Essays on Measuring the Modern Economy

Poquiz, John Lourenze

Awarding institution:
King's College London

Essays on Measuring the Modern Economy

by

John Lourenze Poquiz

Supervisors:
Prof. Mary O'Mahony; Prof. Martin Weale

This dissertation is submitted for the degree of
DOCTOR OF PHILOSOPHY
in Economics

King's College London
Department of Economics

Monday 12th August, 2024

ABSTRACT

The goal of these essays is to provide a set of novel methods for measuring different aspects of the modern economy. In the first essay, we develop a framework for the valuation and accounting of free digital services, such as videoconferencing, email, and online news. In this framework, we relax the barter model, which implies that household sells viewership to platforms in exchange for digital services, by allowing the value of digital services to exceed the value of viewership. We also provide an accounting framework on how to record these transactions in a use table. The second essay provides a methodology to empirically quantify the value of these services. We employ the price of paid digital products as a proxy for the value of their free counterparts. We also use hedonic regression to untangle the shadow price of the "free component" of the service from the set of prices from the paid components. We find that the aggregate value of free digital services in the UK makes up 0.57 to 2 percent of household consumption. We extend this methodology to digital piracy in the third essay. The National Accounts does not discriminate between legal and illegal goods and services. If efforts are being made to measure the value of free digital services, for completeness, these efforts could also be extended to free digital services accessed through illegal means. We find that the value of digital piracy is at a similar level to other illegal products recorded in the UK's National Accounts, namely narcotics and prostitution. In the fourth essay, we develop a novel methodology that exploits the variations in Google Search results to estimate the depreciation of intangibles. In this exercise, we focus on original software and movies. We use the decline in the search activity for each software and movie title as our indicator of the obsolescence of these assets. This assumes that a decline in search activity reflects a decline in the sale of output produced by these assets.

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and the acknowledgments. This dissertation contains fewer than 46,000 words not including appendices, bibliography, footnotes, tables and equations.

John Lourenze Poquiz
May 2024

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my primary supervisor, Professor Mary O'Mahony. Your unwavering support, insightful feedback, and constant encouragement have been invaluable throughout this journey. I am profoundly grateful for your mentorship and the time you have invested in helping me achieve this milestone. I would also like to extend my sincere thanks to my second supervisor Professor Martin Weale. Your critical insights and constructive suggestions have greatly enhanced the quality of this work.

I am thankful to Dr. Paul Schreyer and Professor Jonathan Haskel for examining my thesis to be the examiners of my PhD thesis. I feel incredibly fortunate to have had the opportunity to learn from and work with such distinguished scholars.

I am grateful to Dr. Rachel Soloveichik and Dr. Leaonard Nakamura for their feedback, advice, and support. Many of these papers would not be in the shape they are in today without their help. Professor Kevin Fox and Dr. Carmit Schwartz for all the academic exchange. I look forward to more collaborations with them.

I am also thankful to the participants and organizers of the 2023 NBER Summer Institute - Conference on Research in Income and Wealth for rich discussion and feedback on the earlier draft of my papers; participants of 2022 World KLEMS Conference, most especially Professor Erwin Diewert, for the comments and suggestions on one of my papers; participants of the ESCoE Conferences on Economic Measurement, 2023 Society of Economic Measurement Conference, and The Productivity Institute Conference, notably my discussant Professor Diane Coyle for the generous feedback. I would also like to express special thanks to Professor Erik Brynjolfsson and Dr. David Nguyen for inviting me to discuss my work in their workshop at Stanford University.

Special thanks to Dr. Seyhun Sakalli and Dr. Michele Piffer for taking the time to help me prepare for these conferences, and most especially for being instrumental in my growth as an academic. The PhD Economics reading group for the excellent discussions that I will miss. Dr. John Morrow and Dr. Mehmet Ismail for providing feedback on some of my papers.

I would also like to thank the anonymous referees from the ESCoE Discussion Paper series and The Productivity Institute working paper series. Their comments were invaluable in shaping the essays into their current form.

I am thankful to my PhD colleagues, especially Mia, Nadine, and Richmond, for keeping me sane; Sameedh, Meng, Stylianos, Simon, Nina, and Louisa, for keeping me insane; Linda, Jiyoun, Elodie, Ines, Alice, Kitty, Yi, and others for all the good times, conversations over wine, coffee, hotpot, and Jollibee chicken. I will miss you all.

I am grateful to the King's Business School and the Economic Statistics Centre of Excellence for their funding and academic support. I also want to thank the entire ESCoE team, especially Rebecca Rilley and Sarah Sheppard.

To my family, particularly my parents, Demetrio Poquiz Jr. and Josefina Poquiz.

Finally, I am greatly indebted to my partner and soon-to-be wife, Abbi Llamido. Thank you for your never-ending support, patience, and love.

Further Notes

Part of the first chapter of the thesis, “A Framework for the Accounting of Free Digital Services”, in particular section 2.3.1, is part of a joint work with Dean Villanueva and Faith Baliscan who are consultants of the Asian Development Bank. This does not imply the endorsement of the ADB in relation to the interpretation or analysis of the framework.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

The rapid transition of digitization has brought forth a transformative period, reshaping various facets of modern economies. While the advantages of this digital era are apparent from the perspective of human experience, these benefits are barely noticeable in official statistics, which show that productivity and growth have slowed down for most of the developed world in the past decade. Some economists argue that this discrepancy is partly due to challenges in accounting for new modes of production, consumption, and investment in the modern context (Martin and Riley, 2024; Syverson, 2017).

In the contemporary economy, the profound impact of free digital services, such as videoconferencing and online news, on individuals' lives presents a measurement challenge. Despite their substantial contribution to welfare, these services are not explicitly reflected in official statistics *as separate items in household consumption*. The National Income Accounts present goods and services with market prices. While the costs of providing some of these services are incorporated into the value of advertised products, the current National Accounts structure limits our ability to analyze how households benefit from free digital services. Furthermore, free digital services may offer households additional value beyond those generated by advertising services. As a result, conventional National Accounts estimates may fail to fully capture this added value.

On a related note, efforts to quantify the value of free internet products have predominantly centered on *legal* activities, overlooking the unexplored value of content accessed through digital piracy. These services, though illegal, also contribute to household utility, posing a challenge for policymakers attempting to assess the impact of the digital economy on consumer welfare (Brynjolfsson et al., 2019a; Hulten and Nakamura, 2017; Van Elp and Mushkudiani, 2019). The National Accounts does not discriminate between legal and illegal production. As such, if efforts are being made to estimate the value of free digital services acquired legally, to arrive at a more complete picture of the digital economy, similar efforts should be considered for goods accessed illegally.

Moving on to the measurement of assets, amidst a growing body of research emphasizing the significance of intangible investments in modern economies, the need to comprehensively quantify both intangible investments and assets becomes evident. Various strategies for measuring intangible investments are explored, with Van Criekingen et al. (2022) offering a comprehensive review of modern approaches. However, challenges persist in aspects of the measurement process, such as estimating depreciation rates and price deflators, necessitating innovative solutions for a more accurate representation of intangible contributions to the economy (Corrado et al., 2009; Huang and Diewert, 2011; Soloveichik, 2010).

The *observed* productivity slowdown of many developed economies following the 2008 financial crisis has led many economists to speculate that part of this slowdown is a result of the inability of official statistics to properly capture all the outputs and inputs in the modern economy. This is dubbed in the literature as the mismeasurement hypothesis (Syverson, 2017). This dissertation attempts to address some of these measurement challenges, using both traditional approaches by National Accountants and novel ones such as exploiting big data.

In the first essay, we provide a framework for the valuation and accounting of free digital services. We discuss how efforts to measure the value of free digital products could be centered around measuring the value of the implied barter transaction between households and platforms. In this transaction, households sell viewership to platforms, which provide digital services as payments in kind. This is in line with the barter model proposed by Nakamura and Soloveichik (2015). Their model assumes that the value of viewership is equal to the value of digital services in the transaction.

We show analytically that by imposing the price of paid products that provide similar benefits to free digital services, we are able to relax this restriction set by the barter model, allowing the value of free digital services to exceed the value of viewership. In this case, a surplus is generated for the household, which we interpret as primary income. We provide an accounting framework, focusing on the use table, showing how these values can populate an extended account that aims to capture the aggregate value derived from free digital products.

While the first essay focuses on providing a framework for valuation and accounting, the second paper is centered on developing an empirical methodology to estimate the value of free digital services. For this exercise, we estimate value from the final consumption of three categories of free digital services: videoconferencing, personal email, and online news. As our measurement strategy, we employ the prices of "premium' or paid internet goods as a proxy for the value from their free counterparts. This methodology is called the market-equivalent pricing in the National Accounts and Valuation literature and has been widely applied to various types of non-market products and non-monetary transactions such as agricultural production for own use, services from owner-occupied housing, and barter transactions, to name a few. We also use hedonic regression in order to extract the value of the 'free component' of these goods and untangle them from the value of the premium-exclusive components.

The third essay extends this approach to free digital services accessed illegally. In this essay, we measure the value derived by households from digital piracy. All the previous efforts aimed at measuring the value of free digital products are focused on legal products, ignoring the possible relevance of digital piracy. Similar to the first essay, the measurement strategy employs the price of paid digital goods and services (e.g. Spotify, Netflix) as a proxy for the shadow price of their illegally acquired counterparts. To estimate the number of individuals engaged in digital piracy, we used information from the United Kingdom's Intellectual Property Office annual Piracy Tracker, which provides an estimate of the proportion of the UK population that engages in digital piracy.

In the last paper, we move from the valuation of flows to an examination of how long assets retain their values. In particular, the fourth essay introduces a novel methodology to estimate the depreciation of intangible assets, specifically focusing on software and creative originals, using data from Google Search Volume (GSV), commonly referred to as Google Trends. Depreciation, in this context, is understood as a manifestation of obsolescence. As intangible assets become obsolete, their capacity to generate future output or revenue diminishes. GSV provides a practical means to quantify the popularity of products generated by these assets, as a surge in internet searches indicates their relevance. The decline in search activity over time is directly associated with the concept of obsolescence, aligning with economic depreciation principles. In our analysis, we employ Poisson Pseudo-Maximum-Likelihood (PPML) and negative binomial regressions to estimate the rate of decline of GSV results for a sample of software and movie titles. We also examine the impact of applying our estimates of depreciation to levels of capital stocks and productivity growth for the information and communications industry, a sector where intangible investments have far outstripped tangible investments.

Summary of empirical results

In the essay relating to the valuation of free digital products, our final estimates show that in 2020, the aggregate gross value derived by households from the consumption of the three digital services for the UK was between £7.0 billion and £25.4 billion, which is 0.57 to 2 percent of household final consumption expenditures (HFCE). We also observe that the value derived by households from consuming these goods is growing much faster than HFCE. Our estimates show that in 2020, the initial year of the COVID pandemic, real household final consumption decline would have been 0.07 to 0.13 percentage points slower had the value of the three digital goods been incorporated in the estimates.

In the third essay, we find that the gross value from the piracy of music, video, live sports, software, computer games, and ebooks in the UK was between £3.6 billion and £7.5 billion in 2021. These estimates are not substantially far off from the value of final consumption for other illegal activities in the UK's National Accounts, namely narcotics and prostitution. We also find that while the value from the final consumption of communication services has been consistently rising in the past five years, the gross value from the digital piracy of media has been falling. Back-of-the envelope calculations for other countries also indicate that the share of the value attributed to digital piracy in relation to household consumption is notably higher in countries like Thailand, Sweden, and Hong Kong compared to that of the UK.

Lastly, using search volumes as an indicator for obsolescence we estimate depreciation rates ranging from 13.4 to 19.4 percent for software originals. This is lower than the estimates employed by statistical agencies, which are around 20 to 25 percent for software. In contrast, estimates for movies are comparable to estimates by statistical agencies, notably the US and Germany. Furthermore, our research shows that searches for recently released movies and software exhibit a steeper downward decline compared to earlier releases, highlighting the need to regularly update depreciation rate estimates. We also apply the depreciation rates we generated from this methodology to the estimation of capital stocks. We find that for the information and communications industry, TFP is likely underestimated for non-crisis years, particularly from 1996 to early 2008, and again from 2011 to the end of 2016.

Policy implications

We believe that results from the studies on the valuation of free digital services have substantial implications for statistical policy. The Digitization Task Team of the Inter-secretariat Working Group for the 2025 System of National Accounts update recommended that accounting for the value of free digital services could be explored in a satellite account (Digitalization Task Team, 2022). Statistical agencies can explore the development of satellite accounts to cover the value of free digital products, and our study could be a starting point for the development of such accounts.

The current approach recommended by the task force restricts the value of digital services to the total cost of producing them. As we will discuss later on in the essays, there are some disadvantages to this restriction. First, a possible decline in advertising expenditures does not necessarily translate to a decline in the usage of these services. This will cause a mismatch between the estimates and human experience, as displayed in Nakamura and Soloveichik (2015). Second, employing the cost of production to estimate the value of non-market transactions limits the ability of estimates to accurately reflect overall benefits from these services. This is why there have been many efforts to measure public sector productivity (Dunleavy, 2017; Simpson, 2009; Boyle, 2006), since government services is a prime example of a sector where this approach has been widely used, and criticised. Lastly, there are many instances where free digital products are not financed by advertising and marketing. An example of this is services offered under the *pure* freemium model, where a version of the service is offered for free with the hopes of enticing even a small share of users to register for the paid version (i.e. Zoom, dating apps).

We show that measuring the value of free services independently from the value of viewership overcomes many of these challenges. We also show that it is possible to develop a methodology for measuring the value of these services that is consistent with the core accounting principles of the SNA, ensuring comparability to other macroeconomic aggregates and ease of interpretation.

Our study illustrates that even the limited scope of the three digital products examined significantly impacts household welfare, as indicated by household final consumption. Extending this analysis to a broader range of goods is likely to underscore the greater role digital products play in household consumption, highlighting the reliance on services from companies with the ability to withdraw such services at their discretion. Moreover, this set of statistics would enable policymakers to gain insights into the potential susceptibility of households to supply disruptions associated with these products.

On the results of our fourth study, statistical agencies can consider updating their assumptions on depreciation rates, especially for intangibles, given their increasing significance in the modern economy. Our estimates for the obsolescence of original software are slower than those applied by many statistical agencies. We show applying this change could have an impact on our measures of productivity, and by extension, monetary and fiscal policies and empirical studies on capital and economic growth.

Our methodology can provide flexibility to compilers of official statistics for the estimation of depreciation rates of other assets. In many instances, statistical agencies simply assume the service life of intangibles or rely on old studies that have not been updated since they were first conducted. Our approach has the potential to allow compilers of official statistics to produce more accurate and timely measures of capital stocks that reflect new economic realities, using empirical data.

CHAPTER 2: A Framework for the Accounting of Free Digital Services

ABSTRACT

In this essay, we propose a framework for the accounting of free digital services in the context of the National Accounts. In line with previous studies in this area, we think of free digital services as a product of a barter transaction between households and advertisers. Households sell viewership to advertisers, which in turn, provides digital services as payment in kind. We extend the model by proposing that the value of digital services can exceed the value of viewership. This allows households to earn implicit income in kind from the sale of viewership. In this case, we think of viewership as intermediate inputs. We also discuss how estimates from contingent valuation reflect a different aspect of welfare compared to estimates derived from the barter model. We argue that both approaches are necessary in terms of providing a complete picture of how digital services can improve welfare in the modern economy.

2.1 Introduction

There is a growing literature that suggests that the significance of the digital economy cannot be fully captured by macroeconomic measures such as GDP and Household Final Consumption. For instance, if companies like Google or Microsoft decided to stop offering free email services or if Meta withdrew free access to WhatsApp, how much would these decisions affect consumers? To what degree would taxis and ride-hailing services be affected if Google Maps is suddenly no longer available? These questions cannot be answered by official statistics, ex-ante.

While we can assess the relative importance of traditional market goods by examining their share in total consumption, such an approach is limited for digital services. This is because the National Accounts framework only considers goods and services with market prices[1], leaving out the contribution of products that have no explicit cost to households. This includes many services offered through the internet such as social media, personal email, and search engines, among others. Therefore, it is difficult to examine the impact of the digital economy on overall economic activity using our existing frameworks for official statistics.

There have been many attempts to measure the value received by households from the

[1] With some exceptions such as government services, imputed rent from owner-occupied dwellings, and the extraction of groundwater from wells, among others.

consumption of free digital services. However, few focus on providing a framework for the accounting of these services, which we think is equally important. In this essay, we propose a conceptual and accounting framework for the compilation of the aggregate value of free digital services. The framework would allow for the compilation of aggregates reflecting welfare gains from digital services while requiring little changes to the core accounting principles of the System of National Accounts (SNA).

Currently, there are two main approaches to measuring the value of free digital goods: the contingent valuation approach and the barter approach (or the total cost approach). In both methodologies, the intention is to arrive at the monetary aggregate that represents (sometimes, to a limited extent) the contribution of free internet services to overall economic activity.

The goal of contingent valuation studies is to estimate the value individuals derive from the consumption of free digital services. Often, this involves surveys and experiments where respondents are asked how much they are willing to be paid for abstaining from certain free internet services for a given amount of time. Contingent valuation studies relating to free digital services include those by Corrigan et al. (2018), Brynjolfsson et al. (2019b), Brynjolfsson et al. (2019a), Coyle and Nguyen (2023), and Jamison and Wang (2021). In an analytical framework, Schreyer (2021) explains that estimates from this approach can be interpreted as a form of own-account production of leisure services by households. In this context, households spend time, ICT hardware and software to produce leisure services.

On the other hand, the barter approach imputes the value of free services by employing the expenditures used to produce them. In this approach, the advertising and marketing expenditures that are used to generate free content are recorded as part of final consumption instead of intermediate consumption. Therefore, final demand is augmented by these expenditure items. Research employing the barter approach include the studies by Nakamura and Soloveichik (2015), Nakamura et al. (2017),Van Elp and Mushkudiani (2019) and Van Elp et al. (2022). These studies assume that the provision of free digital services is a product of a barter transaction, wherein households sell their viewership to advertisers in exchange for these services. We will discuss the details of both methods, together with their results, in Chapter 3.

In this essay, we distinguish the value of production derived from the two approaches. We show that they represent different aspects of welfare derived from the presence of free digital services. In particular, estimates from the barter approach measure the value directly associated with these services. Meanwhile, estimates from current contingent valuation studies represent the value of household production of services, such as leisure activities, enabled by the use of digital services. We also provide an expression for the shadow price of free digital services, building from work on platform price theory (Weyl, 2010).

We contribute to the literature in three ways. First, we extend the framework laid out by the barter model by allowing the aggregate value of free digital services to better reflect welfare. While the barter model requires few changes to the core accounting principles of the SNA, proponents of this approach recognize its limited ability to measure welfare. If the marginal cost of producing digital services is zero (or close to zero), estimates employing this approach may fail to incorporate the value received by an additional user of the service. Our framework addresses this issue by distinguishing between the value received by households from the consumption of free digital services from the value of viewership.

For our second contribution, we show how implicit income is generated from the sale of viewership. We think of the value of free digital service as a form of gross output and we think of viewership as the intermediate input required by households for the production of these services. The difference between the value of digital services and the value of viewership can be considered as the implicit primary income of households from the sale of viewership. In the production side of the National Accounts, this would be recorded as the value added of digital services, which is separate from the value added of viewership. We show this through our expression for the unit price of digital services, where household surplus is represented by the difference between the unit value of digital services and the unit value of viewership.

Lastly, we provide recommendations on how these aggregates should be laid out in an accounting framework. The framework presents the role of households and platforms in value generation and presents the links between them. While we discuss the possible accounting framework for both approaches, in the interest of space, this essay will focus more on the barter model.

The outline of this essay is as follows. In section 2.2, we discuss in detail the conceptual framework. We illustrate how this framework can be executed in a simplified supply and use tables in section 2.3. Lastly, we end with some concluding remarks.

2.2 Conceptual framework

A common question raised regarding research aimed at quantifying the value of complimentary digital services is whether there is a risk of double counting. After all, many of these products rely on advertising and marketing expenditures, which are already accounted for in GDP estimates.

Dynan and Sheiner (2018) argue that while some aspects of these services are indeed captured in GDP, we can imagine that there are implied transactions that are not reflected in current measures of National Income (see figure 2.1). For instance, it is possible that digital services provide value independent of the value of advertising and marketing. Additionally, monetary transactions may also fail to account for the value of viewership households are giving up to acquire digital services. If there were efforts to measure the value of free digital services, a reasonable way forward would be to quantify the monetary value of the implied transactions.

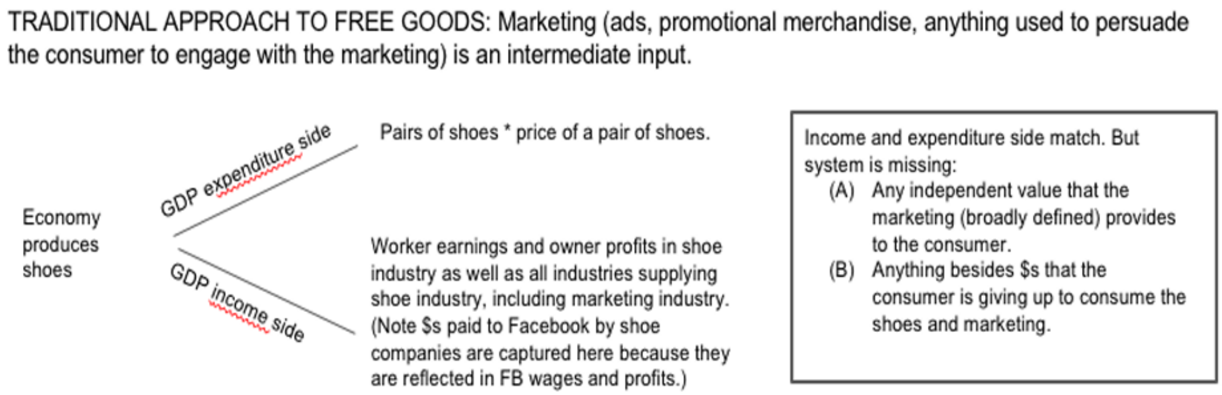


Figure 2.1: Treatment of free digital services in the National Accounts

Note: The figure is taken from Dynan and Sheiner (2018)

General framework

We illustrate this implicit transaction in figure 2.2. To simplify, consider an economy with three actors: market product producers (firms), households, and digital platforms. Transactions between firms and households are straightforward: products (a) are exchanged for monetary payments (b). In this scenario, firms generate primary income from the margin between the product's value and the intermediate costs of production. Primary income includes payments for labor and capital inputs.

Part of these intermediate costs goes towards advertising. Firms pay platforms to promote their products, and these platforms offer digital services to households. The expenses on advertising are often bundled into the product's value, which is why some economists say that the value of digital services is already factored into GDP. Advertising-funded platforms that deliver digital services to households play a significant role in this dynamic, as the value derived from advertising is integrated into the overall value of the market products consumers purchase.

What the traditional framework fails to account for is the implicit transaction between households and platforms. We can imagine households as producers of viewership, which they "sell" to platforms. In exchange for attention, platforms provide digital services as remuneration in kind. This transaction, depicted by the dashed line in Figure 2.2, is barter in nature. Consequently, if we only consider transactions involving monetary payments, GDP estimates would omit the independent value of digital services and viewership exchanged in this interaction.

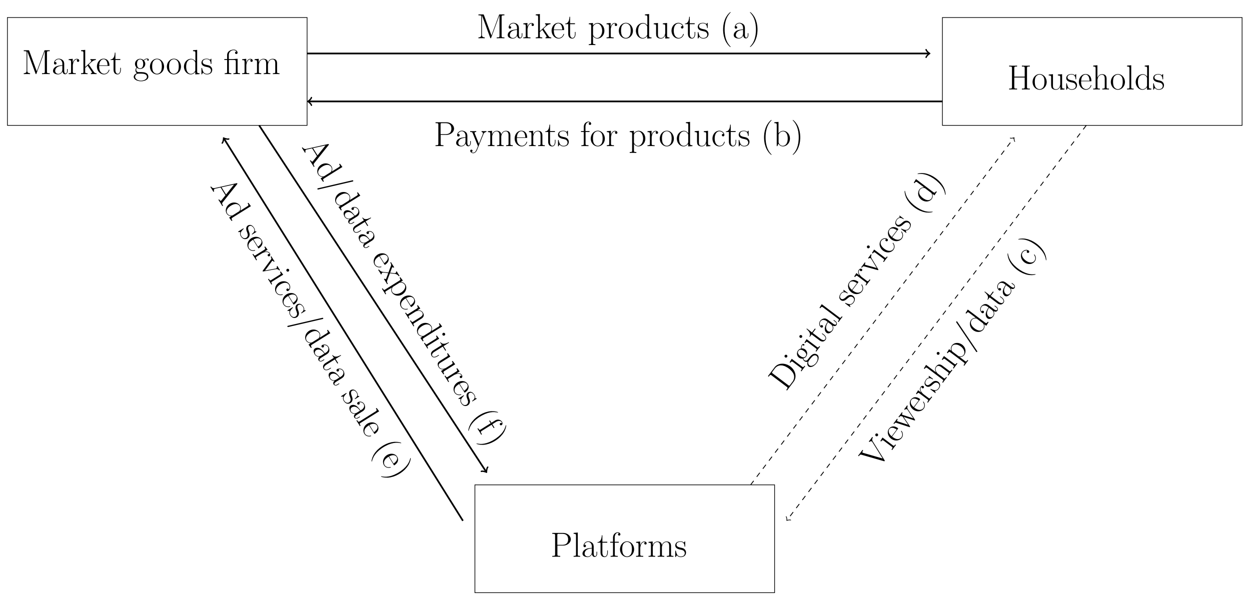


Figure 2.2: General framework

In this section, we discuss how income is generated from this transaction. We show how we can value digital services using existing frameworks employed in National Accounting. We also discuss a possible way to find the unit price of free digital services using existing prices.

2.2.1 Main setup

Consider a representative household that consumes digital goods, q^d, and an aggregate of all other goods, q^p. For simplicity, household preferences are only determined by the consumption of these goods. Household preferences are given by:

$$U = u(q^d, q^p) \tag{2.1}$$

where $u()$ is the household's utility from consumption. We assume that $\frac{\partial u}{\partial q^d} > 0$, $\frac{\partial u}{\partial q^p} > 0$, $\frac{\partial^2 u}{\partial q^d} < 0$, and $\frac{\partial^2 u}{\partial q^p} < 0$ which implies that household's utility is increasing and is strictly concave in the consumption of goods q^d and q^p.

The household earns income Y, which we consider exogenous in this model. P^d and P^p are unit prices of q^d and q^p, respectively. The household's budget constraint is therefore:

$$Y \geq P^d q^d + P^p q^p \tag{2.2}$$

This is represented by the blue line in both panels of figure 2.3. Given the budget constraint, Y_1, a rational household would consume the bundle q_1^d and q_1^p or point A to reach the highest level of utility U_1.

Before we illustrate how to account for the value households receive from free digital services, it helps to think about how nominal GDP is translated to real GDP. The succeeding discussion is based on the cost of living (COLI) framework as described by Dynan and Sheiner (2018). When converting nominal GDP to real GDP, economists are interested in the degree to which consumers are better off from an increase in nominal GDP. Dynan and Sheiner (2018) noted that economists do not ask this question directly because of the difficulty in measuring welfare. Rather, they estimate how much purchasing power improved from period to period. There are two ways to achieve this using the COLI framework. First, by asking how much compensation households require in period 1 to reach the utility in period 2. This is called equivalent variation. The second is to ask how much income we need to take away from households in the second period in order to remain at the same utility as period 1. This is called compensating variation.

Consider a case where technological improvement reduces the price of digital goods, q^d. The household budget constraint would be more flat in period 2. We illustrate this in figure 2.3.

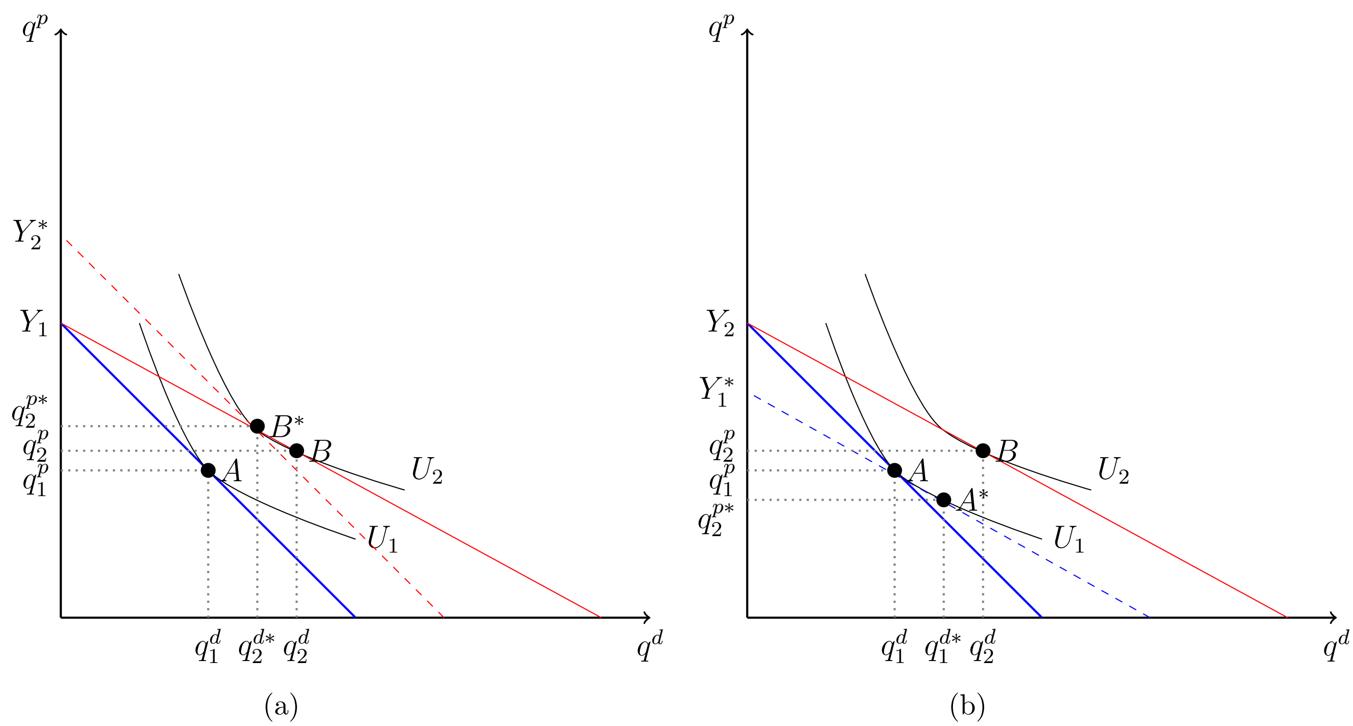


Figure 2.3: Cost of living framework

Suppose relative prices were kept unchanged from period 1, such that $\frac{P_1^d}{P_1^p} = \frac{P_2^{d*}}{P_2^{p*}}$, as represented by the dashed line in figure 2.3a. Then, $Y_2^* - Y_1$ reflects the additional income required in period 1 to reach period 2 utility, U_2, given the price change. This additional income is called equivalent variation.

Similarly, using period 2 prices, such that $\frac{P_2^d}{P_2^p} = \frac{P_1^{d*}}{P_1^{p*}}$, as represented by the dashed line in figure 2.3b. Then, $Y_2 - Y_1^*$ reflects the income needed to be taken from period 2 in order to maintain the same utility in period 1, U_1. The income taken from period 2 is called compensating variation.

The distance $Y_2^* - Y_1$ and $Y_2 - Y_1^*$ are not the same. EV and CV are not expected to be equal because they answer different questions. In welfare analysis, the decision on which measure to use depends on the problem. In National Accounts calculations, the measure used depends on the compiling agency. Most countries use a fixed base year such that for real GDP growth $\frac{Y_t^*}{Y_0}$, prices in period 0 are used in Y_t^*. This is similar to equivalent variation (using previous period prices). The Bureau of Economic Analysis and other developed countries employ chained indices, where the geometric average of the two measures, $\sqrt{\frac{Y_t^*}{Y_0}, \frac{Y_t}{Y_0^*}}$, are used to calculate GDP growth (Dynan and Sheiner, 2018).

Building upon the COLI framework, rather than directly assessing the improvement of welfare resulting from the availability of free digital services, we might explore quantifying the extent to which households' purchasing power has increased during the period when digital services are provided at zero price. In the next sections, we explore how to apply this framework to measuring the implied income from the presence of free digital services. We discuss how each approach measures different aspects of welfare change from digital services. There are valid considerations for adopting either measure, contingent upon the nature of the policy question. We present two cases, the first is for measuring the value of free digital services, and the second is the value of leisure time enabled by digital services.

2.2.2 Case 1: Measuring the value of digital services

Consider a scenario where in period 2, P_2^d drops to 0 (see figure 2.4). In this case, q^d would be provided for free from the point of view of the consumer. The household would increase the consumption q^d from q_1^d to q_2^d. Moreover, he can re-allocate all of his income towards the purchase of q^p, increasing the consumption of the good to q_2^p, such that $Y_2 \geq P_2^p q_2^p$.

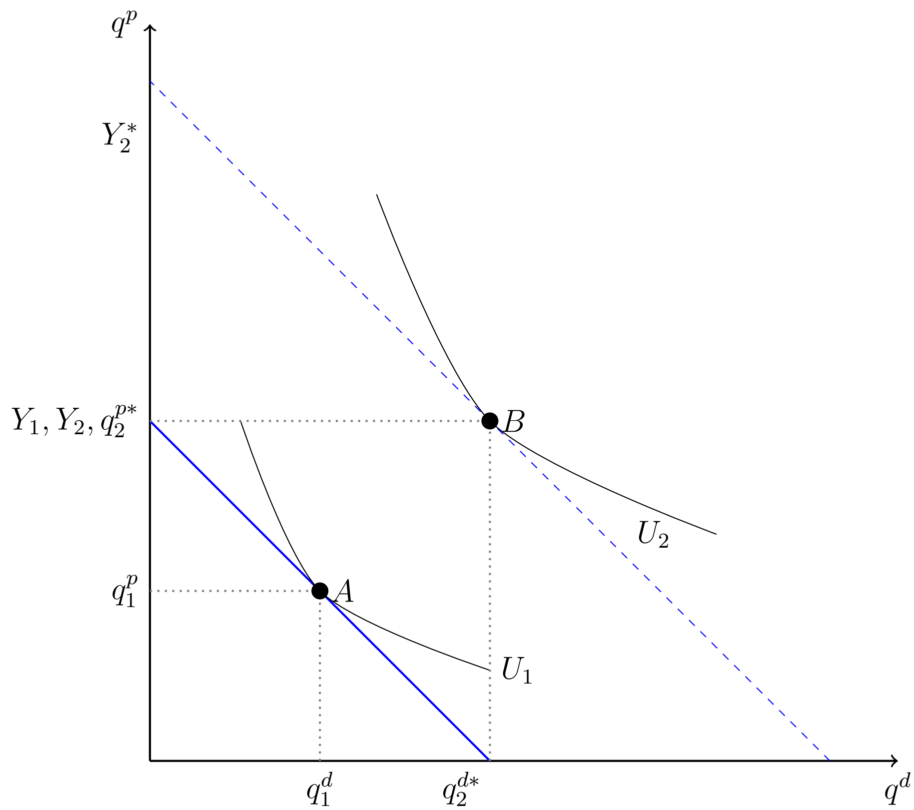


Figure 2.4: Equivalent variation for zero-priced digital products

As the households consume the bundle $\{q_2^d, q_2^p\}$, where $q_2^d > q_1^d$ and $q_2^p > q_1^p$, they reach a higher level of utility, U_2. Conventional measures of nominal GDP would not be able to reflect this increase in utility since, in this case, $Y_1 = Y_2$. However, by keeping prices of q^d constant, we would be able to measure the increase in income required in period 1 to reach the utility level in period 2, where households can consume higher levels of $\{q_2^{d*}, q_2^{p*}\}$ (see figure 2.4). This is similar to equivalent variation in figure 2.3a. GDP growth would then be:

$$\frac{Y_2^*}{Y_1} = \frac{q_2^{p*} + P_1^d q_2^{*d}}{q_1^p + P_1^d q_1^d} \tag{2.3}$$

In practice, this is not this is not straightforward. Most free internet services (for example, Whatsapp, Gmail, etc.) were introduced as "new services" with no paid equivalent in periods before their introduction. As such, it might be necessary to calculate a reservation price for the product. Calculating the reservation price of multiple digital products can be too complex for official statistics, especially from the point of view of communication and policy use. Additionally, for products that have remained free for extended periods, like Gmail and Yahoo Mail, relying on reservation prices from two decades ago can render estimates difficult to interpret, considering the dynamic nature of these services characterized by rapidly changing product quality and shifting consumer preferences.

A reasonable alternative is to identify paid products that offer similar benefits to users as free digital services and use their prices to value the products available at zero price. The rationale behind incorporating previous period prices in the calculation of equivalent variation is to determine the total monetary value of consumption with the new quantity level, maintaining the same value for each unit of consumption of the product. Employing prices of comparable paid products allows us to perform a similar analysis. Assuming equivalence in the benefits derived from both paid and free services, leveraging the price of paid products as a proxy for the value of free services enables us to quantify the overall value of consumption for the product in dollar terms, holding the unit value the same. This is similar to equivalent variation in figure 2.4. Here, instead of comparing income from the previous period to the current period, we would be comparing aggregate income while accounting for the consumption of digital products versus aggregate income without considering them. In practice, this analysis is possible if we are able to apply quality adjustment by removing the value of features available only to paid versions from their prices (i.e. remove the value of domain customization, expanded memory, and other premium features from the price of paid Gmail).

Prices of paid platforms could potentially provide a source of unit valuation for free digital products. Many platforms offer paid service alongside the products' free version. For example, email service providers such as Gmail and Yahoo provide both a free service and a paid version. This is also true of videoconferencing services such as Zoom, music streaming platforms such as Spotify and Amazon Music, and more recently, social media such as Twitter.

Using the price of paid platforms as a proxy for the value of their free versions hinges on the assumption that the prices of paid platforms accurately represent the benefits of using the platform. However, this assumption may oversimplify the intricate nature of platform pricing. Platform pricing is multifaceted, influenced not only by the benefits received by users but also by external factors, such as network effects and cross-subsidies. Therefore, if we employ the price of paid platforms as a proxy for the unit price of free digital products, the question arises: What precisely would we be capturing? Answering this question is important when developing an accounting framework for free digital services.

Platform prices as a source of valuation

The following discussion is based on Weyl (2010)'s price theory on multi-sided platforms, which suggests that platforms internalize externalities in their prices. Platforms are typically defined as service providers that link one set of users to another (Belleflamme and Peitz, 2021). Newspapers are a classic example of a two-sided platform. Advertisers (first group) are linked to readers (second group). It is also possible for platforms to link more than two groups. For instance, providers of freemium services such as Spotify would have three "sides": 1) paying users, those who pay for their subscription to Spotify, 2) free users, those who use the free version supported by advertising, and 3) advertisers. It is common for platforms to subsidize one group through another group. These subsidies are made possible through their unique pricing strategies.

Consider a platform with two sides, I and J. For each side, we can imagine that participants from side I receive a standalone benefit B^I from participating in the platform. Additionally, the users also receive an interaction benefit (or cost) b^I from every user that participates in side J. According to Weyl (2010), users from side I will only participate if the benefits are greater than or equal to the price:

$$B^I + b^I N^J \geq P^I \tag{2.4}$$

where N^J is the number of users for side J. This suggests that platform prices reflect at least the lower bound for the dollar value of the benefits received by their users.

To better illustrate what is accounted for in the price of the platforms, it is important to understand how prices are set. Platforms set prices that maximize their profits by choosing the network size of one side. A two-sided platform's profit function is given by:

$$\pi(N^I, N^J) = (P^I(N^I, N^J) - C^I)N^I + (P^J(N^J, N^I) - C^J)N^J - cN^I N^J \tag{2.5}$$

where C^I and C^J are the marginal costs of providing the service for side I and side J, respectively, and $cN^I N^J$ is the fixed cost of connecting both sides of the platform. The first order condition of the platform's profit-maximizing problem is:

$$\frac{\partial \pi}{\partial N^I} = P^I + P^I_I N^I - C^I + P^J_I N^J + cN^J = 0$$

and a profit-maximizing platform would produce services such that:

$$\underbrace{P^I + P^I_I N^I + P^J_I N^J}_{\text{marginal revenues}} = \underbrace{C^I + cN^J}_{\text{marginal cost}}. \tag{2.6}$$

Re-arranging yield:

$$P^I = C_I + cN^J - P^I_I N^I - P^J_I N^J \tag{2.7}$$

The first three terms represent the firm's marginal cost less the inverse hazard function rate of demand (or market power) $\mu^I = -P^I_I N^I = P^I_I/\epsilon^I$ where ϵ^I is the elasticity of demand (see Weyl (2010)). The final term represents the additional revenues extracted from side J from an additional user of side I. Weyl (2010) describes this as the external benefits from the provision of service on side I. Prices for side I can be expressed as:

$$P^I = \underbrace{C_I + cN^J}_{\text{marginal cost}} + \underbrace{\mu^I}_{\text{market power}} - \underbrace{P_I^J N^J}_{\text{external benefits}}. \tag{2.8}$$

This is how platforms are able to provide services at zero prices. In some cases, external benefits are greater than the combination of cost and market power. Again, newspapers provide a good example. The external benefit of advertisers from the presence of readers outweighs the marginal cost and market power. For the side of advertisers, the $P_I^J N^J$ term is negative since readers receive negative benefits from the increased presence of advertising. As such, the platforms charge higher prices on the side of advertisers by the amount, $P_I^J N^J$. To be more explicit, multiplying a negative value of $P_I^J N$ by the negative sign in the last term of the expression in equation 2.8 results in a positive value. For instance, let's say that the marginal cost of producing a newspaper ad is £0.5 per unit. On the readers' side, the platform loses a £1 of revenues for each unit of advertising. This is because readers are annoyed by the presence of advertising and some readers might not subscribe to the newspaper if there are too many ads. Assuming that market power distortion is 0, the price of a newspaper ad would then be $P = 0.5 - (-1) = 1.5$. Conversely, readers are then subsidized by advertisers since the cost of providing readership is internalized in the price levied on advertisers. Weyl (2010) generalized these results for platforms with more than three sides. The general principle remains the same and the expression for the price of platforms with more than two sides is not materially different from equation 2.8.

We can think of the same mechanism operating for platforms that monetize the provision of their free services through the selling of their data. For example, we can imagine that users of free email and instant messaging apps gain "negative external benefits" from the sale of their data, and this is reflected in the price of their data levied to third parties. On the other hand, these platforms are able to provide email and instant messaging at zero price because the external benefits to third parties are greater than the cost of provision.

To estimate the increase in implicit income resulting from the provision of free digital services as depicted in figure 2.4, we require pricing structures that accurately capture both the marginal benefit derived from these services and the associated production costs. Perhaps we can consider platforms that utilize the freemium model. Under this model, some users are granted access to services for no monetary cost. The funding for the free version typically comes from advertising revenue and the sale of user data to third parties[2]. Additionally, these platforms may offer a paid version alongside the free one, which either remains ad-free or guarantees not to sell user data to third parties. A possible way forward is to exploit the price of the paid version as a proxy for the value of the free version.

Consider a platform that operates under a freemium model. Effectively, the platform would have three sides. Let us define side I as premium or paying users, side J as users of

[2]In some cases, it is the benefit of being linked to a large network side that generates external benefits to paying users

the free version, and side K as advertisers. Prices for premium users would be set as:

$$P^I = \underbrace{C^I + cN^J N^K}_{\text{marginal cost}} + \underbrace{\mu^I}_{\text{market power}} - \underbrace{(P_I^J N^J + P_I^K N^K)}_{\text{external benefits}}. \tag{2.9}$$

(*see proof in appendix 2.B*)

As shown in equation 2.4, we've established that platform prices inherently reflect the benefits garnered from participation. When considering services offered through the freemium model, one may argue that the derived benefits remain comparable if adjustments are made to features exclusive to the paid version, effectively equating the perceived value between paid (I) and free (J) sides: $P^I = B^I + b^I N^J = B^J + b^J N^I = \hat{P}^J$. For example, by removing the unit price of premium-exclusive features from paid Gmail accounts (e.g., expanded storage and personalized domain names), we can assume parity in the core benefits (e.g., email services) between both versions. Hence, from the perspective of benefits, $\hat{P}^J$ could serve as a valid proxy for the price of free digital services. Conceptually, this parallels the approach of market equivalent pricing commonly employed in the valuation of non-market goods and services and non-monetary transactions in the compilation of National Accounts. The key idea is to find prices of similar products for the valuation of the service delivery (or non-monetary transactions in the case of barter and remuneration in kind) where there is no explicit monetary exchange. In our case, this would represent the price of free digital services required for the estimation of equivalent variation in figure 2.4.

For users of the premium versions, the value $P_I^K N^K$ in equation 2.9 would be negative. This negative value represents the *absence of benefits* received by advertisers or third parties. This is because premium users are not exposed to ads and/or their data are not being sold to third parties. In other words, $P_I^K N^K$ reflects the opportunity cost in terms of advertising revenues (or revenues from the sale of data to third parties) in the presence of premium users. For example, if the marginal cost of a paid Spotify service is £5 a month, the Platform could have earned £6 had the service been ad-supported. £6 represents the opportunity cost for the platforms in terms of the additional revenues it could have extracted from the advertising side. The price of Spotify for paying users in this example would be $P = 5 - (-6) = 11$, assuming that there is no market power distortion.

If services from the free side of the platform are similar to the paid side, perhaps we can assume that the unit cost of providing both free and premium services is similar, such that $C^I = C^J$. Here we assume that the provision of email to paid users costs the same (or at least approximately) the cost of providing email to free users. If we *impose* $\hat{P}^J$ as a proxy for the price of free digital services from the side consumed by users in side J, then:

$$\hat{P}^J = C^J + cN^J + \mu^I - P_I^J N^J + P_I^K N^K. \tag{2.10}$$

If we adopt $\hat{P}^J$ as a proxy for the price of free digital services, the expression $P_I^K N^K$ in equation 2.10 would represent the unit value of viewership or data gained from an additional unit of free users. For the paid user, this term reflects the foregone unit value of advertising or revenue generated from data sales. However, free users *are* exposed to advertisements or have their data sold to third parties. Consequently, the final term does not signify the opportunity cost to advertisers and third parties; rather, it represents the actual unit value derived from viewership through advertisements or from the data sold.

We know that from the firm's profit-maximizing decision, the cost of providing the service and markup for side J were already internalized in the price of side I and K. In other words, the cost term, $C^J + cN^J$, was already paid for by users advertisers and paying users. Moreover, one can argue that users of the free version receive no network benefits from paid users. For instance, free users of Spotify do not care about the number of paying Spotify users. In this same way, users of the free version of Gmail do not care about the number of paying users of the platform. Hence, $P_I^J N^J = 0$ in this case. Imposing $\hat{P}^J$ as the price of free digital services generates a residual from the difference between the shadow price and the value of the viewership (or data sold), which we will denote as V^J:

$$\hat{P}^J - P_I^K N^K = V^J$$

We interpret this residual as a surplus received by households from the use of free digital services. As such, the shadow $\hat{P}^J$ can be broken down into:

$$\hat{P}^J = \underbrace{V^J}_{\text{household surplus}} + \underbrace{P_I^K N^K}_{\text{value of viewership or data}} . \tag{2.11}$$

In equation 2.11, we decompose the shadow price of free digital services into two, components[3]. We can think of V^J as an implicit income gained by users of households from the consumption of free digital services. This surplus is gained by selling their viewership or data through a barter transaction. This is conceptually similar to the barter model proposed by Soloveichik (2015) and Nakamura et al. (2017). The only difference here is that we allow the household to generate a surplus above the value of viewership, or the data, that they sold in the barter transaction. As such, the application of this approach would likely yield aggregates that are higher than those derived using the sum of costs method because it takes into account the value households receive from the consumption of free digital products over and above the value advertisers place on their viewership.

In our Spotify example, if we use £11 as a proxy for the unit value of Spotify's free version, then £6 of that would reflect the value of viewership free users forgo to gain access to the service. By extension, the difference between the shadow price of Spotify and the value of

[3]While the notations would be different, the same intuition applies for three-sided platforms, which is more appropriate for the freemium model. We provide the proof in appendix 2.B: the value of viewership or data from free users $P_I^J N^J$, and a surplus V^J, which we interpret as the independent value of the service.

viewership (£11 - £6 = £5) would reflect the surplus gained by free users. A pure barter model would only account for the £6, which would be reflected in advertising expenditures.

This result does not rely on the assumption that $P_I^J N^J = 0$. Even if free users receive some level of value from an additional paid user (for instance, free users of Zoom are probably using the platform because they engaged in meetings and seminars initiated by paying users), the value of $P_I^J N^J$ will only be subsumed in V^J, which we treat as a residual. Therefore $P_I^J N^J$ will be part of the household surplus.

Going back to the general framework in figure 2.2, we can think of $\hat{P}^J$ as the unit price of digital services received by households as payment in kind (d). Meanwhile, viewership, denoted by $P_I^K N^K$, would signify the intermediary expense households incur in generating the attention demanded by platforms. Analogous to producers of market goods, who derive primary income from the difference between the total revenue from their goods and the intermediary expenses required for their production, households similarly generate primary income from the margin between the shadow price of the digital service and the unit value of advertising and data exchanged with platforms in return for digital services.

