (1980) *Economics and Consumer Behavior*, Cambridge University Press.

Edin, K.J. and H.L. Shaefer (2015) *$2.00 a Day: Living on Almost Nothing in America*, (New York: Houghton Mifflin Harcourt).

Jones, C. I. and P. J. Klenow (2016) "Beyond GDP? Welfare Across Countries and Time," *American Economic Review*, Vol. 106, No. 9, pp. 2426–2457.

Jorgenson, D. W. (2018) "Production and Welfare: Progress in Economic Measurement," *Journal of Economic Literature*, Vol. 56, No. 3, pp. 867–919.

Krugman, P. (1994) *The Age of Diminished Expectations*, The MIT Press.

Mishel, L. and J. Bivens (2021) "The Productivity-median Compensation Gap in the United States: the Contribution of Increased Wage Inequality and the Role of Policy Choices," *International Policy Monitor*, No. 41, Fall, pp. 61-97. http://www.csls.ca/ipm/41/IPM_41_Mishel.pdf.

Nolan, B., M. Roser, and S. Thewissen (2018) "GDP per capita versus Median Household Income: What Gives Rise to the Divergence over Time and How Does this Vary Across OECD Countries?" *Review of Income and Wealth*, Vol. 65, No. 3, September, pp. 465-494.

Office for National Statistics (2014) "Economic Well-being: Framework and Indicators." http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/wellbeing/economic-well-being/framework-and-indicators/art-economic-wellbeing.html.

Office for National Statistics (2016) "A Guide to Sources of Data on Earnings and Income Update." https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/aguidetosourcesofdataonearningsandincome.

Office for National Statistics (2017) "Household Disposable Income and Inequality in the UK: Financial Year Ending 2016." https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2016.

Office for National Statistics (2021) "National Life Tables — Life Expectancy in the UK 2018 to 2020." https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020.

Oulton, N. (2004) "Productivity versus Welfare: or, GDP versus Weitzman's NDP," *Review of Income and Wealth*, Vol. 50, No. 3, September, pp. 329-355.

Oulton, N. (2012a) "The Wealth and Poverty of Nations: True PPPs for 141 Countries," Discussion Paper no. 1080, Centre for Economic Performance, January. http://cep.lse.ac.uk/pubs/download/dp1080.pdf.

Oulton, N. (2012b) "Hooray for GDP!," Occasional Paper no. 30, Centre for Economic Performance. http://cep.lse.ac.uk/pubs/download/occasional/op030.pdf.

Oulton, N. (2018) "Challenges to GDP" in Michael O'Sulivan (ed.), *The Future of GDP*, Credit Suisse Research Institute. Also available as: "GDP and the System of National Accounts: Past, Present and Future," Centre for Macroeconomics, London School of Economics, Paper No. CFM-DP-2018-02, June 2018.

Oulton, N. (2019) "The UK and Western Productivity Puzzle: Does Arthur Lewis Hold the Key?" *International Productivity Monitor*, No. 36, Spring 2019, pp. 110-141. http://www.csls.ca/ipm/36/Oulton.pdf and http://www.csls.ca/ipm/36/Oulton_appendix.pdf.

Oulton, N. (2021) "The Effect of Changes in the Terms of Trade on GDP and Welfare: a Divisia Approach to the SNA," Centre for Macroeconomics, London School of Economics, Paper No. CFM-DP2021-26.

Royal Society (2012) *People and the Planet*. http://royalsociety.org/policy/projects/people-planet/report.

Sefton, J. A. and M. R. Weale (2006) "The Concept of Income in a General Equilibrium," *Review of Economic Studies*, Vol. 73, No. 1, pp. 219-249.

Stiglitz, J.E., A. Sen, and J.-P. Fitoussi (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, The Commission. www.stiglitz-sen-fitoussi.fr.

Teichgräber, A. and J. Van Reenen (2021) "Have Productivity and Pay Decoupled in the UK?," *International Productivity Monitor*, No. 41, Fall, pp. 31-560. http://www.csls.ca/ipm/41/IPM_41_Decoupling_IPM_appendix.pdf

Weitzman, M.L. (1976) "On the Welfare Significance of National Product in a Dynamic Economy," *Quarterly Journal of Economics*, Vol. 90, No.1, pp. 156-62.

Wolff, E.N., A. Zacharias, and T. Masterson (2012) "Trends in American Living Standards and Inequality," *Review of Income and Wealth*, Vol. 58, No. 2, June, pp. 197–232.

Appendix

All UK variables were downloaded from the ONS website (ons.gov.uk). Further details on the sources for the UK statistics used in this paper are presented in Appendix Table.

Household Level Variables

The following variables are taken from a spreadsheet entitled hdiireferencetables201920update.xlsx downloaded on 1st June 2021. These data underlie the regular *Statistical Bulletin on Household Income Inequality*. At the time of writing the latest version is at https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householdincomeinequalityfinancial/financialyearending2020.

1. Real median equivalised household disposable income, all people, £ per year, 2019/2020 prices (Sheet "Table 1")
2. Real mean equivalised household disposable income, all people, 2019/2020
3. prices (Sheet "Table 1")
4. Real median equivalised household disposable income of people in the lowest quintile of equivalised income, £ per year, 2019/2020 prices (Sheet "Table 2")
5. Consumer price index including owner-occupiers' housing costs (CPIH) excluding council tax (Sheet "Table 31")

Appendix Table: Variable Sources

Variable	CDID
National Accounts Variables	
GDP, at current market prices , £m	YBHA
GDP, at market prices, CVM, £m, 2018 prices	ABMI
Household disposable income (gross), £m	HABN
Capital consumption of households, £m	HAZH
Population, mid-year, usually resident, number	UKPOP
Labour market variables	
Total weekly hours worked, millions	YBUS
Employment, age 16+, thousands	MGRZ
Unemployment, age 16+, thousands	MGSC
Unemployment rate, age 16+, %	MGSX
Inactive, age 16+, thousands	MGSI
Derived variables	
GDP per hour (productivity)	1000*ABMI/(52*YBUS)
GDP deflator (2018=1)	YBHA/ABMI
Labour force participation rate	(MGRZ+MGSC)/(MGRZ+MGSC+MGSI)
Proportion of population aged 16+	(MGRZ+MGSC+MGSI)/UKPOP

Note: National accounts variables are from the 2020 Blue Book, available for download online as bb20chapter01naataglancefinal-1.xlsx and bb20chapter06hnsectorfinal.xlsx.