This primary income is also what we measure when we calculate $Y^* - Y$ in figure 2.4 upon using the price of premium services to proxy the value of their free version. We can think of it as an implicit income in kind from the use of digital services that have no explicit cost to the household. It would be as if households are a few dollars richer from using a free service that they would have paid for had it not been available for free.

It is entirely possible that the value of the household surplus V_t is zero or negligible. In this scenario, the value of free digital services would simply equal the value of viewership, and the estimates would converge with those derived from the conventional total cost approach of Soloveichik (2015), which is one of the options for a satellite account endorsed by the Digitization Task Team for the 2025 SNA updates (Digitalization Task Team, 2022). It is important to note that even if V_t is zero, there is merit to measuring digital services separately from viewership. This independent estimation process ensures that we do not simply ignore the possibility that households gain value from free digital services above the value of viewership they provide. Intuitively, it is difficult to think of a reason why consumers would value access Youtube content at the same level as advertisers value their attention to adverts. Soloveichik (2015) would therefore be a special case where $V_t = 0$.

2.2.3 Case 2: Measuring the value of digital leisure time

An obvious alternative to using equivalent variation for measuring the value households derive from the use of free digital services is compensating variation. For market goods, the conventional method is to measure welfare changes using this approach, which involves comparing the income change at period 2 prices:

$$CV = Y_2(P_2, U_2) - Y_1^*(P_2, U_1). \tag{2.12}$$

where P_2 is the price vector for period 2. However, if the price of q^d drops to zero in period 2, this comparison becomes unfeasible. One practical approach to addressing this issue is to directly ask individuals about the amount of income they would be willing to forgo in order to maintain the same level of utility as they would have in the absence of q^d in period 2.

A growing number of studies have employed methodologies comparable to the above-mentioned approach. These include works by Corrigan et al. (2018), Brynjolfsson et al. (2019a), Brynjolfsson et al. (2019b), Coyle and Nguyen (2023), and Jamison and Wang (2021). Typically, these studies involve asking participants about their willingness to receive compensation in exchange for foregoing certain digital services for a period. The valuation derived from this approach is akin to compensating variation: individuals are asked how much income they would need to forgo to return to the level of utility they experienced when the digital good was not present.

We mentioned earlier that equivalent variation and compensating variation would not necessarily yield the same estimates of income changes. Current applications to the valuation of digital services involve eliciting survey respondents' willingness to accept. In contrast, household consumption in the National Accounts is valued using exchange prices, the junction between the market's willingness to pay and the marginal cost of providers. As such, applying WTA estimates to value free digital services directly would likely yield inconsistencies with the valuation in the SNA.

Moreover, if individuals were asked how much they need to be paid to give up access to internet service for a month, this exercise would likely produce higher values than their monthly internet expenditures. This was demonstrated in a representative online survey by Coyle and Nguyen (2023), where they found the WTA for certain market services (i.e. Netflix, newspapers, etc.) exceeded monthly prices for these services.

Schreyer (2021) argues that valuations from this approach represent the respondent's leisure time enabled by the digital service. The respondents' willingness to accept captures the value of their time producing leisure services for themselves. Schreyer (2021) considers the production of leisure enabled by digital services as part of household production of services for own consumption.

Following this logic, estimates generated using this approach do not represent the value of the service, but the time used to produce leisure services enabled by the digital services. Digital services, in this case, are inputs to leisure activities, the same way groceries are inputs to the preparation of home-cooked meals. Incorporating this approach is outside the boundary of the SNA, which excludes most household production of services, and more importantly, leisure. The inclusion of estimates from this method would constitute an expansion of the household satellite accounts to include household leisure services.

2.3 Accounting for the value of free digital services

In this section, we illustrate the accounting of free digital services through a simplified supply and use table employing the principles from the previous section. We focus on the use tables for now. First, we show the current treatment of free digital services in the National Accounts. We briefly discuss the barter approach, specifically a simplified version of Soloveichik (2015). Lastly, we illustrate the accounting for free digital services and viewership in the context of our proposed framework.

Consider an economy with three industries, each producing one output: 1) the materials industry, 2) the advertising industry, and 3) the soap industry. The output of the materials industry is used as intermediate input of the soap and advertising industries. The soap industry, meanwhile, sells its output directly to households for final use. The soap industry also spends on advertising to promote its products. The advertising expenditures are used to produce free services (say, free email services) that households enjoy while intermittently displaying advertising content to the viewers. We illustrate using hypothetical values.

Table 2.1: Conventional approach to free digital services

	Materials Industry	Advertising Industry	Soap Industry	Intermediate Demand	Final Demand	Total Use	GDP	Total Output
Materials		10	10	20		20		
Advertising			**20**	20		20		
Soap					60	60		
Intermediate consumption		10	20					
Value Added	20	10	30				60	100
Total Inputs	20	20	60					
GDE					60			
Total Demand					100			

Note: Values are hypothetical and for illustrative purposes only.

In the conventional National Accounts framework (see table 2.1), expenditures on advertising would be recorded in the use table as part of intermediate consumption. In our simplified example, only the gross value of soap is recorded as part of final demand. While advertisers produce content that directly feeds into the household's utility, this is not reflected in the use table as part of household consumption, explicitly. Instead, the value of advertising is embedded in the value of the soap. The problem with this approach is that we are not able to directly examine the value derived by households from the consumption of free digital services.

Going back to our general framework in figure 2.2, the value of products exchanged between producers of market goods (a) and payments to sellers (b) is reflected in the final demand for soap and the value-added of soap and its inputs. This accounting convention ignores the implied transaction between households and platforms, as reflected by the dashed line in this more general framework.

We explore a possible accounting framework for the value of free digital services in the following discussion. As with the previous section, we present two cases. First, we consider only the value of free digital services. The second would be accounting for leisure time enabled by digital services.

2.3.1 Case 1: Measuring the value of digital services

In this section[4], we provide a framework for the accounting of free digital service. The solution proposed by Soloveichik (2015), and later versions of her paper, was to record advertising expenditures as part of final consumption rather than intermediate consumption. To execute this approach, a new industry is introduced, which is household viewership. The approach imagines households as a producer of viewership, which it sells to advertisers in exchange for free digital services.

Because this approach requires few conceptual changes to the core SNA, some statistical agencies opted to adopt this procedure in their experimental estimates for the accounting of free digital services. Examples of this include exercises conducted by researchers from the US (see Nakamura et al. (2017)) and the Netherlands (see Van Elp and Mushkudiani (2019) and Van Elp et al. (2022)).

Our proposal extends this approach by allowing the value of digital services to be greater than the value of viewership. We argued in equation 2.11 that *if* we impose the price of paid services as the unit value of free digital services, then this price would capture not only the value of viewership but a surplus for the households. We can think about this surplus as primary income or the value added gained by households for selling their viewership or data. It is as if households are x $ richer from selling their viewership in exchange for digital services. Digital services, in this context, can be considered as payment in kind, which is consistent with the non-monetary transaction in the National Accounts framework (see 2008 SNA par 3.75). We imagine this as similar to the imputed income households gain from owner-occupied housing.

In our framework, households sell viewership (or data). In exchange, platforms provide payments in kind for the viewership with digital services valued by (d) in figure 2.2. The difference between the value of digital services and the value of viewership would be the value added or implicit income gained by households for selling viewership.

We provide a simplified example of a use table in table 2.2. Using this approach, GDP would be higher by the amount of value added gained by households from the production and sale of viewership and production of digital services. For simplicity, we break down household activity into two: the production of digital services and the production of viewership. Here, viewership is recorded as intermediate consumption for digital service production.

[4]This section of the essay is part of a joint work with Dean Villanueva and Faith Balisacan who are consultants at the National Accounts team of the Asian Development Bank. Views and opinions expressed in this essay are solely of the authors and do not necessarily reflect those of ADB and its management.

Table 2.2: Use table for the proposed approach

	Materials Industry	Advertising Industry	Soap Industry	Household Digital Services	Household Viewership	Intermediate Demand	Final Demand	Total Use	GDP	Total Output
Materials		10	10			20		20		
Advertising			20			20		20		
Soap							60	60		
Digital services							40	40		
Viewership				20		20		20		
Intermediate consumption		10	30	20						
Value Added	20	10	30	20	20				100	160
Total Inputs	20	20	60	40	20					
GDE							100			
Total Demand							160			

Note: Values are hypothetical and for illustrative purposes only.

In this framework, we explicitly account for the value of data as well as any surplus value households receive from the barter transaction. The advantage of this approach is that it allows for the value received by households to increase even as advertising expenditures decline. This is apparent from the earlier work by Nakamura and Soloveichik (2015), which finds that advertising-supported entertainment estimated using the barter approach to be declining relative to GDP from 1980 to 2013, despite the rise of various internet services during that period[5].

If the cost of producing services to an additional user is close to zero, an increase in the user base may not necessarily reflect an increase in the aggregate value of these services. This would be problematic if the goal is the capture of welfare gains from the internet economy. Moreover, in times of crisis, firms usually cut down on advertising and marketing expenditures. This will ultimately cause a reduction in the recorded value of free digital services, which may not be the case. For instance, the world saw a decline in advertising expenditures in the pandemic-induced recession of 2020 (see table 2.C.1) while usage of online platforms increased substantially.

In our simplified illustration, the value of household viewership is equal to the advertising expenditures of the soap industry, and by extension, also equal to the gross output (and total inputs) of the advertising industry. This is no coincidence. One possible way to implement this framework is to value viewership using advertising expenditures, revenues from the sale of data to third parties, and marketing expenditures. By doing so the value of viewership would reflect the advertiser's willingness to pay for the household's attention. For the value of digital services, the price of paid services can be used as a proxy as discussed in the earlier section.

Would this result in double counting? In this framework, the cost of advertising is reflected both as part of the final demand for soap and digital services. Soloveichik (2015), whose barter model operates in a similar way, argues that this is not the case. The consumption of digital services is not predicated on the consumption of advertised goods. As such, the value households place on free digital services is independent of how much they value advertised products, like the soap in our example. She writes:

> *One might argue that advertising-supported media is an intermediate input embedded in final output, and therefore, our experimental methodology double-counts advertising-supported media. However, that argument assumes that consumers can't watch advertising-supported media without buying the products. From a legal standpoint, that's not true. Advertising-supported media is available to everyone without any purchase requirements. The market price for advertised products only covers the products themselves, not the shows on which they're advertised.*

The difficulty with using this framework lies in identifying suitable shadow prices for complimentary digital services. Although we illustrate that market equivalent pricing could

[5] https://ourworldindata.org/internet

potentially resolve the valuation issue[6], not all free digital products have readily accessible paid alternatives. For instance, it is difficult to imagine paid alternatives to search engines, online maps, and social media. Perhaps, the way forward is to consider different tiers of valuation (see appendix 2.A).

2.3.2 Case 2: Measuring the value of digital leisure time

In the previous case, we show how to account for digital services as an output from selling viewership to platforms. For measures of changes in income using compensating variation, what is being valued is the time spent on leisure services rather than the services themselves. This constitutes an extension of the household production satellite account.

Table 2.3: Extended household satellite account

	Output	Inputs	GVA
Housing Services	349	202	147
Transport	189	43	147
Nutrition	196	102	94
Clothing	4	1	3
Laundry	69	14	55
Childcare	212	33	179
Adult Care	45	7	38
Voluntary activity	22	0	22
Digital Leisure	65	40	25
Total without Digital Leisure	1,086	401	685
Total with Digital Leisure	1,151	441	710

We provide a simple illustration of this possible extension in table 2.3. The extension would involve the inclusion of an additional activity, "Digital Leisure", which captures the value from the production of leisure activities enabled by digital services. In this context, the value of digital services would be recorded as inputs to the activity. As proposed by Schreyer (2021), we can use WTA value from surveys and experiments to estimate a shadow wage for households, which would form part of gross value added.

There is no reason for this extension to be limited to digital leisure. Other household activities are also enabled by free digital services. For instance, a large percentage of households refer to Youtube videos or Instagram reels when preparing home-cooked meals. Similarly, many individuals use free online maps to navigate to their destinations. In this context, complimentary digital services can be considered inputs in transportation and nutrition activities.

[6] We will discuss in Chapters 3 and 4 some practical ways how to implement this approach. We provide a detailed methodology as well as some estimates.

2.4 Concluding remarks

In this essay, we propose a set of frameworks for the accounting of free digital services. In the first case, we value the digital services directly, extending the barter framework first proposed by Soloveichik (2015). An alternative would be to measure household production of leisure services enabled by digital services.

We explain how equivalent variation and compensating variation measure different aspects of the benefits of free digital services. The choice of which measure to employ would depend on the policy question the measurement exercise aims to address. Are we interested in the value of digital services themselves or are we interested in the value of household activity enhanced by these services? Are we interested in comparing the value of digital products to the value of goods and services recorded in the National Accounts? Or perhaps we are interested on how these services are improving the productivity of households in producing own-account services? Answering these questions would determine which approach is appropriate.

Future work in this field should also examine the contribution of free digital services to the production of market goods and services. There is little work in this area. As we discuss in section 2.A.2, one possible measurement strategy would be to estimate a production function, which incorporates free digital products as part of inputs. For this endeavor, we imagine that time-use surveys could be used to provide indicators of how much workers spend using free digital services in their work routine.

While there is not much research on this, the linkages between free digital service providers and producers of market goods is very relevant in terms of describing how the internet is shaping human society. It answers the question, to what extent are these businesses vulnerable to shocks in the provision of these services?

Appendix

2.A Valuation for non-market activities

Measuring the value of goods and services without a market—and consequently, outside the production boundary of the National Accounts—is not a new endeavor. A number of non-market activities, such as government services and services from owner-occupied housing–are currently included in standard GDP estimates. Moreover, some satellite accounts focus on extending the production boundary of the SNA to include non-market activities. In this section, we focus on the valuation methods employed by a particular type of non-market account: the environmental satellite account. We then draw parallels with some aspects in the digital economy, highlighting how methods for the valuation of ecosystem services can be used for the valuation of free digital services.

Since the release of the seminal paper by Leontief et al. (1974), many economists and statisticians have expressed interest in measuring the contribution of the environment to overall economic activity. These efforts eventually led to the development of the System of Environmental Economic Accounting (SEEA), a statistical framework designed to produce a set of aggregates describing environmental assets and flows and their relation to the market economy.

A key feature of the SEEA is its consistency with the accounting principles of the core SNA framework. The framework stresses that the valuation of environmental assets and ecosystem activities should follow the same valuation principles employed for other assets and activities in the National Accounts. The rationale for this is to allow for comparability with other macroeconomic aggregates. The SEEA Ecosystem Accounting manual writes:

> *In ecosystem accounting, the primary motivation for monetary valuation using a common monetary unit or numeraire is to be able to make comparisons of different ecosystem services and ecosystem assets that are consistent with standard measures of products and assets as recorded in the national accounts. This requires the use of exchange values. In turn, this facilitates the description of an integrated system of prices and quantities for the economy and the environment that is a core motivation of the SEEA Ecosystem Accounting.* (SEEA Ecosystem Accounting, par 8.2)

Maintaining consistency with the SNA provides the advantage of having a benchmark for analysis. This adds to the usefulness of aggregates generated using the SEEA framework.

2.A.1 Valuation for ecosystem accounting

The SEEA Ecosystem Accounting manual recommends a range of techniques that can be employed for the estimation of the monetary value of these flows. These methods include: the use of prices from similar markets, residual value approach, productivity change method, hedonic regression, replacement cost method, travel cost method, avoided damage cost method, and simulated exchange value method (see United Nations (2014), Atkinson and Obst (2017)). We will not discuss the details of each approach in this paper. However, our goal is to discuss the context in which some of these valuation methods are employed and reflect on their possible applications for the measurement of the digital economy.

SEEA dismisses the role of contingent valuation and stated preference in providing a sound valuation methodology for environmental accounting. It argues that these approaches would incorporate consumer surplus into the estimated value of ecosystem services. Since it is the intention of environmental accounting to maintain consistency with the valuation principle of the National Accounts, SEEA suggests any valuation techniques that may incorporate consumer surplus to the value of ecosystem services should not be considered without appropriate adjustments and a validation.

In a report commissioned by the World Bank, Atkinson and Obst (2017) discusses three channels in which ecosystem services benefit households. They argue, that choice of valuation techniques should depend on these channels. In this section, we describe these three channels, provide some examples, and enumerate the recommended valuation method for each of the channels.

ES#1: As input to production. Here, we can think of ecosystem services as intermediate inputs to the production of market goods and services. For this channel, it is assumed that the value of ecosystem services is embedded in the value of market goods. For instance, natural pollination is required for some agricultural activities. The goal of valuation, in this case, is identifying the contribution of ecosystem inputs to the value of market goods.

Recommended valuation methods: Production function approach, change in productivity.

ES#2: As substitute or a complement for the market goods. Here, ecosystem services can be inferred from related market goods, which can either be a substitute or a complement. For instance, the value of the ecosystem services provided by the beach is complementary to travel expenses. For this example, the idea is that market goods are combined with ecosystem services to produce another product. Another example is when market goods can substitute for ecosystem services. For instance, the value of flood control systems can be used to partially infer the value of mangroves (or at least the flood prevention function of mangroves). For this channel, revealed preferences approaches—such as hedonic regression and travel cost approach—are used for valuation.

Recommended valuation methods: Price of similar products, travel cost method, hedonic regression.

ES#3: As a direct contributor to household utility. Here, ecosystem services are directly consumed by the household. The value of ecosystem services is distinct from market activity.

Recommended valuation methods: Stated preferences, contingent valuation, simulated exchange value.

While the SEEA is cautious about the use of stated preferences for the valuation of ecosystem services, Atkinson and Obst (2017) noted that these methods can be used as a starting point for constructing estimates close to exchange value. They argue that demand curves can be estimated using stated preferences, which is the initial step for simulating a market for ecosystem services.

The main point of Atkinson and Obst (2017) is that context matters for valuation. They argue that the way ecosystem services should be valued depends on the three channels listed above.

2.A.2 Parallels with the digital economy

In this section, we discuss how the taxonomy provided by Atkinson and Obst (2017) can be used in the context of the digital economy. One can find parallels to the challenge faced by ecosystem accounting to those encountered by economists attempting to measure the value of free digital services. First, the goal of both measurement exercises is to develop a set of methodologies, which can provide a shadow price for non-market activities. Second, maintaining consistency with the valuation principles of the SNA will add to the usefulness of the estimates. Users of the data would be able to compare the value of activities from the digital economy with other activities from the broader market economy.

As with ecosystem services, one can argue free digital services can also benefit households through the three channels defined by Atkinson and Obst (2017) and so measured in a similar way. In table 2.A.1, we outline the different channels in which free digital services can have an impact on households. we also provide some examples and possible methods for the estimation of the monetary value of the service.

As with ecosystem services, free digital products can also benefit households as inputs to production (ES#1). Productivity tools such as Google Docs and Google Sheets are being used by many professionals in their daily work activities. Taxi drivers have started using Google Maps to allow them to get to their destinations much faster. For this channel, we can assume that the value of digital services in embedded in the price of market goods and the role of valuation is to estimate the contribution of these services to the total value of the market product.

The value of free digital products can also be inferred from the value of related market goods (ES#2). This is more apparent for the case of substitutes. we will discuss this in detail in the following section.

Table 2.A.1

Channels	**Examples**	**Possible valuation methods**
ES#1 As inputs to production.	Googles maps as an input to transportation services like Taxis, Google Scholar for researchers, Google Docs, Google Sheets for many industries.	Production function; change in productivity
ES#2: As substitute or a complement for the market goods.	Amazon as a trading platform, ride-hailing apps such as Uber, food deliver apps, and online banking, free versions of goods with premium services.	Market substitutes, hedonic regression
ES#3: As a direct contributor to household utility.	Online maps, social media, streaming sites.	Stated preferences and simulated exchange value method

Lastly, some digital services are distinct from the value of market products (ES#3). Social media is a prime example. It is difficult to think of any market product that functions similarly to platforms such as Facebook or Twitter. We can say the same for other forms of digital services such as search engines, review websites (Yelp, etc), and online maps. Since these services are detached from market transactions, this makes the estimation exercise more challenging. As such, perhaps the best way to arrive at the value of these flows would be to ask households directly how much they value these flows through stated preferences or contingent valuation surveys and experiments. Results from stated preferences estimate demand curves that can be used to arrive at simulated exchange value. Though this would also involve making some assumptions on the institutional arrangement for the provision of the digital service.

For this category of services, it may be necessary to employ stated preference methods for the estimation of the value of these flows. One possible way to elicit Willingness to Pay is by asking survey respondents:

"If Meta, the company that owns and operates Facebook, plans to discontinue offering the social networking site for free, what is the acceptable monthly subscription price that you are willing to pay in order to gain access to Facebook as it is now?"

While WTP is not exactly equivalent to exchange value, it can be a starting point for the generation of simulated exchange values. This approach requires information from a demand curve (which can be generated from a stated preference study) and some assumptions on the supply curve to estimate the exchange value of goods and services. Caparrós et al. (2017) demonstrates the use of this approach for the estimation of ecosystem services for certain leisure parks in Italy (see. table 2.A.1). The approach was also recommended by Atkinson and Obst (2017) and SEEA Ecosystems Accounting manual (see United Nations (2014)) for the estimation of the value of ecosystem service flows.

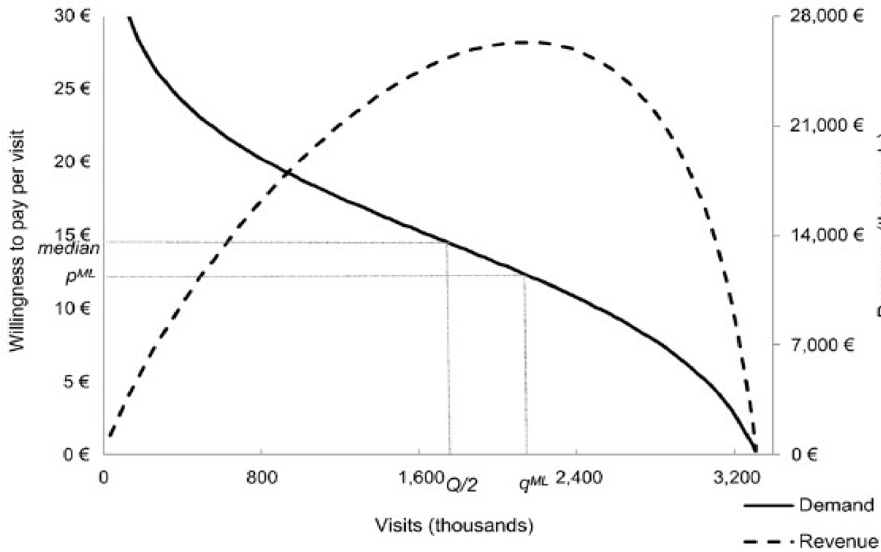


Figure 2.A.1: Simulated exchange value for recreational service by ecosystems in Cazorla

Note: The figure shows the SEV estimates Caparrós et al. (2017) for select nature sites in Cazorla, Italy.

Since surveys and online experiments are expensive endeavours, the use of stated preferences can be limited to digital products that are distinct from market activities. As such, their value cannot be inferred from other market activities. However, this provides a powerful tool for measuring the value of free digital services.

A challenging aspect for free digital services, however, is that some services can cut across channels. In particular, many products that provide utility directly to households (ES#2 and ES#3) can also be used in the production of market goods (ES#1). For instance, while Google Docs is often used for work, the service can also be used for keeping personal notes. Google maps can be used by Taxi drivers, but it can also be used by tourists and commuters. This makes it a bit challenging in terms of identifying the proper valuation method for certain products. As such, stated preference studies which aims to estimate the value of free digital services should take this into consideration in the design of their measurement instrument. It must be clear to the respondents that the goal of the survey (or experiment) is to elicit the value they derive from the *personal* use of these service. Measuring the value of free digital services used in the production of market goods should be a separate endeavour. For such exercise, time-use surveys can be employed to estimate the working hours spent using free digital services.

2.B Price setting for freemium platforms

Consider a platform offering services through the freemium model. The platform would have three sides: 1) side I is defined as the premium or paid users; side J as free users, and side K as advertisers. The platform maximizes its profits by choosing the network size of one side:

$$\begin{aligned}\pi(N^I, N^J, N^K) = (P^I(N^I, N^J, N^K) - C^I)N^I + (P^J(N^I, N^J, N^K) - C^J)N^J + \\ (P^K(N^I, N^J, N^K) - C^K)N^K - cN^I N^J N^K\end{aligned} \tag{2.13}$$

where C^I, C^J, C^K are the marginal costs of providing the service for sides I, J, and K respectively, and $cN^I N^J N^K$ is the fixed cost of connecting both sides of the platform.

Premium users

The first order condition of the platform's profit-maximizing problem is:

$$\frac{\partial \pi}{\partial N^I} = P^I + P_I^I N^I - C^I + P_I^J N^J + P_I^K N^K + cN^J N^K = 0$$

Re-arranging yield:

$$P^I = C^I + cN^J N^K - P_I^I N^I - (P_I^J N^J + P_I^K N^K) \tag{2.14}$$

Similar to two-sided platforms, the first three terms represent the firm's marginal cost less the inverse hazard function rate of demand (or market power) $\mu^I = -P_I^I N^I = P_I^I/\epsilon^I$ where ϵ^I is the elasticity of demand (see Weyl (2010)). The final two terms represent the additional revenues extracted from side J from an additional user of side I and the additional revenues extracted from side K from an additional user of side I. Weyl (2010) describes this as the external benefits from the provision of service on side I. Prices for side I can be expressed as:

$$P^I = \underbrace{C^I + cN^J N^K}_{\text{marginal cost}} + \underbrace{\mu^I}_{\text{market power}} - \underbrace{(P_I^J N^J + P_I^K N^K)}_{\text{external benefits}}. \tag{2.15}$$

Users of the free version

The first order condition of the platform's profit-maximizing problem is:

$$\frac{\partial \pi}{\partial N^J} = P^J + P^J_J N^J - C^J + P^I_J N^J + P^K_J N^K + cN^I N^K = 0$$

Re-arranging yield:

$$P^J = \underbrace{C^J + cN^I N^K}_{\text{marginal cost}} + \underbrace{\mu^J}_{\text{market power}} - \underbrace{(P^I_J N^I + P^K_J N^K)}_{\text{external benefits}}. \tag{2.16}$$

For free users, platform pricing is similar to those for newspapers. The external benefits to advertisers, represented by $P^K_J N$ are larger than the combination of marginal cost and the possible market power distortion the platform could extract. Moreover, the additional revenues platforms extract from the side of paying users from the presence of free users represented by $P^I_J N$ could also lower the price. We can think of this as network benefits. A good example is Zoom, where the wide usage of Zoom allows the company to sell paid versions to some users. Because of the combination of these factors, the market price levied to users of side J is zero.

Advertisers

The first order condition of the platform's profit-maximizing problem is:

$$\frac{\partial \pi}{\partial N^K} = P^K + P^K_K N^K - C^K + P^I_K N^I + P^J_K N^J + cN^I N^J = 0$$

Re-arranging yield:

$$P^K = \underbrace{C^K + cN^I N^J}_{\text{marginal cost}} + \underbrace{\mu^K}_{\text{market power}} - \underbrace{(P^I_K N^I + P^J_J N^K)}_{\text{external benefits}}. \tag{2.17}$$

The external benefits to free users in side J are negative. This marks up the price experienced by advertisers, by the unit amount $P^J_K N^J$.

2.C Additional figures

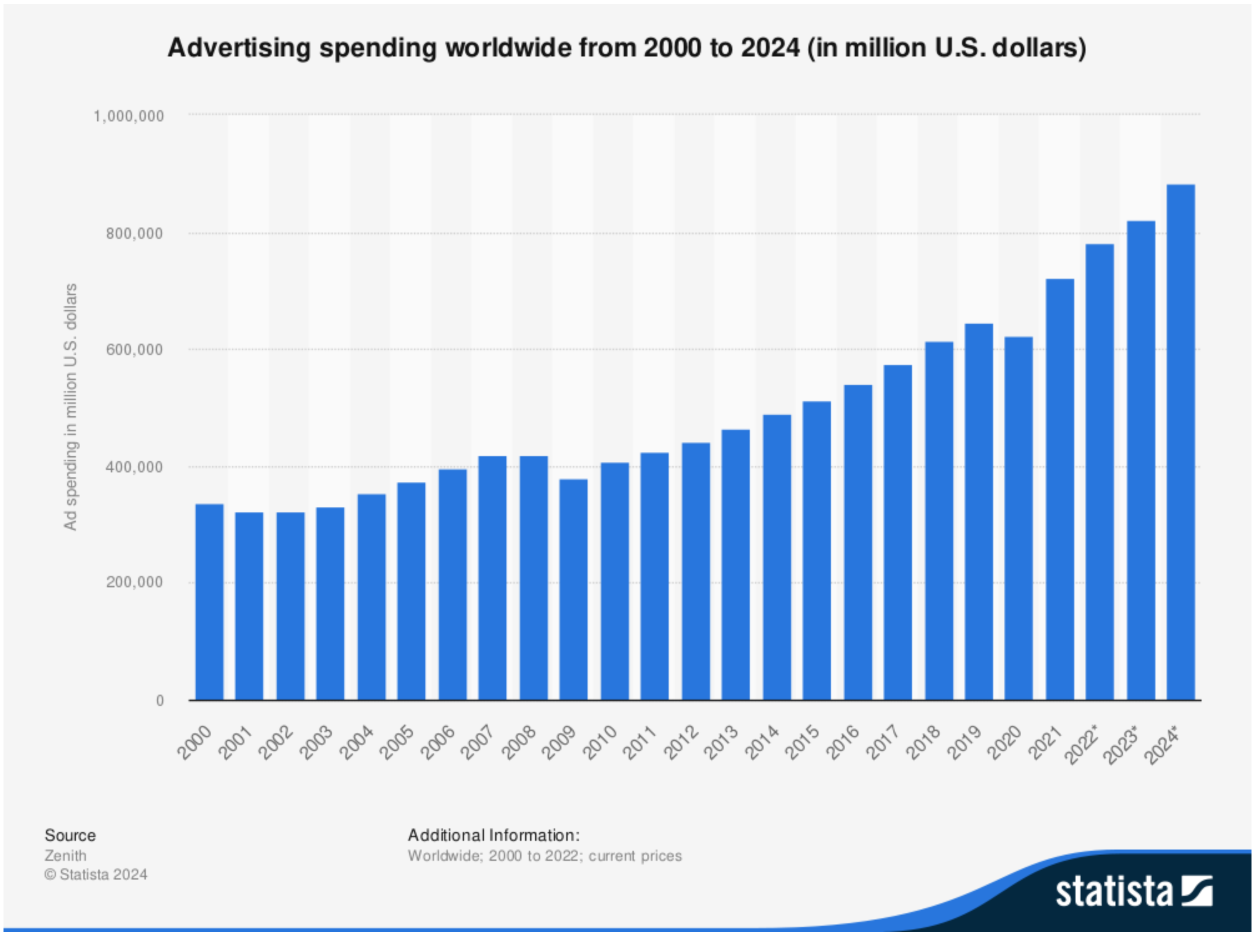


Figure 2.C.1: Global advertising expenditures

Source: Statista.com, accessed 8 April 2024.

Note: Data for the years 2022 to 2024 are based on forecasts.

CHAPTER 3: How much is Videoconferencing Worth? Measuring the Value of Free Digital Services

ABSTRACT

The goal of this study is to estimate and examine the value derived by households from the utilization of free digital services. For this exercise, we estimate value from the final consumption of three forms of free digital services: videoconferencing, personal email, and online news. As our measurement strategy, we employ the prices of "premium" or paid internet services as a proxy for the value of their free counterparts. We also use hedonic regression in order to extract the value of the 'free component' of these products and untangle them from the value of the premium-exclusive components. Our final estimates show that in 2020, the aggregate gross value derived by households from the consumption of the three digital services was between £7.0 billion and £25.4 billion, which is 0.57 to 2 percent of household final consumption expenditures (HFCE). We also observe that the value derived by households from consuming these products is growing much faster than HFCE. Our estimates show that in 2020, the initial year of the COVID pandemic, real household final consumption decline would have been 0.07 to 0.13 percentage points slower had the value of the three digital services been incorporated in the estimates.

3.1 Introduction

While free digital services such as videoconferencing, personal email, and online news have profoundly impacted people's lives, their welfare contributions are not explicitly reflected in official statistics. Existing frameworks for the compilation of macroeconomic aggregates are mostly concerned with the estimation of economic activity with explicit market value[1]. The National Income Accounts present the value of goods and services at market prices. With free digital products, however, it is possible for households to derive utility by using online services that they do not pay for. In this instance, the increase in household utility would not have a corresponding entry in either the production or expenditure side of the National Accounts. As Hulten and Nakamura (2017) put it, "[a]n important implication is that a general increase in the availability of information can increase consumer utility without increasing GDP."

[1] The exceptions being the estimation of the value of ownership of dwellings, own-account production, own-account production of goods, and government services.

Moreover, the substitution between free digital products and traditional market products causes existing estimates of national output to become a misleading indicator of welfare (see Coyle (2019)). A slowdown in GDP growth could be a result of households spending more time on free internet activities rather than market activities (i.e. using Google Maps rather than buying an actual map from the store). This makes it difficult to assess how technological innovations have improved people's lives.

The goal of this research is to estimate and examine the value derived from free digital goods[2] in the context of national income accounting. For this exercise, we estimate the gross value from the consumption of three forms of free digital services: videoconferencing, personal email, and online news. As our measurement strategy, we apply market-equivalent pricing for the valuation of free digital services. In particular, we employ the price of premium versions of digital services as proxies for their free counterparts. For instance, we use the price of paid versions of Zoom as a source of valuation for its free version. We also use hedonic regression in order to extract the value of the free component from these products and untangle them from the value of the premium-exclusive component. Hedonic regression is an econometric approach wherein the price of a good is expressed as a function of its characteristics[3], with the goal of estimating the price, or willingness to pay, for the set of "characteristics" included in the specification.

Our final estimates show that in 2020, the aggregate gross value derived by households from the consumption of the three forms of digital services was between £7 billion and £25.4 billion. This is around s 0.57 to 2 percent of the UK's household final consumption. We also observe that the value derived by households from consuming these products is growing much faster than aggregate household consumption. Our estimates show that in 2020, the initial year of the COVID pandemic, the real household final consumption decline would have been 0.07 to 0.13 percentage points slower had the value of the three digital products been incorporated in the estimates. This tells us that the availability of free internet services was partially able to reduce welfare loss as a result of the lockdown.

Whether GDP can be considered a measure of welfare is a hotly debated topic. In the simplest sense, GDP is regarded as a measure of production, expenditures, and income, but not necessarily welfare. Scholars from the other side of the argument assert that while GDP is not exactly a measure of welfare if output is measured correctly, the application of price deflators transforms GDP into a volume index that represents changes in aggregate utility over time (see Coyle (2015) and Dynan and Sheiner (2018)). For this essay, we do not try to make any normative assertions about whether GDP *should* be used to measure well-being. We understand the limitations of GDP as a welfare measure. Rather, we enter the conversation saying that *if we want GDP to represent welfare better, this is one way to do so.* Moreover, we do not advocate that GDP should be replaced as an official statistic. In line with previous studies (see a discussion by Heys et al. (2019) on expanded welfare measures beyond GDP), we aim to generate a separate set of statistics that complements

[2]While it can be argued that digital products are services rather than goods, for the purposes of this research, we use the terms goods and services interchangeably.

[3]In this context, characteristics are features that describe the good. For cell phones, they can be RAM, storage space, camera quality, etc.

GDP, the same way satellite accounts do.

We also would like to clarify that in the context of this research, *value* relates to the *value of final consumption* as defined in the SNA. Implicitly, *value* would reflect the product of a volume (or quantity) measure and its unit price. In National Accounts terminologies, this is referred to as accounting value. The unit price would also reflect *exchange value* rather than *use value*. Effectively, our definition of value is not the same as consumer surplus or other welfare measures. The reason why we define value this way is to allow for consistency with other National Accounts aggregates. This improves the usefulness of the estimates as we are able to benchmark our figures to the value of other products reported under household final consumption expenditure. In the context of productivity, consistency with the SNA also allows us to compare our estimates with the value of other *produced goods and services* in the economy.

Our estimation methodology is also designed to capture service flows and not the value of the assets used to produce these services. Again, this decision was made to maintain consistency with National Accounts aggregates that we want to benchmark against, such as GDP and household final consumption, all of which represent flows. As such, the value of digital products is not necessarily linked to the company valuation of their service providers (i.e. Google, Meta, etc.).

This study contributes to the growing empirical literature that aims to quantify the economic contribution of free digital products. As of the time of writing of this manuscript, there is no consensus on how to estimate the value of free services. Empirical works on the valuation of free digital services can be classified under two main categories, depending on the approach they take. As discussed in the previous chapter, these are: 1) those involving the contingent valuation approach and 2) those employing the total cost approach.

This study expands the literature in three ways. First, to the best of our knowledge, this is the first study that employs prices of similar market products as a source of valuation for free digital products. While market equivalent pricing has been applied by compilers of the National Accounts to estimate the implicit price of non-market products and non-monetary transactions such as imputed rentals, own-account production of goods by households, and barter transactions, this approach has never been used in the context of free internet products. By doing so, we overcome some of the limitations of both the total cost approach and the contingent valuation studies. Unlike the total cost approach, the aggregate we generate is a function of the number of users of the good, thus, by construction, it guarantees that the value derived by an additional user is reflected in the estimates. Moreover, the use of market prices for similar products is a common approach to valuation in the National Accounts. As such, valuation based on this strategy is fully consistent with the SNA's accounting framework.

Our second contribution is, to our knowledge, the first study to focus on valuing digital services distributed under the "Freemium" model. While the total cost approach aligns well with the core principles of the National Accounts, its application has been limited to free media where production costs can be attributed to advertising and marketing. However, services using the freemium model are sometimes not funded by advertising and marketing, leading to potential underestimation of their output when using the total cost approach.

Our third contribution is our estimation of the aggregate value of free digital services in the UK and its impact on the British economy. To our knowledge, studies that estimate the total value of free digital services are focused only on the US (i.e. Nakamura et al. (2017), Brynjolfsson et al. (2020), and Jamison and Wang (2021)). While the survey of Coyle and Nguyen (2023) was conducted in the UK, they did not estimate the aggregate value of free digital services and their contribution to the UK economy. Our study would be the first to provide insight in this area.

The outline of this essay is as follows. In section 3.2, we discuss and synthesize the empirical literature on the measurement of free internet services. In section 3.3, we detail our estimation strategy. In section 3.4, we show and describe our data. We then discuss our results and preliminary estimates in section 3.5. We end this essay with concluding remarks and our strategies moving forward.

3.2 Empirical Literature

Various studies have been conducted attempting to estimate the economic contribution of free digital services to the economy. We classify these studies into two main categories, depending on the approach they take. These are 1) those involving the contingent valuation approach and 2) those employing the total cost approach. We will discuss each approach and provide a synthesis later on in this essay.

3.2.1 Contingent Valuation Approach

The contingent valuation approach is designed to determine how much *individuals* value free digital services. Since the digital services that these researchers are attempting to value are already free, they are unable to ask them how much they are willing to pay (WTP) in order to gain access to those products. Instead, they try to acquire information on how much compensation individuals are willing to receive for abstaining from these services. This approach is intended to capture the respondents' willingness to accept (WTA), which in theory should be equal to the WTP if close substitutes are available (see Hanemann (1991))

Corrigan et al. (2018) conducted auctions to determine their respondents' WTA for abstaining from the use of the social media website Facebook. The average bid that their auctions generated varied, depending on how long participants were required to deactivate their accounts in order to receive compensation and the population that the respondents belonged to. All groups required at least an annualized WTA of $1,000 to give up the said social network. The students' cohort reported the largest annualized mean WTA at $2,076.

Similarly, Brynjolfsson et al. (2019b,a) conducted incentive-compatible discrete choice experiments in two separate laboratories to determine the value derived by individuals from free digital products. The goal of their exercises was to generate an augmented version of GDP, one that incorporates the benefits of free products. Their first experiment aimed to estimate the contribution of Facebook to the US economy. Participants were asked to either 1) keep their Facebook account, or 2) give up Facebook for a month and get paid $\$E$[4]. To estimate the demand curve for Facebook, they fitted a logistic regression model with the respondent's decision to keep or give up the social media site as an outcome variable and the monetary value (in log scale) as the predictor variable.

The most recent results found that the median WTA of giving up Facebook was about $42.17 a month in 2017. They considered the intercept for the demand curve they fitted as Facebook's reservation price ($2,152) and proceeded to the measurement of the contribution of the social media site to the US economy or welfare. The authors estimate that Facebook contributed an equivalent of $231 billion from 2003 to 2017, or $16 billion a year. They also estimate that US GDP growth would have been 0.11 percentage points faster had the welfare gains from Facebook been accounted for in the estimates.

Brynjolfsson et al. (2019a) also generated estimates of the value of other free products. In a university in the Netherlands, they employed the same methodology to test the valuation of Instagram, Snapchat, Skype, WhatsApp, digital Maps, LinkedIn, Twitter, as well as Facebook. Overall, they were able to analyze the responses from 426 participants. They found that the median WTA they got from the participants in the Netherlands is twice as large as those from the US ($100). The authors estimate the annual percentage contribution of these products to welfare growth would have been as follow: 0.82 percentage points for Whatsapp, 0.11 percentage points for Facebook, 0.07 percentage points for digital maps, and 0.01 percentage points for Instagram.

Jamison and Wang (2021) employed the same approach in the US following the Coronavirus Disease (COVID-19) pandemic to arrive at WTAs for various online services, namely internet search, email, maps, video, e-commerce, social media, music, instant messaging, and video conferencing[5]. Similar to the Brynjolfsson et al. (2019a) study, their experiment involves asking their respondents whether they are willing to abstain from the use of the said internet services for an amount X, where X is a randomly generated price point. They also employed a logistic regression model to estimate the demand curves for each good. For 2020, the highest mean valuation they arrived at was from email services ($2,095) and the lowest was from Zoom ($44.93). They also found that the mean WTA for all internet services covered by their study increased following the pandemic.

Coyle and Nguyen (2023) conducted an online survey to estimate the WTA of online products in the United Kingdom. Employing a YouGov online panel, the researchers asked their participants how much they were willing to be compensated in order to give up a variety of products, which includes both traditional products and free services. The services that were identified in the survey include email, search engines, online banking, online maps, radio, TV, traditional and online news, streaming services, and social media, among others. In terms of comprehensiveness, the study covers far more products than the study of Brynjolfsson

[4]They randomly assigned participants to discrete price points: E = (1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000).

[5]With the exception of Zoom, their research design did not involve asking individuals to abstain from using specific service providers. Rather, they classified these services into ad-hoc groups of internet activities.

et al. (2019a) and Jamison and Wang (2021). However, none of the participants were actually compensated for giving up the services mentioned in the survey and random selection was not considered. Moreover, there was no randomization procedure involved in the data acquisition.

Coyle and Nguyen (2023)'s survey was conducted in two rounds, one in February 2020 and another in May 2020. The first round had a sample size of 10,500 adults while the second round had a smaller sample size of 1,600 adults. Coincidentally, the gap between the two rounds coincided with the national lockdown in the UK. They found that following the lockdown, the value attached by individuals to free products such as Facebook and WhatsApp generally increased.

Coyle and Nguyen (2023) constructed demand curves from the percentage of their participants who were willing to give up the respective services for the given compensation level. The authors did not, however, construct measures of aggregate monetary values of welfare from free services.

Table 3.2.1 compares the WTA derived by the different studies employing the contingent valuation approach. It can be noticed the WTA values from the Coyle and Nguyen (2023) are substantially larger than other estimates in this literature.

Table 3.2.1: Comparison of WTA estimates for select digital products

	Brynjolfsson et al. (2019b)	Corrigan et al.	Brynjolfsson et al. (2019a)		Coyle and Nguyen (2023)		Jamison and Wang
	(2017)	(2018)	US	Netherlands	May (Mean)	May (Median)	(2021)
Facebook	–	155.3	42.2	77.8	78	6.6-13	–
Instagram	–	–	–	5.1	36.9	0-0	–
LinkdIn	–	-	–	1.3	10.1	0-0	–
Twitter	–	–	–	<1	20.9	0-0	–
Snapchat	–		–	1.6	16.6	0-0	–
Social Media	26.8	–	–	–	–	–	140.3
Whatsapp	–	–	–	384.1	103.8	13-32.5	–
Messenger	–	–	–	–	64.1	0.1-6.5	–
Instant Messengers	12.9	–	–	–	–	–	310.7
Maps	304	–	–	47.6	57.7	6.6-13	1,157.6
Skype	–	–		0.2	16.3	0-0	–
Zoom	–	–		–	–	–	44.9
Videoconferencing	–	–	–	–	–	–	337.5
Search engines	1,460.8	–	–	–	180.2	65-129.8	8,703.3
Personal email	701.2	–	–	–	192	129.9-324.5	2,095.7

Note: Table compares the estimated monthly WTA by Brynjolfsson et al. (2019b), Corrigan et al. (2018), Brynjolfsson et al. (2019a), Coyle and Nguyen (2023) and Jamison and Wang (2021). Figures are in USD. For comparability, WTA estimates by Brynjolfsson et al. (2019a) and Coyle and Nguyen (2023) were converted using their respective yearly average exchange rates during the period of data gathering. WTA estimates were acquired at different time periods. In particular, data by Brynjolfsson et al. (2019a) for the US (Facebook) were acquired in 2017; data from the Netherlands were acquired at different dates: January 2009 for Facebook, February 2004 for Whatsapp, October 2010 for Instagram, May 2003 for LindkIn, August 2003 for Skype, and March 2006 for Twitter. No inflation adjustments were applied to the data. Annual data from Brynjolfsson et al. (2019b) and Coyle and Nguyen (2023) were divided by 12 to generate an approximation of the monthly value.

3.2.2 Total Cost Approach

The total cost approach employs the cost of producing free services to represent the value consumers derive from them. Many services that are freely available on the Internet are financed either through advertising or marketing expenses. For instance, YouTube is largely financed by advertisers. In the National Accounts, these expenditures are recorded as part of intermediate consumption. In the total cost approach, these expenditures would be recorded under Final Consumption Expenditure to reflect the welfare gained by households from consuming these services.

Soloveichik (2015) argued that the provision of free digital services is a product of a barter transaction between advertisers and internet users. She states that households are producers of data and, as unincorporated enterprises, they sell their viewership to advertisers through a barter transaction. Advertisers finance the production of the free digital services that households consume. Using this concept, she developed experimental estimates of US GDP, which considers advertising expenditures as part of household consumption. Because of the decline in advertising spending since 2000, the estimated GDP growth for the experimental estimates was smaller by 0.001 percentage points compared to the original estimates. Nakamura and Soloveichik (2015) extended this measurement strategy to include other countries. They found that globally, advertising-supported media accounts for less than 0.5 percent of GDP. Their results also show the global GDP growth would be faster by 0.019 percentage points per year, had advertising-supported media been included as part of household consumption.

The full implementation of the above concept was executed by Nakamura et al. (2017), where they also imputed the value of the viewership households sell to advertisers. They also included free services that were financed through marketing expenditures in their estimates. These expenditures include corporate spending on content and other promotional material that are not part of advertising. Expenditures on free mobile apps fall into this category. Their estimates show that GDP growth estimates for the US would be faster by 1.53 percentage points in the period 2005 to 2015.

Van Elp and Mushkudiani (2019) of Statistics Netherlands applied the same principles to estimate the contribution of free services to the Dutch economy. They used advertising expenditures to represent the value households derived from free digital services. They found that free digital services would account for 1 to 3.4 percent of the Dutch GDP and 2.3 to 7.8 percent of its household final consumption. In a presentation to the Economic Statistics Centre of Excellence 2021 conference, Heys and Taylor (2021) noted that the Office of National Statistics is also attempting to employ this approach to measure the contribution of free internet platforms to the UK economy.

In a recent paper, Van Elp et al. (2022) introduced the concept "final consumption by business", which incorporates the free services provided by firms to households, as part of the firms' marketing strategy. If included as part of final consumption, these services would be around 3.0 to 4.7 percent of GDP in 2019. The inclusion of these services would also cause year-on-year GDP growth to be faster by 0.3 to 0.5 percentage points.

3.2.3 Synthesis of related works

While there have been a number of attempts to measure the value of free digital services, there is still no consensus as to how their economic impacts should be measured. Quantifying the degree to which these services are having an impact on welfare and productivity is becoming increasingly relevant, especially during the recent pandemic when many countries enforced lockdown measures to contain the virus, and much of the world's population was forced to work from home.

The goal for most of the empirical studies is to generate a statistical aggregate representing the value derived by households from the consumption of free digital services. This is analogous to household final consumption in the expenditure side of the National Accounts. Moreover, most of these studies (i.e. Nakamura et al. (2017), Brynjolfsson et al. (2019a), Jamison and Wang (2021), and Van Elp and Mushkudiani (2019)) constructed augmented GDP statistics, those that incorporated the value of free digital services.

The main advantage of the total cost approach is that it requires little change to the core accounting principles of the SNA. Measuring non-market outputs in terms of the total cost of producing them follows the practice for other non-market goods and services that are currently being recorded as part of GDP. This includes output from governments and non-profit institutions serving households (NPISH). However, they also suffer the same disadvantages, in that they have a limited ability to reflect welfare changes[6] (see Bean (2016)).

If the goal of developing an augmented set of accounts is to estimate the welfare gains from free services, then the total cost approach would be lacking for such an endeavor. For many digital services, the marginal cost of production for every unit of consumption is close to (if not equal to) zero. Welfare gains from every increment of usage would be likely to be understated by this approach.

Nakamura et al. (2017) noted this in their paper saying: "[W]e do not capture a welfare measure of the value of Google Maps [and other free gods], but only measure the cost of providing it. This could be viewed as an underestimate of the contribution of this 'free' content to output and productivity—but it is consistent with the standard national accounting methodologies for estimating industry output and input."

An alternative to this is the use of contingent valuation or stated preferences approach to estimate the welfare benefits of free digital services. A growing number of studies are employing contingent valuation techniques to estimate the value individuals derive from free services. Since valuation is based on the individual, it should be easy to generate a measure of aggregate value to consumers by multiplying the value per user by the number of users. The advantage of this approach is that it captures the incremental level of welfare received from an additional user of the service in the aggregate. The main disadvantage of this approach, however, is that it would introduce inconsistencies with the core accounting principles of the SNA if aggregated with estimates of national output.

The valuation of goods and services in the SNA is based on exchange value. Paragraph

[6]This is because output = input.

3.118 of the 2008 SNA writes:

> *The power of the SNA as an analytical tool stems largely from its ability to link numerous, very varied economic phenomena by expressing them in a single accounting unit. The SNA does not attempt to determine the utility of the flows and stocks that come within its scope. Rather, it measures the current exchange value of the entries in the accounts in money terms, that is, the values at which goods, services, labor or assets are in fact exchanged or else could be exchanged for cash (currency or transferable deposits).* (see United Nations and others (2009))

Exchange value is a product of the intersection between the supply curve and the demand curve, which is the market's WTP at each level of quantity demanded. Standard economic theory predicts that with small income effects WTA and WTP should be equivalent or at least, close to each other (see Willig (1976) and Randall and Stoll (1980)). However, Hanemann (1991) showed in an analytic framework that the values are not equal, with WTA being greater than WTP, for goods with little to no close substitutes. This was corroborated by a randomized control trial by Shogren et al. (1994).

Nordhaus (2006) also cautions about the use of subjective measures such as contingent valuation. He argued that valuing non-market activities based on this approach will introduce inconsistencies in an expanded set of accounts He writes:

> *National-income accountants generally prefer valuation techniques that have an objective behavioral component, whether in market prices or individual actions. Valuation techniques that are largely subjective and based only on survey information alone—such as contingent valuation—are difficult to validate and should be avoided where possible, but may be needed in some areas.*

Moreover, considering how integral free digital services are to people's lives, it is not surprising that the WTA of these services is substantially high. To put this in perspective, consider the case of electricity. The final consumption of electricity is recorded in GDP in terms of its volume (kilowatt hour consumption) multiplied by its price ($ per kilowatt hour). If you ask individuals how much they are willing to be compensated to give up electricity for a month, people would likely provide values that are higher than the amount they pay for electricity in a given month.

This is also reflected in the findings of Coyle and Nguyen (2023), where they also asked their respondents for their WTA for traditional products. The authors found that (for May 2020) the mean WTA for paid goods and services like printed newspapers (£430), Cinema (£589), and Netfix (£1,373) are substantially higher than their market price, which is the valuation used in the SNA. As such, one can argue that aggregates generated from contingent valuation may not be truly consistent with GDP and other aggregates compiled using the same accounting framework. In a way, estimates from this approach may reflect the level of welfare individuals receive from having access to these services, but they are not necessarily comparable with estimates of final consumption of the product, as measured by National Accounts[7].

While it can be argued that the value of free digital services can be recorded in a satellite account (as recommended by Schreyer (2021)), even satellite accounts attempt to preserve the core accounting principles of the SNA. For instance, the System of Environmental-Economic Accounting (SEEA), the international standard for the compilation of satellite accounts for the environment, recommends the use of valuation based on exchange (see paragraph 9.22 of United Nations (2012)). Other methods–such as direct surveys and binary choice experiments–are not recommended by the SEEA without validation or some form of adjustment (see paragraph 9.24 of United Nations (2012)). The reason for this restriction is to maintain the internal consistency of the account with the core national account estimates. If the goal is to measure the contribution of non-market output, such as ecosystem services, to total human activity, it is necessary that the accounting principle for non-market transactions should be consistent (or at least similar) to the accounting principles applied to market transactions in order to ensure comparability.

Schreyer (2021) argues that estimates from these approaches can be interpreted as part of household production of services. We can think of this as similar to the value of childcare done by members of the family or the preparation of home-cooked meals. For these exercises, respondents value the time they spent engaging in household leisure enabled by digital services. As such, valuations derived from their approach do not reflect the value of the services themselves but the time spent on leisure activities. Digital services, in this context, are inputs to the production of leisure services, in the same way grocery items are inputs to the preparation of meals.

While estimates from contingent valuation studies provide valuable insight into welfare and the value of home production, this approach is limited in terms of providing estimates comparable to the value of other services to the National Accounts aggregates[8]. Moreover, Schreyer (2021) argues that the median WTA likely reflects the value added from leisure services. A complete set of household accounts on digital leisure would require the value of intermediate inputs as well to come up with gross output.

What is missing from the literature is an estimation strategy that captures the welfare gains from the consumption of free services, and is consistent with the accounting framework of the SNA. We address this gap in the literature by employing the price of premium versions of free goods as a source of valuation for their free counterparts.

[7]As an example, aggregate welfare estimates from this approach cannot be compared with estimates the gross value from the consumption of hotels and restaurants, as reflected in household final consumption expenditures.

[8]Estimates from this approach are more comparable to estimates of home production in the Household Satellite Accounts

3.3 Estimation Strategy

In this study, we apply market-equivalent pricing for the valuation of free digital services. In particular, we employ the price of premium versions of digital services as proxies for their free counterparts. This strategy is not new for the valuation of non-market products and non-monetary transactions. The SNA suggests the use of prices of products from similar markets as a source of valuation when prices cannot be explicitly observed. Paragraph 3.123 of the 2008 SNA states:

> *When market prices for transactions are not observable, valuation according to market-price-equivalents provides an approximation to market prices. In such cases, market prices of the same or similar items when such prices exist will provide a good basis for applying the principle of market prices. Generally, market prices should be taken from the markets where the same or similar items are traded currently in sufficient numbers and in similar circumstances. If there is no appropriate market in which a particular good or service is currently traded, the valuation of a transaction involving that good or service may be derived from the market prices of similar goods and services by making adjustments for quality and other differences.* (see United Nations and others (2009))

Compilers of National Accounts statistics often use the market price of similar products to impute the value of certain non-market goods and services and non-monetary transactions, such as services from owner-occupied housing, extraction of groundwater, agricultural products for own consumption, barter transactions, remuneration in kind, among others. Compilers of the Household Satellite Accounts also use this strategy to value household services such as childcare. As discussed in the previous Chapter, the use of price of similar market goods is grounded on the assumption that these products provide the same benefit as the free digital product.

Subscribers to premium services would have access to the services provided by the free version with the addition of other features exclusive to the premium version. One can argue that the price of the premium versions p_p would have two components: a 'freely-available' p_f and a premium component p_z. If the relationship of the two components is additive, the price of premium services can be expressed as,

$$\underbrace{p_p}_{\text{price of premium service}} = \underbrace{p_f}_{\text{'freely-available' component}} + \underbrace{p_z}_{\text{premium-exclusive component}} . \tag{3.1}$$

The component 'free component' can be interpreted in two ways. From the producer's perspective, the free component would represent the cost of producing services that are also available for free, if one chooses to consume it separately from the bundle of premium-exclusive services. Meanwhile, from the perspective of consumers, this represents the value derived by households from the consumption of the services that they can also acquire through the free version of that good. It reflects the household's WTP for the services that are likewise present in the free version.

There are certain advantages to valuing free services this way. First, it avoids the problem of inconsistencies with the measurement principles of the National Accounts. This problem is typically encountered by contingent valuation studies, where they utilize WTA as a proxy for WTP. Moreover, this approach would only produce values no greater than how much consumers would actually be willing to pay for the purchase of digital services.

The second stems from how aggregates can be derived using implicit prices. As with traditional services, gross value can be calculated by multiplying the implicit price of free services p_f with a measure of volume q_f. This volume measure can be represented by the number of individuals employing the free service. As such, the calculated aggregate value from free services would increase with the number of users enjoying the service. This is in contrast to the total cost approach. As discussed in the previous section, if the marginal cost of producing the free good is zero (or close to zero), then the additional unit of consumption for that good would not be recorded in the aggregate calculated with the total cost approach. Since the aggregate to be calculated is explicitly a function of volume, it is easier to argue that the gross value generated by this approach would be closer to the concept of welfare.

The challenge is to isolate the prices attributable only to the services present in the free versions. We employ hedonic regression to disentangle the price attributable to free services from the price of their premium versions. This strategy effectively limits the scope of our estimation to services having paid counterparts.

3.3.1 Hedonic Regression

The Lancaster (1966) model suggests that households derive utility from "characteristics" rather than the products per se. For instance, individuals do not consume houses, but the characteristics associated with houses such as their ability to shield from the elements, security from other people, and the overall aesthetics of the structure, to name a few. Hedonic regression applies this principle by allowing for the estimation of how characteristics are able to contribute to the value of goods (see Groshen et al. (2017)). This method has been used to generate quality-adjusted price indexes (see Triplett (2006), de Haan and Diewert (2013), Groshen et al. (2017)) and the estimation of the willingness to pay for producing particular characteristics of goods (see de Haan and Diewert (2013)).

For this research, we employ hedononic regression to estimate the implicit price of free digital products using prices of their "premium service" counterparts. In particular, we limit the scope of this exercise to videoconferencing services, personal email, and online news, although these are not the only free services that have paid counterparts. We assume that premium versions of these services are imperfect substitutes for the free versions. As such, the price of premium versions would reflect the willingness to pay for the utility derived from the consumption of the services. In this case, the price of the paid version of free digital products would reflect the marginal utility of these products as a characteristic, which is also present in their free version. However, we cannot simply use the market price of premium services as a proxy for free services because the former also incorporates the marginal value attached to characteristics that are present in premium versions but are not present in free versions. For instance, Zoom and Microsoft Teams allow for the creation of breakout rooms in their premium versions but not in the free versions of their services. Their prices reflect this and employing these prices to impute the value of free products would yield biased estimates. Hedonic regression allows us to control for these characteristics and estimate the price of these services once premium-exclusive characteristics are removed.

The hedonic regression approach assumes that the price p_p of a good p can be expressed as a function of its characteristics z_{in} and a random error term ε_i. Thus, We have,

$$p_p = f(z_{i1}, ..., z_{i,n}, \varepsilon_i) \tag{3.2}$$

for a good with n characteristics. The marginal contribution of each characteristic can be estimated through a regression framework. In this study, we employ the logarithmic-linear (or semi-log) model[9]. In this exercise, we employ a modified time dummy variable model given by:

$$log(p_{i,j}^t) = \sum_{j=1}^{J} \sum_{t=1}^{T} (\delta_j \times \tau^t) + \sum_{k=1}^{K} \beta_k Z_{i,j} + \varepsilon_{i,j} \tag{3.3}$$

where $log(p_{ij}^t)$ represents the the natural log of the prices at year t. The index i indicates the plan type (Standard, Pro, Business, etc) while the index j represents the service provider (Zoom, Cisco Webex, Microsoft Teams, etc). The list of service providers and their respective pages are listed in the appendix. These prices are regressed against a set of characteristics contained in matrix Z_{ij} and a set of service provider fixed effects δ_j. Details on the characteristics are described in section 3.4. In our specification, the term $(\delta_j \times \tau^t)$ represents the interaction term between the service provider dummies δ_j and year dummies τ^t. This ultimately generates separate intercept terms for each service provider for each year. We interpret each of the intercept terms as the quality-adjusted price for each service provider j for time t. The error term ε_{ij} is assumed to be normally distributed with mean 0 and constant variance. For this study, we follow the technical guidance of Aizcorbe (2014) and those of de Haan and Diewert (2013). One of the difficulties in implementing the hedonic approach is its sensitivity to the regression specification. It is critical that all characteristics that determine the price should be included as explanatory variables, otherwise this would lead to omitted variable bias. We incorporate all features advertised on the service provider's

[9]An alternative is the linear specification where the levels of prices are used as the dependent variable. Diewert (2003) noted that it is more appropriate to employ the log-linear model for technological products since it often mitigates the problem of heteroskedasticity as their prices tend to have a log-normally distribution.

websites that are not multicollinear with the service provider fixed effects.

In a typical regression framework[10], the intercept term a_0 represents the expected value of the dependent variable if the value of all explanatory variables are zero. In the context of this research, this parameter represents the average (log) price of free products if the value of all premium-exclusive characteristics is netted out. As such, $exp(a_0)$ would reflect the shadow price of the free version of digital products. Since we allow each service provider to have a different intercept for each point in time, our hedonic regression equation would produce separate quality-adjusted price indices for each service provider. To impute the price of free digital services, we take the average of these quality-adjusted price indices for the specific year (see table 3.4). For videoconferencing and email, we also include continuous variables as regressors that cannot be assumed to be zero. These are the number of participants in the case of videoconferencing and the mail storage in the case of email. We assume a certain value for these variables (z_1) in the prediction model and multiply them by their coefficient. Lastly, the expectation of the error term $E(log(\varepsilon_{ij}))$ should be taken into consideration in the estimation of the price, otherwise, the estimates would be biased. The standard correction suggested by the literature (see Pakes (2003); Aizcorbe (2014); Erickson (2016)) is the inclusion of the term $exp(0.5Var(\varepsilon_{ij}))$ for a log-linear model. The imputed price of free videoconferencing can be calculated by the expression:

$$\hat{p}^t = \left[\frac{1}{J}\sum_{j=1}^{J} exp(\hat{\delta}_j \times \hat{\tau}^t) \times exp(\hat{\beta}_1 log(z_1))\right] \times exp(0.5Var(\hat{\varepsilon_{ij}})) \tag{3.4}$$

In the area of official statistics, this approach has been adopted to generate quality-adjusted price indices for technological products by the US Bureau of Economic Analysis (Groshen et al., 2017) and the UK's Office of National Statistics (Office for National Statistics, 2018, 2019b,a).

One limitation of our approach is that we were only able to control for characteristics that were stated on the service providers' websites. It is possible that other characteristics–such as speed, size of the subscriber network, and aesthetics of the interface, to name a few–would affect prices but are not explicitly indicated as a feature of the service as stated in their websites. Moreover, traits like the subscriber network are often undisclosed, and the aesthetics of the interface are difficult to quantify. We try to address this by incorporating service-provider fixed effects δ_j, which are intended to control for these differences. It is assumed that characteristics such as those mentioned earlier are specific to the providers of the service and their marginal contribution to prices should be absorbed by dummy variables.

[10]Consider a typical regression equation $y_i = a_0 + \sum_{k=1}^{K} \beta_k X_i + \varepsilon_i$ where outcome y_i is expressed as a linear function explanatory variables contained in matrix X_i

3.4 Data

Hedonic price imputation requires information on subscription prices and characteristics. Data on 22 videoconferencing service providers, 13 email service providers, and 10 news sites were acquired. For videoconferencing and email, the providers were identified by entering the keywords "paid videoconferencing services", and "paid email service", which returns websites that list top providers of these services. For online news, the list of providers included in the study was based on the report by the Office of Communications (Ofcom) for News Consumption in 2020[11]. There are platforms that allow for videoconferencing but do not offer premium (paid) services, such as Discord, Facebook Messenger, and WhatsApp, among others. The same goes for email and news. These providers were not included in our data set since we require price data for the regressions.

We used the website Internet Archive (www.archive.org), a US-based digital library, to acquire data for the years 2017 to 2020. The website allows for public access to past versions of websites, allowing us to acquire information on prices and characteristics from previous years. A description of the panel structure for the hedonic regressions is discussed in appendix 3.A.

On average, the price of videoconferencing appears to be increasing over time (see table 3.4.1). The average price of videoconferencing in 2017 was \$25.5. This increased to \$46.4 in 2021. However, the average number of participants each call can accommodate increased as well. In 2017, the average number of participants for videoconferencing services was 72.5 participants. This increased to 183.2 participants in 2021. If we normalize the prize to the number of participants, prices actually declined from \$0.8 in 2017 to \$0.4 in 2021. This could reflect improvements in technology, which we see in the trend of other information goods (see Roser and Ritchie (2013)).

The data also shows the average price of email services is increasing over time from \$7.8 in 2017 to \$9.7 in 2020 until prices fell to \$7.2 in 2021. The range of prices was stable between \$1.0 to \$57.0 from 2017 to 2020. If we normalize the price to the amount of mail storage, prices are actually stable (hovering between \$0.7 and \$0.8) from 2017 to 2020, until they fell to \$0.5 in 2021.

The data shows that the price of online news is increasing over time. The average subscription price in 2017 was \$13.4. This increased to \$19.2 in 2021. While the maximum subscription price increased to \$67.0 in 2021 from \$34.0 in 2017, the minimum price stayed the same at \$3.1 for all years in the panel.

3.5 Results

In this section, we describe the price estimates generated by our model and compare them to the WTA value from other studies. For a detailed discussion of the regression results, see

[11] https://www.ofcom.org.uk

Table 3.4.1: Descriptive statistics over time

	2017	2018	2019	2020	2021
			Videoconferencing		
Ave Price (in USD)	25.5	38.7	34.6	48.5	46.6
Ave Participants	72.5	228.4	237.6	173.3	183.2
Ave Price per Participant	0.8	0.6	0.4	0.4	0.4
Total Plan Types	33	37	45	59	69
Number of Providers	12	12	15	20	22
			Email		
Ave Price (in USD)	7.8	22.3	22.9	10.0	7.7
Ave Mail Storage (in GB)	58.8	40.0	25.6	23.0	23.2
Ave price per GB	0.5	0.9	0.9	0.7	0.5
Total Plan Types	26	30	32	37	37
Number of providers	9	10	10	13	14
			Online News		
Ave Price (in USD)	13.4	13.0	12.1	18.7	19.2
Total Plan Types	14	13	13	14	14
Number of providers	10	10	10	10	10

Note: The table shows the mean prices for paid versions of videoconferencing services, personal email, and online news from 2017 to 2021. The table also shows the average number of participants and the average price per participant for videoconferencing, the average mail storage space and price per storage space for email, as well as the total plan types and number of providers for each year in the data set. All prices are in $and mail storage is expressed in gigabytes (GB). A detailed discussion of the data can be found in appendix 3.B.

appendix 3.B.4.

3.5.1 Shadow price of free digital services

We impute the price of free digital services using equation 3.4. We used the confidence interval for the coefficient of the number of participants $[\beta_1^U \hat{,} \beta_1^L]$ to generate our upper and lower bound estimates of the price[12]. For online news, however, this is not possible because the regression does not incorporate any continuous variable. As such, we only take the average of the upper and lower bound estimates of the quality-adjusted price indices, δ_j for email in order to generate interval estimates of its shadow price.

The specification in equation 3.4 requires us to assume a value for the continuous variable, z_1. For videoconferencing, this is the number of participants. The number of participants for free videoconferencing services is different for each provider. The top three messenging apps that offer free videoconferencing features are Whatsapp, Facebook Messenger, and We Chat[13]. The maximum number of participants for both Whatsapp and Facebook Messenger is 8 while the maximum for We Chat is 9 participants. The maximum number of participants for other popular applications offering free videoconferencing services also varies: for Viber, it is 20 participants, for Discord, 25, and for Telegram the maximum is 1,000. To be on the conservative side, we adopt the assumption of 8 participants for the generation of our price estimates. This is consistent with the maximum number of participants for Whatsapp and Facebook Messenger, the two largest providers of free videoconferencing at the time of writing.

Since mail storage is also a continuous variable, the prediction model requires us to assume a level of storage space for the price estimation. Similar to videoconferencing, the mail storage limit is different for each provider. In terms of users, the top three providers of personal email are Gmail (Google), Outlook (Microsoft), and Yahoo. Both Gmail and Outlook offer 15 GB of storage while the storage limit of Yahoo is 1 terabyte. In this exercise, we assumed storage space of 15 GB, which is based on the storage space per person of Gmail and Outlook.

[12] While it is also possible to generate the confidence interval for the quality-controlled price indices of each service provider δ_j, the standard errors for the said coefficients are small. So much so that the difference between the upper and lower bound estimates would be immaterial.

[13] Statista Research Department (2021b)

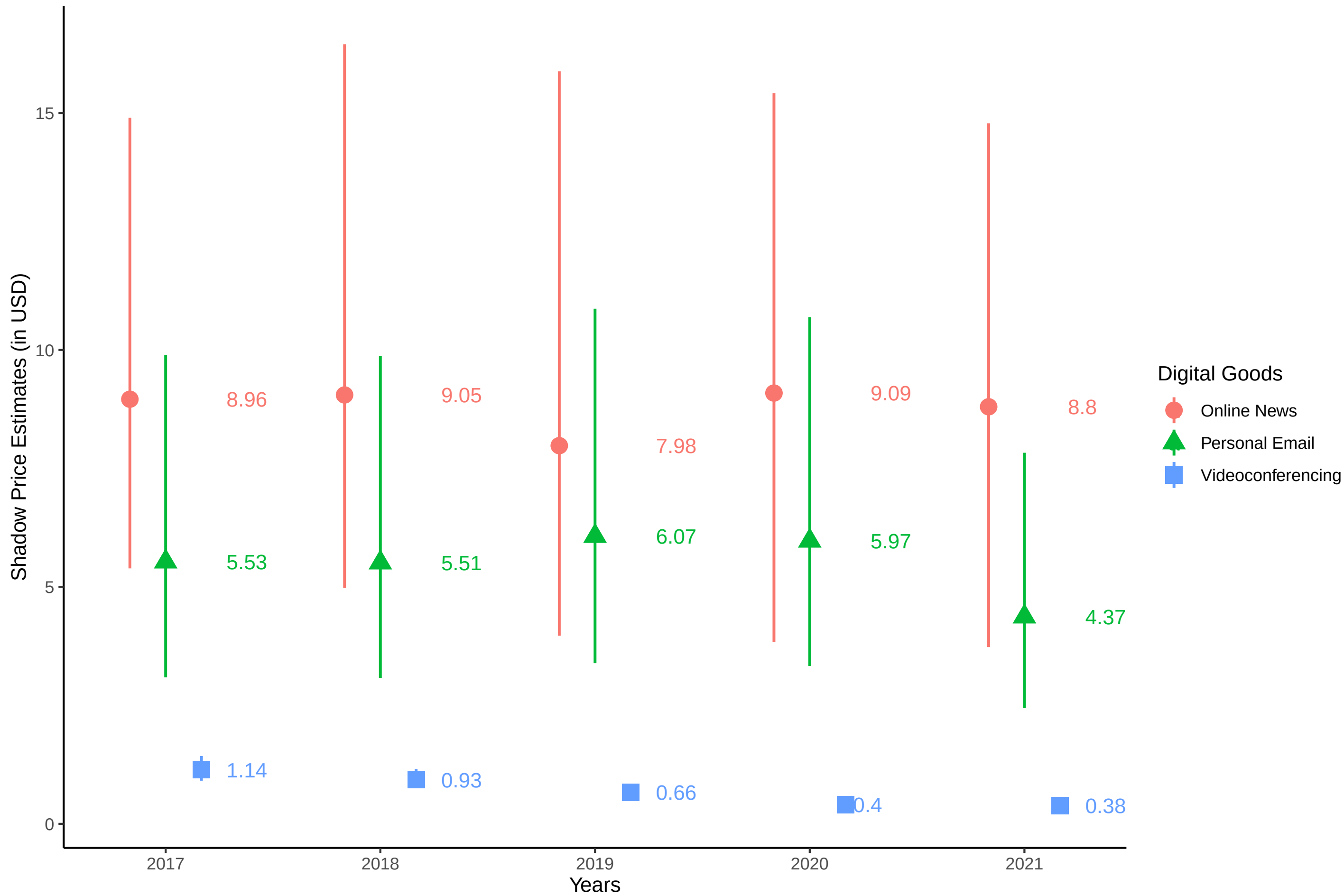


Figure 3.5.1: Imputed price of videoconferencing, personal email, and online news

Note: The figure shows and shadow price estimates for videoconferencing, personal email, and online news, generated by the prediction model in equation 3.4. The price estimates, upper and lower bound estimates can be viewed in appendix 3.F.

We present our estimates for the shadow price of videoconferencing, personal email, and online news in figure 3.5.1 (see appendix 3.F for the table showing the interval estimates of the price levels for each year). Of the three forms of digital services covered in this study, our price estimate for videoconferencing was the lowest. Our estimates show that the shadow price for videoconferencing was approximately $0.40 in 2020, lower compared to $1.14 in 2017. This decline is consistent with the decline in the price per participant that we observe in table 3.4.1. For personal email, we estimate a shadow price of about $6.0 in 2020, slightly higher than the 2017 estimate of $5.5. Based on 95 percent prediction intervals, however, estimates across years for personal email are not statistically different from one another. As such, we cannot make any conclusive conjecture on the price trend for email. The largest price estimate we generated across the years is for online news. Our estimates show that the shadow price of online news is approximately $9.1. As with email, we observe no apparent trend for the price movement of online news across time.

3.5.2 Validity and Robustness Check

Several efforts were made to test the validity of our estimation procedure. First, we compare our shadow price estimates to the estimated price of each characteristic, with the assumption that the majority of the value of each service would come from the core feature of the service, which can be found in the free version. Second, we use our model to predict the average price of the premium version and compare these estimates with the observed mean prices from the data. We find that these estimates satisfy our validity criteria. We discuss the details of this test in appendix 3.G.

We show in tables 3.G.1, 3.G.2 that predictions from the hedonic regression model are all within the intervals of the observed mean price. To us, this implies that our model is effective in approximating the price of paid platforms. We also find that the implied value of digital services accounts for the largest share of the predicted price. We believe that makes sense. If the value of premium-exclusive features dominates the price, then that implies what users are paying for when they subscribe to the premium version does not include the value of the service that is also available in the free version. However, our validation exercise suggests that the estimated value of services available in the free version accounts for the largest share of the estimated price for email (see table 3.G.5) and online news (see table 3.G.6), and constitutes the second largest share in the case of videoconferencing (see table 3.G.4).

Price estimates from hedonic regressions can also be sensitive to characteristics included in the specification. It is possible that the inclusion or exclusion of any explanatory variable in the regression can result in substantial changes in the estimates. Therefore, we examine the degree to which our price estimates would change given varying specifications. Our selection criteria for the inclusion of explanatory variables in the hedonic regression was aimed at maximizing the number of characteristics we can incorporate in the empirical model given the information set published by the service providers on their websites. The goal of this approach is to minimize omitted variable bias. We wanted to assume with confidence that all steps were taken in order to incorporate all *observable* characteristics in the baseline specification. One of the pitfalls of this approach is that it could result in overfitting. The coefficient estimates for some of the characteristics in regression results displayed in the results tables are not statistically significant. A more parsimonious model could be a better fit for our purposes. As such, we perform forward, backward, and stepwise regressions to test the robustness of our price estimates to changes in specifications. We find that for most specifications, our shadow price estimates are not materially different from those generated by other models. We find that our estimates do not deviate too much from the prices generated by more parsimonious models.

A full discussion on the robustness checks is in appendix 3.H.

3.5.3 Comparison with other studies

We compare our imputed prices of videoconferencing, personal email, and online news to the WTA estimates by other authors, namely Brynjolfsson et al. (2019a), Coyle and Nguyen

(2023), and Jamison and Wang (2021). For videoconferencing, the comparison can be viewed from figure 3.5.2 (table 3.5.2).

Brynjolfsson et al. (2019a) and Coyle and Nguyen (2023) did not ask their respondents about their WTA for videoconferencing as a general service. Rather, they asked the participants for their WTA for Skype (which, for the longest time, was almost synonymous with videoconferencing). We compare their estimates for Skype to the estimates from the hedonic regression, considering that they are the closest to videoconferencing, conceptually. Coyle and Nguyen (2023) also asked their respondents about their valuation for Whatsapp and Facebook Messenger, the two most popular providers of videoconferencing service in the UK. We include these estimates in our comparison. Another important note is that the experiment described by Brynjolfsson et al. (2019a) was carried out in 2003. In our attempt to make the estimate as comparable as possible, we inflate their estimates using the Dutch CPI inflation from 2003 to 2020.

The mean WTA estimates from the online survey of Coyle and Nguyen (2023) for Skype, Messenger, and Whatsapp are higher than those generated by the hedonic regression. Their median bands are substantially lower compared to the mean estimates. For Messenger the mean estimate is 243 folds higher than the median, while for Whatsapp, the mean estimate is higher by a factor of 261. [14] This suggests that many individuals reported extreme value in their response to the survey, causing the mean estimates to be very high.

The estimates by Jamison and Wang (2021) for videoconferencing is also considerably higher than the estimates from the hedonic regression. It is interesting to note that for Jamison and Wang (2021), their median WTA for videoconferencing is 7.5 fold greater than that of Zoom, which is one of the most popular service providers at the time of their study. WTA estimates for videoconferencing, as a general service, was between \$228.9 to \$446.2. This is the highest estimate recorded for this category of digital services.

Not counting the median band for messenger by Coyle and Nguyen (2023), only the estimates of Brynjolfsson et al. (2019a) coincide with the interval estimates from the market-equivalent pricing. This is true even when the estimates are inflated to 2020 price levels. It is possible that during the time when the experiment was conducted, videoconferencing was not as essential to daily work activity. Therefore, people valued it less. It would be interesting to know if they would arrive at the same value if they conducted the exercise today.

The same pattern can be observed for both personal email and online news. We compare the estimates to those from Coyle and Nguyen (2023) and Jamison and Wang (2021) in figure 3.5.3 (table 3.I.2). For both services, estimates from the hedonic imputation is substantially lower than the WTA estimates from the contingent valuation studies. We offer two explanations for this observation.

[14] The median band for Skype by Coyle and Nguyen (2023), which we assume meant that it is less than zero. Therefore, we did not include it in the comparison.

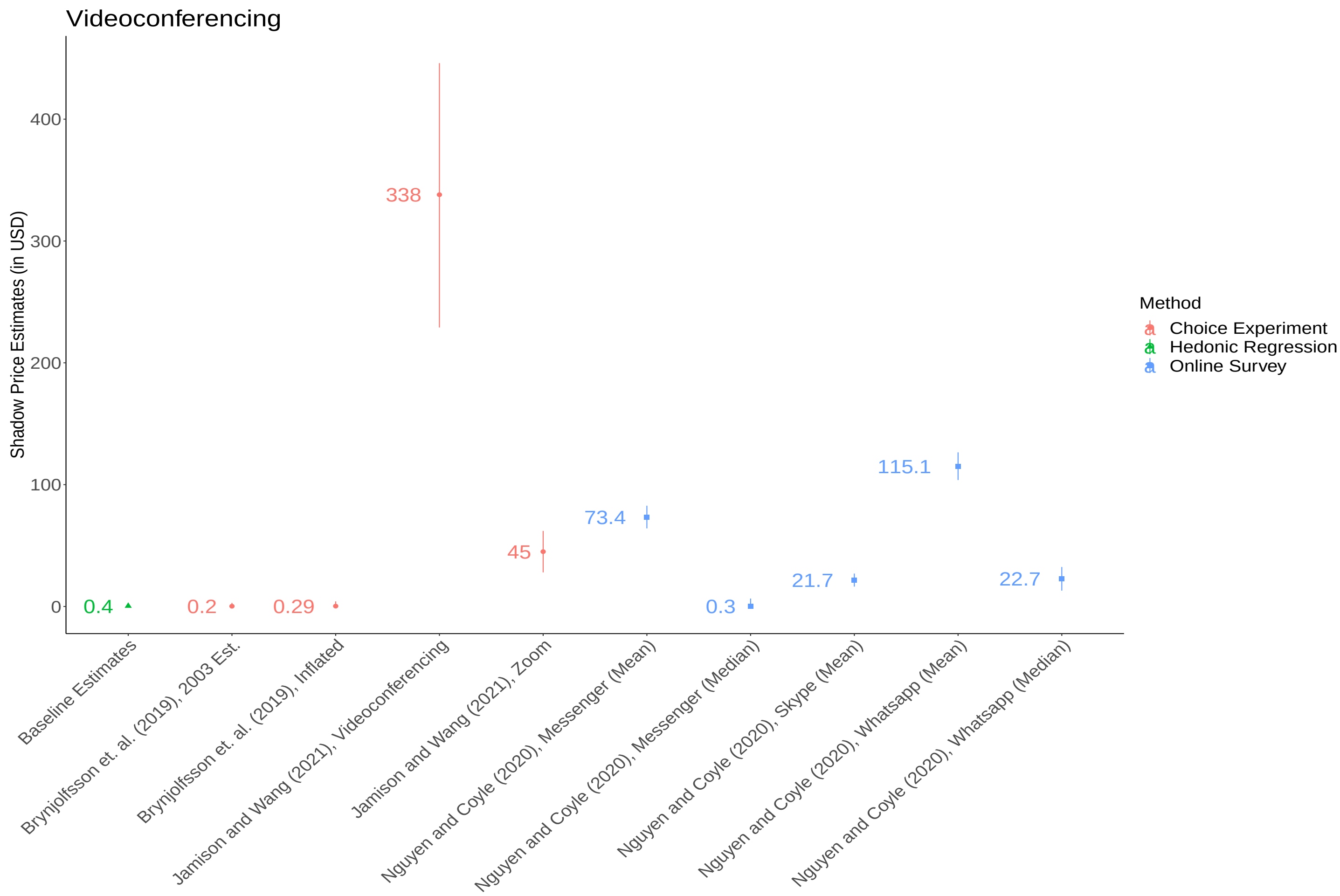


Figure 3.5.2: Comparison of WTA values with the price imputations for videoconferencing

Note: The figure compares the WTA estimates from Brynjolfsson et al. (2019a), Coyle and Nguyen (2023), and Jamison and Wang (2021) with the price estimates from the hedonic regression. Estimates by Brynjolfsson et al. (2019a) in column (2) were inflated to 2020 prices using Dutch CPI inflation from 2017 to 2020. Figures can be viewed from table .

It is possible that respondents from the contingent valuation studies are not able to internalize the available substitutes. For instance, when they are asked how much they are willing to be paid to give up online news, they are not thinking that they can purchase printed news as a substitute for online news when they make their choice. As such, the individual's willingness to pay for the consumption of news service, in that case, would not necessarily equate with how much they are willing to be compensated for giving up access to the service entirely. Furthermore, we see from the Coyle and Nguyen (2023) study that these discrepancies extend to traditional goods and services where their WTA are substantially larger than their market equivalent.

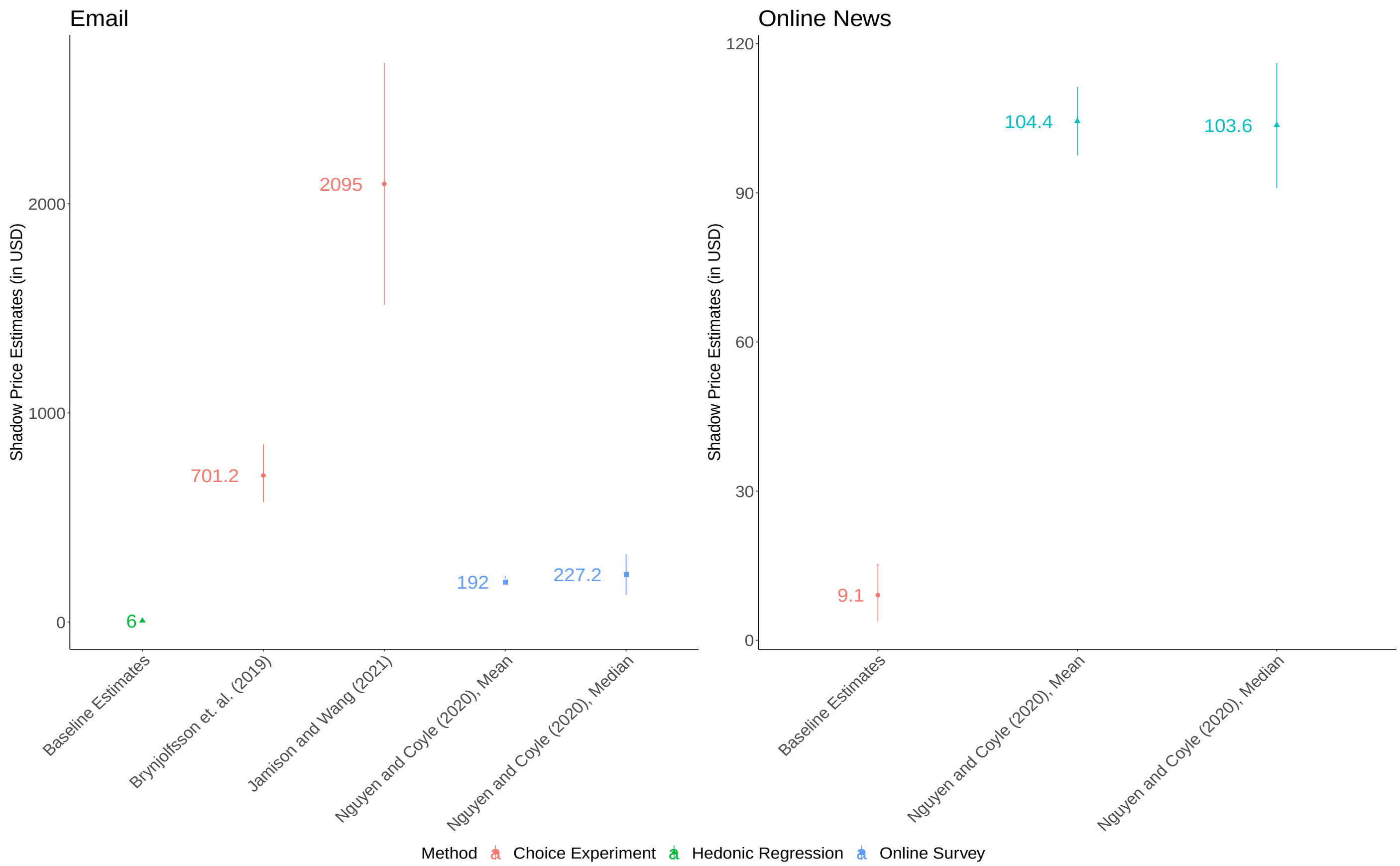


Figure 3.5.3: Comparison of WTA values with the price imputations for email and online news

Note: The figure compares the WTA estimates from Coyle and Nguyen (2023) and Jamison and Wang (2021) with the price estimates from the hedonic regression. Figures can be viewed from 3.I.2.

This is not surprising. As discussed in the previous chapter, Schreyer (2021) argues that instruments that elicit WTA likely capture the value of the household activity (leisure in the case of Facebook) rather than the digital products themselves. In this framework, digital services are inputs to the activity. Hence, it makes sense that the unit value of the activity would be higher than the unit value of their inputs.

3.6 Gross value of free digital services

To estimate the aggregate willingness to pay for free digital services, we multiply its imputed price from equation 3.4 by a volume measure. The total monetary value of free services V^t can be expressed as,

$$V^t = \sum_{f=1}^{F} \hat{p}_f q_f^t \tag{3.5}$$

where p_f is the shadow price[15] of free digital good f and q_f^t is a measure of its volume (or quantity). The expression V^t would represent the aggregate value derived by individuals from the consumption of free internet services and could be part of household final consumption. This begs the question, what is the most appropriate measure of volume for our purposes?

There are two ways one can think about volume when it comes to digital services 1) the number of times an individual accesses a specific service (every time a person opens or uses the application), and 2) simply having access to the service (subscription). The first is more intuitive. It assumes that utility is derived from the direct consumption of the good (i.e. when a person eats at a restaurant). The second, one assumes that utility is derived simply by having access to the service, whether they use it or not. An example of this is a gym membership.

TABLE 5: INTERNET ACTIVITIES, 2007 TO 2020

Within the last 3 months %

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Sending/receiving emails	57	62	68	69	:	73	75	75	75	79	82	84	86	85
Finding information about goods or services	58	59	59	58	62	67	66	73	70	76	71	77	78	81
Internet banking	30	35	41	42	44	47	50	53	55	60	63	69	73	76
Using instant messaging services (eg Skype or WhatsApp)	:	:	:	:	:	:	:	: :		:	:	:	72	71
Social networking (eg Facebook or Twitter)	:	:	:	:	45	48	53	54	57	63	66	65	68	70
Reading online news, newspapers or magazines	20	34	39	39	42	47	55	55	61	60	64	:	66	70
Watching video content from sharing services such as YouTube	:	:	:	:	:	:	:	: :		47	:	62	:	66
Listening to or downloading music	:	:	:	:	:	:	:	: :		:	:	:	:	62
Looking for health-related information (eg injury, disease, nutrition, improving health e	18	24	32	30	34	:	43	:	48	51	53	54	63	60
Watching internet streamed live or catch-up TV	:	:	:	:	:	:	:	: :		43	:	56	:	59
Watching Video on Demand from commercial services	:	:	:	:	:	:	:	: :		29	:	46	:	56
Making video or voice calls over the internet (eg via Skype or Facetime)	8	:	16	18	17	32	25	:	36	43	46	45	50	49
Playing or downloading games	:	:	:	:	:	:	:	: :		32	:	31	:	41
Selling goods or services over the internet	12	13	14	16	25	22	28	23	21	18	19	25	29	21
Making an appointment with a medical practitioner via a website or app	:	:	:	:	:	10	:	10 :		15	:	13	:	21
Using other online health services via a website or app instead of having to go to the hospital or visit a doctor, for example getting a prescription or a consultation online	:	:	:	:	:	:	:	:	:	:	:	:	:	15
Accessing personal health records online	:	:	:	:	:	:	:	:	:	:	:	:	:	8
Listening to music	:	:	:	:	:	:	:	:	:	49	:	58	65	:

Base: Adults (aged 16+) in Great Britain.
: Data not available.

Source: Office for National Statistics

Figure 3.6.1

Note: The figure shows a screenshot of Table 5 Internet Access survey of the ONS, UK.

For our application, the only feasible course of action is to adopt the second case since the only information we have on prices is based on subscriptions. The task of acquiring reliable data on the number of subscribers to free services is not straightforward. This type of information is not readily available from any source that we know of at this point. As such, we estimate the number of individuals who have access to videoconferencing and video calls

[15] The price we generated from the hedonic regression was based on monthly subscriptions. To arrive at the annual price, we multiply the imputed monthly price by 12.

using the ONS' Internet Access Survey and population statistics. In particular, we employ table 5 of the ONS survey (see figure 3.6.1). We multiply the proportion of adults with access to certain internet activities–which in our case is, "Making voice and video calls", "Sending and receiving emails", and "Reading online news"–by the estimated number of individuals aged 18 years and above based on the ONS' population projection data set. We arrive at the gross value of free services by multiplying our estimated number of subscribers for each activity by their respective implied prices.

It is important to note that there would probably be double counting in the estimates when we aggregate them with HFCE and/or GDP levels. The volume measure we used q_f^t includes *both* free and paying users of the good. In order to appropriately aggregate these estimates with official statistics, it is important either to identify the number of free users or to net out the value derived by paying users. A counterargument to this is that individuals who subscribe to paid services often subscribe to their free counterparts as well. For instance, people who read the news through the paid version of Telegraph also read the news from free sources such as the BBC or CNN. Therefore, while double counting may be a problem, its effects on our estimates may not be so severe. Moreover, it is important to note that free users account for the vast majority of digital service users. In many cases, premium users only account for 1 percent[16] of the total user base. As such, free users would account for the vast majority of what is being captured by the ONS survey.

3.6.1 Estimates of the gross value of free digital services

At this stage, we interpret our estimates as measures of the gross value of free digital services. As such, we consider our estimates as part of the consumption side of GDP rather than the production side. The current price estimates of the gross value of digital services are shown in figure 3.6.2. The initial figures that we generated were in USD. In order to be comparable with the UK's National Accounts data, we convert the estimates to GBP. We apply only one exchange rate (which is the average exchange rate from 2017 to 2020), in order to avoid having foreign exchange fluctuations affect our results.

Based on our estimates, the point estimate for the gross value of these services is £5.9 billion, higher by 9.2 percent than the £5.4 billion in 2017. The data shows that the gross value of free digital services makes up 1.1 percent of Household Final Consumption Expenditures and 0.6 percent of GDP in 2020. The interval shows that the gross value of free digital services is between £3.2 billion to £11.1 billion in 2020.

[16]For Zoom, only 470,100 of their 200 millions users are subscribed to their paid service in 2020: https://www.businessofapps.com/data/zoom-statistics/

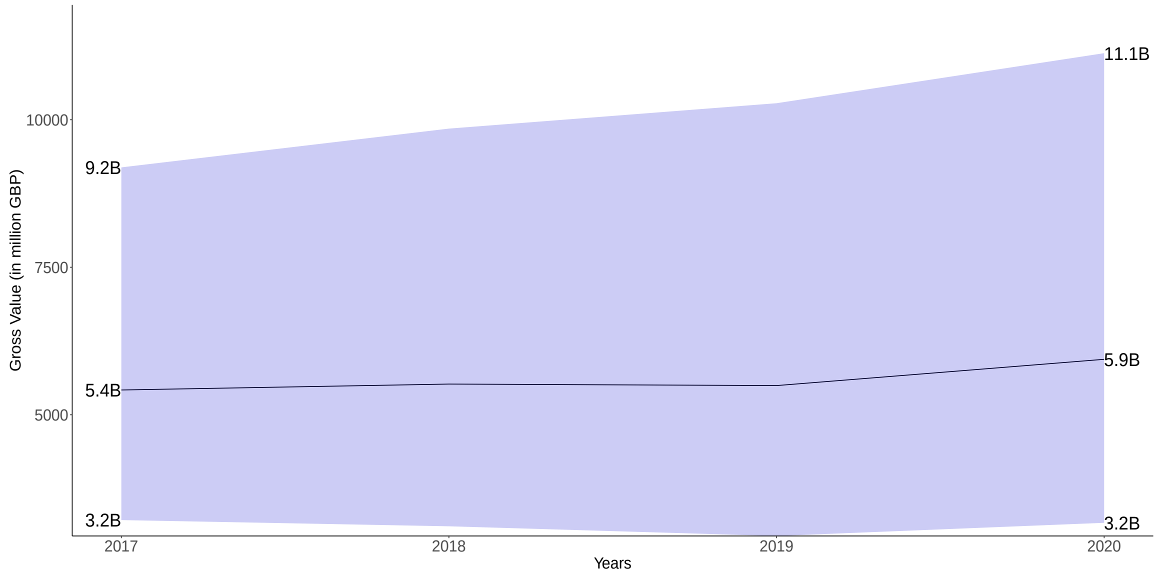


Figure 3.6.2: Gross Value of free digital services

Note: the interval estimate of the aggregate gross value for the three digital services, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS. All estimates are in million £. Figures can be viewed from table 3.J.1 of appendix 3.J.

We also generate constant price estimates (see appendix 3.J) by deflating the nominal figures with an implicit Laspeyres price index. We chose 2019 as the reference year in order to be consistent with the ONS. We add the constant price estimates to the chain volume measure estimates of the UK's HFCE and GDP to generate "expanded HFCE" and "expanded GDP" measures that include the consumption of these three digital products. We show the growth rates in table 3.6.1.

For both 2018 and 2019, the gross value of free digital services has grown substantially faster than both aggregate household consumption and GDP. Our estimates show that with the inclusion of the three digital services, the decline in HFCE would have been slower by 0.03 to 0.1 percentage points in 2020. The decline in GDP for 2020, meanwhile, would be 0.02 to 0.09 percentage points slower with the inclusion of the value of digital services.

Table 3.6.1: Growth rates of digital services and household consumption

	2017-2018	2018-2019	2019-2020
HFCE	2.08	0.99	-12.94
GDP	1.71	1.60	-11.03
Digital services			
Point Estimate	2.42	2.78	2.90
Lower	2.38	2.94	2.63
Upper	2.44	2.67	3.13
HFCE + digital services			
Point Estimate	2.08	1.00	-12.88
Lower	2.08	0.99	-12.91
Upper	2.09	1.00	-12.83
GDP + digital services			
Point Estimate	1.71	1.61	-11.00
Lower	1.71	1.61	-11.01
Upper	1.71	1.61	-10.97

Note: The table shows the growth rates of the household final consumption expenditure and gross domestic product chain volume measure estimates of the ONS, constant price estimates of the gross value of digital services, HFCE + digital services, and GDP + digital services. Figures are in percent.

As mentioned earlier, we employ the Internet Access Survey of the ONS for our baseline estimates of gross value. The problem with this approach is that the survey was conducted between January and February 2020, before the UK government announcement of the lockdown on 23 March. As such, it would not be able to capture any change in internet consumption patterns during the pandemic. In order to assess the value derived by individuals from the consumption of free services at the time of the national lockdown, we employ a different set of indicators for our volume measure.

3.6.2 Effect of the pandemic on the gross value of free digital services

Since the Internet Access survey of the ONS was no longer representative of internet consumption behavior during the COVID lockdown, we decided to employ data from the "2021 Online Nation" report of the UK's Office of Communications (Ofcom). The report for this year includes data on the share of the UK population engaged in certain internet activities, such as video calling and email. Unfortunately, past reports do not contain the same information. Therefore, linking estimates using figures from Ofcom with estimates derived using the ONS data would produce a series that is not fully comparable. However, we feel that this adjustment is necessary and more appropriate than simply employing the ONS data from 2020, which we know is not representative of the pandemic year.

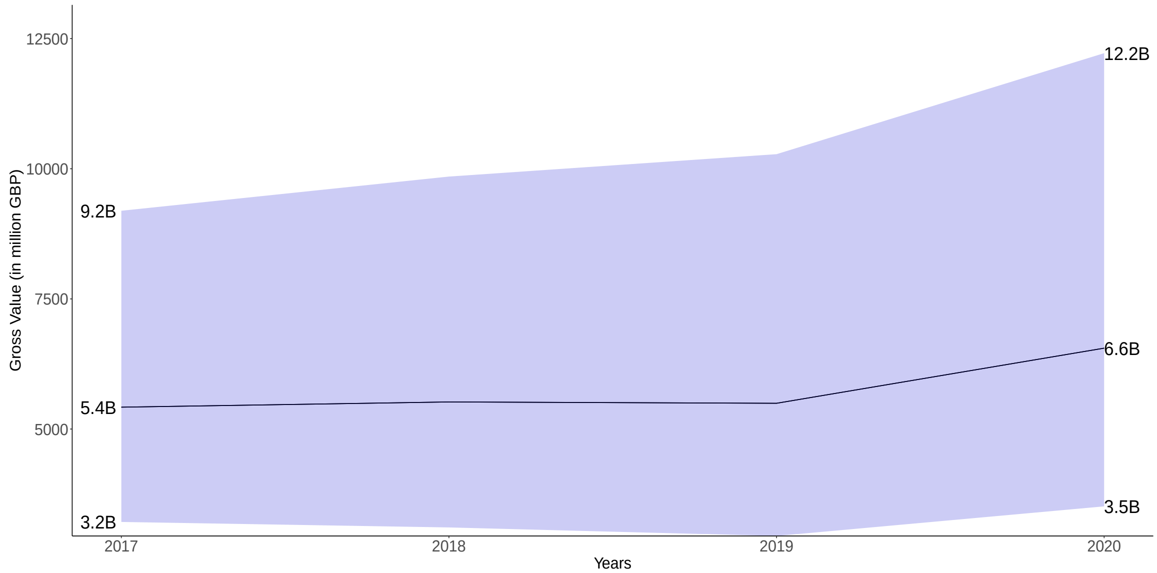


Figure 3.6.3: Gross value of digital services adjusted for Ofcom data, at current prices

Note: the interval estimate of the aggregate gross value (at current prices) for the three digital services, videoconferencing, personal email, and online news, after estimates in 2020 is adjusted using Ofcom data. All estimates are in million £. Figures can be viewed from table 3.J.3 of appendix 3.J.

Our estimates show that the point estimate for the gross value of free digital services was at £6.6 billion in 2020, higher by 22.2 percent compared to the 2017 figures. The interval shows that the gross value of free digital services is between £3.5 billion to £12.2 billion in 2020.

We present the growth rates of the gross value of digital services, calculated using volume measures from the Ofcom report in table 3.6.3. In contrast to the earlier estimates, the gross value from free digital services during 2020 grew by double digits. The impact on HFCE is more substantial compared to earlier. Household consumption decline was slower by 0.06 to 0.19 percentage points, when including the free digital services. The estimates also show that the inclusion of free digital services in GDP would have slowed its decline by 0.03 to 0.11 percentage points.

Table 3.6.2: Growth rates of digital services and household consumption using Ofcom volume indicators

	2017-2018	2018-2019	2019-2020
HFCE	2.08	0.99	-12.94
GDP	1.71	1.6	-11.03
Digital services			
Point	2.42	2.78	14.41
Lower	2.38	2.94	14.79
Upper	2.44	2.67	14.04
HFCE + digital services			
Point	2.08	1.00	-12.84
Lower	2.08	0.99	-12.88
Upper	2.09	1.00	-12.75
GDP + digital services			
Point	1.71	1.61	-10.97
Lower	1.71	1.61	-11.00
Upper	1.71	1.61	-10.92

Note: The table shows the growth rates of the household final consumption expenditure and gross domestic product chain volume measure estimates of the ONS, constant price estimates of the gross value of digital services, HFCE + digital services, and GDP + digital services. Figures are in percent.

3.6.3 Accounting for multiple provider usage

In our earlier estimates, we measured volume in terms of the number of individuals that utilize certain categories of the free digital services we are concerned with. We take the share of the population engaged in the activity (as reported by the ONS and the Ofcom surveys) and multiply these figures with the population belonging to the age range covered by the surveys. As such, a user of free digital services would be counted only once regardless of how many providers of that service he or she employs.

In reality, people often use multiple service providers for the same purpose. For instance, it is common that a person who uses WhatsApp for video calls would also engage the services of other videoconferencing providers such as Facebook Messenger or WeChat. One can argue that the utility received by individuals from the use of one service provider is separate from the utility it derives from another provider[17]. In the case of market services, if a person subscribes to both Netflix and Disney Plus, a subscription to the two services would be counted separately in GDP and HFCE.

We generate a separate set of estimates, which accounts for the use of multiple providers.

[17]For videoconferencing, Whatsapp probably allows a person access to a network of people separate from the network provided by WeChat.

Ideally, the best way to achieve this is by employing the number of users for each service provider. Unfortunately, precise data on the number of users are not readily available.

The top two providers of videoconferencing service (in terms of user share) in the UK are Whatsapp and Facebook Messenger (see Statista Research Department (2021b)). The top two most downloaded videoconferencing applications in the UK are Whatsapp and Telegram (see Statista Research Department (2021a)). We employ the number of Facebook Messenger users published by Statista Research Department (2022) and Statista Research Department (2021c) for the number of Whatsapp users. We impute the number of telegram users by taking the proportion of Telgram downloads to Whatsapp downloads in Statista Research Department (2021a) and applying the ratio to the number of Whatsapp users for each year.

For online news, we estimate the number of individuals who read the news from the web pages of the following news sources: BBC, Sky News, The Guardian, Daily Mail, Google News, Youtube, Local Newspaper, Huffington Post, ITV, BuzzFeed, MSN, LADbible, Yahoo News, The Sun, and The Metro. We use the data on the percentage of individuals who identify as viewers for the respective source from Ofcom's 2021 News Consumption report, conducted by Jigsaw Research (2021). We multiply the share of news viewers/readers per news source with the population estimates from the ONS in order to arrive at the number of viewers/readers for each news source.

To arrive at the estimates of gross value, we multiply our indicators for the number of subscribers for free videoconferencing and online news to the price estimates we generated in section 3.5. Unfortunately, we are not able to find any data on the number of users for Gmail, Outlook, or Yahoo Mail (the top three providers of free email services in the UK). As such, we maintain our earlier estimates for email.

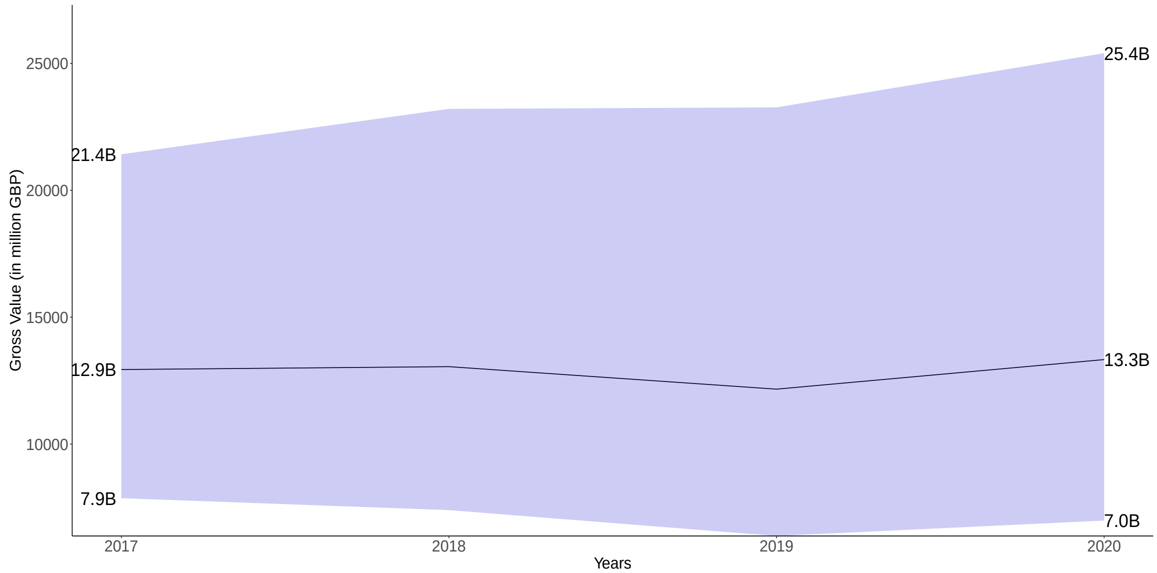


Figure 3.6.4: Gross value of digital services accounting for multiple service provider usage, at current prices

Note: The table shows the interval estimate of the aggregate gross value (at current prices) for the three digital services, videoconferencing, personal email, and online news, accounting for multiple service provider use. Figures can be viewed from table 3.J.5 of appendix 3.J.

Accounting for multiple service provider use, the estimates for the gross value of digital services was more than doubled compared to the baseline figures. Our estimates show that the gross value of digital services is around £13.3 billion in 2020. The interval estimates show that the gross value of free digital services is between £7.0 billion to £25.4 billion in 2020 (see figure 3.6.4).

Table 3.6.3: Growth rates of digital services and household consumption for multiple service provider usage, at constant prices

	2017-2018	2018-2019	2019-2020
HFCE	2.08	0.99	-12.94
GDP	1.71	1.60	-11.03
Digital services			
Point Estimate	1.53	2.32	1.64
Lower	1.67	2.22	1.95
Upper	1.43	2.39	1.46
HFCE + digital services			
Point Estimate	2.08	1.00	-12.82
Lower	2.08	1.00	-12.88
Upper	2.07	1.01	-12.71
GDP + digital services			
Point Estimate	1.70	1.6	-10.96
Lower	1.70	1.60	-10.99
Upper	1.70	1.60	-10.90

Note: The table shows the growth rates of the household final consumption expenditure and gross domestic product chain volume measure estimates of the ONS, constant price estimates of the gross value of digital services, HFCE + digital services, and GDP + digital services. Figures are in percent.

Not surprisingly, the percentage points impact on GDP growth rates is also larger. Our estimates show that the impact on real HFCE decline in 2020 was between 0.06 to 0.23 percentage points. For GDP, the impact on real GDP decline in 2020 was between 0.04 to 0.13 percentage points.

3.6.4 Discussion

Even accounting for multiple service provider use, the estimates that our methodology generated are small relative to the UK economy. Based on our results, free digital services account for 0.57 to 2 percent of the UK's HFCE in 2020 and 0.3 to 1.2 percent of the UK's GDP in the same year.

The figures that we generated, however, are likely conservative estimates of the true value of free digital services for two reasons. First, we are unable to account for multiple service providers use of email. It is possible that many internet users hold multiple accounts from different free email providers. Second, we only accounted for the users of the top three videoconferencing providers. Due to data constraints, our estimates do not include users of Facetime, WeChat, Skype, and even Zoom. Both of these reasons are likely to cause our estimates to have a downward bias. A third reason is that we are only accounting for three categories of free digital services. Given that this study does not include other widely-used services like online maps, search engines, instant messaging, and social media, we argue that a comprehensive analysis covering all free services would reveal an even more significant impact on the UK economy.

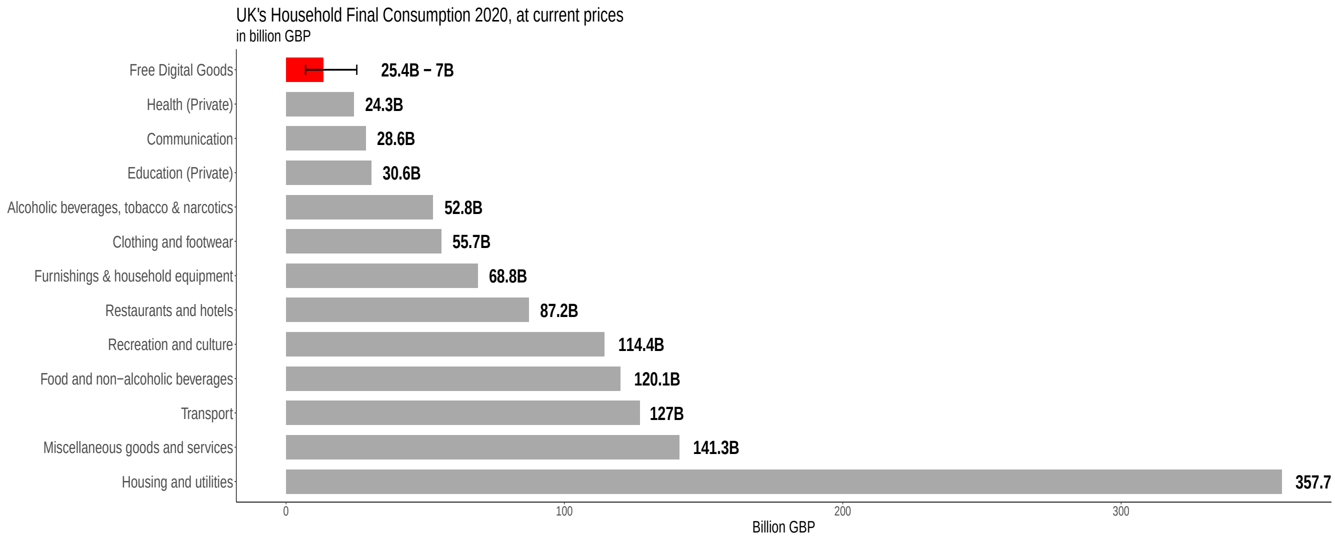


Figure 3.6.5: Comparison with other expenditure items

Note: The figure compares the current price estimates of gross value of free digital services (accounting for multiple provider use) in table 3.J.5 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.

Despite this, we argue that our estimates for the value of free digital services are economically significant. Our estimates show that the gross value of free digital services was between £7 billion to £25.4 billion in 2020. For context, the lower limit of our estimates is already 30 percent of the total final consumption expenditures for communications, which is at £28.6 billion (see figure 3.6.5). Meanwhile, the upper limit almost exceeds the value of the same expenditure item.

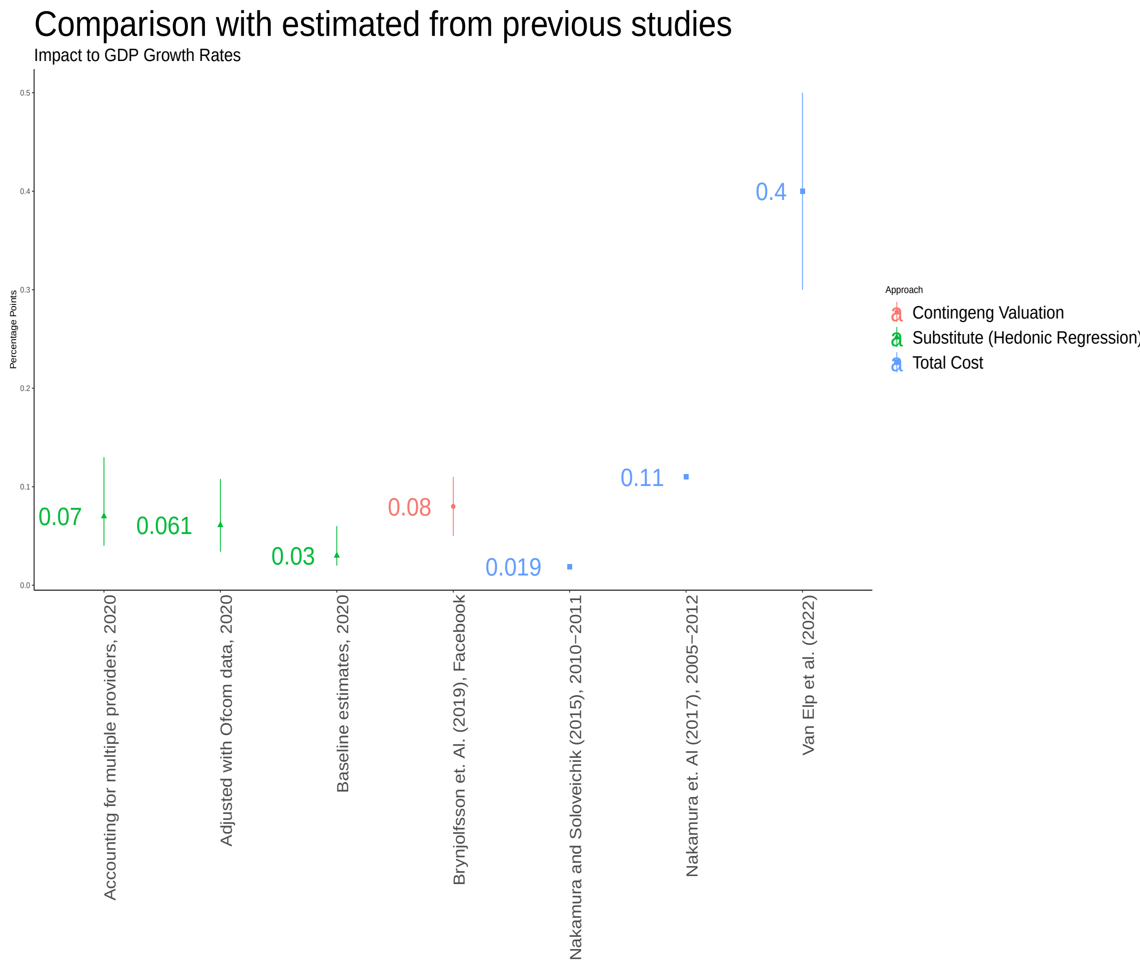


Figure 3.6.6: Comparing estimated impact to GDP with estimates from other studies

Note: The figure compares the estimated impact of digital services to GDP growth rates to the estimates by Brynjolfsson et al. (2019a), Nakamura and Soloveichik (2015), Nakamura et al. (2017), and Van Elp et al. (2022). All figures are expressed in percentage points.

We also compare our estimates to the findings of other authors (see figure 3.6.6). In particular, we compare the impact on GDP growth rates to the estimated impact by Nakamura and Soloveichik (2015), Nakamura et al. (2017), Brynjolfsson et al. (2019a), and Van Elp et al. (2022). It is important to note that these estimates cover different time periods and that the four studies are concerned with the impact of free digital services on the US economy (in the case of Nakamura and Soloveichik (2015), Nakamura et al. (2017), Brynjolfsson et al. (2019a)) and the Dutch economy (in the case of Van Elp et al. (2022)), as opposed to the UK economy, which is the focus of this study. However, we believe that this comparison would still provide valuable insight into how our approach differs compared to others.

The estimated impact on GDP growth rates of the three categories of free digital services is close to the estimated impact of Brynjolfsson et al. (2019a) for Facebook. Their study finds that the inclusion of Facebook would cause GDP growth to be about 0.5 to 0.11 percentage points faster, annually. Meanwhile, we find that the inclusion of three digital services in national production would contribute 0.07 percentage points to GDP growth (at the maximum, 0.12 percentage points).

Our estimated impact on GDP growth is also close to those generated by Nakamura and Soloveichik (2015) and Nakamura et al. (2017), both of which employ the total cost approach. So far, Van Elp et al. (2022) recorded the largest impact on GDP growth (0.3 to 0.5 percentage points). It should be noted that the three studies intended to cover *all* advertising-finance (and marketing-financed free media in the case of Nakamura et al. (2017) and Van Elp et al. (2022)). Meanwhile, the estimates that we generate only cover three forms of digital services. One can make an argument that our approach complements the estimates of the total cost approach, since we cover freemium services that are often not financed by advertising.

3.7 Conclusion and way forward

We demonstrate that the gross value of free digital services, such as videoconferencing, personal email, and online news can be estimated using observable data. Our estimation strategy overcomes some of the drawbacks encountered by previous research. First, unlike contingent valuation studies, our approach does not introduce inconsistencies with the core accounting principles of the National Accounts. As such, it would be possible to compare the imputed value of free services to other aggregates such as household consumption (and subcategories of consumption). Second, our estimation strategy does not suffer the limitations of the total cost approach, since gross value, in our case, is linked to volume. If the marginal cost of producing digital services is close to zero, an additional subscriber would not generate incremental value for the economy when estimates are derived using the total cost approach, unlike our chosen method.

Our estimates show that prior to the pandemic, the inclusion of the gross value of videoconferencing, personal email, and online news, did not make any substantial change to the growth of household consumption aggregates. During the pandemic year, however, the inclusion of these services in consumption would slow its decline by 0.07 to 0.2 percentage points. This suggests that welfare, as measured by aggregate consumption, would have been worse had it not been for the presence of these free services. While these impacts are relatively small, it is important to note that we are measuring the value of only three categories of internet services for this exercise. The inclusion of other internet services could have a substantial impact on household consumption statistics and GDP.

The goal of this effort is to develop an initial template that other researchers can use to estimate the contribution of free services to aggregate welfare. The natural extension of our research is to apply the same methodology to other internet activities with paid counterparts.

Since our approach employs the price of premium services to derive the value of their free counterpart, the method effectively limits our application to digital services with paid versions. It is possible that in the subject of measuring the value of free services, multiple approaches are needed to generate a complete picture.

At this stage, we interpret our estimates as the value of consumption from free digital services. With this interpretation, we think of households gaining utility directly from the consumption of these services. Perhaps another interpretation of our results would be that these services are also inputs to household production. For instance, videoconferencing and email services are inputs to communication or leisure, as part of household production. Moving forward, it could be interesting to see how the value of these services fits as intermediate consumption and contributes to productivity gains in household production, as measured in the household satellite accounts (see discussion in Chapter 2).

At the moment, the confidence intervals of our estimates are large, which limits the usefulness of this approach. The large confidence interval is mostly from email and online news (see figure 3.5.1), where we have a limited number of observations, especially in the case of online news (see appendix 3.A.1). Expanding the samples by including less popular news sites and providers could resolve this issue. In the methodology aspect, we also hope to test different specifications for equation 3.4 to see how estimates change with different adjustments to the error term.

While we understand that the estimates we generate are not perfect at measuring the aggregate welfare value derived by households from the consumption of free services, we believe they can serve certain objectives. First, it provides a source of external validity for other studies aimed at generating estimates of the individual's willingness to pay for free services. Second, from a time series perspective, the aggregate generated by the estimation methodology can serve as an indicator of how fast the value provided by free digital services is growing. Lastly, the methodology employed is simple enough to allow for the regular updating of the estimates, with little need for additional resources (as opposed to surveys and randomized experiments). As such, estimates can be updated frequently, which will be advantageous if these indicators are employed to guide short-term policies.

Appendix

3.A Panel Structure

Table 3.A.1 shows the number of plan types for each service provider for each year. Each provider often offers more than one plan type. The number of plan types in our data set is greater for the years 2020 and 2021 than for the previous year. There are two reasons for this. First, some of these service providers only started operations in recent years. Second, it is possible that for some providers, their websites were not archived in past years. Their services exist but there is no approach that we can think of that would allow us access to their data. This might cause some bias in our estimates. However, we will show in the robustness check that it would be better to maintain an unbalanced panel rather than to drop the service providers where information cannot be acquired in all years in the data set.

Table 3.A.1: Number of plan types for each videoconferencing provider for each year

	2017	2018	2019	2020	2021	All Years
3CX (assuming 2)	-	-	2	2	2	6
Adobe Connect	-	-	-	2	2	4
Blackboard	-	-	-	2	1	3
Bluejeans	2	2	2	2	3	11
Circuit	3	3	3	3	3	15
ClickMeetings	6	7	7	12	12	44
Element	-	-	-	3	4	7
Eyeson	-	-	-	1	1	2
GoToMeeting	2	3	3	3	3	14
Google	3	3	3	2	2	13
HiBox	-	-	2	2	2	6
Lifesize	2	3	3	3	4	15
PGI	1	1	2	2	2	8
Proficonf	-	-	-	2	2	4
Ring Central	3	3	4	4	4	18
Microsoft Teams	-	-	-	-	2	2
Uber	1	1	1	1	1	5
UMeeting	-	-	-	-	4	4
Cisco Webex	3	3	3	3	2	14
Whereby	-	-	2	2	2	6
Zoho	4	5	5	5	8	27
Zoom	3	3	3	3	3	15
Total Plan Types	33	37	45	59	69	243
Number of Providers	12	12	15	20	22	22

Note: The table shows the number of plan types for each service provider for each year.

Table 3.A.2: Number of plan types for each email provider for each year

	2017	2018	2019	2020	2021	All Years
Ctemplar	-	6	7	4	4	21
Hey	-	-	-	1	2	3
Hushmail	-	-	-	3	3	6
Kolab	2	3	2	2	2	11
Mailbox	6	5	7	7	3	28
Mailfence	2	2	2	3	3	12
Outlook	-	-	-	2	2	4
Pesteo	1	1	1	1	1	5
Rickspace	-	-	-	-	3	3
Runbox	4	4	4	4	4	20
Soverin	1	1	1	1	1	5
Thexyz	3	3	3	3	3	15
Tutanota	4	2	2	2	2	12
Zoho	3	3	3	4	4	17
Total Plan Types	26	30	32	37	37	162
Number of Providers	9	10	10	13	14	14

Note: The table shows the number of plan types for each service provider for each year.

Table 3.A.3: Number of plan types for each online news provider for each Year

	2017	2018	2019	2020	2021	All Years
Bloomberg	1	1	1	1	1	5
Daily Mail	2	2	2	2	2	10
FT	2	2	2	2	2	10
Independent	1	1	1	2	2	7
NYT	2	1	1	1	1	6
Telegraph	2	2	2	2	2	10
The Economist	1	1	1	1	1	5
The Guardian	1	1	1	1	1	5
The Times	1	1	1	1	1	5
WSJ	1	1	1	1	1	5
Total Plan Types	14	13	13	14	14	68
Number of Providers	10	10	10	10	10	10

Note: The table shows the number of plan types for each service provider for each year.

3.B Detailed description of the data

In this section, we describe in detail the data employed for the study. We show the mean, standard deviation, minimum and maximum values for each provider. We also show the count and corresponding share of each characteristic included in the regression.

3.B.1 Videoconferencing

The standard descriptive statistics of the monthly price for each provider of videoconferencing services are shown in table 3.B.1. From the descriptive statistics, it can be noticed that the range between the minimum and maximum prices is large [$1.0 to £750]. The standard deviation is relatively large as well, which in this case is $88.4 (more than double the average monthly price). The average price for the pooled data set is at $40.7. The likely reason why this is so is because of the presence of services that are dedicated and optimized for webinars and large online conferences.

Videoconferencing service providers cater to two types of customers 1) those that require a venue for online meetings and 2) those needing to reach a broader audience (with 100 or more participants). Zoom, one of the most popular videoconferencing service providers at the time of the writing of this manuscript, was able to cater to both types of customers. However, some service providers opted to specialize and cater to the second type of customers[18]. These services are often priced higher than those that are targeted for smaller online meetings.

[18] In the case of Zoho and ezTalks, they offer separate plan types for the two sets of customers.

Table 3.B.1: Summary statistics of monthly subscription price by videoconferencing provider

	Ave	SD	Min	Max
3CX (assuming 2)	1.2	0.2	1.0	1.5
Adobe Connect	90.0	46.2	50.0	130.0
Blackboard	508.3	418.6	25.0	750.0
Bluejeans	19.9	21.1	10.0	83.0
Circuit	9.8	5.5	4.5	17.1
ClickMeetings	109.5	108.4	25.0	500.0
Element	2.9	0.9	2.0	4.0
Eyeson	9.0	0.0	9.0	9.0
GoToMeeting	14.7	7.1	7.4	26.5
Google	11.2	7.0	4.1	25.0
HiBox	6.0	2.2	4.0	8.0
Lifesize	20.2	8.0	12.5	44.0
PGI	16.5	6.2	12.0	24.0
Proficonf	18.5	7.5	12.0	25.0
Ring Central	12.2	4.8	6.2	19.5
Microsoft Teams	8.5	4.9	5.0	12.0
Uber	15.0	0.0	15.0	15.0
UMeeting	56.2	65.0	10.0	150.0
Cisco Webex	22.1	7.8	13.5	39.0
Whereby	18.7	20.4	7.0	60.0
Zoho	24.6	17.5	2.5	63.0
Zoom	17.9	2.7	12.4	20.0
Total	40.7	88.4	1.0	750.0

Note: Table shows the mean, standard deviation, and the range for the monthly price for each of the videoconferencing service providers. Price data are expressed in USD.

In table 3.B.2, the sample is divided into two segments: plan types that focus on participants requiring large audiences or "webinar-focused" plans, and those that do not. We identified 21 webinar-focused plan types in our data sets. As anticipated, the monthly average price is higher for the services that are focused on webinars compared to those that are not. Plan types that are not focused on webinars have a mean monthly price of $14.6, substantially lower than the average for the pooled data set. Due to this source of heterogeneity, we control for whether the plan type is webinar-focused in the hedonic regressions.

Table 3.B.2: Comparison between webinar and non-webinar providers

	Ave	SD	Min	Max
Webinar focused				
Price	112.2	149.1	10.0	750.0
Participants	374.6	885.0	25.0	5,000.0
Non-Webinar focused				
Price	14.6	9.6	1.0	60.0
Participants	112.7	107.3	5.0	1,000.0
Total				
Price	40.7	88.4	1.0	750.0
Participants	182.7	478.6	5.0	5,000.0

Note: Table shows mean, standard deviation, and the range for the variable's monthly price and the number of participants. The first panel is restricted to webinar-focused plans. The second panel is restricted to non-webinar focused plan. The third panel comprised all plans in the data set. Price data are expressed in USD.

Twenty-six characteristics were identified for inclusion in the hedonic regression. These are: Number of participants, the ability to download recording, digital whiteboard, screen sharing, media/file sharing/storage, breakout rooms poll/Q&A/raise hand, virtual background, admin control, share control, transcription, multiple hosts, single sign-on, streaming, analytics/statistics/reporting, custom domain, branding, local and international calls, translations, Microsoft integration, encryption, HD quality noise/echo cancellation, multishare, permanent meeting rooms, calendar. In this set, only the number of participants is a continuous variable. The rest are dummy variables that take the value of 1 if the characteristic is present, zero otherwise.

Table 3.B.4 presents the number of observations possessing each of the respective characteristics in the data set, as well as the percent share of the observation with said characteristics. As mentioned earlier (with the exception of the number of participants), each characteristic would be represented by a dummy variable with a value of 1 if the characteristic is present in the particular plan, 0 otherwise. Information on the characteristics of these services was acquired through the scraping of the service providers' websites. One limitation of this approach is that our information on characteristics is dependent on whether the service provider was able to accurately reflect the features of their services on their websites. In order to address this, we subscribed to the trial versions of these services in order to validate the presence (or absence) of the said characteristics for each of the providers.

Table 3.B.3: Summary statistics of videoconferencing data over time

	Ave	SD	Min	Max
2017				
Price	25.5	25.6	4.5	145.0
Participants	72.5	95.2	5.0	500.0
Price per Participant	0.8	0.7	0.1	2.8
2018				
Price	38.7	83.2	4.5	500.0
Participants	228.4	812.0	5.0	5000.0
Price per Participant	0.6	0.7	0.0	2.8
2019				
Price	34.6	76.3	1.2	500.0
Participants	237.6	743.6	6.0	5000.0
Price per Participant	0.4	0.5	0.0	2.8
2020				
Price	48.5	107.0	1.1	750.0
Participants	173.3	229.4	5.0	1500.0
Price per Participant	0.4	0.6	0.0	2.8
2021				
Price	46.4	100.5	1.0	750.0
Participants	183.2	223.6	5.0	1500.0
Price per Participant	0.4	0.6	0.0	2.8

Note: Table shows the mean, standard deviation, and the range for the variables' monthly price, the number of participants, and the monthly price per participant ($price^t/participants^t$) for each year from 2017 to 2021. Data on prices and price per participant are expressed in USD.

Table 3.B.4 shows the number of plan types possessing each of the respective characteristics in Panel A and their proportion to the total number of plan types in Panel B. In terms of their count and share of the total number of participants, videoconferencing services in recent years possess more characteristics than past years. This may be a manifestation of how technology is improving these services. Moreover, the pandemic might have contributed to this, forcing some service providers to offer more services because the demand for videoconferencing substantially rose during the lockdowns.

Table 3.B.4: Frequency and share of premium-exclusive features from videoconferencing across time

	Panel A: Count						Panel B: Percent Share					
	2017	2018	2019	2020	2021	All Years	2017	2018	2019	2020	2021	All Years
Download Recording	25	28	33	47	55	188	75.8	75.7	73.3	79.7	79.7	77.4
Whiteboard	16	15	15	31	39	116	48.5	40.5	33.3	52.5	56.5	47.7
Screen sharing	32	36	40	56	66	230	97	97.3	88.9	94.9	95.7	94.7
Media/File Sharing/Storage	17	19	31	48	55	170	51.5	51.4	68.9	81.4	79.7	70
Breakout rooms	9	11	11	28	30	89	27.3	29.7	24.4	47.5	43.5	36.6
Poll/QnA/Raise hand	12	13	19	33	39	116	36.4	35.1	42.2	55.9	56.5	47.7
Virtual Background	3	3	3	6	13	28	9.1	8.1	6.7	10.2	18.8	11.5
Admin Control	12	9	18	43	55	137	36.4	24.3	40	72.9	79.7	56.4
Share control	7	7	7	21	23	65	21.2	18.9	15.6	35.6	33.3	26.7
Transcription	8	9	13	26	33	89	24.2	24.3	28.9	44.1	47.8	36.6
Multiple hosts	2	2	5	5	7	21	6.1	5.4	11.1	8.5	10.1	8.6
Single Sign On	6	7	7	13	18	51	18.2	18.9	15.6	22	26.1	21
Streaming	11	13	11	17	19	71	33.3	35.1	24.4	28.8	27.5	29.2
Analytic/Statistics/Reporting	13	14	21	36	41	125	39.4	37.8	46.7	61	59.4	51.4
Custom Domain	2	2	6	8	16	34	6.1	5.4	13.3	13.6	23.2	14
Branding	12	15	18	27	29	101	36.4	40.5	40	45.8	42	41.6
Local and International Calls	16	18	22	29	36	121	48.5	48.6	48.9	49.2	52.2	49.8
Translations	6	7	7	12	14	46	18.2	18.9	15.6	20.3	20.3	18.9
Microsoft Integration	2	2	6	10	16	36	6.1	5.4	13.3	16.9	23.2	14.8
Encryption	11	15	13	18	28	85	33.3	40.5	28.9	30.5	40.6	35
HD Quality	18	19	22	31	29	119	54.5	51.4	48.9	52.5	42	49
Noise/Echo Cancellation	0	0	0	0	2	2	0	0	0	0	2.9	0.8
Multishare	3	3	3	3	8	20	9.1	8.1	6.7	5.1	11.6	8.2
Permanent meeting rooms	4	8	9	12	16	49	12.1	21.6	20	20.3	23.2	20.2
Calendar	18	21	18	21	36	114	54.5	56.8	40	35.6	52.2	46.9

Note: Table shows the number of plan types possessing each of the respective characteristics in Panel A and their proportion to the total number of plan types in Panel B.

Note that these are not apples-to-apples comparisons of the data. Data on some providers are only available in recent years, either because the past versions of their websites were not archived or they only began operations recently.

3.B.2 Email

The standard set of descriptive statistics was also generated for the price data on personal email (see table 3.B.5). Similar to videoconferencing, the range for the pooled data set is noticeably large [$0.8 to $57.0]. The mean of the pooled data set is $7.8 while the standard deviation is at $10.1.

Table 3.B.5: Summary statistics of monthly subscription price by email provider

	Ave	SD	Min	Max
Ctemplar	15.2	12.0	8.0	50.0
Hey	5.8	4.2	1.0	8.3
Hushmail	6.7	2.7	4.2	10.0
Kolab	5.5	1.8	4.2	9.6
Mailbox	11.0	15.4	1.1	57.0
Mailfence	9.5	9.3	2.9	28.5
Pesteo	1.1	0.0	1.1	1.1
Rickspace	4.7	2.1	3.0	7.0
Runbox	3.8	1.9	1.7	6.7
Soverin	3.7	0.0	3.7	3.7
Thexyz	7.1	5.3	1.9	14.9
Tutanota	8.2	15.7	1.1	57.0
Zoho	3.5	2.3	0.8	8.0
Total	7.8	10.1	0.8	57.0

Note: Table shows the mean, standard deviation, and the range for the monthly price for each email service provider. Price data are expressed in USD.

We identified 10 characteristics that can be included in the hedonic regression. These are: mail storage space, calendar, the availability of a mobile application specific to the email provider, data encryption, domain customization, virus and malware filters, availability of aliases, availability of email templates, VPN function, and chat functions. Of the 10 characteristics, only mail storage is a continuous variable. The rest are dummy variables that take the value of 1 when the characteristic is present. Similar to videoconferencing, email services have more premium-exclusive features in later years. The most common feature is Custom Domain domain, which is present in 89.2 percent of the plan types. The least common feature is Email Template, which is present in only 7.4 percent of the plan types.

Table 3.B.6: Summary statistics of email data over time

	Ave	SD	Min	Max
		2017		
Price	7.8	11.6	1.1	57.0
Storage	57.6	193.9	1.0	1000.0
Price per GB	0.7	0.7	0.1	2.6
		2018		
Price	6.1	3.9	1.1	14.9
Storage	50.3	181.0	1.0	1000.0
Price per GB	0.8	0.7	0.0	2.6
		2019		
Price	8.2	10.6	1.0	57.0
Storage	19.3	24.9	1.0	1000.0
Price per GB	0.8	0.7	0.1	2.6
		2020		
Price	9.7	12.9	1.0	57.0
Storage	26.4	30.7	1.0	1000.0
Price per GB	0.7	0.6	0.1	2.6
		2021		
Price	7.2	9.3	0.8	50.0
Storage	25.3	27.4	1.0	1000.0
Price per GB	0.5	0.5	0.0	1.7

Note: Table shows mean, standard deviation, and the range for the variables monthly price, mail storage capacity, and the monthly price per gigabyte of storage ($price^t/storage^t$) for each year from 2017 to 2021. Data on prices and prices per participant are expressed in USD. Data on storage capacity is expressed in gigabytes.

Table 3.B.7: Frequency and share of premium-exclusive features from email services across time

	Panel A: Count						Panel B: Percent Share					
	2017	2018	2019	2020	2021	All Years	2017	2018	2019	2020	2021	All Years
Calendar	17	17	18	23	23	98	65.4	56.7	56.3	62.2	62.2	60.5
Mobile App	9	8	10	13	14	54	34.6	26.7	31.3	35.1	37.8	33.3
Encryption	18	23	24	28	26	119	69.2	76.7	75	75.7	70.3	73.5
Custom Domain	23	28	30	33	26	140	88.5	93.3	93.8	89.2	70.3	86.4
Virus Filters	9	12	14	16	15	66	34.6	40	43.8	43.2	40.5	40.7
Aliases	20	24	27	27	23	121	76.9	80	84.4	73	62.2	74.7
Email Template	4	2	2	2	2	12	15.4	6.7	6.3	5.4	5.4	7.4
VPN	1	1	1	5	5	13	3.8	3.3	3.1	13.5	13.5	8
Chat Function	2	2	2	3	8	17	7.7	6.7	6.3	8.1	21.6	10.5

Note: Table shows the number of plan types possessing each of the respective characteristics in Panel A and their proportion to the total number of plan types in Panel B.

3.B.3 Online News

The standard descriptive statistics for the price of online news are shown in table 3.B.8. For the pooled data set, the prices of online news subscription ranges between \$3.1 to \$67.0. The average price of news subscription is \$15.4. It can be noticed that the average price of business and financial news providers such as the Wall Street Journal, Bloomberg, and the Financial Times are generally higher than those of the other providers. Because of this, we include a dummy variable for business-focused news providers.

Table 3.B.8: Summary statistics of monthly subscription price by online news provider

	Ave	SD	Min	Max
Daily Mail	14.4	6.2	8.6	20.9
Independent	10.5	2.4	7.0	15.0
New York Times	8.0	5.4	4.0	17.5
The Telegraph	10.5	5.2	6.2	18.7
The Guardian	20.0	–	20.0	20.0
The Times	3.1	–	3.1	3.1
The Economist	4.0	–	4.0	4.0
The Wall Street Journal	19.1	0.5	18.5	19.5
Bloomberg	35.8	2.4	34.0	40.0
Financial Times	26.5	24.9	6.5	67.0
Total	15.4	13.4	3.1	67.0

Note: Table shows mean, standard deviation, and the range for the variable's monthly price of paid online news providers. Price data are expressed in USD.

Only eight characteristics were identified for inclusion in the hedonic regression for online news. These are: perks and freebies, access to games and puzzles, live feed of breaking news, access to multimedia content, access to the weekly newsletter, access to the digital version of the paper, and access to premium content. Unlike videoconferencing and email, all of the variables for the hedonic regression are categorical.

Table 3.B.9: Summary statistics of online news prices over time

	Ave	SD	Min	Max
2017	13.4	8.5	3.1	34.0
2018	13.0	9.0	3.1	35.0
2019	12.1	9.2	3.1	35.0
2020	18.7	17.8	3.1	67.0
2021	19.2	18.3	3.1	67.0

Note: Table shows the mean, standard deviation, and the range for the variable's monthly prices of online news for each year from 2017 to 2021. Data on prices are expressed in USD.

As with videoconferencing and email, the data shows a gradual increase in premium exclusive characteristics over the years. Among the news sites, the most common premium-exclusive feature is access to games and puzzles, which appeared in 45.6 percent of observations. The least common is access to the weekly newsletter and digital versions of the paper, both of which appeared in 19.1 percent of the observations.

Table 3.B.10: Frequency and share of premium-exclusive features from online news services across time

	Panel A: Count						Panel B: Percent Share					
	2017	2018	2019	2020	2021	All Years	2017	2018	2019	2020	2021	All Years
Perks	6	5	5	7	7	30	42.9	38.5	38.5	50	50	44.1
Games and Puzzles	5	5	6	7	8	31	35.7	38.5	46.2	50	57.1	45.6
Breaking News	1	1	2	2	4	10	7.1	7.7	15.4	14.3	28.6	14.7
Multi Media	4	4	4	6	6	24	28.6	30.8	30.8	42.9	42.9	35.3
Newsletters	3	3	3	3	1	13	21.4	23.1	23.1	21.4	7.1	19.1
Digital Paper	2	2	2	3	4	13	14.3	15.4	15.4	21.4	28.6	19.1
Premium Content	4	5	5	5	5	24	28.6	38.5	38.5	35.7	35.7	35.3

Note: Table shows the number of plan types possessing each of the respective characteristics in Panel A and their proportion to the total number of plan types in Panel B.

3.B.4 Hedonic Regression Results

We estimate the hedonic regression model in equation 3.3 using ordinary least squares. We show the regression results in the section. Each coefficient estimate for the hedonic regression represents the marginal contribution of each characteristic to the log price of videoconferencing. The exponential of each coefficient can also be interpreted as the WTP for the said characteristic.

For videoconferencing (see figure 3.B.1), it can be noticed that not all coefficients are statistically significant. This suggests that the presence of some characteristics probably does not contribute substantially to the variations in prices across plan types and/or across service providers. We also observe the presence of negative coefficients that are statistically significant. If we were to interpret each coefficient as the marginal contribution of each characteristic to the price, it stands to reason that none of the variables should have a negative value for their coefficients. We offer two likely explanations for this. First, it is important to note that the coefficient estimates are partial elasticities and that we can only arrive at the marginal contribution of each characteristic by applying exponential transformations to the coefficients. In this case, the transformation would yield a positive value that is close to zero. Second, Erickson (2016) shows that it is possible for hedonic regressions to generate negative coefficient estimates if there are trade-offs between the characteristic with the negative coefficient and other characteristics in the regression. For instance, the trade-off between horsepower and mileage could result in negative coefficient estimates in a hedonic regression for automobiles. In the case of this exercise, only *Encryption* yielded a negative coefficient that is statistically significant. An examination of the correlation between covariates (see appendix 3.D) shows that the presence of encryption is negatively correlated with some of the statistically significant explanatory variables in the hedonic regression[19]. One can argue that the presence of these features makes it difficult to make calls more secure.

A major limitation of the panel hedonic regression is that it assumes that the marginal values of characteristics are fixed over time. It is possible that this assumption may not be true. From the descriptive statistics in table 3.4.1, we show that the average price per participant varies across years. We generate a second regression where we interact the time dummies with the natural log of the maximum participants (z_1). This effectively generates a separate coefficient for the log of participants for each year.

Allowing the coefficient for log participants to vary over time does not cause any substantial changes to the values of the other coefficients (see left panel on of figure 3.B.1.). Moreover, the yearly coefficient for the log of participants does not seem to be statistically different from the coefficient estimate of the said variable in our regression where it is kept fixed for all years (see right panel on of figure 3.B.1.). This implies that having fixed coefficients might be sufficient for imputing the shadow price of videoconferencing.

[19]These variables and their respective correlation coefficients with respect to Encryption are: File Sharing (-0.33), Breakout Rooms (-0.25), HD Quality (-0.13), and Log Participants (-0.11). See appendix for the correlation matrix.

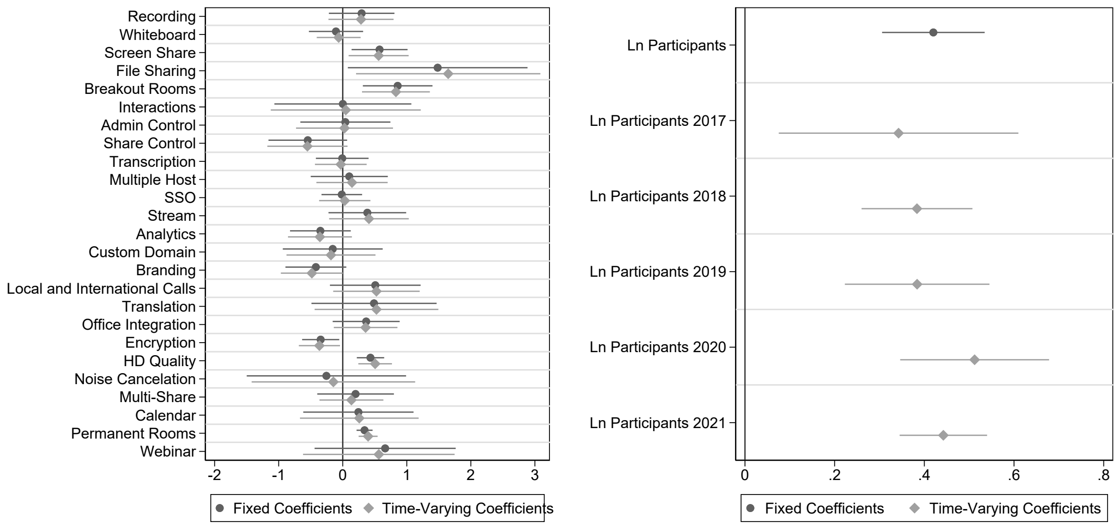


Figure 3.B.1: Coefficient plot for videoconferencing

Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the hedonic regression in equation 3.4. The coefficients represent column (3) of tables 3.C.1 and 3.C.2 in appendix 3.C.

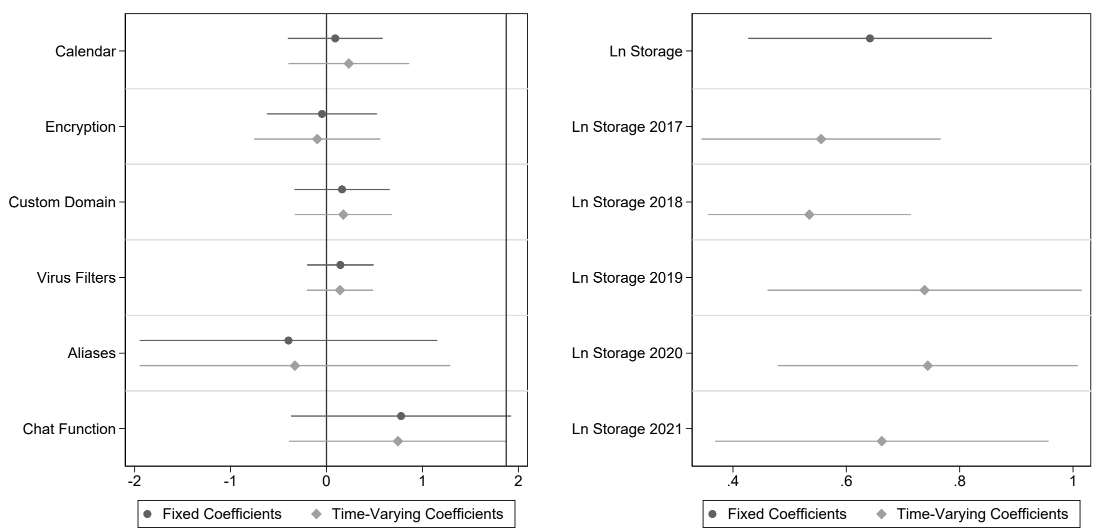


Figure 3.B.2: Coefficient plot for email

Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the hedonic regression in equation 3.4. The coefficients represent column (2) of table 3.C.3 in appendix 3.C.

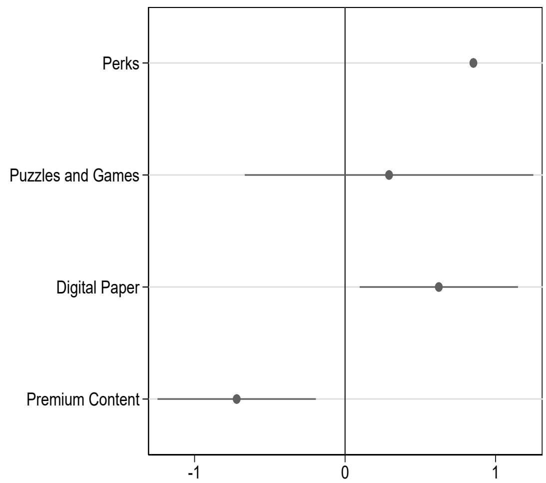


Figure 3.B.3: Coefficient plot for online news

Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the hedonic regression in equation 3.4. The coefficients represent column (2) of table 3.C.4 in appendix 3.C.

With the exception of the natural log of mail storage (in gigabytes), none of the characteristics included in the hedonic regression for email was statistically significant (see figure 3.B.2). The presence of chat functions generated a relatively large coefficient, however, its confidence interval still incorporated zero. As with videoconferencing, we interact the continuous variable (mail storage) with the time dummies in order to determine whether the coefficient would materially vary across time. Allowing the coefficient for mail storage to vary across time does not cause the estimates for the other coefficients to change. Moreover, the yearly estimates for the coefficients are not statistically different from the coefficients generated by the regression assuming that the parameter is fixed across time.

For online news, only the availability of perks (freebies and other offers) and the digital version of the paper were statistically different from zero. The coefficient for the availability of premium content is negative. We offer the same argument earlier regarding negative coefficients.

3.C Hedonic regression coefficients

Table 3.C.1: Hedonic regression results for videoconferencing

	(1)	(2)	(3)	(4)
Log Participants	0.335***	0.249***	0.435***	0.420***
	(0.062)	(0.079)	(0.053)	(0.055)
Recording	0.086	0.292	0.287	0.295
	(0.236)	(0.229)	(0.248)	(0.246)
Whiteboard	0.170	−0.055	0.067	−0.105
	(0.312)	(0.227)	(0.179)	(0.204)
Screen Share	0.623**	0.668***	0.533**	0.575**
	(0.238)	(0.224)	(0.195)	(0.210)
File Sharing	−0.214	−0.105	1.679**	1.482**
	(0.354)	(0.256)	(0.773)	(0.674)
Breakout Rooms	0.618	0.724**	0.802***	0.858***
	(0.376)	(0.282)	(0.249)	(0.261)
Interactions	0.462	−0.161	0.235	0.002
	(0.331)	(0.261)	(0.585)	(0.514)
Virtual Background	−0.965**	−0.529	0.000	0.000
	(0.352)	(0.363)	(.)	(.)
Admin Control	0.036	−0.009	0.016	0.041
	(0.232)	(0.223)	(0.364)	(0.338)
Share Control	−0.296	−0.113	−0.773*	−0.546*
	(0.229)	(0.174)	(0.384)	(0.295)
Transcription	−0.328	−0.573	0.018	−0.008
	(0.386)	(0.350)	(0.266)	(0.198)
Multiple Host	−0.522	−0.041	0.076	0.102
	(0.452)	(0.270)	(0.307)	(0.289)
SSO	0.372	0.584**	−0.060	−0.015
	(0.225)	(0.216)	(0.164)	(0.152)
Stream	0.327	0.178	0.385	0.384
	(0.269)	(0.211)	(0.281)	(0.292)
Analytics	−0.077	−0.508**	−0.250	−0.349
	(0.336)	(0.214)	(0.191)	(0.228)
Custom Domain	−0.093	−0.021	0.062	−0.156
	(0.286)	(0.248)	(0.257)	(0.375)
Branding	0.416	0.065	−0.428*	−0.420*
	(0.333)	(0.247)	(0.238)	(0.228)
Local and International Calls	0.057	0.070	0.389	0.508
	(0.231)	(0.201)	(0.404)	(0.341)
Translation	0.154	−0.268	0.346	0.489
	(0.540)	(0.388)	(0.371)	(0.470)
Office Integration	−0.831	−0.450	0.340	0.366
	(0.498)	(0.320)	(0.227)	(0.251)
Encryption	0.162	−0.053	−0.352**	−0.345**
	(0.302)	(0.195)	(0.145)	(0.139)
HD Quality	0.306	0.378**	0.407***	0.433***
	(0.242)	(0.174)	(0.107)	(0.103)
Noise Cancelation	0.052	−0.001	−0.522	−0.255
	(0.753)	(0.320)	(0.550)	(0.598)
Multi-Share	0.317	0.406	0.364	0.200
	(0.436)	(0.306)	(0.388)	(0.288)
Calendar	−0.114	−0.018	0.238	0.244
	(0.307)	(0.188)	(0.401)	(0.414)
Permanent Rooms	0.031	0.241	0.350***	0.341***
	(0.213)	(0.211)	(0.076)	(0.060)
Webinar		2.099***		0.662
		(0.402)		(0.530)
Observations	243	243	243	243
Adjusted R^2	0.948	0.965	0.985	0.985
Service provider fixed effects × time dummy	No	No	Yes	Yes

Note: Table shows the results of the hedonic regression in equation 3.3. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) show the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider's fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.C.2: Hedonic regression results the coefficient for the number of participants varying over time

	(1)	(2)	(3)	(4)
Ln Participants 2017	0.131	0.045	0.347**	0.343**
	(0.092)	(0.089)	(0.125)	(0.128)
Ln Participants 2018	0.227***	0.185**	0.388***	0.384***
	(0.064)	(0.075)	(0.060)	(0.059)
Ln Participants 2019	0.265***	0.223**	0.387***	0.384***
	(0.066)	(0.082)	(0.078)	(0.078)
Ln Participants 2020	0.437***	0.337**	0.530***	0.512***
	(0.125)	(0.130)	(0.073)	(0.080)
Ln Participants 2021	0.497***	0.353***	0.470***	0.442***
	(0.098)	(0.101)	(0.053)	(0.047)
Observations	243	243	243	243
Adjusted R^2	0.948	0.965	0.985	0.985
Service provider fixed effects × time dummy	No	No	Yes	Yes

Note: Table shows the results of the hedonic regression in equation 3.3. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) show the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 3.C.3: Hedonic regression results for personal email

	(1)	(2)
Ln Storage	0.404***	0.642***
	(0.131)	(0.099)
Calendar	−0.229	0.091
	(0.405)	(0.227)
Mobile App	−0.074	0.000
	(0.387)	(.)
Encryption	0.724**	−0.047
	(0.292)	(0.264)
Custom Domain	0.163	0.162
	(0.264)	(0.228)
Virus Filters	−0.201	0.145
	(0.295)	(0.159)
Aliases	−0.219	−0.396
	(0.485)	(0.712)
Email Template	0.284	0.000
	(0.420)	(.)
VPN	0.448	0.000
	(0.397)	(.)
Chat Function	−0.141	0.778
	(0.310)	(0.527)
Observations	158	158
Adjusted R^2	0.851	0.932
Service provider fixed effects × time dummy	No	Yes

Note: Table shows the results of the hedonic regression in equation 3.3. Column (1) shows the results of the classical time dummy variable model. Column (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 3.C.4: Hedonic regression results for online news

	(1)	(2)
Puzzles and Games	0.126	0.292
	(0.260)	(0.424)
Breaking News	0.225	0.000
	(0.269)	(.)
Multimedia Content	0.245	0.000
	(0.295)	(.)
Newsletter	−0.109	0.000
	(0.197)	(.)
Share Subscription	0.175	0.000
	(0.465)	(.)
Digital Paper	0.546***	0.623**
	(0.150)	(0.233)
Premium Content	−1.054***	−0.720**
	(0.221)	(0.233)
Business	0.696**	0.000
	(0.304)	(.)
Observations	158	158
Adjusted R^2	0.851	0.932
Service provider fixed effects × time dummy	No	Yes

Note: Table shows the results of the hedonic regression in equation 3.3. Column (1) shows the results of the classical time dummy variable model. Column (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficient estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.D Correlation among characteristics

The coefficient estimates of each of the explanatory variables in the hedonic regression can be viewed as the marginal contribution of each of the respective characteristics to the price of the good being modeled. Another interpretation of the coefficients is that they represent the household's willingness to pay for the specific characteristics (see de Haan and Diewert (2013)).

Given the nature of hedonic regressions, one would expect that all dummy variables representing the presence of a characteristic should be positive. After all, the presence of an additional feature to any good or service can only contribute positively to its price. However, this is not what we observe from the results of the hedonic regression in table 3.C.1. The coefficient estimate for one of the variables (share control) was negative and significant at a 10 percent level. Moreover, the variable representing the presence of call encryption was also negative and is significant at a 5 percent level. This is in contrast with the intuition of interpreting coefficients for hedonic regressions. An individual cannot possibly have a negative value for their willingness to pay.

We offer two explanations for this. First, coefficient estimates for the semi-log specification represent partial elasticities and the WTP is derived by applying an exponential transformation to the estimate. The resulting WTP for negative coefficients would, in turn, be positive but close to zero, holding everything else constant. However, they would still have the ability to pull the predicted price when compounded with other characteristics.

Erickson (2016) shows that negative coefficient estimates are possible if there is a trade-off between other characteristics. An examination of the correlation between covariates shows that the presence of Encryption is negatively correlated with some of the statistically significant explanatory variables in the hedonic regression. These variables and their respective correlation coefficients with respect to Encryption are: File Sharing (-0.33), Breakout Rooms (-0.25), HD Quality (-0.13), and Log Participants (-0.11).

Table 3.D.1: Correlation matrix of characteristics for videoconferencing

(1)

	Encryption	Share Control	Log Participants	Recording	Whiteboard	Screen Share	File Sharing	Breakout Rooms	Interactions	Virtual Background	Admin Control	Share Control
Encryption	1											
Share Control	-0.112	1	0.154*	0.282***	0.353***	0.144*	0.0512	0.351***	0.260***	0.248***	0.344***	1
Log Participants	-0.110	0.154*	1									0.154*
Recording	-0.0157	0.282***	0.220***	1								0.282***
Whiteboard	-0.200**	0.353***	0.348***	0.399***	1							0.353***
Screen Share	0.174**	0.144*	-0.0558	0.352***	0.227***	1						0.144*
File Sharing	-0.329***	0.0512	0.0568	0.0746	0.303***	0.00378	1					0.0512
Breakout Rooms	-0.253***	0.351***	0.266***	0.370***	0.676***	0.181**	0.293***	1				0.351***
Interactions	-0.286***	0.260***	0.0781	0.399***	0.439***	0.227***	0.608***	0.607***	1			0.260***
Virtual Background	0.0596	0.248***	0.183**	0.164*	0.352***	0.0858	0.124	0.421***	0.274***	1		0.248***
Admin Control	0.0884	0.344***	0.100	0.357***	0.160*	0.270***	0.292***	0.273***	0.342***	0.317***	1	0.344***
Transcription	-0.0740	0.312***	0.255***	0.411***	0.471***	0.181**	0.442***	0.539***	0.659***	0.180**	0.255***	0.312***
Multiple Host	0.174**	0.211***	0.184**	0.166**	0.322***	0.0731	0.138*	0.313***	0.205**	0.669***	0.271***	0.211***
SSO	0.215***	0.0995	0.274***	0.0856	0.0739	0.123	-0.0811	0.0697	-0.0475	0.320***	0.127*	0.0995
Stream	-0.0348	0.245***	0.180**	0.304***	0.509***	0.153*	0.303***	0.451***	0.328***	-0.00513	-0.000526	0.245***
Analytics	-0.133*	0.252***	0.446***	0.360***	0.352***	0.172**	0.154*	0.482***	0.434***	0.170**	0.308***	0.252***
Custom Domain	0.00266	0.0779	0.119	0.0197	-0.0530	-0.220***	0.109	0.0873	0.0658	0.300***	0.0677	0.0779
Branding	-0.00576	0.283***	0.139*	0.356***	0.214***	0.201**	0.279***	0.329***	0.481***	0.00947	0.153*	0.283***
Local and International Calls	-0.0229	0.254***	0.217***	0.303***	0.0863	-0.0193	0.0960	0.268***	0.301***	0.0273	0.196**	0.254***
Translation	-0.310***	0.301***	0.179**	0.261***	0.506***	0.115	0.294***	0.570***	0.485***	-0.141*	0.00139	0.301***
Office Integration	0.0828	-0.0950	0.216***	-0.0513	-0.0739	-0.107	-0.131*	-0.149*	-0.190**	-0.0779	-0.147*	-0.0950
HD Quality	-0.132*	0.542***	0.295***	0.215***	0.349***	0.123	0.0134	0.280***	0.0856	0.214***	0.247***	0.542***
Noise Cancelation	-0.0668	-0.0550	-0.00905	0.0493	0.0953	0.0217	-0.139*	0.120	-0.0871	0.252***	0.0801	-0.0550
Multi-Share	-0.125	0.360***	0.132*	0.162*	0.283***	0.0712	0.131*	0.301***	0.253***	0.783***	0.263***	0.360***
Calendar	0.469***	-0.102	-0.221***	0.0158	-0.222***	0.223***	-0.229***	-0.167**	-0.139*	0.126	-0.00452	-0.102
Permanent Rooms	0.105	0.0207	-0.130*	-0.0713	-0.172**	0.119	-0.208**	-0.190**	-0.0491	-0.0208	-0.0336	0.0207
Observations	243											

	Transcription	Multiple Host	SSO	Stream	Analytics	Custom Domain	Branding	Local and International Calls	Translation	Office Integration	HD Quality	Noise Cancelation	Multi-Share	Calendar
Transcription	1													
Multiple Host	0.283***	1												
SSO	0.0487	0.345***	1											
Stream	0.507***	0.157*	-0.0422	1										
Analytics	0.534***	0.0644	0.157*	0.262***	1									
Custom Domain	0.186**	0.298***	0.200**	-0.207**	0.0121	1								
Branding	0.624***	0.0675	-0.0656	0.376***	0.335***	0.238***	1							
Local and International Calls	0.593***	0.133*	-0.0484	0.247***	0.556***	0.00166	0.513***	1						
Translation	0.636***	-0.111	-0.197**	0.752***	0.469***	-0.165*	0.573***	0.464***	1					
Office Integration	-0.149*	-0.0870	0.382***	-0.115	0.197**	-0.101	-0.305***	-0.0215	-0.172**	1				
HD Quality	0.195**	0.314***	0.122	0.384***	0.293***	-0.158*	0.0591	0.325***	0.283***	0.00858	1			
Noise Cancelation	-0.0693	-0.0280	-0.0470	-0.0585	0.0885	-0.0367	-0.0768	-0.0907	-0.0440	-0.0380	-0.0892	1		
Multi-Share	0.114	0.494***	0.250***	-0.0936	0.231***	0.268***	0.112	0.0611	-0.106	-0.125	0.186**	0.304***	1	
Calendar	-0.338***	0.151*	0.224***	-0.133*	-0.341***	0.0249	-0.174**	-0.326***	-0.391***	0.0490	-0.228***	0.0969	0.0185	1
Permanent Rooms	-0.254***	-0.00856	0.220***	-0.210***	-0.209**	-0.173**	-0.237***	-0.295***	-0.190**	-0.00749	-0.0820	-0.0458	-0.00123	0.370***

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.D.2: Correlation matrix of characteristics for personal email

	(1)								
	Calendar	Mobile App	Encryption	Custom Domain	Virus Filters	Aliases	Email Template	VPN	Chat Function
Calendar	1								
Mobile App	0.478***	1							
Encryption	0.0748	0.324***	1						
Custom Domain	-0.0339	0.195*	0.370***	1					
Virus Filters	0.267***	0.763***	0.433***	0.211**	1				
Aliases	0.0613	0.183*	0.434***	0.598***	0.230**	1			
Email Template	-0.250**	-0.195*	-0.469***	0.115	-0.230**	0.159*	1		
VPN	-0.128	-0.204*	-0.0757	-0.212**	-0.241**	-0.106	-0.0858	1	
Chat Function	0.162*	0.203*	0.0288	-0.214**	0.265***	0.0473	-0.0995	-0.104	1
Observations	158								

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.D.3: Correlation matrix of characteristics for online news

	(1)								
	Perks	Puzzles and Games	Breaking News	Multimedia Content	Newsletter	Share Subscription	Digital Paper	Premium Content	Bussiness
Perks	1								
Puzzles and Games	0.376**	1							
Breaking News	-0.202	-0.0466	1						
Multimedia Content	-0.408***	-0.244*	0.388**	1					
Newsletter	0.171	-0.445***	0.326**	0.0322	1				
Share Subscription	0.196	0.190	0.173	-0.129	-0.0846	1			
Digital Paper	0.171	-0.145	-0.0963	-0.281*	0.144	0.137	1		
Premium Content	-0.408***	-0.120	-0.307*	0.0985	-0.359**	-0.129	0.0322	1	
Bussiness	-0.677***	-0.698***	0.114	0.394***	0.0171	-0.133	0.0171	0.394***	1
Observations	68								

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.E Quality-Adjusted Price Indices

In a typical cross-section hedonic regression, the intercept term would represent the quality-adjusted price index[20] of the good being analyzed (see de Haan and Diewert (2013)). An alternative to the cross-section hedonic regression is the time dummy variable model, which allows for the intercept term to vary over time. The model can be written as,

$$log(p^t_{i,j}) = \sum_{t=1}^{T} \tau^t + \sum_{k=1}^{K} \beta_k Z_{i,j} + \varepsilon_{i,j} \tag{3.6}$$

where τ represents the quality-adjusted price index for the given year t, Z_i is a matrix of characteristics that affects the price p^t_i and the error term ε_i is assumed to be normally distributed with zero mean and constant variance.

In this study, however, we emphasized a modified time dummy variable model, which allows for the intercept term to vary for different service provider at different points in time. The specification for the model is given by equation 3.3. We present the quality-adjusted price indices for both specifications in tables 3.E.1, 3.E.2, and 3.E.3.

[20]For hedonic regressions that employ the log of price as the outcome variable, it is the exponential of the intercept term that represents the quality-adjusted price index.

Table 3.E.1: Coefficient estimates of the service provider dummies for videoconferencing

	(1)	(2)	(3)	(4)
year=2017	0.497	0.749		
	(0.367)	(0.437)		
year=2018	0.378	0.643		
	(0.349)	(0.401)		
year=2019	0.354	0.533		
	(0.323)	(0.424)		
year=2020	0.197	0.403		
	(0.331)	(0.394)		
year=2021	0.332	0.476		
	(0.325)	(0.397)		
3CX (assuming 2) 2019			−2.663	−2.643
			(0.625)	(0.607)
3CX (assuming 2) 2020			−5.324	−4.718
			(0.805)	(0.816)
3CX (assuming 2) 2021			−5.218	−4.593
			(0.623)	(0.755)
Adobe Connect 2020			−1.168	−0.816
			(0.603)	(0.657)
Adobe Connect 2021			−1.168	−0.816
			(0.603)	(0.657)
Blackboard 2020			−0.023	−0.236
			(1.078)	(1.011)
Blackboard 2021			0.851	0.884
			(0.795)	(0.786)
Bluejeans 2017			0.143	−0.131
			(0.667)	(0.690)
Bluejeans 2018			0.142	−0.131
			(0.667)	(0.690)
Bluejeans 2019			0.034	0.058
			(0.525)	(0.511)
Bluejeans 2020			−0.589	−0.718
			(0.552)	(0.561)
Bluejeans 2021			−1.211	−1.505
			(0.722)	(0.709)
Circuit 2017			−1.528	−1.143
			(0.901)	(0.961)
Circuit 2018			−2.146	−1.770
			(0.909)	(0.965)
Circuit 2019			−2.146	−1.770
			(0.909)	(0.965)
Circuit 2020			−2.146	−1.770
			(0.909)	(0.965)
Circuit 2021			−2.146	−1.770
			(0.909)	(0.965)
ClickMeetings 2017			−2.295	−2.554
			(0.976)	(1.024)
ClickMeetings 2018			−2.201	−2.454
			(1.021)	(1.070)
ClickMeetings 2019			−2.167	−2.419
			(0.992)	(1.039)
ClickMeetings 2020			−1.352	−1.869
			(0.945)	(1.027)
ClickMeetings 2021			−1.218	−1.727
			(0.941)	(1.020)
Service provider fixed effects × time dummy	No	No	Yes	Yes

	(1)	(2)	(3)	(4)
Element 2020			−1.899	−1.736
			(0.800)	(0.697)
Element 2021			−1.895	−1.736
			(0.807)	(0.703)
Eyeson 2020			−2.689	−2.474
			(1.024)	(0.887)
Eyeson 2021			−2.721	−2.530
			(1.072)	(0.946)
GoToMeeting 2017			0.686	0.621
			(0.211)	(0.218)
GoToMeeting 2018			0.097	0.053
			(0.222)	(0.233)
GoToMeeting 2019			0.097	0.053
			(0.222)	(0.233)
GoToMeeting 2020			−1.182	−1.372
			(0.691)	(0.680)
GoToMeeting 2021			−1.805	−1.978
			(0.717)	(0.742)
Google 2017			−1.894	−1.698
			(1.115)	(1.044)
Google 2018			−2.491	−2.271
			(1.200)	(1.115)
Google 2019			−2.491	−2.271
			(1.200)	(1.115)
Google 2020			−3.301	−2.712
			(1.124)	(1.128)
Google 2021			−3.329	−2.843
			(1.085)	(1.110)
HiBox 2019			−2.016	−1.533
			(0.904)	(0.770)
HiBox 2020			−2.016	−1.533
			(0.904)	(0.770)
HiBox 2021			−2.016	−1.533
			(0.904)	(0.770)
Lifesize 2017			1.244	1.396
			(0.470)	(0.437)
Lifesize 2018			0.526	0.500
			(0.513)	(0.512)
Lifesize 2019			0.526	0.500
			(0.513)	(0.512)
Lifesize 2020			−0.681	−0.663
			(0.582)	(0.578)
Lifesize 2021			−0.697	−0.583
			(0.591)	(0.567)
PGI 2017			−1.446	−1.473
			(0.526)	(0.527)
PGI 2018			−1.446	−1.473
			(0.526)	(0.527)
PGI 2019			−2.007	−1.899
			(1.038)	(1.019)
PGI 2020			−1.977	−1.892
			(1.023)	(1.007)
PGI 2021			−1.909	−1.643
			(1.019)	(1.034)
Service provider fixed effects × time dummy	No	No	Yes	Yes

	(1)	(2)	(3)	(4)
Proficonf 2020			−2.673	−2.297
			(0.995)	(0.802)
Proficonf 2021			−2.138	−1.757
			(1.090)	(0.916)
Ring Central 2017			−0.799	−0.845
			(0.580)	(0.550)
Ring Central 2018			−0.564	−0.615
			(0.555)	(0.524)
Ring Central 2019			−2.803	−2.701
			(1.078)	(0.966)
Ring Central 2020			−2.803	−2.701
			(1.078)	(0.966)
Ring Central 2021			−2.789	−2.692
			(1.010)	(0.916)
Microsoft Teams 2021			−4.029	−3.243
			(1.030)	(1.150)
Uber 2017			−0.677	−0.730
			(0.573)	(0.544)
Uber 2018			−0.677	−0.730
			(0.573)	(0.544)
Uber 2019			−0.677	−0.730
			(0.573)	(0.544)
Uber 2020			−0.677	−0.730
			(0.573)	(0.544)
Uber 2021			−0.677	−0.730
			(0.573)	(0.544)
Cisco Webex 2017			−1.450	−1.159
			(1.324)	(1.192)
Cisco Webex 2018			−1.584	−1.271
			(1.287)	(1.150)
Cisco Webex 2019			−2.348	−2.066
			(1.454)	(1.290)
Cisco Webex 2020			−3.423	−2.866
			(1.064)	(1.110)
Cisco Webex 2021			−4.042	−3.676
			(1.189)	(1.221)
Whereby 2019			1.378	1.585
			(0.397)	(0.480)
Whereby 2020			−1.373	−1.209
			(0.662)	(0.684)
Whereby 2021			−0.380	−0.145
			(0.416)	(0.419)
Service provider fixed effects × time dummy	No	No	Yes	Yes

	(1)	(2)	(3)	(4)
Whereby 2019			1.378	1.585
			(0.397)	(0.480)
Whereby 2020			−1.373	−1.209
			(0.662)	(0.684)
Whereby 2021			−0.380	−0.145
			(0.416)	(0.419)
Zoho 2017			1.160	1.009
			(0.602)	(0.582)
Zoho 2018			1.173	1.026
			(0.609)	(0.588)
Zoho 2019			−0.725	−0.956
			(0.691)	(0.705)
Zoho 2020			−0.570	−0.847
			(0.698)	(0.715)
Zoho 2021			−1.310	−1.320
			(0.911)	(0.877)
Zoom 2017			−2.566	−1.995
			(0.918)	(0.970)
Zoom 2018			−2.609	−2.027
			(0.697)	(0.762)
Zoom 2019			−2.842	−2.252
			(0.706)	(0.768)
Zoom 2020			−2.668	−2.083
			(0.699)	(0.763)
Zoom 2021			−3.969	−3.540
			(1.138)	(1.166)
Service provider fixed effects × time dummy	No	No	Yes	Yes

Note: Table shows the coefficient estimates for the year fixed effects and service provider dummies equation 3.3. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) show the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider's fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.E.2: Coefficient estimates of the service provider dummies for personal email

	(1)	(2)
year=2017	0.251	
	(0.548)	
year=2018	0.288	
	(0.558)	
year=2019	0.419	
	(0.510)	
year=2020	0.349	
	(0.501)	
year=2021	0.144	
	(0.491)	
Ctemplar 2018		1.318
		(0.763)
Ctemplar 2019		1.327
		(0.763)
Ctemplar 2020		1.413*
		(0.767)
Ctemplar 2021		1.413*
		(0.767)
Hey 2020		−0.845*
		(0.454)
Hey 2021		−1.783***
		(0.497)
Hushmail 2020		0.300
		(0.368)
Hushmail 2021		0.300
		(0.368)
Kolab 2017		0.940**
		(0.313)
Kolab 2018		0.941**
		(0.354)
Kolab 2019		0.925**
		(0.313)
Kolab 2020		0.925**
		(0.313)
Kolab 2021		0.785
		(0.440)
Mailbox 2017		0.183
		(0.799)
Mailbox 2018		−0.047
		(0.793)
Mailbox 2019		0.234
		(0.810)
Mailbox 2020		0.234
		(0.810)
Mailbox 2021		−0.055
		(0.789)
Mailfence 2017		0.347
		(0.797)
Mailfence 2018		0.308
		(0.797)
Mailfence 2019		0.308
		(0.797)
Mailfence 2020		0.549
		(0.813)
Mailfence 2021		0.549
		(0.813)
Pesteo 2017		−0.405
		(0.295)
Pesteo 2018		−0.405
		(0.295)
Pesteo 2019		−0.405
		(0.295)
Pesteo 2020		−0.405
		(0.295)
Pesteo 2021		−0.405
		(0.295)

	(1)	(2)
Rickspace 2021		-1.498^{*}
		(0.783)
Runbox 2017		0.302
		(0.605)
Runbox 2018		0.252
		(0.752)
Runbox 2019		0.252
		(0.752)
Runbox 2020		0.216
		(0.751)
Runbox 2021		-0.356
		(0.804)
Soverin 2017		-0.918^{**}
		(0.384)
Soverin 2018		-0.918^{**}
		(0.384)
Soverin 2019		-0.918^{**}
		(0.384)
Soverin 2020		-0.918^{**}
		(0.384)
Soverin 2021		-0.918^{**}
		(0.384)
Thexyz 2017		-1.350
		(0.785)
Thexyz 2018		-1.352
		(0.785)
Thexyz 2019		-1.350
		(0.785)
Thexyz 2020		-1.080
		(0.827)
Thexyz 2021		-1.348
		(0.785)
Tutanota 2017		-0.124
		(0.715)
Tutanota 2018		0.431
		(0.708)
Tutanota 2019		0.502
		(0.708)
Tutanota 2020		0.456
		(0.702)
Tutanota 2021		0.274
		(0.702)
Zoho 2017		0.133
		(0.798)
Zoho 2018		-1.540
		(0.905)
Zoho 2019		-0.922
		(0.776)
Zoho 2020		-1.130
		(0.758)
Zoho 2021		-1.619^{**}
		(0.682)

Note: Table shows the coefficient estimates for the year fixed effects and service provider dummies. Column (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.E.3: Coefficient estimates of the service provider dummies for online news

	(1)	(2)
year=2017	2.358***	
	(0.240)	
year=2018	2.380***	
	(0.270)	
year=2019	2.254***	
	(0.264)	
year=2020	2.501***	
	(0.403)	
year=2021	2.403***	
	(0.438)	
Perks	−0.243	0.853***
	(0.289)	(0.000)
Bloomberg 2017		3.526
		(.)
Bloomberg 2018		3.555
		(.)
Bloomberg 2019		3.555
		(.)
Bloomberg 2020		3.555
		(.)
Bloomberg 2021		3.689
		(.)
Daily Mail 2017		1.855***
		(0.424)
Daily Mail 2018		1.855***
		(0.424)
Daily Mail 2019		1.855***
		(0.424)
Daily Mail 2020		1.428***
		(0.424)
Daily Mail 2021		1.136**
		(0.355)
FT 2017		2.144***
		(0.000)
FT 2018		2.168***
		(0.000)
FT 2019		2.191***
		(0.000)
FT 2020		3.989***
		(0.000)
FT 2021		3.995***
		(0.000)
Independent 2017		1.169**
		(0.424)
Independent 2018		1.169**
		(0.424)
Independent 2019		1.169**
		(0.424)
Independent 2020		1.096**
		(0.355)
Independent 2021		0.674*
		(0.355)

	(1)	(2)
NYT 2018		2.507***
		(0.606)
NYT 2019		1.814**
		(0.606)
NYT 2020		1.814**
		(0.606)
NYT 2021		1.814**
		(0.606)
Telegraph 2017		1.215***
		(0.116)
Telegraph 2018		1.215***
		(0.116)
Telegraph 2019		0.868***
		(0.116)
Telegraph 2020		0.868***
		(0.116)
Telegraph 2021		1.124***
		(0.158)
The Economist 2017		2.106***
		(0.233)
The Economist 2018		2.106***
		(0.233)
The Economist 2019		2.106***
		(0.233)
The Economist 2020		2.106***
		(0.233)
The Economist 2021		2.106***
		(0.233)
The Guardian 2017		2.995
		(.)
The Guardian 2018		2.995
		(.)
The Guardian 2019		2.703***
		(0.424)
The Guardian 2020		2.703***
		(0.424)
The Guardian 2021		2.703***
		(0.424)
The Times 2017		1.564**
		(0.606)
The Times 2018		0.712
		(0.606)
The Times 2019		0.712
		(0.606)
The Times 2020		0.712
		(0.606)
The Times 2021		0.712
		(0.606)
WSJ 2017		3.638***
		(0.233)
WSJ 2018		3.638***
		(0.233)
WSJ 2019		3.691***
		(0.233)
WSJ 2020		3.691***
		(0.233)
WSJ 2021		3.691***
		(0.233)

Note: Table shows the coefficient estimates for the year fixed effects and service provider dummies. Column (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.F Price Estimates

3.F.1 Price estimates for hedonic regression with service provider fixed effects

Table 3.F.1: Imputed price estimates for videoconferencing

	2017		2018		2019		2020		2021	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\hat{p}^t$	1.11	1.14	0.91	0.93	0.61	0.66	0.35	0.4	0.32	0.38
Lower $\hat{p}^t$ (CI 95%)	0.89	0.91	0.74	0.75	0.49	0.53	0.29	0.32	0.26	0.3
Upper $\hat{p}^t$ (CI 95%)	1.37	1.43	1.13	1.16	0.75	0.83	0.44	0.51	0.39	0.47
Webinar Dummy	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Note: The table shows the price estimates for each year. Panel A estimates the prices using equation 3.4. Panel B presents an alternative model, where the variable for the log of participants interacted with the time dummies. All estimates are in USD.

Table 3.F.2: Imputed price estimates for personal email

	2017	2018	2019	2020	2021
$\hat{p}^t$	5.5	5.5	6.1	6.0	4.4
Lower $\hat{p}^t$ (CI 95%)	3.1	3.1	3.4	3.3	2.4
Upper $\hat{p}^t$ (CI 95%)	9.9	9.9	10.9	10.7	7.8

Note: The table shows the implicit price estimates of personal email for each year. All estimates are in USD.

Table 3.F.3: Imputed price estimates for online news

	2017	2018	2019	2020	2021
$\hat{p}^t$	9.0	9.0	8.0	9.1	8.8
Lower $\hat{p}^t$ (CI 95%)	5.4	5.0	4.0	3.8	3.7
Upper $\hat{p}^t$ (CI 95%)	14.9	16.4	15.9	15.4	14.8

*Note:*The table shows the shadow price estimates of online news for each year. All estimates are in USD.

3.F.2 Time Dummy Variable Model

We generate imputations for the price of digital goods using the classical time dummy variable model. The exponential of the coefficients for time dummies would represent the quality-adjusted price indices for each year. The prediction model is expressed as follows.

$$\hat{p}^t = exp(\tau^t) \times exp(\beta_1 log(z_1)) \times exp(0.5Var(\varepsilon_{ij})). \tag{3.7}$$

One can observe that the price estimates using this model are larger compared to those generated when we control for service provider fixed effects.

Table 3.F.4: Price estimates using the time dummy variable model for videoconferencing

	2017		2018		2019		2020		2021	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\hat{p}^t$	3.3	3.6	2.9	3.2	2.9	2.9	2.4	2.5	2.8	2.7
Lower $\hat{p}^t$ (CI 95%)	2.6	0.6	2.3	2.3	2.2	2.1	1.9	1.8	2.2	2
Upper $\hat{p}^t$ (CI 95%)	4.2	4.9	3.8	4.4	3.7	4.0	3.1	3.5	3.6	3.7
Webinar Dummy	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Note: The table shows the price estimates for each year using the time dummy variable model in equation 3.7 All estimates are in USD.

Table 3.F.5: Price estimates using the time dummy variable model for personal email

	2017	2018	2019	2020	2021
$\hat{p}^t$	4.82	5.0	5.7	5.31	4.33
Lower $\hat{p}^t$ (CI 95%)	2.23	2.32	2.64	2.46	2.01
Upper $\hat{p}^t$ (CI 95%))	10.42	10.82	12.33	11.5	9.37

Note: The table shows the price estimates for each year using the time dummy variable model in equation 3.7 All estimates are in USD.

Table 3.F.6: Price estimates using the time dummy variable model for online news

	2017	2018	2019	2020	2021
$\hat{p}^t$	12.2	12.5	11	14.1	12.8
Lower $\hat{p}^t$ (CI 95%)	7.1	6.8	6.1	5.7	4.7
Upper $\hat{p}^t$ (CI 95%)	21	23	20	35.1	34.4

Note: The table shows the price estimates for each year using the time dummy variable model in equation 3.7 All estimates are in USD.

3.G Validity check

The validity of our imputed price estimates is contingent on the validity of our hedonic regression model. We assume that the regression parameters should approximate *true* WTP for each characteristic. To test this, we generated a predicted price of premium versions of the digital goods, i using the prediction model in equation 3.8. Here, we include all characteristics in the regression model to generate a estimated price $\hat{p}_i^t$. If our predicted price approximates the observed price, then it should be fair to argue that our model is valid.

$$\hat{p}_i^t = \left[exp\left(\frac{1}{J}\sum_{j=1}^{J}\delta_j^t\right) \times exp\left(\sum_{k=1}^{K}\beta_k log(z_k)\right)\right] \times exp(0.5Var(\varepsilon_{ij})) \tag{3.8}$$

In figure 3.G.1 we compare the estimated premium price of videoconferencing (blue line) with the average price of paid videoconferencing from our data (dashed line) with its corresponding prediction interval[21]. We observe that the predicted price is within interval estimates of the average price for all years. However, the predicted price does not exhibit the same upward trend apparent in the mean price data. Moreover, we also observe that the prediction interval for our estimated price (blue-shaded region) is smaller compared to the interval of the mean price. These discrepancies probably arose because the hedonic regression model we employed assumes that the marginal contributions to the price of all premium-exclusive characteristics are fixed across time. By keeping the coefficients fixed over time, the model does not capture some of the time variations we see in the data. Allowing every characteristic to vary over time by interacting them with time dummies would likely capture this dynamic. Since the inclusion of additional variables increases the standard error, the said action would also likely broaden the prediction interval. However, we choose not to do this for two reasons. First, the price estimates are still within the intervals of the observed data, implying that our estimates are reasonable. Second, having large intervals may be a good thing when the goal is to generate unbiased estimates of parameters (for example, when examining relationships). But for the purposes of estimation, this is not the case. Large intervals are often not useful when generating a prediction, as it reflects a large degree of uncertainty for the estimates. Estimates with large intervals are often not useful for policy purposes as well.

An alternative way to estimate the price of digital goods using hedonic regression is by including only variables that are statistically significant (at $\alpha = 0.05$). One can argue that characteristics that are not statistically significant do not contribute materially to the price of the good and could be ignored. Dropping all the variables that are not statistically significant would cause the estimated price to drop substantially for videoconferencing (see right panel of figure 3.G.1). Interestingly, this is not the case for email. Dropping variables that are

[21] Since we are plotting the average price per year, the prediction intervals in our figure is based on the standard confidence interval (CI) for the mean ($\bar{x}$), where $CI = \bar{x} \pm z_{0.025} \times \sigma/n$, where σ and n represent the standard deviation and a number of observations for each year, respectively.

not statistically significant causes the estimates to better align the observed data (see figure 3.G.2). This is likely to be because most of the coefficients for email were not significant in the first place.

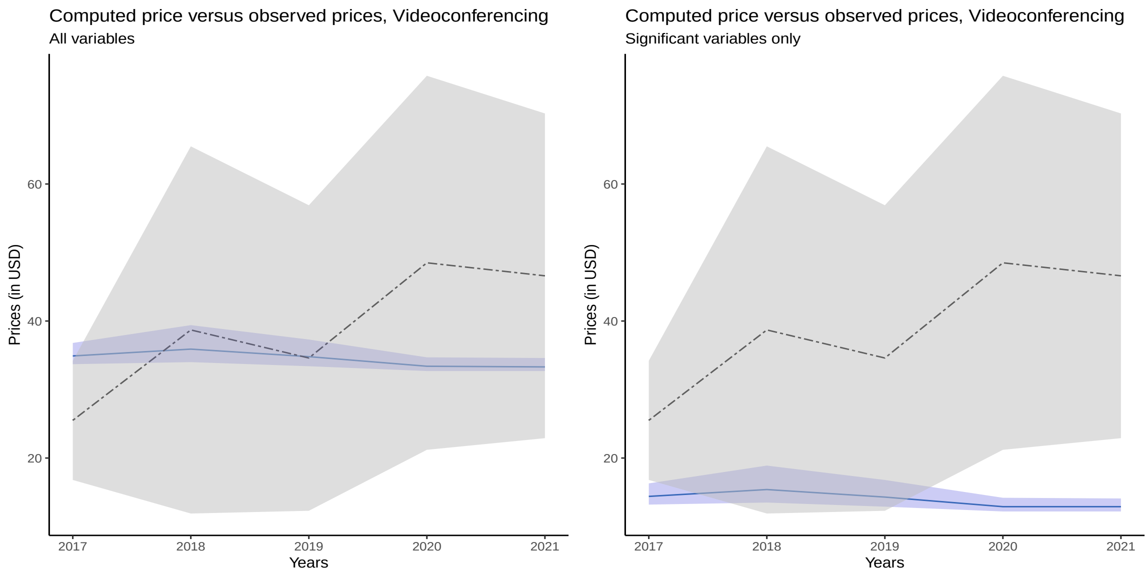


Figure 3.G.1: Estimated price versus observed price for videoconferencing

Note: The figure shows the predicted monthly price of paid videoconferencing (blue line), its corresponding prediction intervals (blue area), the average monthly price of videoconferencing (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 3.8. The left panel shows all the price estimates which employ all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 were incorporated in the prediction model. All figures are in USD.

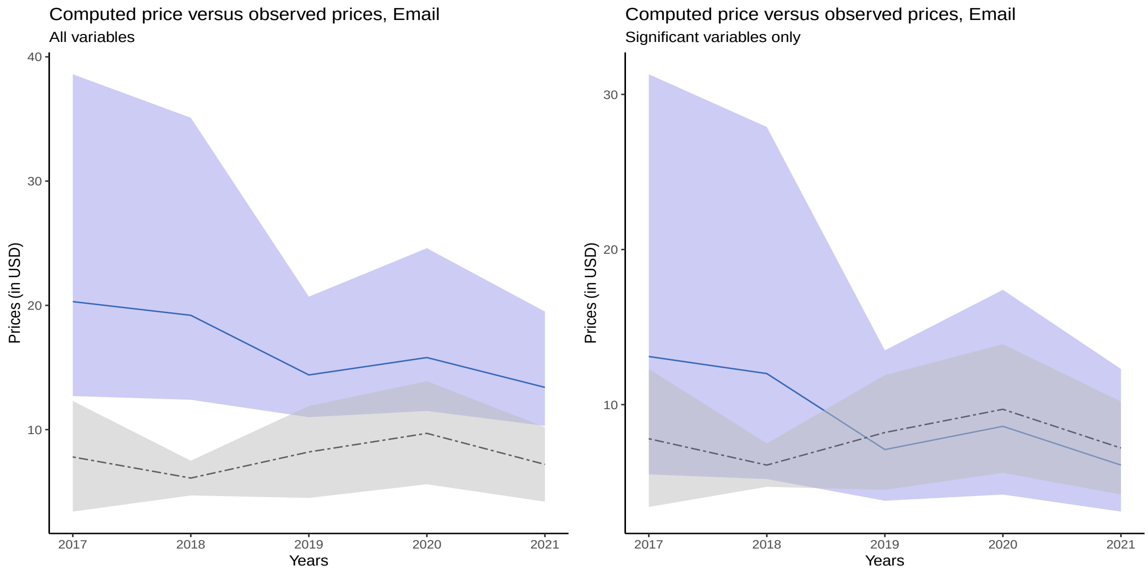


Figure 3.G.2: Estimated price versus observed price for email

Note: The figure shows the predicted monthly price of paid email (blue line), its corresponding prediction intervals (blue area), the average monthly price of email (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 3.8. The left panel shows all the price estimates which employ all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 were incorporated in the prediction model. All figures are in USD.

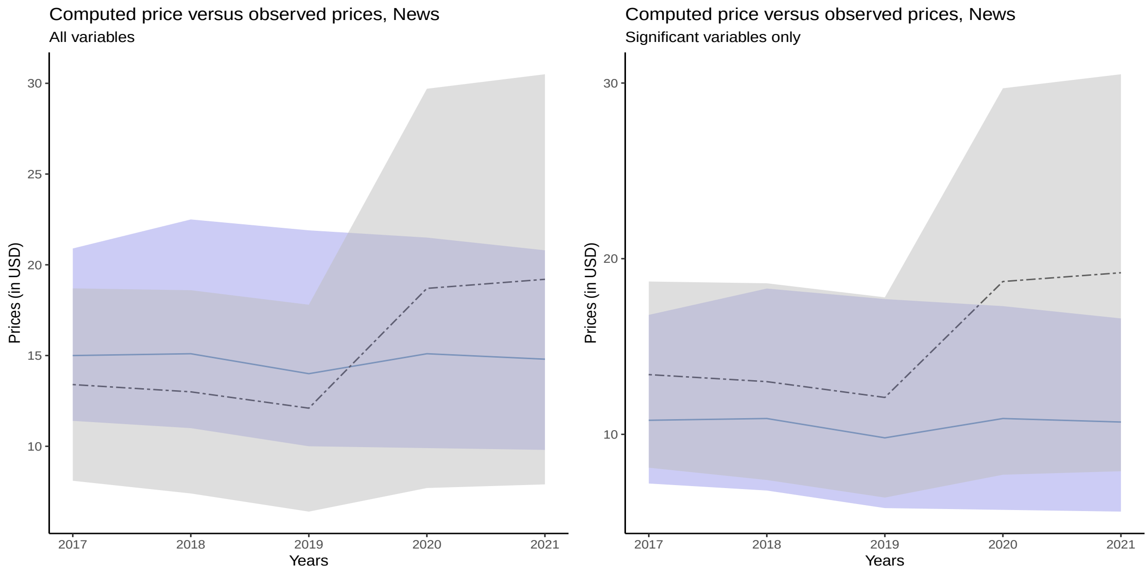


Figure 3.G.3: Estimated price versus observed price for videoconferencing

Note: The figure shows the predicted monthly price of paid online news (blue line), its corresponding prediction intervals (blue area), the average monthly price of online news (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 3.8. The left panel shows all the price estimates which employ all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 were incorporated in the prediction model. All figures are in USD.

If all insignificant variables are dropped, the predicted price of email becomes more aligned with the observed data for more recent years. For 2017 and 2018, the hedonic regression model appears to overestimate the mean price. However, the mean price of email is still within the prediction interval for those years. This should not matter for our purposes, since we are only interested in the value of free email.

For online news, both the predicted price and the observed mean price are within the 95 percent prediction intervals of one another. One thing to note is that the price increase for the predicted price of online news from 2019 to 2020 is less pronounced compared with the jump in prices seen in the data.

For all three forms of digital goods, the price estimates generated by the hedonic regression do not substantially deviate from the observed price. It can be noticed that the predicted prices are more stable over time, which could be a result of the assumption that prices for each characteristic are fixed. Therefore, our model would probably underestimate inflation, which is one limitation resulting from our chosen specification. For our purposes, however, this may not matter as much. One of the goals of this study is to measure the welfare contribution of free digital goods. Welfare changes are often reflected by real growth in gross value, as opposed to nominal growth. Real growth in gross value is achieved by keeping prices fixed over time, allowing volume changes to dominate that change in value. Because of this, we argue that our estimates would probably serve the purpose of tracking welfare changes over time.

It would also be interesting to see how much the "free component" of digital goods contributes to the overall price. Our prior is that the free component should account for the majority of the value of the overall price. When you subscribe to the paid version of Outlook, most of the value you derive from the subscription would probably come from the email service rather than the other features. As such, we take the percentage share s_k^t of each component z_k relative to the predicted price $\hat{p}_i^t$ using:

$$s_k^t = \frac{exp(\beta_k log(z_k))}{\hat{p}_i^t} \times 100. \tag{3.9}$$

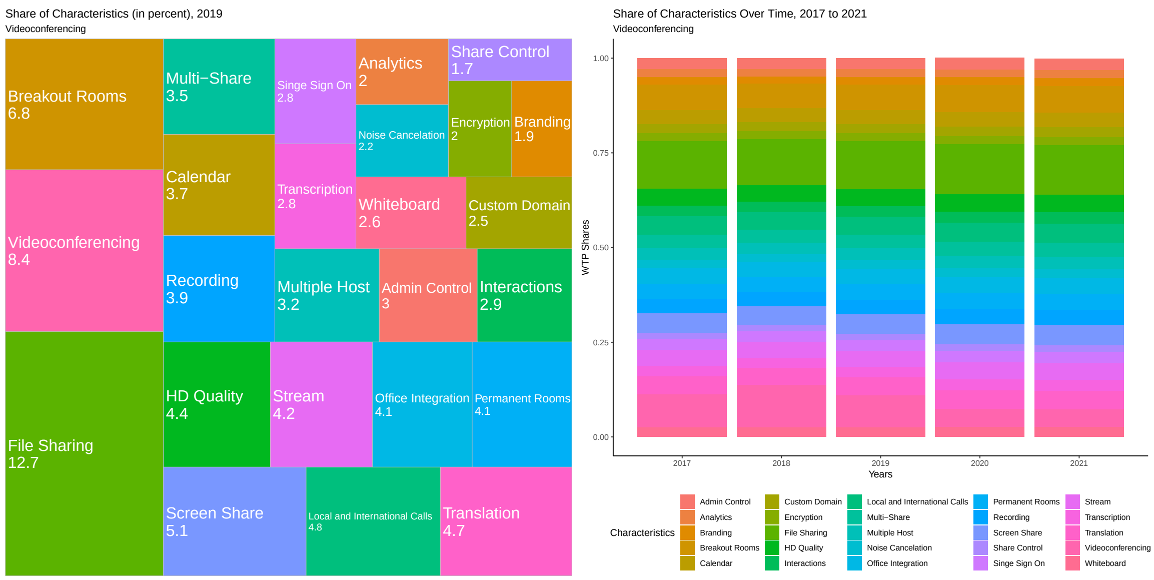


Figure 3.G.4: Percentage share of characteristics to the predicted price, videoconferencing

Note: The figure on the left shows the percentage share of each characteristic to the predicted price for videoconferencing in 2019, as computed in equation 3.9. The figure on the right shows the share of each characteristic of videoconferencing from 2017 to 2021.

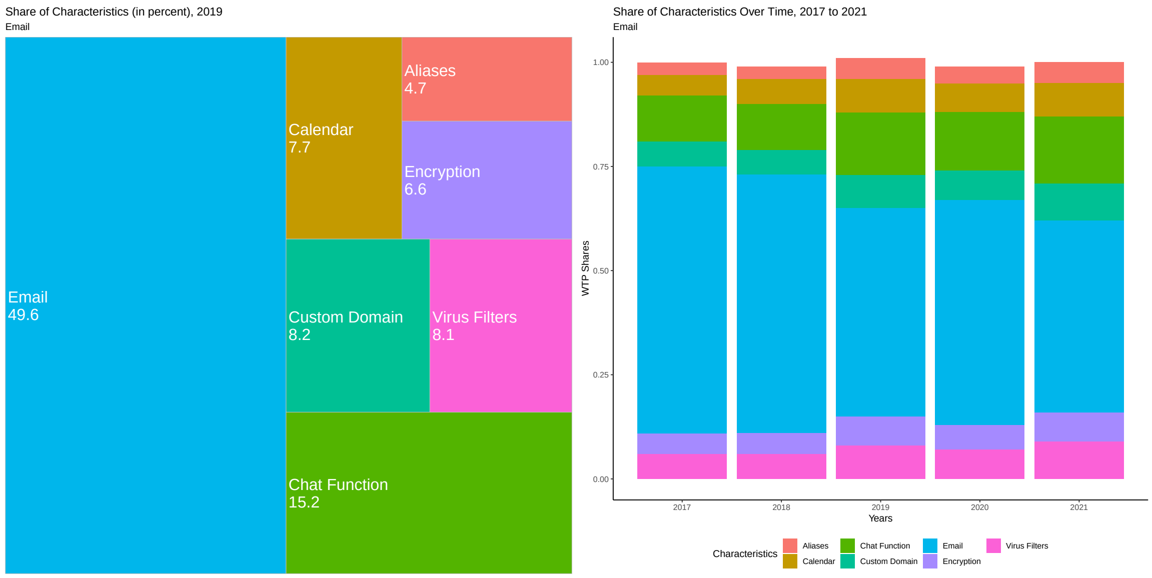


Figure 3.G.5: Percentage share of characteristics to the predicted price, email

Note: The figure in the left shows the percentage share of each characteristic to the predicted price for email in 2019, as computed in equation 3.9. The figure on the right shows the share of each characteristic of email from 2017 to 2021.

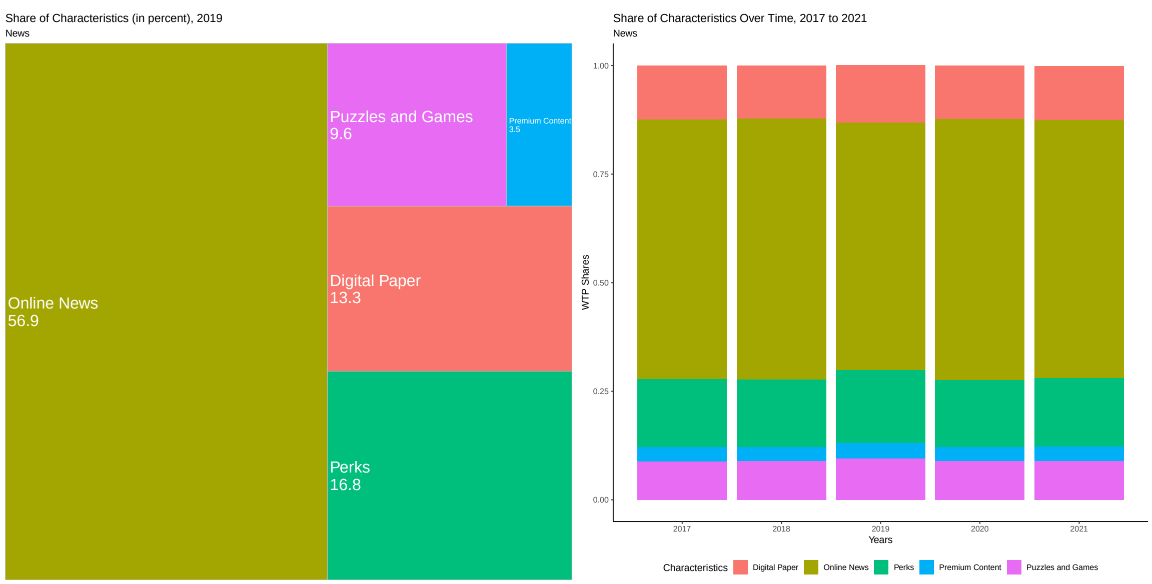


Figure 3.G.6: Percentage share of characteristics to the predicted price, online news

Note: The figure on the left shows the percentage share of each characteristic to the predicted price for online news in 2019, as computed in equation 3.9. The figure on the right shows the share of each characteristic of online news from 2017 to 2021.

We report that percentage share of the estimated WTP for all characteristics in figures 3.G.4, 3.G.5, and 3.G.6. For both email and online news, the estimated share of the "free component" accounts for about half of the predicted price. For videoconferencing, it accounts for the second largest share of its predicted price. This is consistent with our prior. The shares also appear to be consistent over time.

3.H Robustness check

To test whether or not our price estimates are robust to changes in model specification, we employ forward, backward, and stepwise selection. Forward selection begins by running an empty model (a model containing only the intercept term) and proceeds by including regressors (in this case, characteristics) that are significant at a certain p-value threshold (in this case, 0.2 and 0.1). Backward selection is the opposite approach. It begins by running a regression with all regressors. Regressors with p-values less than the set thresholds are dropped from the model. Stepwise selection combines both forward and backward selection. The resulting regression equations from these selection models would be more parsimonious than the baseline specification.

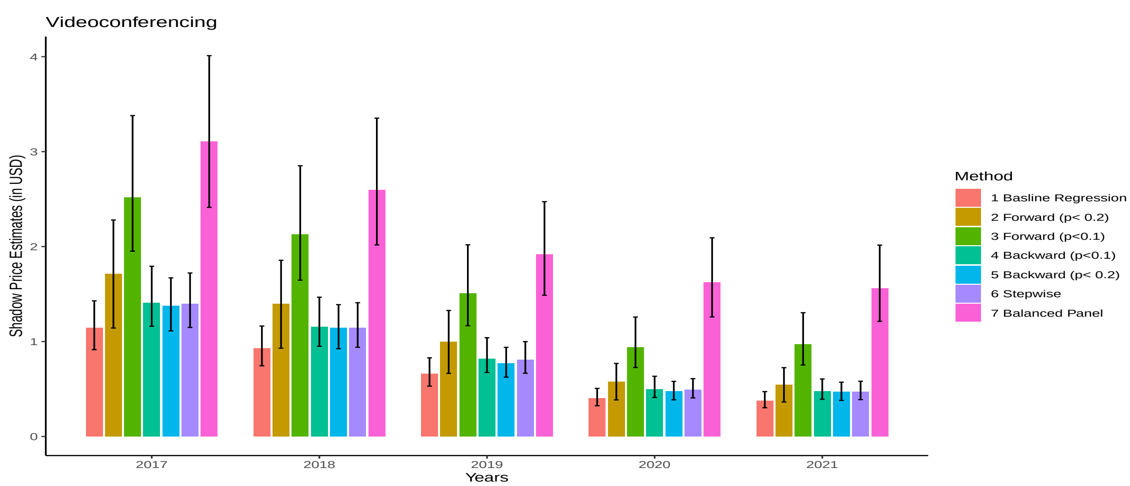


Figure 3.H.1: Robustness check, videoconferencing

Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for videoconferencing. The figure also the interval estimates for the price of videoconferencing if service providers that are not present in all years are dropped from the data set. Figures can be viewed in table 3.H.1.

For videoconferencing, the shadow price estimates from the forward, backward, and stepwise selection models are shown in figure 3.H.1 (table 3.H.1). Estimates from the forward selection are slightly higher than the baseline regression. Estimates from both the backward and stepwise selection models were closer to those generated from the baseline hedonic regression. Only the estimates from the restrictive 0.1 p-value threshold of the forward selection were outside the interval estimates from the baseline regression. Even so, the deviation is arguably not that substantial.

Lastly, we earlier noted that the panel data set we used for our regressions is not balanced. As shown in table 3.A.1, we do not have data for all service providers for all of the years covered in the panel. This is either because the service provider had not started operating in those years, or that data simply cannot be acquired. To test how much attrition affects our estimates, we run the hedonic regression is equation 3.3, dropping all service providers with incomplete data. The results are shown in table 3.H.1 column (7).

We notice that the price estimates are higher when we drop the service providers with incomplete data. In some years, the estimates for the balanced panel are twice as large as the baseline regression. This difference though is likely to be due to survivorship bias. It is possible that service providers whose data sets are more complete are likely to be offering their services at a higher price than other providers.

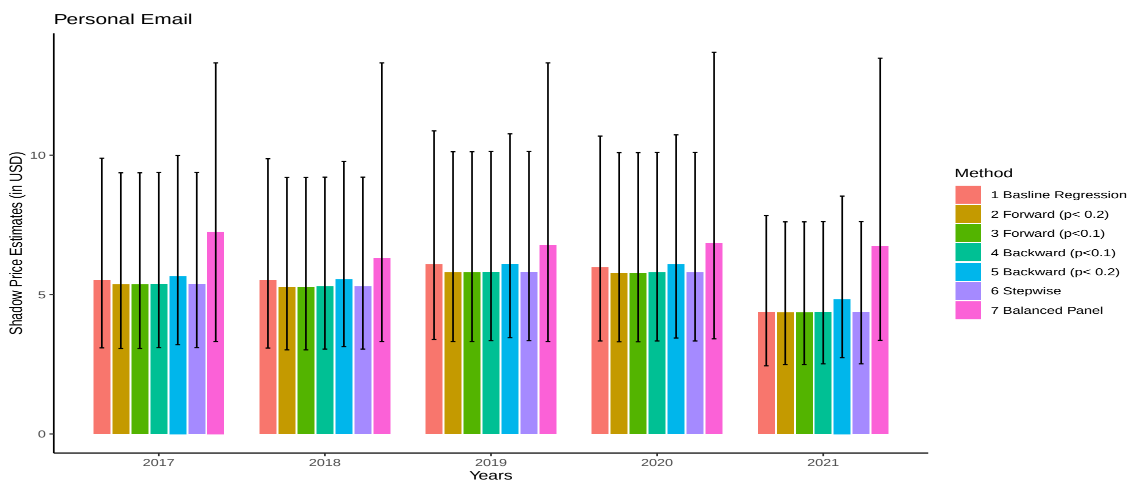


Figure 3.H.2: Robustness check, personal email

Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for personal email. The figure also the interval estimates for the price of personal email if service providers that are not present in all years are dropped from the data set. Figures can be viewed in table 3.H.2.

The results are more robust for email. From figure 3.H.2 (table 3.H.2), we can observe that the shadow price estimates from the baseline specification are noticeably similar to the estimates generated from the forward, backward, and stepwise selection models. Moreover, the intervals for all estimates overlap, implying that there is no statistical difference between the estimates for all six models.

Similar to what we saw with videoconferencing, the price estimates from the balanced panel were higher compared to the estimates from the baseline specification. Nonetheless, the intervals of both the baseline estimates and the estimates from the balanced panel also overlap, implying that there is no statistical difference between with two.

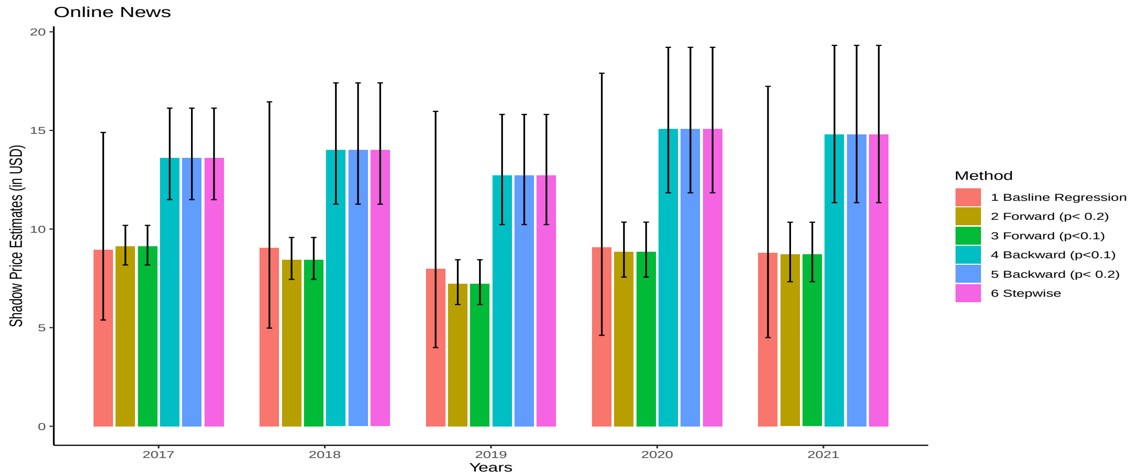


Figure 3.H.3: Robustness check, online news

Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for online news. Figures can be viewed in table 3.H.3

For online news, shadow price estimates from the forward, backward, and stepwise estimations are within the prediction interval of the baseline specification (see figure 3.H.3 and table 3.H.3). Estimates from the forward estimation tend to be lower and estimates from the backward estimation tend to be higher.

Table 3.H.1: Robustness Check of Price Estimates for Videoconferencing

	Baseline Specification	Forward ($p < 0.2$)	Forward ($p < 0.1$)	Backward ($p < 0.2$)	Backward ($p < 0.1$)	Stepwise Selection	Balanced Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Point Estimate							
2017	1.14	1.71	2.52	1.41	1.38	1.4	3.11
2018	0.93	1.39	2.13	1.15	1.14	1.14	2.6
2019	0.66	1	1.51	0.82	0.77	0.81	1.92
2020	0.4	0.58	0.94	0.5	0.48	0.49	1.62
2021	0.38	0.54	0.97	0.48	0.47	0.47	1.56
Panel B: Lower (CI 95%)							
2017	0.91	1.14	1.95	1.16	1.11	1.15	2.1
2018	0.75	0.93	1.65	0.95	0.92	0.94	2.1
2019	0.53	0.67	1.17	0.67	0.63	0.67	2.1
2020	0.32	0.39	0.73	0.41	0.39	0.41	1.26
2021	0.3	0.36	0.75	0.39	0.38	0.39	1.21
Panel C: Upper (CI 95%)							
2017	1.43	2.28	3.38	1.79	1.67	1.72	3.49
2018	1.16	1.86	2.85	1.47	1.39	1.41	3.49
2019	0.83	1.33	2.02	1.04	0.94	1	3.49
2020	0.51	0.77	1.26	0.63	0.58	0.61	2.09
2021	0.47	0.73	1.3	0.61	0.57	0.58	2.02

Note: The table shows the price estimates for each year using different specifications. Column (1) shows the price estimates from the baseline hedonic estimate where all characteristics were included as explanatory variables in the regression. Columns (2) and (3) show the price estimates from the hedonic regression using forward selection, where regressors are added once they are significant at (p<0.2) and (p<0.1), respectively. Columns (4) and (5) show the price estimates from the hedonic regression using backward selection, where regressors are removed when they are not significant at (p<0.2) and (p<0.1), respectively. Column (6) shows the price estimate for the Stepwise regression with a backward cut-off of p<0.2 and a forward cut-off of p<0.1. Column (7) shows the hedonic regression using all characteristics as explanatory variables but retaining only service providers where prices are observed for all years. All estimates are in USD.

Table 3.H.2: Robustness Check of Price Estimates for Personal Email

	Baseline Specification	Forward ($p < 0.2$)	Forward ($p < 0.1$)	Backward ($p < 0.2$)	Backward ($p < 0.1$)	Stepwise Selection	Balanced Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Point Estimate							
2017	5.53	5.36	5.36	5.38	5.65	5.38	7.25
2018	5.51	5.27	5.27	5.29	5.53	5.29	6.32
2019	6.07	5.79	5.79	5.81	6.1	5.81	6.78
2020	5.97	5.78	5.78	5.79	6.08	5.79	6.84
2021	4.37	4.35	4.35	4.37	4.83	4.37	6.74
Panel B: Lower (CI 95%)							
2017	3.09	3.07	3.07	3.1	3.2	3.1	3.32
2018	3.08	3.02	3.02	3.04	3.13	3.04	3.32
2019	3.39	3.32	3.32	3.35	3.45	3.35	3.32
2020	3.33	3.31	3.31	3.33	3.44	3.33	3.41
2021	2.44	2.49	2.49	2.52	2.74	2.52	3.36
Panel C: Upper (CI 95%)							
2017	9.89	9.37	9.37	9.38	9.98	9.38	13.31
2018	9.87	9.2	9.2	9.21	9.77	9.21	13.31
2019	10.87	10.12	10.12	10.13	10.77	10.13	13.31
2020	10.69	10.09	10.09	10.09	10.73	10.09	13.69
2021	7.83	7.61	7.61	7.61	8.53	7.61	13.48

Note: The table shows the price estimates for each year using different specifications. Column (1) shows the price estimates from the baseline hedonic estimate where all characteristics were included as explanatory variables in the regression. Columns (2) and (3) show the price estimates from the hedonic regression using forward selection, where regressors are added once they are significant at (p<0.2) and (p<0.1), respectively. Columns (4) and (5) show the price estimates from the hedonic regression using backward selection, where regressors are removed when they are not significant at (p<0.2) and (p<0.1), respectively. Column (6) shows the price estimate for the Stepwise regression with a backward cut-off of p<0.2 and a forward cut-off of p<0.1. Column (7) shows the hedonic regression using all characteristics as explanatory variables but retaining only service providers where prices are observed for all years. All estimates are in USD.

Table 3.H.3: Robustness Check of Price Estimates for Online News

	Baseline Specification	Forward ($p < 0.2$)	Forward ($p < 0.1$)	Backward ($p < 0.2$)	Backward ($p < 0.1$)	Stepwise Selection
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Point Estimate			
2017	9.0	9.1	9.1	13.6	13.6	13.6
2018	9.0	8.4	8.4	14.0	14.0	14.0
2019	8.0	7.2	7.2	12.7	12.7	12.7
2020	9.1	8.9	8.9	15.1	15.1	15.1
2021	8.8	8.7	8.7	14.8	14.8	14.8
			Panel B: Lower (CI 95%)			
2017	5.4	8.2	8.2	11.5	11.5	11.5
2018	5.0	7.4	7.4	11.3	11.3	11.3
2019	4.0	6.2	6.2	10.2	10.2	10.2
2020	4.6	7.6	7.6	11.8	11.8	11.8
2021	4.5	7.3	7.3	11.3	11.3	11.3
			Panel C: Upper (CI 95%)			
2017	14.9	10.2	10.2	16.1	16.1	16.1
2018	16.4	9.6	9.6	17.4	17.4	17.4
2019	16.0	8.4	8.4	15.8	15.8	15.8
2020	17.9	10.4	10.4	19.2	19.2	19.2
2021	17.2	10.3	10.3	19.3	19.3	19.3

Note: The table shows the price estimates for each year using different specifications. Column (1) shows the price estimates from the baseline hedonic estimate where all characteristics were included as explanatory variables in the regression. Columns (2) and (3) show the price estimates from the hedonic regression using forward selection, where regressors are added once they are significant at ($p<0.2$) and ($p<0.1$), respectively. Columns (4) and (5) show the price estimates from the hedonic regression using backward selection, where regressors are removed when they are not significant at ($p<0.2$) and ($p<0.1$), respectively. Column (6) shows the price estimate for the Stepwise regression with a backward cut-off of $p<0.2$ and forward cut-off of $p<0.1$. All estimates are in USD.

3.I Comparison with other studies estimates

Table 3.I.2: Comparison of WTA values with the price imputations for email and online news

	Hedonic Regression	Brynjolfsson (2019)	Coyle and Nguyen (2023)		Jamison and Wang (2021)
	May 2020 (1)	2017 (2)	Mean 2020 (3)	Median 2020 (4)	March 2020 (5)
Personal email					
Point	5.97	701	192	227	2095
Lower	3.33	574	206	130	1517
Upper	10.69	852	221	324	2673
Online news					
Point	9.09	–	81	81	–
Lower	3.84	–	76	71	–
Upper	15.42	–	87	90	–

Note: The table compares the WTA estimates from Brynjolfsson et al. (2019b), Coyle and Nguyen (2023) and Jamison and Wang (2021) with the price estimates from the hedonic regression. Estimates from Coyle and Nguyen (2023) were based on their May 2020 data collection. All estimates are in USD.

3.J Gross Value of free digital goods, levels

3.J.1 Baseline estimates

Table 3.J.1: Gross value of digital goods and Household Final Consumption Expenditures, at current prices

	2017	2018	2019	2020
Point Estimate	5,420	5,520	5,495	5,940
Lower	3,215	3,110	2,946	3,169
Upper	9,193	9,850	10,280	11,128
HFCE	1,301,142	1352042	1387664	1,214,474
GDP	2,085,008	2,157,410	2,238,348	2,109,594

Note: The table shows the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS. All estimates are in million £.

Table 3.J.2: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2019)

	2017	2018	2019	2020
Point Estimate	5,220	5,346	5,495	5,654
Lower	2,795	2,862	2,946	3,023
Upper	9,774	10,012	10,280	10,602
HFCE	1,346,008	1,374,051	1,387,664	1,208,053
GDP	2,166,073	2,203,005	2,238,348	1,991,439

Note: The table shows the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS. All estimates are in million £.

3.J.2 Adjusted using Ofcom data

Table 3.J.3: Gross value of digital goods adjusted for Ofcom data, at current prices

	2017	2018	2019	2020
Point Estimate	5,420	5,520	5,495	6,554
Lower	3,215	3,110	2,946	3,515
Upper	9,193	9,850	10,280	12,220

Note: The table shows the interval estimate of the aggregate gross value (at current prices) for the three digital goods, videoconferencing, personal email, and online news after estimates in 2020 is adjusted using Ofcom data. All estimates are in million £.

Table 3.J.4: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2019)

	2017	2018	2019	2020
Point Estimate	5,220	5,346	5,495	6,286
Lower	2,795	2,862	2,946	3,382
Upper	9,774	10,012	10,280	11,723

Note: The table shows the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS, at constant prices. All estimates are in million £.

3.J.3 Accounting for multiple provider use

Table 3.J.5: Gross value of digital goods and Household Final Consumption Expenditures, at current prices

	2017	2018	2019	2020
Point Estimate	12,937	13,051	12,164	13,329
Lower	7,869	7,399	6,380	6,986
Upper	21,417	23,206	23,266	25,401

Note: The table shows the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS, at current prices. All estimates are in million £.

Table 3.J.6: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2019)

	2017	2018	2019	2020
Point Estimate	11,709	11,888	12,164	12,364
Lower	6,139	6241	6380	6,504
Upper	22,403	22,723	23,266	23,604

Note: The table shows the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS, at constant prices. All estimates are in million £.

3.K Comparison with other expenditure items

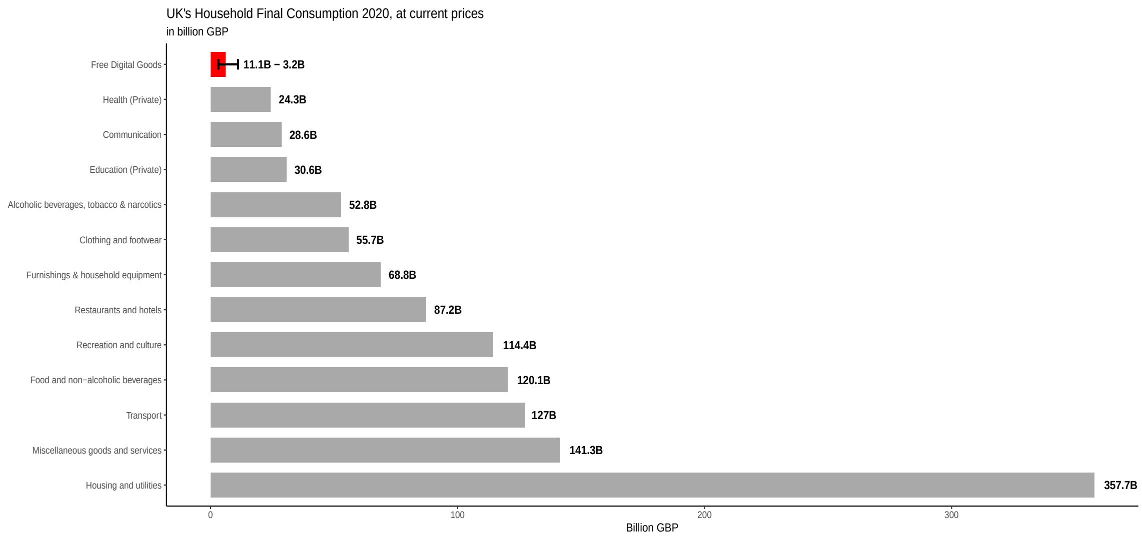


Figure 3.K.1

Note: The figure compares the current price estimates of the gross value of free digital goods in table 3.J.1 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.

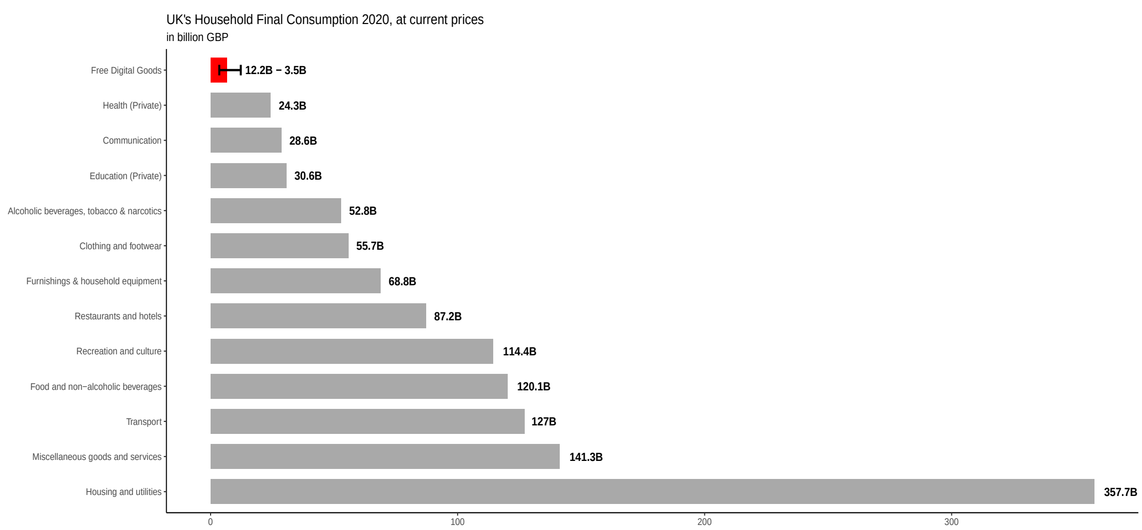


Figure 3.K.2

Note: The figure compares the current price estimates of the gross value of free digital goods (adjusted using Ofcom data) in table 3.J.5 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.

CHAPTER 4: How much is Illegal Streaming Worth? Measuring the Value of Digital Piracy

ABSTRACT

The goal of this study is to measure the value derived by households from digital piracy. Our measurement strategy employs the price of paid digital services and services (e.g. Spotify, Netflix) as a proxy for the shadow price of their illegally acquired counterparts. This is a common method for the valuation of non-market activity and non-monetary transactions for national accounting purposes. To estimate the number of individuals engaged in digital piracy, we used information from the United Kingdom's Intellectual Property Office annual Piracy Tracker, which provides an estimate of the proportion of the UK population that engages in digital piracy. We find that, on average, the gross value from the piracy of music, video, live sports, software, computer games, and ebooks was between £3.6 billion and £7.5 billion in 2021. We also find that while the value from the final consumption of communication services has been consistently rising in the past five years, the inflation-adjusted gross value from the digital piracy of media has been falling. Our data shows that the growth in the value of final consumption from communication services would have been slower by 0.2 percentage points in 2021 had the value from the consumption of pirated content been accounted for in the estimates.

4.1 Introduction

Since macroeconomic aggregates such as Gross Domestic Product only account for economic activities with market prices, this set of statistics tends to miss the benefits households receive from the consumption of free internet services such as personal email, chat, and online news, among others. As such, various efforts have been made by economists (and statisticians) to quantify the value individuals derive from free Internet products. Most of the research in this area focuses on social media (Brynjolfsson et al., 2019a; Schreyer, 2021) and advertising-supported services such as online maps and search engines (Nakamura et al., 2017; Van Elp and Mushkudiani, 2019). None of the studies, however, tackle the value of the consumption of content accessed through digital piracy.

We can view the consumption of content from digital piracy as a form of "free" service provided through the internet. One of the reasons why researchers are interested in measuring the value of free digital services is due to their ability to provide households with utility, even though the consumption of these services is not explicitly reflected in official statistics. One can argue that products and services acquired through digital piracy—while illegal—also improve household utility, despite being missed by National Accounts estimates as part of household consumption. While some of these websites earn revenues by displaying advertisements, and the revenues from these websites are already included in GDP, we believe these earnings underestimate the value derived by households from the consumption of these products. Considering that the marginal cost of production of each content is close to zero, the value derived from an additional user of these sites may be understated in GDP. Moreover, since these sites generate revenues from advertising, their income would be accounted as intermediate consumption of the products they advertise and not household consumption. As such, policymakers would not be able to examine how the production of content affects consumer welfare.

The goal of this study is to measure and examine the value derived by households from digital piracy. As our measurement strategy, we employ the price of paid digital services (e.g. Spotify, Netflix, etc.) as a proxy for the shadow price of their illegally acquired counterparts (e.g. illegal streaming, torrent, etc.). Market substitute pricing, or market-equivalent pricing, has been applied by compilers of the National Accounts to impute for the implicit price of other non-market activity and non-monetary transactions such as the rental from owner-occupied housing, and agricultural production for own consumption, and barter transactions. This strategy ensures that our estimates would be consistent with the accounting principles of the SNA.

To estimate the number of individuals engaged in digital piracy, we used information from the UK's Intellectual Property Office annual Piracy Tracker. The report provides estimates on the share of the UK's population that engages in digital piracy. This dataset is unique to the UK and is based on a survey employing a representative sample of the country's online population. The survey is conducted annually and allows us to generate estimates that are consistent across time.

We find that the gross value from the piracy of music, video (TV series and movies), live sports, software, computer games, and ebooks was between £3.6 billion and £7.5 billion in 2021. We also find that while the value from the final consumption of communication services has been consistently rising in the past five years, the gross value from the digital piracy of media is falling. Our data shows that the growth in the value of the final consumption from communication services would have been slower by 0.2 percentage points in 2021 had the value from the consumption of pirated content been accounted for in the estimates. Therefore, it is likely that official statistics are overstating the growth in welfare from communication services since they do not account for the value of digital piracy.

While the SNA mostly considers market goods and services as part of GDP, it also recommends that a select group of non-market services, such as government services and services from owner-occupied housing, should be considered as part of production. That said, the majority of non-market production such as household production of services (except services from owner-occupied housing) and ecosystem services are still not considered as part of GDP. The value derived by households from the consumption of these non-market services is compiled in separate sets of statistics called Satellite Accounts. As such, this essay does not recommend that the value of pirated content should be included in GDP. Rather, our goal is to develop a methodology for a possible Satellite Account that incorporates the value of free digital services.

This study contributes to three areas of research. First, it contributes to the growing body of research that aims to measure the value of free digital services. Most of the studies in this area focused on services that can be accessed legally, such as social media, search engines, and messenging apps (Corrigan et al., 2018; Brynjolfsson et al., 2019a; Coyle and Nguyen, 2023) as well as advertising and marketing-supported content in general (Nakamura and Soloveichik, 2015; Nakamura et al., 2017; Van Elp et al., 2022). This study extends the measurement literature by focusing on the value of media accessed through digital piracy. To the best of our knowledge, none of the earlier studies have paid attention to internet services accessed through illegal means. This study also extends the measurement literature by employing the price of paid digital products as a proxy for the value of their pirated counterparts, which are often acquired with little or no explicit cost to the user. While this strategy has been applied in the core National Accounts and environmental accounting, to our knowledge, this approach has seldom been used for the valuation of free services.

This study also contributes to the literature on the measurement of illegal activities. According to the SNA, all production activity—whether legal or not—should be included as part of the National Accounts estimates. Many of these activities fall under production by households as an unincorporated enterprise. Various efforts have been made to incorporate the value of illegal activity in the National Accounts. Countries that have engaged in these attempts include the United Kingdom (Abramsky and Drew, 2014), The Balkans (Blades, 2011), The Netherlands and Romania (Dragusin, 2015), and the United States (Soloveichik, 2019), among others. This study extends the literature on the measurement of illegal activities by generating estimates for the value of digital piracy in the UK.

This study also contributes to the empirical literature that attempts to measure the impact of piracy. Most of the studies in this area focus on the displacement of sales in creative industries. The studies by Peitz and Waelbroeck (2004), Zentner (2005), and Liebowitz (2008) found that piracy of music negatively impacts album sales. Meanwhile, Rob and Waldfogel (2006) argued that illegal access to media actually improves welfare by reducing deadweight loss. This study contributes to the literature on the impact of digital piracy by quantifying the dollar-value of welfare received by households from the consumption of pirated media. None of the earlier research has examined the aggregate welfare impact of digital piracy, especially in the context of the National Accounts.

The outline of this essay is as follows. The next section explains the rationale for measuring the value of digital piracy based on the SNA framework. Section 4.3 summarizes the literature on the impact of piracy. Section 4.4 discusses our empirical strategy. In section 4.5, we present the data, and in section 4.6, we discuss our results. We end this essay with our conclusions and ways forward.

4.2 Why measure the value of digital piracy?

The internet provides a number of free services that substitute for market activities (Coyle, 2015). For instance, instead of buying maps, travelers would rather access Google Maps or Citymapper to find their way around. In such cases, travelers receive the same (or even better) level of benefit without incurring explicit costs. However, this weakens the ability of official statistics, such as GDP and Household Final Consumption, to reflect welfare changes, since households are gaining utility from the consumption of goods and services that do not have a market and are therefore not explicitly reflected in the National Accounts.

The popularity of free digital services has motivated a strand of the measurement literature to account for the value derived by households from these services (Corrigan et al., 2018; Brynjolfsson et al., 2019b,a; Coyle and Nguyen, 2023; Jamison and Wang, 2021; Nakamura and Soloveichik, 2015; Nakamura et al., 2017; Van Elp and Mushkudiani, 2019; Van Elp et al., 2022). The goal of these studies is to generate aggregates that capture the value of these services and partially address the current weakness of GDP estimates in terms of serving as an indicator of welfare.

One can argue that digital piracy can also cause this substitution. Dejean (2009) observed that music sales have been falling since the late 1990s. A number of studies attribute this decline to the rise of file-sharing networks that make it easier to download songs instead of buying albums (Peitz and Waelbroeck, 2004; Zentner, 2005; Liebowitz, 2008).

By imputing for the value derived by households from the consumption of media accessed through piracy, it would be possible to better measure welfare changes contributed by the creative industries such as film and music. But is it even proper to measure services from digital piracy as part of output and consumption? For one, they are illegal. Second, they cause negative externalities such as revenue losses for industries.

The National Accounts do not discriminate between legal and illegal consumption. Even though digital piracy is a form of intellectual property infringement and is therefore illegal in most countries, final consumption of a service still transpires. Paragraph 25.25 of the 2008 SNA writes:

> *[I]n principle, the fact that an activity may be illegal is not a reason to exclude it from the production boundary. In some countries, the difficulties of capturing illegal activities may mean that they are either not well covered or deliberately ignored on pragmatic grounds. However, some countries ignoring the production of drugs, for instance, would seriously underestimate the overall level of economic activity* (United Nations and others, 2009).

There have been substantial efforts aimed at improving the measurement of illegal activities in National Accounts estimates (Abramsky and Drew, 2014; Blades, 2011; Sturgess et al., 2018; Kazemier et al., 2013; Dragusin, 2015; Soloveichik, 2019). For the UK, two categories of illegal activities are included by the Office for National Statistics in their GDP estimates. These are illegal drugs and prostitution (Abramsky and Drew, 2014).

Furthermore, while digital piracy could cause negative externalities such as loss in revenues to creative industries (see the literature review of Dejean (2009)), the National Accounts statistics have long included other transactions that also cause negative effects on individuals and society. For instance, even though cigarettes have been known to cause negative effects on a person's health, the consumption of cigarettes is still included in GDP as part of household final consumption.

It stands to reason that if there are efforts to estimate the value from the consumption of free digital products, the value from digital piracy should be taken into consideration as well. Based on the accounting principles of the SNA, there is no reason why the value from digital piracy should be excluded in attempts to measure welfare from the internet. Generating such measures would also provide us with a more complete picture of welfare changes in the digital era.

Not all illegal activities are considered part of the production boundary in the National Accounts. For example, the SNA (see paragraph 3.97) argues that theft should not be regarded as a form of production because 1) it does not produce any value; assets are simply transferred from one entity to another, and 2) theft is not considered a "voluntary transaction." The same argument applies to embezzlement. While most legal systems consider digital piracy as "intellectual property theft," we argue that a transaction occurs when households consume pirated content. When individuals watch a movie via illegal streaming sites, value is created for the viewer. Unlike theft, which only involves asset transfer, in this case, value is generated, similar to how entertainment services are created when individuals watch Netflix.

To put this in perspective, when a person steals another person's handbag, the act of stealing does not constitute a productive activity because it is considered as transfer in kind. But once the thief sells the handbag, then there is an explicit transaction between a seller and a buyer. We argue that this is what we are capturing when we measure the value household receive from digital piracy. While the theft of intellectual property does not constitute production in the context of the National Accounts, the provision of content in exchange for advertising viewership in illegal streaming sites does.

4.3 Literature on the impact of digital piracy

Various studies have been conducted aimed at evaluating the impact of digital piracy on society. Most of them, however, were focused on assessing its impact on sales in the creative industry.

Dejean (2009) explains that there are two types of piracy. The first is called *hard goods piracy*. This involves the duplication of the physical medium containing intangible assets, such as CDs and DVDs. The second type of piracy involves the replication of intellectual property through the internet using file sharing. Dejean (2009) referred to this as *digital piracy* or piracy through *peer-to-peer* (P2P) networks. Because of how the expansion of broadband access has improved internet download speeds globally, most of the recent empirical studies are aimed at examining the impact of piracy from P2P networks.

Due to the lack of data on infringement in the 1980s and 1990s, early studies on the effects of piracy were mostly focused on developing theoretical models that predict the impact of copying. In an analytic model, Johnson (1985) showed that copying, in the context of hard piracy, reduces social welfare. Restrictions that are successful in inhibiting piracy, on the other hand, improve overall surplus. The theoretical literature notes two ways in which economic welfare is reduced due to piracy. First, the cost of originals could increase in order to internalize the cost of infringement (Ordover and Willig, 1978; Liebowitz, 1985; Besen and Kirby, 1989). Second, due to network effects, piracy could cause the value of originals to increase, following a rise in demand (Besen and Kirby, 1989; Takeyama, 1994).

Some of the first empirical studies that aimed to evaluate the effects of digital piracy on outcomes employed the spread of broadband networks as a proxy for IP infringement activity. Again, this is due to the lack of reliable data at the time. Country-level studies by Peitz and Waelbroeck (2004) and Zentner (2005) found that digital piracy played a relevant role in the decline in CD sales. A city-level analysis by Liebowitz (2008) corroborated these results, concluding that file sharing is responsible for the drop in music industry sales. At the time of these studies, low-cost streaming services, such as Spotify, were not as widespread as they are today. As such, the decline in CD sales from an increase in paid music streaming is not likely.

While research employing broadband usage as a proxy painted digital piracy as the cause of the decline in the output of creative industries, results from studies using direct measures of IP infringement are less conclusive. Oberholzer-Gee and Strumpf (2007) examines the impact of the number of illegal album downloads from the portal OpenNap on their sales. In order to address endogeneity, the authors used data on international school holidays as an instrument for illegal downloads. They found that the number of downloads from the portal does not cause a significant change in album sales. Meanwhile, Blackburn (2007) argued that P2P piracy has two potential externalities. First, digital copies are able to substitute for originals, causing declines in sales. Second, file sharing causes a penetration effect that allows for sales of creative output to increase. The study found that the substitution effect is greater for well-known artists. Meanwhile, the penetration effect benefits artists that are less known. Since total sales in the music industry are driven by sales of popular artists, the decline in overall sales is likely due to the substitution effect.

Meanwhile, results from micro-level studies provide evidence that file sharing negatively impacts sales in the creative industry. Zentner (2006) employed internet sophistication and internet speed as an instrument for P2P network usage. They found that the use of these networks reduces the likelihood of music purchases by 30 percent. Using data from the US Consumer Expenditure Survey, Michel (2006) and Seung-Hyun (2005) found that P2P networks negatively impact creative industry sales.

Rob and Waldfogel (2006) takes a more nuanced analysis by examining the welfare effects of digital piracy. They argue that illegal download of media actually improves consumer welfare by reducing deadweight loss (see Figure 4.3.1). In their survey of 500 students, they found that respondents generally download albums that are more expensive than their willingness to pay. P2P networks allowed these students access to music that they would not have been able to purchase in the first place. They conclude their findings by saying: "*[w]hile downloading reduces expenditure (on hit albums, 1999–2003) by $25 per capita in the subsample for which we perform a direct welfare analysis of downloading, it raises sample consumers' welfare associated with these albums by $70 per capita.*"

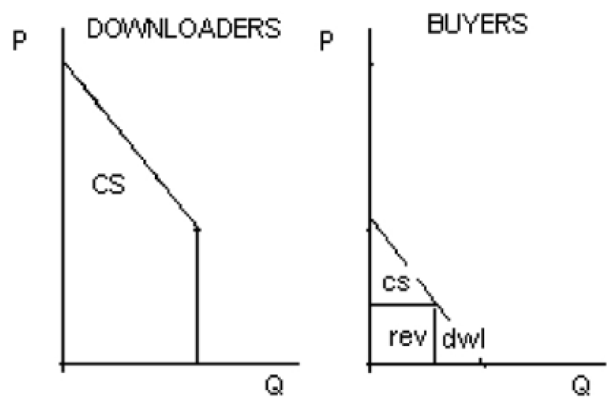


Figure 4.3.1: Reduction of deadweight loss from digital piracy

Notes: Figure sourced from Rob and Waldfogel (2006).

Some scholars have also generated estimates of the dollar-value of industry loss due to piracy. Most of these estimates are based on displacement rates from regressions that evaluate the effects of illegal downloads and industry sales. Blackburn (2007) estimated that reducing P2P network activities by 30 percent would enable the music industry to gain around US$260 million in revenues. Meanwhile, De Vany and Walls (2007) found that the US movie industry loses US$41.7 million annually from piracy. These estimates are generated using back-of-the-envelope calculations, based on a certain hypothesized relationship between piracy and revenues. However, they do not provide an estimate of the value of illegally accessed creative content.

A report released by Blackburn et al. (2019) calculated both industry loss from piracy and the value of illegally accessed content in the US. Based on a displacement rate of 14 percent, they estimated that piracy cost the country's film and TV industry between US$29.2 billion to US$71.0 billion. Their estimates also show that the value of digital content was US$166.6 billion for film content and US$200 billion for TV content. They arrived at these estimates by calculating the average industry revenue per movie and per TV episode, and then multiplying the average revenues by the number of illegally accessed movies and TV episodes.

While the report by Blackburn et al. (2019) was mostly focused on the industry revenues displaced by piracy, it is, to the best of our knowledge, the only study that attempts to generate the aggregate value of pirated content. The main drawback of their methodology is how they value each unit of pirated content. TV and film companies receive revenues from sources other than payment for content. For instance, film companies also generate revenues from merchandise and advert placement, which is on top of their earnings from ticket sales and DVD sales. This may cause the average revenues per content to be a biased estimator of the exchange value for creative output. We address this gap in the literature by using the price of paid products as a proxy of the value derived by households from the consumption of illegally acquired media. In the following section, we discuss the measurement strategy we employ for the estimation.

4.4 Measurement strategy

For this study, we employ the price of similar market products as our source of valuation of digital piracy products. While the use of cost production (see discussion in Chapter 3) is also a valid measurement strategy for non-market valuation, this approach is limited in terms of capturing the welfare improvements if the marginal cost of producing the service is zero (or close to zero). Meanwhile, the use of market equivalent pricing is an alternative method to measure the value from digital piracy since similar products can easily be identified with explicit monetary value. Moreover, the valuation would be consistent with National Accounting practice, as discussed in Chapter 3. This method also aligns with the concept of measuring the increase in households' implied income from the availability of free digital services, as detailed in Chapter 2.

In this study, we measure the value of digital piracy for seven types of products, namely music, live sports, software, computer games, ebooks, film, and TV. We employ the paid version of these products as a proxy for the value of their pirated counterparts. This strategy is similar to the valuation approach in the previous chapter, which employs the price of paid versions of digital services as a proxy for their free counterparts.

For TV, film, music, and sports, we employ the price of paid media content as a proxy for the shadow price of pirated media. Paid media is priced differently depending on the popularity of the content and/or the artist, making it difficult to determine an average price that represents access to certain media. The recent popularity of streaming services makes this task simpler. Services such as Spotify, Netflix, and Amazon Prime allow users to access unlimited content for a fixed monthly fee. This fee represents the value consumers derive from having the option to watch or listen to creative content for a certain period. One can argue that P2P networks provide similar services. For instance, an individual can watch unlimited movies and TV series from illegal streaming sites such as 123Movies, FMovies, and similar portals. As such, we employ the price of these streaming services as a proxy for the shadow price of illegally accessed media. Here, we assume that paid streaming services and illegal streaming services are substitutes. However, we must note that this strategy would likely provide *conservative* estimates of the value of digital content. The libraries of paid streaming services such as Netflix and Disney Plus are limited, in contrast to illegal streaming sites, which often host a materially larger library of content. We argue, however, that this does not invalidate our strategy because the limitation of paid streaming is likely offset by the disutility from accessing illegal media, which we will discuss in the following section.

To estimate software, computer games, and ebook piracy, we employ information from the Personal and Household Finances survey of the UK's Office for National Statistics. In particular, we use data in Workbook 1 (Detailed Expenditure and Trends), which includes information on weekly household expenditures for certain items. We use expenditures on software, videogames, and books as a proxy for the value of software, videogames, and ebook piracy, respectively. We multiply the weekly value to estimate monthly expenditures per household. Since these expenditure values are for the household, we divide the estimated monthly expenditures by the average household size (based on ONS data) to arrive at the value for each individual.

To arrive at the implicit price of illegally accessed music, TV and film, and live sports, $\hat{p}^d$, we take the weighted average of each paid streaming service, p^p:

$$\hat{p}^d_{it} = \sum_{j=1}^{J} (p^p_{ijt} \times \theta_{ijt}) \tag{4.1}$$

where t is an index representing the year, i is an index representing the type of content (video, music, sports), j is an index representing providers of paid content (Netflix, Disney Plus, etc), and θ_{ij} representing the market share of each content provider. We derive the market share of each provider by taking the ratio of their subscribers to the total number of subscribers across all providers. To arrive at the gross value of illegal media consumption, we multiply the estimate of the shadow price of pirated content in equation 4.1 to a measure of volume. Since the price that we used for the computation of the shadow price is based on the price per user, by extension, the measure of volume we need for the estimation of gross value should also be based on the number of users. As such, we assume that households derived utility just by having access to illegal streaming. The gross value of pirated content can be expressed as,

$$\hat{V}^d_{it} = \hat{p}^d_{it} \hat{q}^d_{it} \tag{4.2}$$

where the gross value V_t is the product of the shadow price $\hat{p}^d_{it}$ of the free media i and its measure of volume $\hat{q}^d_{it}$ at time t. Since these estimates makes no reference to intermediate consumption, we interpret these figures as the gross value of consumption in the expenditure side of the National Accounts.

We apply the same principles to derive the value of illegally-downloaded software, computer games, and ebooks. Instead of using a price as an indicator of value, we employ the monthly expenditures of households from the Personal and Household Finances survey. This places a revealed preference assumption on our estimation strategy. We assume that the average monthly expenditures reflect the value derived by households from the consumption of these products.

4.5 Data

In this section, we present the data employed for the estimation of the gross value from digital piracy. The goal is to generate estimates for the years 2016 to 2021, though this methodology could extend the estimates further back in time (as long as data on prices and the volume measure for infringement are available).

To estimate the gross value from the consumption of digital piracy, we multiply the shadow price of illegal media consumption with a measure of volume, $\hat{q}^{d}_{it}$. Since the proxy price we employ in this exercise is based on the number of users, one can argue that the appropriate measure of volume is the number of users of digital piracy.

0%	Wave 5 (2015)	Wave 6 (2016)	Wave 7 (2017)	Wave 8 (2018)	Wave 9 (2019)	Wave 10 (2020)	Wave 11 (2021)
Music	27%	31%	32%	33%	40%	37%	42%
TV programmes	32%	33%	35%	31%	42%	42%	44%
Films	20%	22%	26%	28%	34%	42%	45%
Video games	8%	8%	9%	10%	11%	11%	13%
Software	6%	6%	8%	8%	9%	8%	10%
Live sports					14%	8%	15%
E-books	6%	7%	7%	6%		8%	9%
E-publishing					12%		
Digital magazines						5%	7%
Audiobooks						5%	6%

Figure 4.5.1: Online Copyright Infringement Tracker Report of the UK's IPO

Note: The table is sourced from the 2022 Online Copyright Infringement Tracker Report of the UK's IPO. The table shows the percentage of survey respondents who said that they engaged in infringement activities for certain media.

We employ information from the 2021 Online Copyright Infringement Tracker Report of the UK's Intellectual Property Office (IPO). The survey aims to track the consumer behavior of individuals 12 years old and above in relation to online copyright infringement. The survey is conducted annually. The report released in January 2022 is the 11th installment of the survey. For the latest report, the survey employed a sample of 5,000 individuals, representative of the UK's 15+ population.[1]

[1] While we preferred to use the population of individuals 12 for our estimates, this is not available in the ONS data. As such, it is possible that our final estimates would have some downward bias.

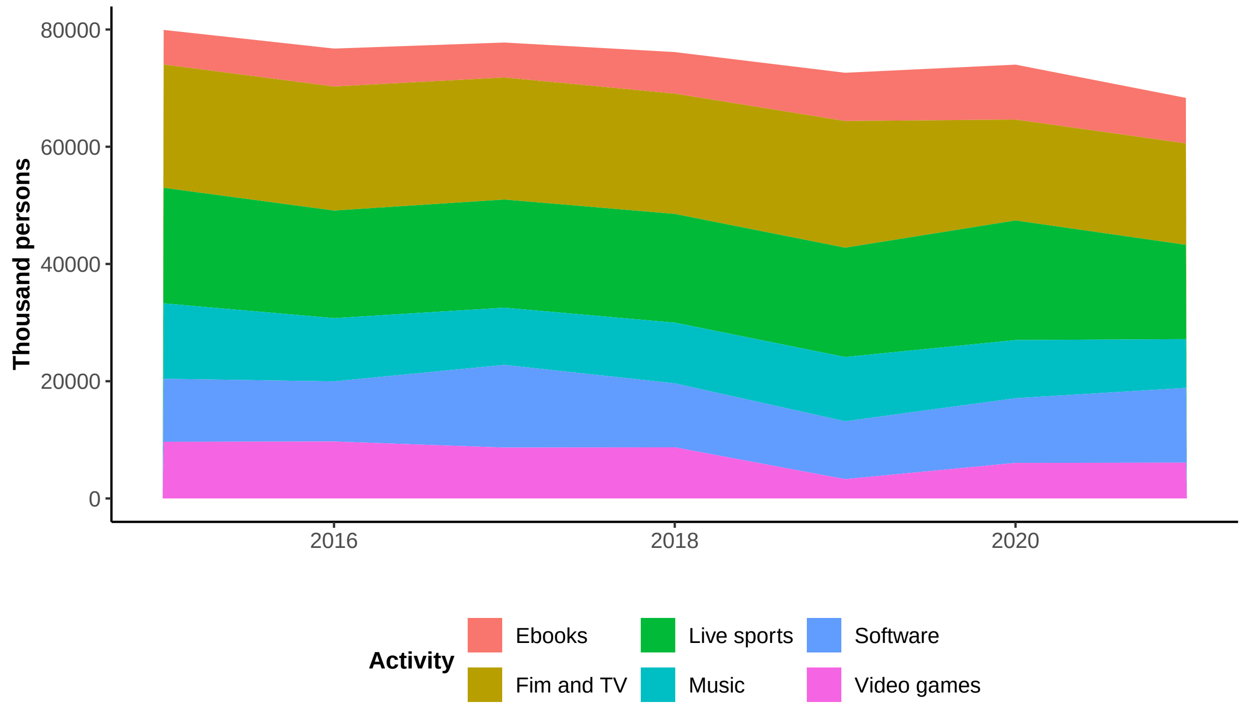


Figure 4.5.2: Estimated number of individuals engaged in digital piracy

Note: The figure shows the estimated number of individuals engaged in digital piracy. The sources of basic data are the UK's IPO and the ONS.

We plot the estimated number of individuals engaged in digital piracy in Figure 4.5.2. Video (TV and film) and live sports make up the majority of the estimates. On average, video and sports accounted for 29.8 percent and 27.7 percent of the total, respectively. We also observe that the estimated number of individuals engaged in piracy is generally declining over time.

To derive the shadow value of products from digital piracy, we use information on both the subscription price of equivalent products and the average expenditures of similar products. For music, we take the weighted average price of the five most popular paid streaming platforms, namely, Spotify, Apple Music, Amazon, YouTube, and Soundcloud Go+. These platforms offer two sets of prices: standard and a student discount price. We take the weighted average of these prices. We used the population of individuals from 16 to 21 to arrive at the weight for the student price. To acquire the prices for previous years, we accessed the price for the previous versions of the website using the portal *Internet Archive* (www.archive.org).

For video streaming, we employ the average price of popular streaming sites from 2016 to 2021. We use price data from Netflix, Amazon Prime Video, Now Entertainment, Apple TV, and Disney Plus. For sports, we use the subscription prices of Now Sports and BT Sports. We generate two types of price estimates for video and sports: the basic price and the price per screen. The basic price reflects the lowest subscription price available on the provider's website. Meanwhile, we derive the price per screen by dividing each available subscription price by the number of screens users can potentially watch for each subscription. For instance, the basic subscription to Disney Plus allows users to watch on four screens simultaneously.

Table 4.5.1: Weighted average prices in GBP, monthly

		2016	**2017**	**2018**	**2019**	**2020**	**2021**
Music		8.7	8.7	8.7	8.7	8.7	8.7
Film and TV	Basic	7.3	7.4	7.4	7.9	7.6	7.9
	Per Screen	3.2	3.2	3.4	3.7	3.3	3.4
Sports	Basic	6.0	7.2	7.7	8.0	18.7	26.3
	Per Screen	3.7	4.9	5.1	6.0	8.5	11.0
Computer software (including games)	Household	2.4	2.8	2.8	3.2	2.8	3.6
	Individual	1.0	1.2	1.2	1.4	1.2	1.5
Books	Household	4.8	4.8	5.2	4.8	4.8	5.6
	Individual	2.0	2.0	2.2	2.0	2.0	2.4

Note: The table shows the average prices of popular streaming platforms from 2016 to 2021, as well as the estimated monthly expenditures of households for computer software and books. Basic data on household expenditures were sourced from the Household Finances survey of the UK's ONS. Estimates of individual-level expenditures are derived by dividing the household expenditures by the average household size for each year. Prices of streaming services are found in appendix 4.A. All figures are in GBP.

For both computer games and books, we employ user data on the monthly expenditures of households from the Personal and Household Finances survey of the ONS. We have two options for this approach. The survey tracks weekly expenditures of *households*. However, the volume measure we derived from the IPO report is based on the number of *individuals* involved in digital piracy. Because of this, we estimate the expenditures for each individual by dividing the household expenditures by the average household size in the UK[2].

We show the average prices over time in Table 4.5.1. The average price of music is flat from 2016 to 2021. Moreover, we find no cross-sectional variation in the price. The subscription price of music is the same for all providers. The mean subscription price for video and sports is more variable. The average subscription prices for the two categories are both increasing over time. The average subscription price for live sports increased the fastest for the period 2016 to 2021 from £6.0 per month to £26 per month.

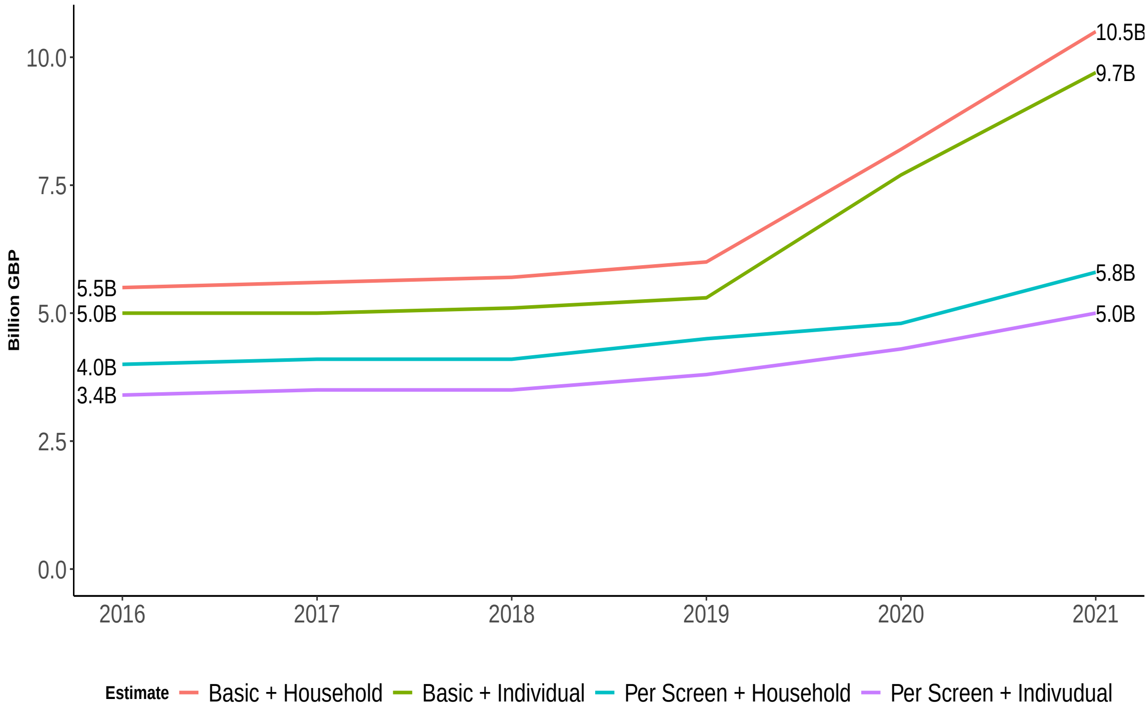
The average expenditures for computer software (including games) and books are likewise increasing. At the household level, the expenditures for software increased from £2.4 per month in 2016 to £3.6 in 2021. Meanwhile, the average expenditure for books increased from £4.8 per month to £5.6 per month.

[2]Data on the average household size is sourced from the ONS. The data varies each year.

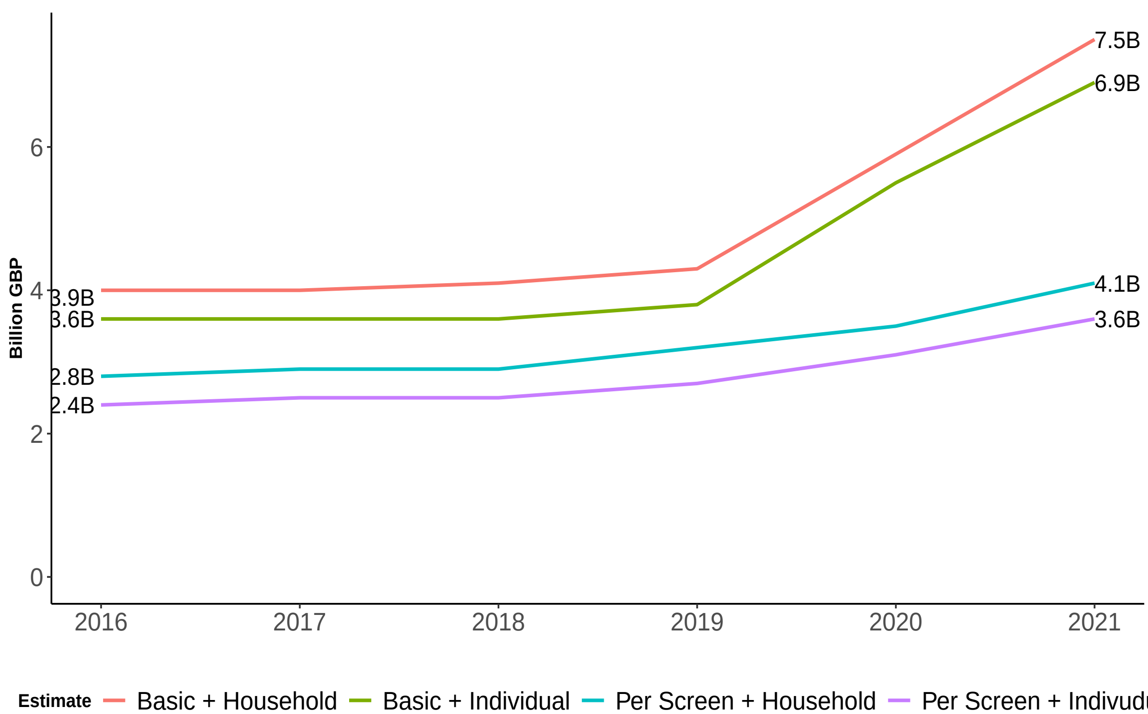
4.6 The gross value of digital piracy

In this section, we report our estimates for the gross value of Digital Piracy. We use the formula in equation 4.1 to generate our estimates.

We show the baseline estimates in Figure 4.6.1a. We report the four iterations of the estimates. In the first iteration, we employed the basic price of streaming services and the household-level average expenditures for books, computer software, and games. For the second set of estimates, we used the basic price of streaming services and the estimated individual-level expenditures for books, software, and games. The third iteration employs the average price per screen that we calculate for streaming services and household-level expenditure values for books, software, and games. In the last set of estimates, we use the price per screen and the estimated individual-level expenditures for books, software, and games. The fourth set is our preferred set of estimates because it provides us with conservative values. As such, we can interpret them as the lowest possible estimates for the aggregate value of digital piracy. For discussion purposes, we interpret the first set of estimates as the *upper bound* figures and the fourth set of estimates as the *lower bound*.



(a) Baseline



(b) Adjusted

Figure 4.6.1: Gross value of digital piracy, at current prices

Note: The figure shows the estimates for the gross value of digital piracy. Left figure shows the baseline estimates from equation 4.2. Right figure shows the estimates adjusting for the disutility of advertising, where the prices are derived using equation 4.3.

The estimates show that in 2016, the gross value of digital piracy is between £3.4 billion and £5.5 billion. The data shows that the nominal value of piracy is generally increasing over time. The lower bound estimate grew to £5.0 billion in 2021. Meanwhile, the upper bound estimate jumped to £10.5 billion for the same period. The sharp increase in the upper bound estimates is driven by the increase in the average basic price of live sports subscriptions.

Our estimates show that live sports accounts for the largest share of the gross value of digital piracy (35 percent). Music and video accounts for the second and third largest share at 28.3 percent and 23.9 percent, respectively. With the exception of 2021, the percentage distribution across products appears to be stable over time. We hypothesize that the increase in the value of live sports that year was a result of the Euro 2020 football tournament.

One can argue that these figures overestimate the value of digital piracy. Because we employ the price of paid market products as a proxy for the illegally downloaded products, there is an argument to be made that these valuations are not equivalent. One clear distinction is the presence of advertising in illegal streaming sites. Most paid streaming services are ad-free while illegal streaming sites often subject their users to excessive advertising. This is also often the case for websites that allow for the illegal download of software, computer games, and ebooks. We control for this difference by introducing an adjustment term:

$$\hat{p}i^{aj} = \hat{p}i^{d}\gamma, \tag{4.3}$$

where $\hat{p}_i^{aj}$ is the adjusted shadow value of digital piracy. In equation 4.3, we introduce a disutility parameter γ which takes the value $0 < \gamma < 1$. There are many possible reasons why the experience of consuming content from paid sources is better than consuming content from illegal avenues. Among these is the ease of access, security, and assurance that accessing media will not result in accidental download of malware. Ideally, γ should be able to capture the disutility from all the reasons mentioned. Because of the lack of data, we are only able to adjust for the disutility from advertising. Recently, the streaming company Netflix offered an advertising-supported tier. This allows users from selected markets to gain access to a cheaper (30 percent cheaper than the standard subscription) version of Netflix, but with intermittent advertising while viewing. We use the ratio of the price for the ad-supported tier over the standard subscription as our basis for γ. While this ratio is not perfect in capturing all factors contributing to the difference between the value of illegally accessed digital products and their paid counterpart, this is the best empirical proxy we are able to employ thus far.

We show the results of the adjusted gross value of digital piracy in Figure 4.6.3. The upper bound estimate for 2021 is at £7.5 billion and the lower bound is at £3.6 billion.

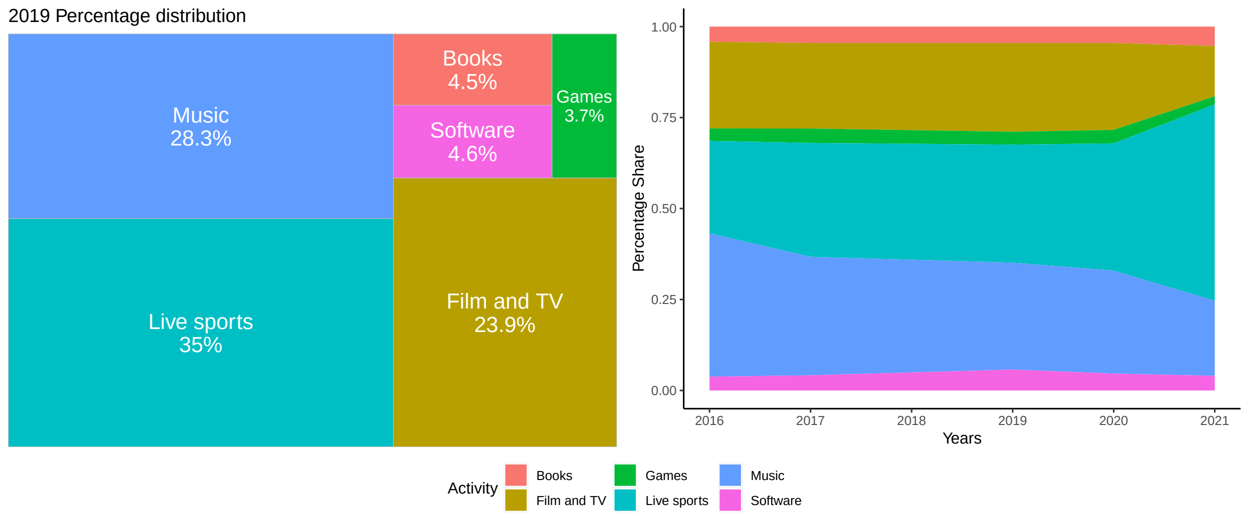


Figure 4.6.2: Percent distribution of digital piracy products

Note: The figure shows the percent distribution of the gross value of each illegally accessed digital service. The left figure shows the share of each product in 2019. The right figure shows the percent distribution over time.

We also generate constant price estimates for the gross value of digital piracy. We first construct a Laspeyres index by extrapolating the base year prices with the growth rates of the shadow price of each product (see equation 4.4). We chose 2019 as our base year to be consistent with the National Accounts estimates of the ONS. We then deflate our nominal value estimates using the price index (see equation 4.5) to arrive at the constant price estimates:

$$Index_i^t = Index_i^{t-1} \times (\hat{p}_i^t / \hat{p}_i^{t-1}) \tag{4.4}$$

$$\hat{V}i^{R,t} = \frac{\hat{V}i^{N,t}}{(Index_i^t/100)} \tag{4.5}$$

Here, $Index_i^t$ denotes the price index at time t, $\hat{V}i^{N,t}$ represents the estimates for the gross value of digital piracy in nominal terms, and $\hat{V}i^{R,t}$ represents the gross value in real terms.

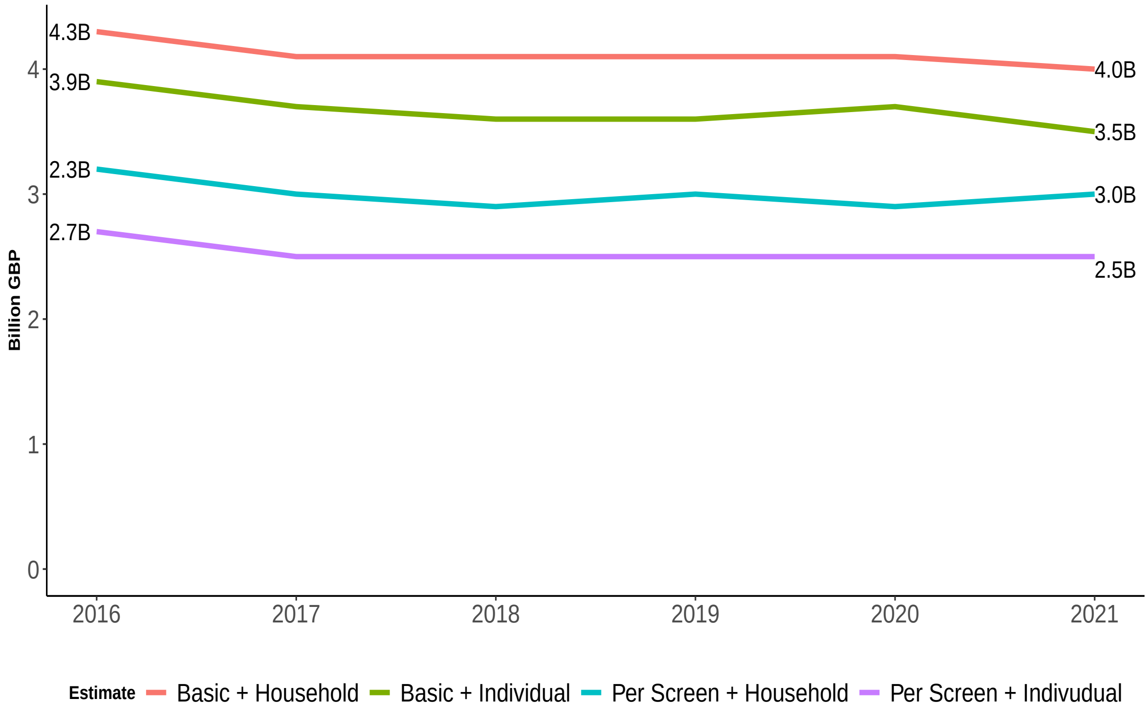


Figure 4.6.3: Gross value of digital piracy at constant (100 = 2018) prices

Note: The figure shows the constant price estimates of gross value of digital piracy derived using equation 4.5.

Our estimates suggest that the constant price estimates of the gross value of digital piracy, with 2019 as the base year, have generally declined from 2016 to 2020. This decline reflects the downward trend in the number of individuals engaged in digital piracy.

4.7 Discussions and applications

The advantage of generating estimates consistent with the accounting principles is that it allows for the estimates to be linked with the core National Accounts data, providing a benchmark for interpretation.

In Figure 4.7.1, we present the share of digital piracy (at current prices) in certain household consumption items in 2019. We find that digital piracy accounts for a small share (0.2 percent) of total household final consumption expenditures. Relative to total communication expenditures—which includes paid counterparts to video and music streaming, computer software, and computer games—digital piracy makes up 8.8 percent. This implies that in aggregate terms, households derive a relatively small share of their total consumer welfare from digital piracy.

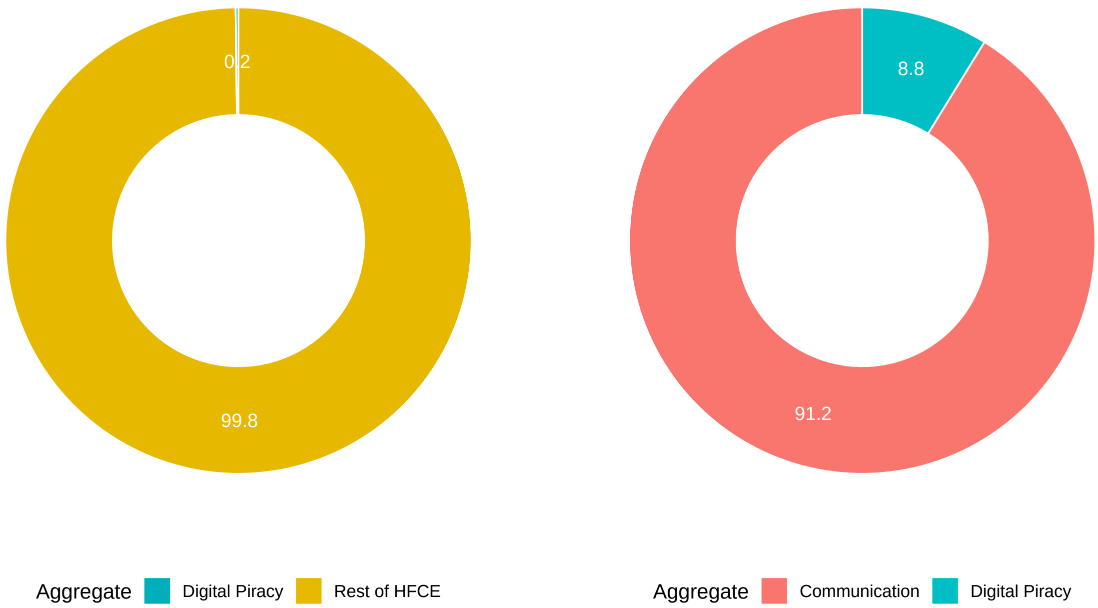


Figure 4.7.1: Share of the value of piracy in household consumption

Note: The figure shows the share of the gross value of digital piracy (at current prices) in Household Final Consumption Expenditures (left figure) and the gross value of Household Consumption Expenditures in Communication (right figure). Source of basic data on HFCE is the UK's National Accounts from the ONS.

We can also analyze the pace at which the consumption of digital piracy is growing relative to other goods and services consumed by households. We show the growth rates in Figure 4.7.2. The gross value of communication expenditures is consistently growing from the period 2016 to 2021. Meanwhile, the gross value of piracy has declined in three of the five years in the reference period. Summing the value of digital piracy to the expenditures in communications, we find that this aggregate is growing at a slower rate compared with the gross value of communications as reflected in the National Accounts (see right figure). One interpretation that we can give to this is that the National Accounts tends to overestimate the value from the consumption of services such as video and music streaming, sports, software purchases, among others. Households were already consuming a high level of these services through piracy. The welfare derived from this consumption, however, is not reflected in the National Accounts estimates. Perhaps with the advent of cheaper streaming services combined with better enforcement of intellectual property laws, households are switching to the consumption of paid services, artificially increasing the consumption of communication products as reflected in the National Accounts.

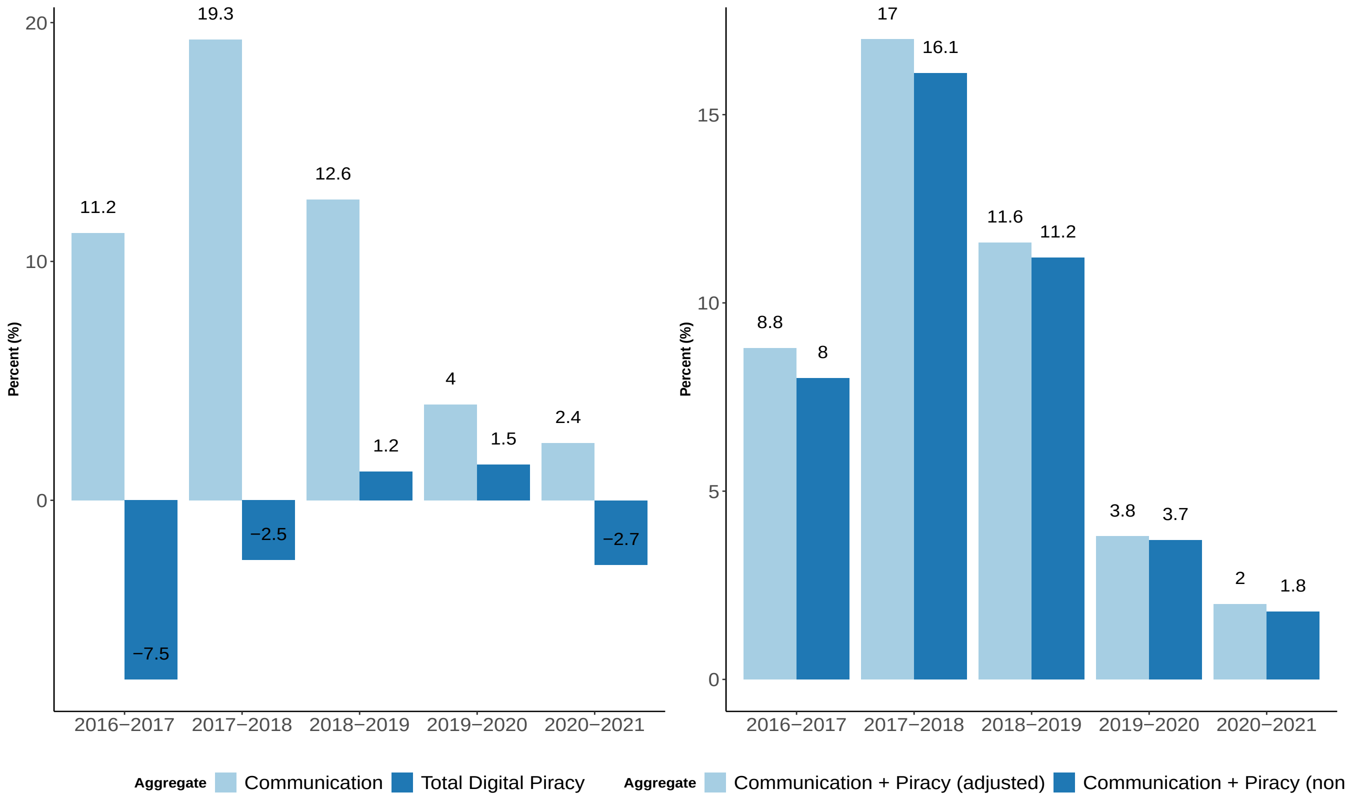


Figure 4.7.2: Growth rates of digital piracy consumption of communication services

Note: The figure shows the growth rates of the gross value of digital piracy and the value consumption of communication services. All figures are at constant prices. The source of basic data for communication expenditures is the UK's ONS.

It would also be interesting to compare the value of digital piracy to the value of consumption of other illegal services in the UK. We compare the gross value of digital piracy with the gross value of narcotics and prostitution, the only illegal products recorded in the UK's National Accounts (see Figure 4.7.3). We find that the value of digital piracy is comparable with the value of narcotics but below the value of prostitution. It must be noted that not all narcotics and prostitution are illegal in the UK (the ONS does not provide a breakdown between legal consumption and illegal consumption). As such, it is possible that the value of consumption of illegal narcotics and illegal prostitution is comparable or lower than the value of digital piracy. If so, this suggests piracy is more prevalent than the other two.

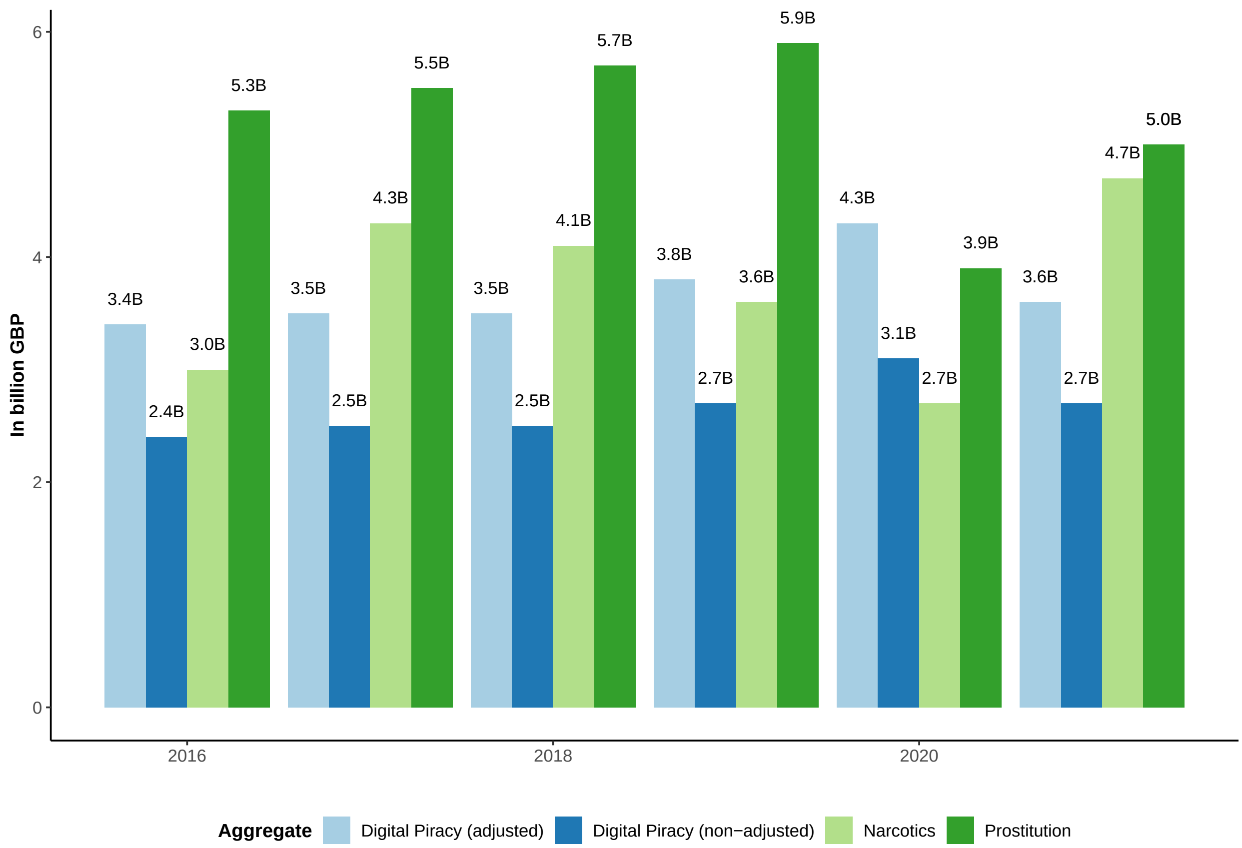
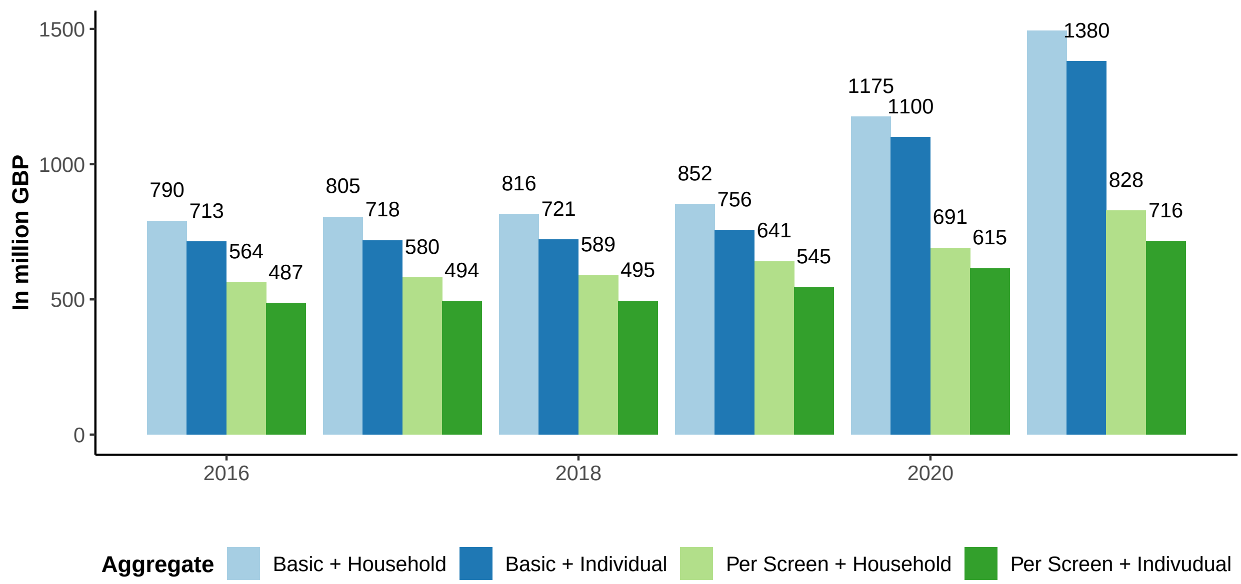


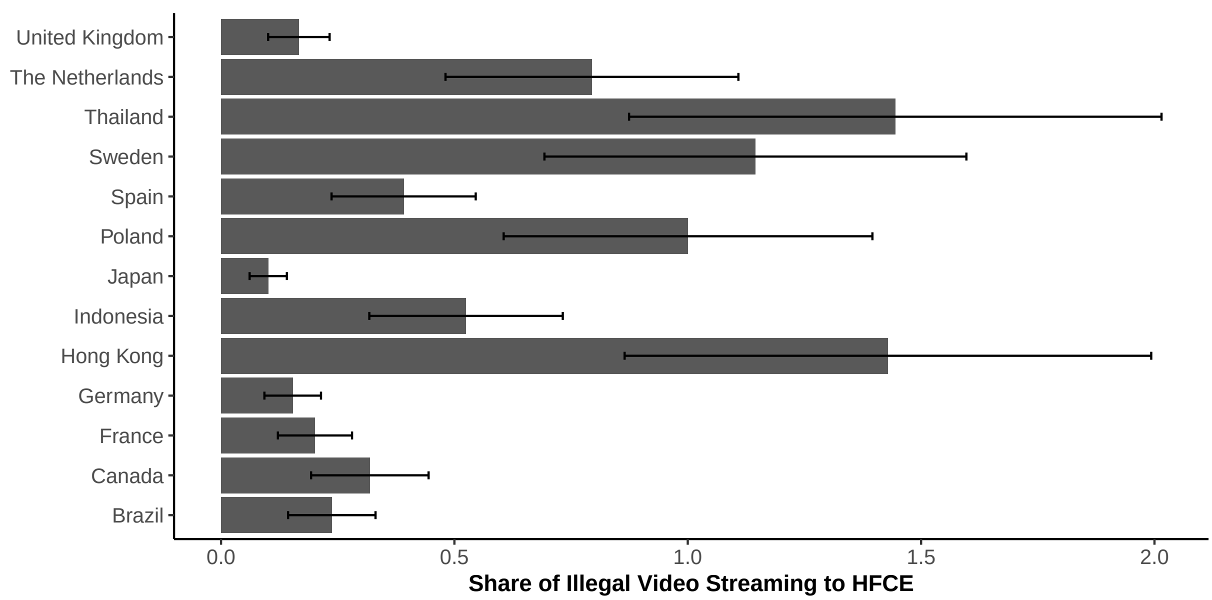
Figure 4.7.3: Gross value of illegal goods and services and digital piracy in the UK

Note: The figure shows the gross value of narcotics, prostitution, and digital piracy (at constant prices) in the UK. Source of basic data is the UK's National Accounts from the ONS.

Estimates on the gross value of digital piracy can have various policy applications. For instance, we can generate estimates of "forgone" government revenues as a result of piracy. We can interpret this as the revenues that the government could have raised from taxes had these services been paid and subject to levies. This is different from displacement, which requires information on the elasticity of paid services to their illegal counterpart. We show the estimates in figure 4.7.4a. Assuming that a 20 percent VAT could be levied on these products had they been legal, our estimate is that in 2021, the government could have raised *at least* £700 million. To put this in perspective, this is the estimated budget shortfall for the City of London in 2023 (see Chantler-Hicks (2022)).



(a) Estimated VAT revenues



(b) Cross-country comparison

Figure 4.7.4: Other applications

Note: The figures show estimates on the foregone VAT revenues from piracy (top figure) and estimates of the gross value of video streaming piracy in selected countries (bottom figure).

Multilateral institutions and academics may also be interested in comparing the value of piracy across countries. We estimate the gross value of illegal video streaming for other countries using data from Quintais and Poort (2018). The results are shown in Figure 4.7.4b. We find that the value of piracy in the UK is actually lower compared to other countries.

Based on our estimates, Thailand received the largest value. One can argue piracy is perhaps one avenue that households in developing countries use to "equalize" welfare to some degree. While many individuals from developing countries are not able to subscribe to streaming services such as Netflix and Spotify, they are still able to access the same content through piracy, and these estimates can aid in analyzing this phenomenon.

4.8 Concluding remarks and recommendations

In this exercise, we show that the value derived by households from the illegal consumption of media and other digital products can be estimated using existing data. It is interesting to see that the value of digital piracy is declining over time. We offer two likely explanations for this. First, it is possible that the government's crackdown on copyright infringement is causing a decline in digital piracy. While this is a positive development for content producers, we can see from the estimates that this also has an effect on the level of welfare derived by households from the consumption of content. Second, the popularity of low-cost streaming could also be a factor. While the price of Netflix has increased over the years, it is still relatively low. As such, it is possible that more and more households are opting for paid subscriptions as opposed to incurring the cost of piracy.

There are three caveats to our findings. First, we acknowledge that estimates employing the price of paid media as a proxy for illegally accessed media would provide an overestimate of the value of the latter. The logical way forward is to adjust the estimates to address this bias. Ideally, we need to adjust for all sources of disutility from the consumption of pirated media and other products. So far, we are only able to partially adjust for the presence of advertising. It can be argued that there are other reasons why the experience of consuming pirated content is different from the consumption of its paid version. Future research can focus on addressing the challenge of accounting for these other sources of disutility.

Second, we estimate the number of individuals engaged in digital piracy using the population projections for those above 15 years old. The proportion of digital piracy users in the IPO survey, however, is a representative sample for individuals 12 years old and above. As such, this might cause a downward bias in the estimates. We hope to address this by employing the appropriate population levels for the estimation of individuals engaged in digital piracy.

Lastly, the volume metric that we use in the measurement strategy is the estimated number of individuals engaged in digital piracy. As such, an individual who streams pirated content from two websites would only be counted once. This would likely result in an underestimate of the value of digital piracy since individuals should derive separate utility from the two sites. In the National Accounts, if an individual is subscribed to both Netflix and Disney Plus, the value they derived from the two streaming services would be counted separately.

While the methodology we employed is not perfect, we believe that our estimates are still able to provide useful insights on the size and evolution of the value derived by households from the consumption of pirated digital services. Future research in this area could attempt to address the caveats mentioned above.

Appendix

4.A Prices of Online Streaming

Table 4.A.1: Music streaming prices

	Headline Prices		Average Prices	
	Basic	Student	Simple Ave	Weighted Ave
2016	9.99	4.99	7.49	8.69
2017	9.99	4.99	7.49	8.69
2018	9.99	4.99	7.49	8.69
2019	9.99	4.99	7.49	8.69
2020	9.99	4.99	7.49	8.69
2021	9.99	4.99	7.49	8.69

Note: The table shows the monthly standard prices and monthly student prices of music streaming platforms. Prices are in GBP.

Table 4.A.2: Live sports streaming prices

	Headline Prices		Average Prices	
	Basic	Per Screen	Ave	Per Screen
2016	6.99	2.33	5	5
2017	6.99	2.33	7.5	7.5
2018	7.99	2.66	7.5	7.5
2019	5.99	2	10	10
2020	9.99	3.33	27.5	13.75
2021	25	8.33	27.5	13.75

	Netflix			Now Entertainment		Amazon Video		Apple TV		Disney Plus	
	Basic	Per Screen	Implied Price	Basic	Per Screen	Basic	Per Screen	Basic	Per Screen	Basic	Per Screen
2016	7.99	3.93	6.09	6.99	2.33	5.99	2				
2017	7.99	3.93	5.69	7.99	2.66	5.99	2				
2018	7.99	4.2	5.09	7.99	2.66	5.99	2				
2019	8.99	4.83	4.74	7.99	2.66	5.99	2	4.99	0.83		
2020	8.99	4.83	4.51	8.99	3	5.99	2	4.99	0.83	5.99	1.5
2021	8.99	4.83	4.17	9.99	3.33	5.99	2	4.99	0.83	7.99	2

Note: The table shows the monthly standard prices and monthly prices per screen of video streaming platforms. Prices are in GBP.

CHAPTER 5: What's the Depreciation Rate of Microsoft Office? Measuring the Depreciation of Intangible Assets

ABSTRACT

This study introduces a novel methodology to estimate the depreciation of intangible assets, specifically focusing on software and creative originals, using data from Google Search Volume (GSV), commonly referred to as Google Trends. Depreciation, in this context, is understood as a manifestation of obsolescence. As intangible assets become obsolete, their capacity to generate future output or revenue diminishes. GSV provides a practical means to gauge the popularity of products generated by these assets, as a surge in internet searches indicates their relevance. The decline in search activity over time is directly associated with the concept of obsolescence, aligning with economic depreciation principles. In our analysis, we employ Poisson Pseudo-Maximum-Likelihood (PPML) and negative binomial regressions to estimate the rate of decline of GSV results for a sample of software and movie titles. Our findings reveal a depreciation rate for software originals ranging from 13.4 to 19.4 percent. This is lower than the estimates employed by statistical agencies, which is around 20 to 25 percent. Estimates for movies are comparable to estimates by statistical agencies, notably the US and Germany. We also find that if we apply the depreciation rates we generated from this methodology to the estimation of capital stock, TFP growth for the Information and Communication industry would be higher particularly from 1996 to early 2008, and again from 2011 to the end of 2016.

5.1 Introduction

One of the reasons why statistical agencies present Gross Domestic Product instead of Net Domestic Product as their headline measure of economic growth is the difficulty of measuring depreciation. While net figures arguably better reflect welfare changes as compared to gross figures (O'Mahony and Weale, 2021), statistical agencies often find it difficult to separate the value of depreciation from gross operating surplus. This is true for many categories of physical capital assets, but the challenge is more pronounced for intangible assets such as research and development, creative originals, and software, where the physical decline in the condition of the asset cannot be observed.

There is growing evidence that suggests that advanced economies are strongly reliant on intangible investments as a source of growth and productivity. Corrado et al. (2009)

demonstrated that integrating intangibles into economic measures enhanced observed output per worker growth in the US. Additionally, Corrado et al. (2014) revealed that intangibles' impact on labor productivity growth surpasses labor quality's contribution. Research also suggests that intangible capital expansion correlates with total factor productivity gains (Corrado et al., 2022; Haskel et al., 2018). Moreover, the data shows that for sectors such as the Information and Communication Industry (ISIC code J), investments in intangibles have outstripped investments in tangible capital for some industries in the past two decades (see figure 5.B.2).

In light of the increasing body of research underscoring the crucial role of intangible investments in contemporary economies, the necessity to comprehensively quantify both intangible investments and the accumulation of intangible assets is becoming apparent. There are various strategies for measuring intangible investments. Van Criekingen et al. (2022) provides a comprehensive review of the modern approaches for the measurement of intangible assets. This review also highlights intricate challenges inherent to aspects of the measurement process, notably pertaining to issues such as depreciation and price deflators. For depreciation, it is the difficulty of observing the asset service life that makes the measurement exercise particularly challenging. Current approaches involve making assumptions on the service life of these assets (Corrado et al., 2009), surveys, or using projected future cash flows from production (Huang and Diewert, 2011; Soloveichik, 2010), which present their own difficulties.

Simulations conducted by Pionnier et al. (2023) show that the choice of depreciation rate for the estimation exercise largely affects the estimates of the capital stocks, and by extension, other aggregates such as productivity statistics. As such, it is imperative that the challenge of measuring depreciation should be addressed in order to accurately account for growth in the modern economy.

In this study, we develop a methodology for measuring the depreciation of intangible assets using data from Internet sources. In particular, we employ information from Google Search Volume (GSV), popularly known as Google Trends (trends.google.com), to generate estimates for the depreciation rates of software and creative originals. One can view depreciation as a form of obsolesce. As software and creative originals become obsolete, their ability to generate future output/revenues diminishes. GSV is potentially a practical way of gauging the popularity of intangible assets. For instance, if a large number of internet users search for the name of a particular movie, then we can assume that the movie is popular. As the number of searches fades over time, it follows that the movie's popularity is fading as well, along with its ability to produce future revenues. Often, the number of searches would peak at the month and year of a movie's release. Searches fade gradually following its release.

We assume that the decay in the search index is directly tied to the concept of obsolescence, which can be equated to the economic principle of depreciation. In our analysis, we employ Poisson Pseudo-Maximum-Likelihood (PPML) and negative binomial regressions to estimate the rate of decline of GSV for a sample of software and movie titles. Our findings reveal a depreciation rate for software originals ranging from 13.4 to 19.4 percent. This is lower than the estimates typically employed by statistical agencies, which is around 20 to 25 percent. In contrast, estimates for movies are comparable to estimates by statistical agencies, notably the US and Germany. Furthermore, our research shows that searches for recently released movies and software exhibit a steeper downward trajectory compared to movies released earlier, highlighting the need to regularly update depreciation rate estimates. We also find that if we apply the depreciation rates we generated from this methodology to estimate capital stock, TFP growth would be higher for the period between 2003 to 2019, and again from 2011 to 2016, than the original estimates.

This study contributes to the literature on addressing the challenges related to the accounting of intangible assets. There have been various proposals on how to capture intangible capital (Van Criekingen et al., 2022; Martin, 2019; Soloveichik, 2010; Corrado et al., 2009). Corrado et al. (2009) examines how the inclusion of intangibles would affect Macroeconomic aggregates in the US. This exercise incorporates non-SNA intangibles such as brands and firm-specific resources. Soloveichik (2010) provides a set of methodologies for the estimation of investments in artistic originals in the US. Her paper includes a methodology to estimate the depreciation of creative originals using the decline in their net present value. This approach requires an assumption of the asset's future revenue streams. Nadiri and Prucha (1996) proposes methodologies for measuring depreciation rates for Research and Development. The work outlined by Martin (2019) details the initiatives pursued by the UK's Office for National Statistics to address the measurement challenges associated with intricate intangibles like in-house branding investments, employer-sponsored training outlays, and in-house investments targeting organizational capital. These assets are currently not captured by conventional National Accounts estimates.

This study extends the literature in three ways. First, we provide a methodology for the estimation of depreciation using readily available data. Except for Soloveichik (2010) and survey-based approaches, most efforts rely on making assumptions about asset lives. Our methodology instead assumes the Google Trends results correlate with popularity, an indicator of the asset's ability to generate revenue streams for the asset owner, and consumer value for the households. We evaluate this hypothesis by testing the predictive power of GSV results on movie revenues. We also provide a methodology that can regularly be updated to adjust to changing preferences and economic conditions. Many statistical agencies and researchers rely on static depreciation rates that are rarely updated. Our methodology would allow for the estimation of depreciation rates for assets released at different periods. Changes in technology such as the availability of low-cost streaming and piracy have drastically changed how consumers experience and access media. This has led to many changes in the industry, including the shortening of the theatrical window (Ahouraian, 2021). The impact of this change is likely not reflected by depreciation rates that are rarely updated. Our methodology also has the advantage of being flexible for the adoption of different dimensions such as geographic coverage and asset classes.

Second, our methodology captures the depreciation of original software for reproduction. Survey-based approaches, where firms are asked the expected service lives of the software they employ in production likely capture the depreciation of software copies. The literature and current practices provide little information on the depreciation of original software (i.e. the decline in the value of the Microsoft Office program to Microsoft). In an age where software as a service is becoming more prominent[1], further exploration into the dynamics of original software depreciation is essential for understanding its economic implications and informing more accurate accounting practices.

Third, we also contribute to the literature on the mismeasurement hypothesis on the productivity puzzle. The continued slowdown in productivity in most developed countries following the 2008 financial crisis has been widely documented using both macro and micro-level data (Riley et al., 2015; Barnett et al., 2014; Goodridge et al., 2018). The failure to measure outputs and inputs correctly has been regularly cited as one of the possible reasons for the *observed* productivity slowdown (Goodridge et al., 2013a; Riley et al., 2018; Fernald and Inklaar, 2022; Roth, 2019). We examine the impact of changing the asset life on estimates of capital stocks and total factor productivity.

This essay is structured as follows. In the next section, we discuss how the depreciation of intangibles is characterized in the System of National Accounts and current approaches to estimating their value. We discuss our framework in section 5.3. We proceed by elaborating our empirical methodology in section 5.4. We describe our data in section 5.5 and discuss our results in section 5.6. We present the impact on capital stock estimates and productivity figures in section 5.7. We end with some applications and concluding remarks and ways forward.

5.2 Depreciation of intangibles

In this section, we discuss the current approaches to measuring the depreciation of intangibles. We begin by characterizing intangible assets and the challenges associated with the measurement of intangibles and their depreciation rates. We also discuss methods developed by other researchers and the estimates employed by statistical offices.

Depreciation estimates are central to measuring key economic aggregates such as capital stock (see appendix 5.A), which is an input to the estimation of productivity statistics and the estimation of net figures in the National Accounts.

The 2008 SNA refers to depreciation as the consumption of fixed capital (CFC). Conceptually, CFC captures the economic cost of expected physical wear and tear and anticipated obsolescence (Schreyer, 2004; OECD, 2009). Unanticipated reductions in the asset's value—for instance, the damages due to natural calamities—are not recorded as part of CFC. Rather, these changes in the book value are recorded as "other changes in volume assets".

Intangibles, unlike tangible assets, do not undergo physical wear and tear. Consequently, their CFC is predominantly linked to obsolescence, (OECD, 2010; Del Rio and Sampayo, 2014; Görzig and Gornig, 2015). Various perspectives exist regarding the concept of obsolescence. Diewert et al. (2006) describes obsolescence as a result of *demand shifts*. This occurs when products produced requiring the asset are no longer demanded by the market. On a

[1] Noted by the Harvard Business Review as the fastest-growing business model for tech entrepreneurs: https://hbr.org/2023/04/the-rebirth-of-software-as-a-service

similar note, obsolescence is also perceived as the decline in the asset's ability to generate private wealth or profits for the asset owner (Pakes and Schankerman, 1984; Nakamura, 2010; De Rassenfosse and Jaffe, 2017; Li, 2014; Li and Hall, 2020).

In practice, quantifying the rate at which the value of intangibles declines due to obsolescence can be a complex endeavor and was subject to many scholarly investigations. For R&D, approaches range from the use revenues attributable to patents (De Rassenfosse and Jaffe, 2017), the estimation of a profit model (Li and Hall, 2020), the estimation of a production function (Hall, 2007; Huang and Diewert, 2011), amortization (Lev and Sougiannis, 1996; Ballester et al., 2003), and market valuation (Hall, 2007; Warusawitharana, 2015).

The strong interest in measuring the obsolescence of R&D stems from the notion that depreciation rates for this asset also reflect the rate of technological progress. Perhaps a more dramatic way of phrasing it would be, the rate at which ideas die. There is relatively less scholarly attention directed towards other forms of intangible assets, notably software and artistic originals, notwithstanding their substantial economic importance. In the year 2020, software, which covers both purchased and own-account software, constituted half of the intangible assets incorporated into the core National Accounts of the UK. Moreover, artistic originals constituted 10 percent of the total intangible assets within the National Accounts for the same year. Most depreciation rate estimates are implied from the assumed service lives of the assets.

Despite accounting for the majority of SNA intangibles in most countries, there is little scholarly work on the depreciation of software. OECD (2010) recommends making assumptions on the service life of software to estimate their depreciation rates. The manual states that some of the ways to empirically inform the assumptions include: surveying software users, surveying software suppliers, and consulting software consultants. This approach may involve asking users about their expectations of their software's service lives or the service life of their recently retired software. However, this approach would likely be more informative on the service life of software copies to asset users and less informative on the ability to generate revenues for developers of the software.

For artistic originals, OECD (2010) recommends the use of empirical data such as the net present values of royalties to estimate their service lives. The manual also notes that the depreciation function must reflect relatively rapid depreciation in the first few years of an asset's life.

The focus of this essay is to estimate the obsolescence of software and creative originals. Table 5.2.1 shows the depreciation rates employed by selected OECD member countries for software and creative originals as published by Pionnier et al. (2023). In this table, software was classified into three categories, packaged software, custom software, and own account software. France, Germany, Italy, and the UK employ the same depreciation rate for all three categories. Meanwhile, the US and Canada apply a higher rate for packaged software.

Interestingly, Canada applies a depreciation rate of 1.00 for theatrical movies and long-lived TV programs. This implies that it does not recognize movies and TV shows as assets but as intermediate consumption. France, Italy, and the UK apply the same depreciation rate for all categories of creative originals.

Table 5.2.1: Depreciation rate for Software and Artistic Originals by country

	US	Canada	France	Germany	Italy	UK
Packaged software	0.550	0.550	0.244	0.359	0.325	0.256
Custom software	0.330	0.330	0.244	0.359	0.325	0.256
Own account software	0.330	0.330	0.244	0.359	0.325	0.256
Theatrical movies	0.093	1.000	0.331	0.110	0.172	0.183
Long-lived TV programs	0.168	1.000	0.331	0.181	0.172	0.183
Books	0.121	0.121	0.331	0.137	0.172	0.183
Music	0.267	0.267	0.331	0.273	0.172	0.183
Other entertainment originals	0.109	0.109	0.331	0.125	0.172	0.183

Source: Pionnier et al. (2023)

The available documentation regarding the calculation methodologies for these figures by statistical agencies is limited, However, certain resources are accessible for reference.Calderón et al. (2022) notes that estimates by the BEA for the depreciation of software are based on assumed service lives. According to the authors, the service life for software is determined through estimates of the correlation between computer and software expenditures. Moreover, they also consider anecdotal evidence regarding the typical duration of software use prior to replacement, as well as the service lives of software as defined by tax laws. The study also includes an informal survey of business software usage.

Soloveichik (2010) provides a detailed methodology on how the depreciation of artistic originals can be estimated using net present value (NPV). Her approach requires calculating the NPV of each original as revenues less the non-art cost of the asset, plus the discounted future NPV of the asset:

$$NPV_{t=0} = revenue_0 - nonartcost_0 + \frac{NPV_{t=1}}{1+\rho}$$

$$NPV_{t=1} = revenue_1 - nonartcost_1 + \frac{NPV_{t=2}}{1+\rho}$$

$$NPV_t = revenue_t - nonartcost_t + \frac{NPV_{t+1}}{1+\rho}$$

where ρ is the discount rate. The methodology and data sources for each set of creative originals are detailed in Soloveichik et al. (2013c), Soloveichik et al. (2013a), Soloveichik et al. (2013b), Soloveichik (2014). The approach is highly data-intensive requiring a wealth of information on revenues, retail sales, production costs, and assumptions on the discount rate.

For the UK, depreciation rates are determined by the assumed service life of an asset. The initial asset lives for software were determined through a survey by Awano et al. (2010) and a second survey by Field et al. (2012). The methodology for the estimation of capital stocks for creative originals was developed by Goodridge et al. (2013b). However, this paper did not specify how the service lives were determined. The assumed asset lives and effective depreciation rates currently being used by the ONS are detailed in Rincon-Aznar et al. (2017). These estimates were determined using analysis of depreciation from company accounts, consultation with UK industry experts, and comparisons with other countries' experiences.

As noted in this section, traditional methods for estimating depreciation often rely on survey-based approaches or assumptions about asset service lives. However, these methods may lack precision and can be subjective. Moreover, intangible assets, such as software and creative originals, are inherently dynamic and subject to rapid changes in technology and consumer preferences. Existing methods may struggle to capture these dynamics effectively. In the next section, we discuss a framework on how we can possibly use data on Internet searches as an objective and readily-available source of information on the depreciation of intangibles.

5.3 Framework

In this section, we present a framework for measuring the depreciation of intangibles using GSV results. We follow the framework by De Rassenfosse and Jaffe (2017), which sets out to estimate the depreciation rate of R&D using data on patents and firm revenues. According to the authors, depreciation of an intangible asset such as R&D can be measured by estimating the decline in revenues attributed to the asset:

$$V_{k,t} = V_{k,0}e^{-\delta t} \tag{5.1}$$

The expression presented above is a two-period framework that expresses the current-period revenues associated with asset $V_{k,t}$ during period t. While predicting the evolution of revenues from innovation over time is challenging, De Rassenfosse and Jaffe (2017) asserts that V steadily diminishes with age from its initial value $V_{k,0}$ at a constant rate denoted by δ. A constant rate of decline for the value of an asset is consistent with depreciation that follows a geometric pattern. They assume that the rate of revenue decline corresponds to the rate of asset value depreciation. This characterization of depreciation was first noted by Pakes and Schankerman (1984) and aligns with Nakamura (2010), which states that the depreciation of intangible assets should mirror the reduction in the asset's capacity to generate private wealth.

While this model was initially constructed to represent R&D, we argue that the same principles are applicable to other categories of intangible assets, such as software and creative originals. Although these assets may utilize a physical medium (e.g., CDs or hard drives), their intrinsic value is not inherently tied to this medium but rather stems from their ability to generate revenue for their owners.

For this exercise, we extend De Rassenfosse and Jaffe (2017) by assuming that revenues attributable to intangible assets such as software and creative originals are also determined by the popularity of search words. In the modern world, interest in a product is often accompanied by an internet search of that product. Prior to viewing a film, we would often initiate an online query for said film. While not all searches result in a purchase, one can argue that the total number of searches for a product could indicate the "potential demand" for the product. Assuming that associated revenues are proportional to potential demand, we extend the model by expressing present revenues as a function of search results:

$$V_{k,t}\left(\sum_i g_{i,k,t}, \gamma_k, \mu_{k,t}\right) = \bar{g}_{k,t}^{\beta_k}\gamma_k\mu_{k,t} \tag{5.2}$$

where $\sum_i g_{i,k,t}$ signifies the total search results for asset k across all individuals i. We also argue that revenues can be explained by time-invariant attributes of the asset, denoted as γ_k, in addition to various other influencing factors captured within the error term $\mu_{k,t}$. The error term captures all factors affecting revenues that are not captured by the *expected* obsolescence of an asset. To maintain simplicity, we assume a multiplicative relationship between search results and the other factors that impact revenues. Without loss of generality, we can also express searches as a normalized index, denoted as $\bar{g}_{k,t}$, which concurrently represents the popularity of these searches. β_k is a parameter, whose values range from 0 to 1, representing the degree to which internet searches impact the revenues derived from the asset.

Combining equation 5.1 and 5.2 and rearranging expression yields:

$$\bar{g}_{k,t}^{\beta_k} = \frac{V_{k,0}e^{-\delta t}}{\gamma_k\mu_{k,t}}. \tag{5.3}$$

For simplicity, we can take the log of equation 5.3:

$$log(\bar{g}_{k,t}) = \frac{1}{\beta_k}\left[log(V_{k,0}) - \delta t - log(\gamma_k) - log(\mu_{k,t})\right]. \tag{5.4}$$

By transforming the equation in this way, we can potentially simplify the relationship between the popularity of search terms and the various factors influencing it, making it easier to analyze statistically and draw insights from the data. We could control for the asset-specific factors, γ_k, and the initial revenues derived from the asset, $log(V_{k,0})$, through a set of fixed effects. Note that this only holds when $\beta_k > 0$. Moreover, if the relationship between internet searches and revenues is not strong, then we cannot use GSV to estimate the obsolescence of intangibles.

In the next section, we discuss our empirical approach to estimate these parameters.

5.4 Econometric strategy

Our approach is predicated on the assumption that the GSV results represent a form of obsolescence. We further assume that the high search level (the period when the index takes the value 100) represents the period when the asset is capitalized. For instance, when a movie is released, searches for the movie peak during the month of release. Searches decay over time, following a decrease in the popularity of the movie. We assume that decay in the search index corresponds to the rate of obsolescence for the particular movie. The same assumption is made for software. We estimate the rate of decline in the GSV index using a log-linear model as follows:

$$log(\bar{G}_{k,l,t}) = \delta\tau_t + \Gamma_k + \theta_l + \varepsilon_{k,r,t} \tag{5.5}$$

where $\bar{G}_{k,l,t}$ are GSV results at time t for keyword (movie/software titles) k released on year r; τ is a linear time trend; Γ_k and θ_l are fixed effects for the keywords and release dates, respectively. Since we are omitting the constant term from the empirical model, the asset-specific fixed effects, Γ_k, would absorb both the initial value of the asset V_{k0} and other time-invariant characteristics of the asset, γ_k. Lastly, the error term $\varepsilon_{k,r,t}$ can be decomposed into the error attributable to $\mu^r_{k,t}$ the relationship between searches and revenues, and a pure error term $\tilde{\varepsilon}_{k,r,t}$.

The dataset consists of an unbalanced panel comprising GSV results related to various movie and software titles. We employ truncation to initiate the dataset precisely when the index takes the value of 100. This point in time signifies the period at which we presume the asset begins to be capitalized. For instance, this may represent the month when the movie is shown in theaters or when software is released for distribution. We then estimate the semi-elasticity parameter δ in equation 5.5.

In many instances, GSV takes the value of zero. This could mean many things, including the possibility that searches for that month did not reach the threshold to be included in the GSV sample. This adds a complication to our specification in 5.5 since we can only take logs of positive numbers. As such, we add an arbitrary value, Δ, to all in order to apply the log transformation. As an alternative, we also employ other estimators such as the Poisson Pseudo Maximum Likelihood (PPML) by Silva and Tenreyro (2006), which allows for the estimation of semi-elasticities for observations with zero values. This is a common methodology in the empirical trade literature. We also employ Negative binomial regression since PPML can be restrictive in the sense that it assumes that the mean and the variance are equal. The Negative Binomial model is a generalization of the Poisson model that allows for overdispersion (variance greater than the mean). While the PPML is considered a more restrictive model over the Negative Binomial because of this assumption, simulations find the this model performs better in the presence of heteroskedasticity due to zero observation (Martin and Pham, 2020) and is widely used in the trade literature.

5.5 Data and descriptives

Two sets of information are required for the exercise: a list of movies and software, and GSV results. In this section, we discuss the data that we use, particularly how the data was obtained and managed.

5.5.1 Data sources

To facilitate the extraction of GVS results, it is necessary to compile a comprehensive catalog of titles to be employed as keywords within our search queries. Our objective is to maximize the representativeness of our estimate by covering a wide array of titles. Consequently, we undertook extensive efforts to assemble an exhaustive list of keywords. Our primary source for this exercise was the Internet Movie Database (IMDb) for theatrical movies and Wikipedia for software-related titles.

IMDb (imdb.com), a subsidiary of Amazon, serves as an extensive online repository encompassing movies, television series, video games, and related forms of media. This platform offers a wealth of information on each title, including details such as release year, ratings, run time, box office earnings, directorial credits, cast members, genre classifications, and reviews, among other pertinent attributes. Information from IMDb has been used in various research in economics and marketing, particularly those involving sentiment analysis (Shaukat et al., 2020; Harish et al., 2019; Topal and Ozsoyoglu, 2016) and the prediction of movie ratings (Hsu et al., 2014; Oghina et al., 2012; Gogineni and Pimpalshende, 2020).

Meanwhile, Wikipedia serves as a limited source of data. Although the platform has the potential to offer a substantial amount of information for each software entry, the majority of this data is within the individual articles dedicated to each software. We are not able to systematically extract this information for practical reasons. Furthermore, the consistency of information across these entries is notably variable. Revenue-related data is frequently absent. Consequently, our utilization of this resource is primarily limited to extracting the names of software titles, as well as the categories each titles were tagged with.

For the exercise, we confine our analytical scope to movies that have garnered box office revenues exceeding \$1 million. This criterion led to the inclusion of 4,623 movie titles and 1,089 software titles in our dataset. To formulate precise search queries, we combine the movie title with its respective year of release (e.g., "Avengers: End Game 2019"). This naming convention aligns with the typical referencing of movie titles and helps mitigate any potential confusion between the movie title and unrelated search results. Conversely, for software, we employ the software title verbatim as our search query. To avoid confusion with common terms, we include the word "software" in the search query, and in some cases, the category of the software. For example, we use the queries, "Vala (programming language)" and "Songbird (software)". Queries such as these account for less than 5 percent of our total software queries.

The main dataset we employ for this exercise is Google Search Volume (Google Trends) results. GSV provides a real-time index for a *random sample* of queries in Google's search engine. According to Google News Lab, the sample is unbiased (Rogers, 2016), though Google does not disclose the specifics of its sampling design.

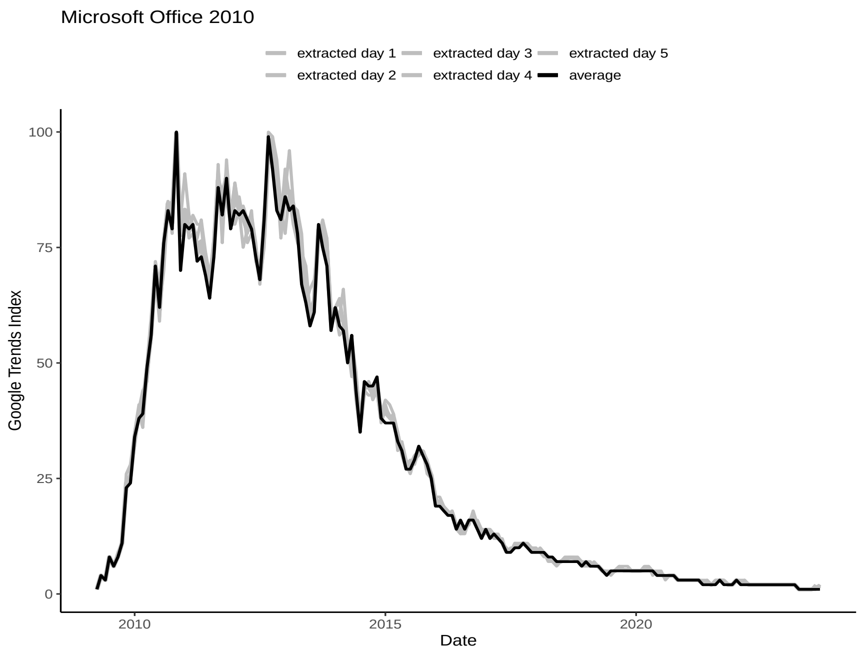
Users of Google Trends can filter the data in two ways: real-time and non-real-time. Within the real-time filtration, a random selection of queries from the preceding seven days is employed. Meanwhile, the non-real-time approach involves the use of a randomized subset of search queries extracted from the Google dataset, which covers data points spanning from 2004 up to roughly 36 hours prior to the search. Moreover, Google Trends allows users to compare five search queries simultaneously.

One of the major limitations of GSV data is that it does not present the number of searchers for each keyword. Instead, GSV reports an index that undergoes a two-tier standardization process. Initially, all search volumes are standardized in relation to the total number of searchers during a reference period. This adjustment compensates for the substantial variations in Google's user base over time, such as the significant increase in users since 2004. The number of searches for a particular query is divided by the number of searches during that specific year (or month) to render the index comparable. Secondly, this index is further normalized to a scale ranging from 0 to 100, where 100 signifies the highest frequency of search queries for a given topic and zero signifies the lowest. Consequently, GSV results are conventionally interpreted as indicative of the relative popularity of specific search queries.

Numerous researchers have employed GSV to illustrate its potential in forecasting macroeconomic indicators, including GDP and unemployment (Woloszko, 2021; Kohns and Bhattacharjee, 2023; Ferrara and Simoni, 2022), as well as in predicting trends in tourism (Havranek and Zeynalov, 2021). Furthermore, GSV has been utilized in diverse domains, such as monitoring COVID prevalence during the pandemic (Hamulka et al., 2020; Effenberger et al., 2020; Cervellin et al., 2017; Zattoni et al., 2021).

For this research, we would employ non-real-time Google Trends results. Given the large number of keywords that we intend to use in the analysis, manually extracting individual Google Trends data for each keyword within a reasonable timeframe is impractical. As such, we employ an R package designed to systematically retrieve Google Trends data for all keywords in our lists. We pool the data for all GSV results for all keywords to construct an unbalanced panel that we used for the analysis.

Figure 5.5.1



Note: The figure shows Google trends results for the query "Microsoft Office 2010". The figure overlays the results extracted on five (5) separate days (grey line) and the average of the cross-sectional average of five results.

Since the index reported by GSV is based on a sample, the results extracted today could be different from the index extracted the next day and the days following (Choi and Varian, 2012; Cervellin et al., 2017). The literature recommends extracting GSV results for the same keyword on multiple days and calculating the cross-sectional mean across different samples at t (McLaren and Shanbhogue, 2011; Carrière-Swallow and Labbé, 2013; Eichenauer et al., 2022).

At present, there is no standard set by the literature on the number of samples required to construct a reliable index. Some papers recommend a sample of 7 (McLaren and Shanbhogue, 2011), some 50 (McLaren and Shanbhogue, 2011). Extracting a large volume of GSV results is not straightforward. Google blocks the IP address of users after too many queries within a short period. Carrière-Swallow and Labbé (2013) explains that in practice, the number of samples extracted is a result of a balancing act between the reduction of the sampling problem and stressing the Google API. Given that our cross-section is substantially large (4,624 movie titles and 1,089 software titles), we were only able to draw 4 to 7 times for each keyword before Google blocked our IP address.

To illustrate, we show in figure 5.D.4 GSV results after 5 draws, as well as the average for each draw. While there are some variations between draws, we can see that the general oscillation tends to move in the same direction. We will use the cross-sectional average across different draws as our dependent variable in estimating equation 5.5.

5.5.2 Descriptive statistics

In this section, we present some descriptive statistics on the software and movie titles we employ in our sample. For software, we were able to classify the titles into broad categories. We exploit the category tags at the bottom of each Wikipedia entry as the basis of our classification. We explain in detail how we were to come up with these categories in appendix 5.C. For movies, we present some summary statistics from the data extracted from the IMDb website.

Table 5.5.1: Count and share by software type

	Operating System	Media Players	Office Tools	Web Browser	Social Media
n	212	24	19	47	42
Share (%)	21	2.4	1.9	4.7	4.2

Note: The table shows the number of entries and share of selected software types in the sample.

Around 21 percent of our sample is a form of operating system. Around 2.4 percent were categorized as media players, 1.9 percent were office tools, 4.7 percent were web browsers, and 4.2 percent were tagged as social media. Note that the shares do not add to 100 percent. This is because of the way we have categorized the software entries. We have only highlighted the categories we find most interesting and intuitive, acknowledging that there are countless other ways to categorize software.

Table 5.5.2: Count and share by operating system

	Windows	Android	IOS	MacOS	Linux
n	257	172	164	133	169
Share (%)	25.5	17	16.3	13.2	16.7

Note: The table shows the number of entries and share of software titles by operating system in the sample.

More than a quarter of the software in our sample can run on the Windows operating system, while 13.2 percent can run using MacOS. About 17 percent of our sample can run in Android OS, while 16.3 percent can run in iOS.

Table 5.5.3: Count and share of free and open-source software

	With Free Version	Open-Source	Strictly Open-Source
n	401	83	13
Share (%)	39.7	8.2	1.3

Note: The table shows the number of entries and share of free and open-source software in the sample.

According to the category tags, around 40 percent of the software in our sample has a free version. Only 8.2 percent of the sample was classified as open-source and only 1.3 percent of the sample was tagged as open-source but were not tagged as having a free version.

Table 5.5.4: Descriptive Statistics for Revenues and Run Time

Variables	Mean	SD	Min	Max
Revenues (in million US$)	23.4	59.5	1.0	936.7
Run time (in minutes)	104.8	20.3	8.0	321.0

Note: The table shows the descriptive statistics of revenues and run time of movie titles in the sample.

For movies, the average revenues we calculated for our sample was at $23.4 million. Variation in our sample is notably high. The standard deviation for revenues at $59.5 million, was more than double the mean. The maximum revenue in our sample is $936.7 million, while the minimum is $1 million by construction.

In terms of the run time, the average run time in our sample is 1.8 hours. The standard deviation is only 20 minutes. The shortest movie in our sample is only 8 minutes and the longest movie in the sample is 5.3 hours long.

5.6 Empirical measures of depreciation

In this section, we present the estimates for the depreciation rates of movies and software. We present the estimates using OLS, PPML, and the negative binomial regression. We also compare our estimates with the depreciation rates being employed by a select set of countries. Lastly, we present the implied service lives arising from the depreciation rates.

Estimating equation 5.5 requires taking the natural log of GSV results. However, there are instances when GSV results take the value of zero. As such, log transformation is not possible. The typical solution is to add a constant, Δ. In our OLS specification, we choose $\Delta = 1$. We suspect that the OLS results are biased because of the heteroskedasticity from zero observations.

5.6.1 Software

We present the results for software in table 5.B.2. We interpret these semi-elasticities as monthly depreciation rates or the percentage decline in the value of the asset each month. The coefficient from using OLS is smaller than those from PPML and negative binomial. For software, however, the estimates from the latter two are almost identical.

Table 5.6.1: Regression results for software

	OLS	PPML	Negative Binomial
	(1)	(2)	(3)
δ	-0.006^{***}	-0.008^{***}	-0.009^{***}
	(0.0003)	(0.00001)	(0.0001)
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

We note that our sample also includes software titles that are either distributed without any explicit monetary cost to their users or can be classified as open-source. Among these software titles, certain ones function as loss leaders and were conceived not solely with the aim of generating revenue for their developers. As such, these titles would not qualify under the capitalization criteria of the National Accounts. We show in table 5.6.2 that removing software titles with free versions, as well as those classified as open-source, does not make any changes to the estimates.

Table 5.6.2: Results for PPML model removing free and open-source software

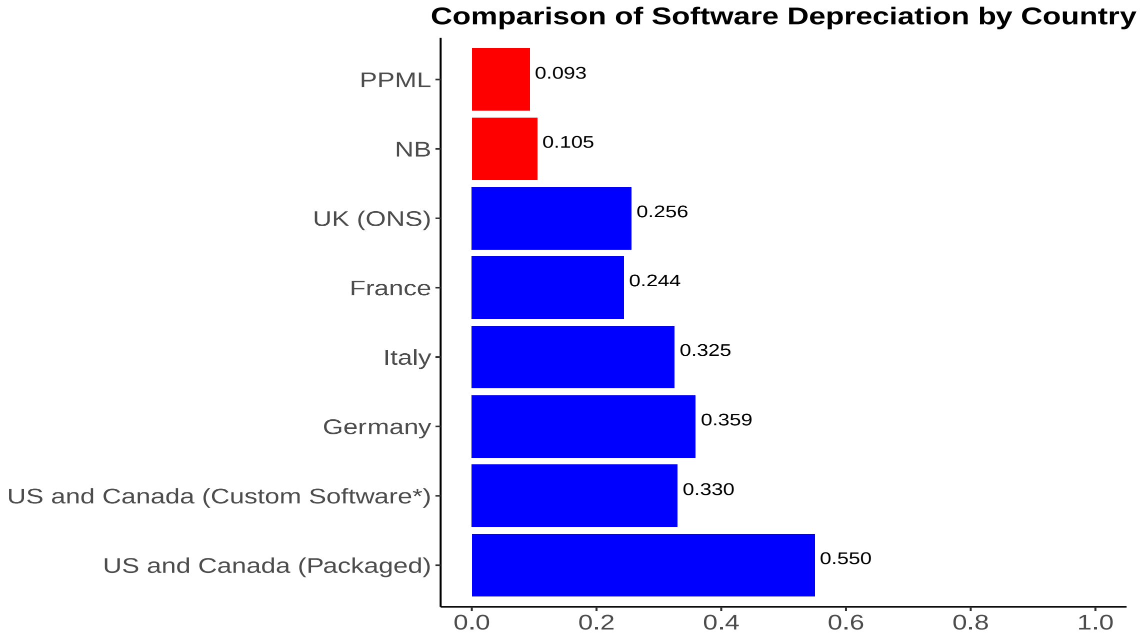
	All Software	Removing those tagged "Free & Open Source"	Removing those tagged Open Source only
	(1)	(2)	(3)
δ	-0.008^{***}	-0.008^{***}	-0.008^{***}
	(0.00001)	(0.00002)	(0.00001)
n	161,735	135,186	144,453
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

We calculate the annual depreciation[2] and compare them to the estimates employed by a select set of statistical offices, as reported by Pionnier et al. (2023). All our estimates are materially smaller than the depreciation rates employed by other countries (see figure 5.6.1). In particular, the estimates are approximately half the depreciation rates employed by statistical offices.

[2]Assuming a geometric pattern, we calculate the annual depreciation rate (δ^a) from the monthly depreciation rate (δ^m) as:

$$\delta^a = ((1 + \delta^m)^{12}) - 1$$

Figure 5.6.1: Comparing the annual depreciation rates estimates for software with those from other countries



Note: The figure compares results from PPML and negative binomial (NB) model for software with the annual depreciation rates employed by a select set of countries. The monthly estimates from table 5.B.2 were annualized assuming a geometric pattern. The depreciation rates employed by other countries were sourced from Pionnier et al. (2023) *Custom software also includes own-account software.

We calculate the implied service life as $1/\delta$. The difference between the estimates using GSV and depreciation rates employed by statistical agencies is also reflected in the implied service life. The implied service lives from our estimates for software are more than double those assumed by statistical agencies.

Table 5.6.3: Implied Service Life of Software by Country

Country	Implied Service Life
PPML	10.8
NB	9.5
UK (ONS)	3.9
US and Canada (Packaged)	1.8
US and Canada (Custom Software*)	3.0
France	4.1
Germany	2.8
Italy	3.1

*Custom software also includes own-account software.

There are two possibilities that we can draw from these results. First, current levels of capital stock are grossly underestimated. Our estimates suggest that the existing stocks of software assets are possibly double the current estimates. Doubling the estimates for software stocks would have substantial implications for both growth accounting and estimates of net domestic product. Second, it is possible what we are capturing is not the depreciation rates for *all forms* of software assets.

To explain the second possibility, it is important to note that software assets can be classified under two categories: *original software* and *software copies* (OECD, 2010). Original software can be further broken down into *originals for reproduction* and *other originals* (OECD, 2010). The first subcategory refers to software intended to be reproduced for sale or leased. Other software originals refer to custom-made software produced by a firm intended for the production of other goods and services.

Statistical agencies estimate software depreciation either by making an assumption on the asset's service life (as documented by Calderón et al. (2022) and recommended by OECD (2010)) or, as with the UK, by conducting a survey on firms about the assumed service life of their existing software (Awano et al., 2010; Field et al., 2012).

The use of surveys appears to be the more scientific of the two approaches since it requires the use of empirical data. However, the estimated depreciation rate employing this approach (as demonstrated by the UK) is still double the estimates from the approach in our study. Since the respondents of these surveys are firms that use software to produce their own products, we suspect that the service life that they estimate generally captures the asset lives of software copies and other software originals. These surveys were not designed to draw from a representative sample for each category. Considering that originals for reproduction are "generally produced by specialist software companies", it is unlikely that surveys would be representative enough to account for this category.

On the other hand, the rate at which search volumes decay likely represents the obsolescence of originals for reproduction. We can think of it as the value of the codes behind software titles such as Microsoft Office or Zoom. We interpret our estimates as the depreciation rates for the assets by software developers such as Microsoft or Zoom. As such, our estimate reflects a different category compared to those captured by surveys on the asset life of software employed by firms.

Software developers could earn revenues from the master copy of software far longer than the average service life of a software copy for firms that purchase software copies. This could explain why our estimates are substantially slower than those being employed by statistical agencies. Little is known about the depreciation rate of originals for reproduction. Software companies do not usually present revenues from specific products they release. Moreover, since software-as-service is becoming more popular as a business model for developers, one can argue that in the future, companies will be less reliant on capitalizing software copies. Most of the capitalization would occur with software developers and using GSV would likely be a more appropriate strategy in measuring the obsolescence of this software class.

5.6.2 Original software

While original software is capitalized in the National Accounts estimates by all countries subscribing to the 2008 version of the SNA, the asset category could be different for each country. Following BEA's National Accounts update in 2018, software originals were reclassified to *software R&D* rather than software investments (Chute et al., 2018). Meanwhile, most European countries, including the UK, purposely remove expenditures for software development from R&D investments, following the European System of Accounts (ESA) Guidelines 2010. Investments in original software are classified under own-account software under the software development industry (under SIC sector J).

To effectively capture the depreciation of original software, we must argue that the decline in Google GSV results corresponds directly to the decrease in revenues experienced by software developers. However, we explain later that this assumption is only partially true. In this section, we would argue that search queries are indicative of software usage, thereby warranting careful consideration in our analytical framework.

We show in the figure 5.6.2 a simplified illustration of a software's lifespan. Development of software begins at time t_0 and ends at t_1. In this simplified version of a software's life cycle, we assume the sales begin after development at time t_1 and end until the last copy is sold at time t_2. From the perspective of the asset owner of the original software (the software developer), the original is fully depreciated at time t_2 because it no longer profits from the sale of copies of the software. However, usage of the software extends up to the time that the last copy bought is retired, which is at time t_3.

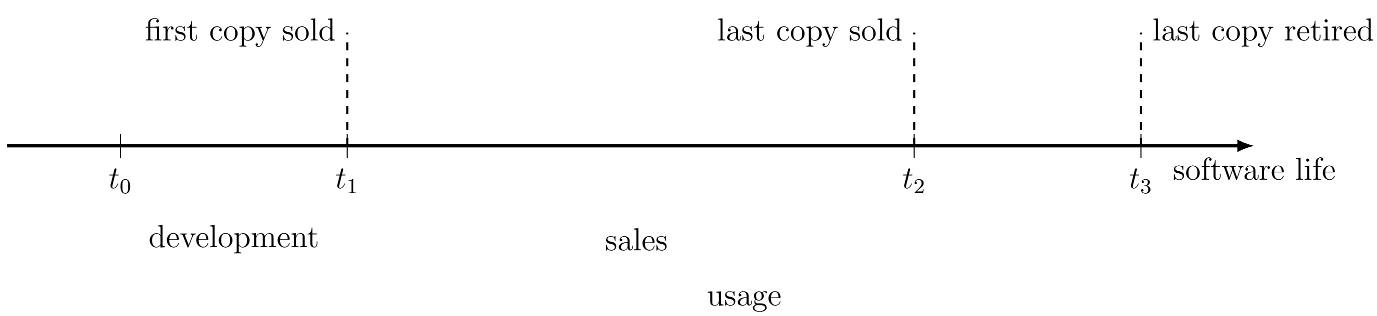


Figure 5.6.2

The service life of the software original is reflected by the distance between t_1 and t_2. Meanwhile, Google Trends is likely capturing usage, which is from t_1 to t_3. Given that the last copy is sold at t_2 and retired at t_3, this distance would likely reflect the service life of software copies. As such, if we are confident about estimates of the depreciation of software copies, then we can work out the depreciation of software originals.

We subtract our estimates of the service life for software originals from the service life of software copies in table 5.6.3. On average, we find that the difference between our estimate and the estimates by statistical agencies is about 3 years for PPML and only 0.8 years for

Negative Binomial. While our approach still implies that the services life of software R&D is overstated, the difference is only by a maximum of 4 years.

Table 5.6.4: Implied Service Life of Software Originals by Country

	PPML	NB
UK	6.9	6.0
US	7.8	5.0
Canada	7.8	5.9
France	6.7	5.0
Germany	8.0	4.5
Italy	7.7	4.5
Ave Diff	3.1	0.8
Ave	7.5	5.2
δ	0.134	0.194

As mentioned earlier, European countries account for software originals as part of software assets. The service life they employed are the same as those shown in table 5.6.3. Our estimates for the asset life of original software in table 5.6.4 is longer than the asset life of software employed by statistical agencies in Europe. The US, which records original software as part of R&D, employs a service life of 4.5 years (Pionnier et al., 2023) for original software, which is closer to the lower bound of our estimates.

How do we know if these estimates are valid? We do not observe the actual service life of original software and in most cases, software companies do not present revenues that can be traced to specific software titles. This makes it easier to verify another form of intangibles, specifically, theatrical movies. We present estimates for movies in the next section.

5.6.3 Movies

Unlike software, it is relatively easier to determine the revenues attributable to specific movie titles. Box office revenues (which account for the majority of the movie revenues) and DVD sales are tracked and published on various websites such as IMDb.com and TheNumbers.com. Therefore, it is possible to empirically estimate equation 5.2 by tracking how much revenues decline over time. Soloveichik et al. (2013c) takes this further by estimating the net present value of movies. She calculates the depreciation rates of movies from the rate of decline in their net present value.

This is the approach taken by the BEA, which, we believe to be the appropriate way to estimate the depreciation of such assets. This is also the approach recommended by OECD (2010). If estimates using our approach match those from the BEA, this could support the validity of our methodology.

We are able to distinguish the residency of the studio producing each movie. This allows us to estimate different depreciation rates for different countries. We show the regression results for the US and Germany in table 5.6.5. We chose to present results for these countries because it would be easier to compare our estimates to official data later on.

Table 5.6.5: Regression results for movies

	OLS	PPML	Negative Binomial
		All Countries	
	(1)	(2)	(3)
δ^{All}	-0.004^{***}	-0.011^{***}	-0.008^{***}
	(0.0001)	(0.00001)	(0.00002)
		United States	
	(4)	(5)	(6)
δ^{US}	-0.004^{***}	-0.010^{***}	-0.007^{***}
	(0.0001)	(0.00001)	(0.00003)
		Germany	
	(7)	(8)	(9)
δ^{DE}	-0.004^{***}	-0.009^{***}	-0.008^{***}
	(0.0002)	(0.00003))	(0.0001)
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

Results from OLS are substantially smaller compared to those from PPML and the Negative Binomial models. We suspect that the large difference stems from the fact that most zero observations occur at the tail end of the GSV results. Typically, four to five years following a movie's release, search activity tends to diminish to a point where it falls below Google's inclusion threshold, resulting in these instances registering as zero entries. To test this hypothesis, we employ a truncation approach, wherein all observations within three years of a movie's release are removed from the dataset. Results in appendix 5.B.1 show that OLS and PPML estimates are similar if the value of delta is small. However, we do not use these results because in many cases, we observe sparse spikes for movies released many years after their debut. This could be driven by many factors including seasonality (for instance, people watching Top Gun every Christmas) or sequels (people watching the Tobey Maguire Spider-Man movies before the release of No Way Home).

From the regression results, we also note estimates for the US and Germany are similar, with Germany's being slightly faster. This could imply that American films depreciate slower than German films.

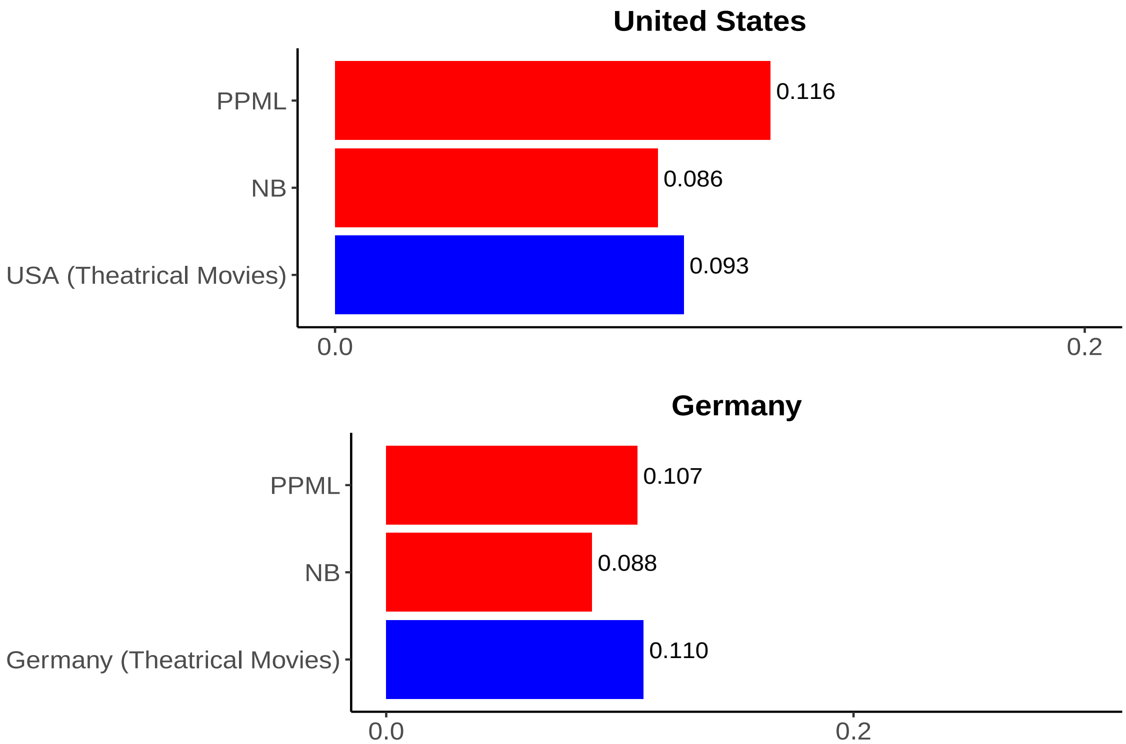


Figure 5.6.3: Comparing the annual depreciation rates estimates for movies with those from other countries

Note: The figure compares results from PPML negative binomial (NB) model with the annual depreciation rates for theatrical movies employed by the United States and Germany. The monthly estimates from table 5.6.5 were annualized assuming a geometric pattern. The depreciation rates employed by the US and Germany are sourced from Pionnier et al. (2023).

While we would prefer to compare our estimates to the estimates for the depreciation rates by other statistical offices, we can only find estimates for movies from the US and Germany. Other countries present estimates at the aggregate level, as part of ”artistic originals” (see table 5.2.1). We compare our estimates to the depreciation rates employed by the US and Germany. We show the comparison in figure 5.6.3. Estimates for the UK, France, and Italy are shown in appendix table 5.B.1.

For the US, our estimates are not materially different from those employed by BEA. Our estimated asset life for movies (see table 5.6.6) of 8.6 to 11.6 years is close to the estimates of BEA, at 11 years. Similarly, estimates for Germany also approximate the official data. Our estimated service life for German films of 9.3 to 11.4 years is only slightly above the official estimates of 9.1 years.

Our methodology presents a notable advantage in its flexibility for frequent updates and the ability to compute depreciation rates for specific time intervals. As demonstrated in table 5.6.7, we partitioned the dataset into two segments: one that includes movies released between 2004 and 2010 and another comprising movies released from 2011 to 2016. Although the depreciation rate exhibits only a subtle disparity in the monthly rates, this translates to

Table 5.6.6: Implied Service Life by Country

Country	Implied Service Life
United States	
PPML	8.6
NB	11.6
Official Estimate	10.8
Germany	
PPML	9.3
NB	11.4
Official Estimate	9.1

Table 5.6.7: Estimates for movies by release dates

	Monthly	Annual	Service Life
All Release Dates	1.1	12.3	8.2
Released 2004 to 2010	0.9	10.0	10.0
Released 2011 to 2016	1.7	18.6	5.4

a substantial difference in the implied asset life.

Our estimates show that the service life of films released from 2011 to 2016 is almost half of that of movies released between 2004 to 2006. Technological advancements, like the rise of affordable streaming services and piracy, have significantly altered how consumers interact with and obtain media. This transformation has brought about various industry shifts, notably the reduction of the theatrical release window (Ahouraian, 2021). Traditional depreciation rates, which are not frequently revised, may not accurately capture the repercussions of this shift.

5.6.4 Validity check

There is a consensus that the obsolescence of intangibles is generally linked to the asset's ability to generate revenues (Pakes and Schankerman, 1984; Nakamura, 2010; De Rassenfosse and Jaffe, 2017; Li, 2014; Li and Hall, 2020). There are limited avenues on how to test this. In the case of software, for instance, firms rarely break down revenue streams attributable to specific software titles. The case is not the same for movies, however. Data is available on box office revenues from IMDb.com. To a limited extent, DVD sales data is also available from the-numbers.com. These were the data sets employed by Soloveichik et al. (2013c) and Goodridge et al. (2013b). By combining these datasets with GSV results, we can test whether there is a statistically significant relationship between Google Trends results and movie revenues.

For the first part of this validation exercise, we construct variable $\bar{V}^B_{k,l,d,t}$, which represents lifetime box office revenues for movie k, released on year l, by taking the sum of all future revenues[3] of the movie from time, t up to the end of the theatrical window at time, T:

$$\bar{V}^B_{k,l,d,t} = \sum_{t=1}^{T} V^B_{k,l,d,t}. \tag{5.6}$$

Since daily data on streaming revenues and DVD sales are scarce, lifetime revenues, in this case, only cover revenues earned from the theatrical release. However, the majority of movie revenues are generated from the box office[4]. As such, we believe that our analysis is still valid despite this limitation.

We estimate the model:

$$log(\bar{V}^B_{k,l,d,t}) = \tilde{\beta} \cdot log(\bar{G}_{k,l,d,t}) + \tilde{\delta}\tau_t + \Gamma_k + \theta_l + \omega_d + \tilde{\mu}_{k,r,d,t} \tag{5.7}$$

where $\bar{G}_{k,l,t}$ are GSV results, τ_t is a time trend; Γ_k, ω_d are day of the week fixed effects, and θ_l are movie title and release date (year) fixed effects, respectively; and $\tilde{\mu}_{k,r,d,t}$ is a random error term.

We are interested in the parameter $\tilde{\beta}$, which is expected to be positive, implying a direct relationship between GSV results and box office revenues. Box office revenues generally decline following the first day of the release of the movie. As such, regressing box office revenues with any variable that is also trending downward will generate significant results. This part of the exercise aims to uncover whether $\bar{G}_{k,l,t}$ can explain some of the variations in the decline in revenues, *on top of what can be absorbed by the time trend.*

[3]Data on box office revenues were sourced from the-numbers.com.

[4]https://www.statista.com/statistics/1194522/box-office-home-and-mobile-video-entertainment-revenue-worldwide/

Table 5.6.8: Testing the relationship between Box Office revenues and Google Trends Results

	Dependent Variable:			
	Google Trends	Lifetime Box Office Revenues		
	(1)	(2)	(3)	(4)
$\tilde{\delta}$ (time trend)	-0.005^{***}	-0.072^{***}		-0.071^{***}
	(0.0003)	(0.001)		(0.001)
$\tilde{\beta}$ (Log Google Trends)			0.291^{***}	0.046^{***}
			(0.017)	(0.005)
AdjR2	0.347	0.900	0.451	0.900
n	118,050	118,050	118,050	118,050
Keyword FE	Yes	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes	Yes

The results are presented in table 5.6.8. We observe from the first two columns that box office revenues decline faster than the Google Trends index. This is not surprising since the lifespan of movies often exceeds their screening dates. As such, there would still be some search activities for movie titles even though they are no longer shown in cinemas. In the last column, the coefficient for GSV results is positive and significant, confirming that there is a direct relationship between Google Trends results and box office revenues, other than what can be explained by the normal passage of time.

We can also extend this analysis by exploring the dynamics between GSV and box office revenues. Equation 5.8 incorporates lagged terms for GSV results, acknowledging the potential scenario wherein individuals frequently conduct internet searches for movies prior to viewing them:

$$log(\bar{V}^{B}_{k,l,d,t}) = \sum_{n=0}^{N} \Big(\tilde{\beta}_{t+n} \cdot log(\bar{G}_{k,l,d,t-n})\Big) + \tilde{\delta}\tau_t + \Gamma_k + \theta_l + \omega_d + \tilde{\mu}_{k,r,d,t} \tag{5.8}$$

Results shown in table 5.6.9 show that lagged GSV results explain some of the variations in box office revenues, on top of what can be absorbed by the time trend. The relationship is also positive, suggesting that both variables move in the same direction. We believe that our results provide evidence of a direct relationship between Google Trends and movie revenues.

Table 5.6.9: Testing the relationship between box office revenues and Google Trends Results

	Dependant Variable: Lifetime Box Office Revenues				
	(1)	(2)	(3)	(4)	(5)
$\tilde{\delta}$ (time trend)	-0.071^{***}	-0.071^{***}	-0.070^{***}	-0.070^{***}	-0.070^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\tilde{\beta}_{t+n}$ Log Google Trends					
Lag 0	0.040^{***}	0.035^{***}	0.031^{***}	0.027^{***}	0.025^{***}
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Lag 1	0.038^{***}	0.034^{***}	0.030^{***}	0.027^{***}	0.023^{***}
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)
Lag 2		0.033^{***}	0.030^{***}	0.027^{***}	0.024^{***}
		(0.004)	(0.004)	(0.004)	(0.003)
Lag 3			0.029^{***}	0.027^{***}	0.025^{***}
			(0.004)	(0.004)	(0.003)
Lag 4				0.025^{***}	0.023^{***}
				(0.004)	(0.004)
Lag 5					0.023^{***}
					(0.004)
AdjR2	0.902	0.903	0.904	0.905	0.905
n	114,142	110,889	107,750	105,069	102,563
Keyword FE	Yes	Yes	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes	Yes	Yes

5.7 Impact on industry aggregates

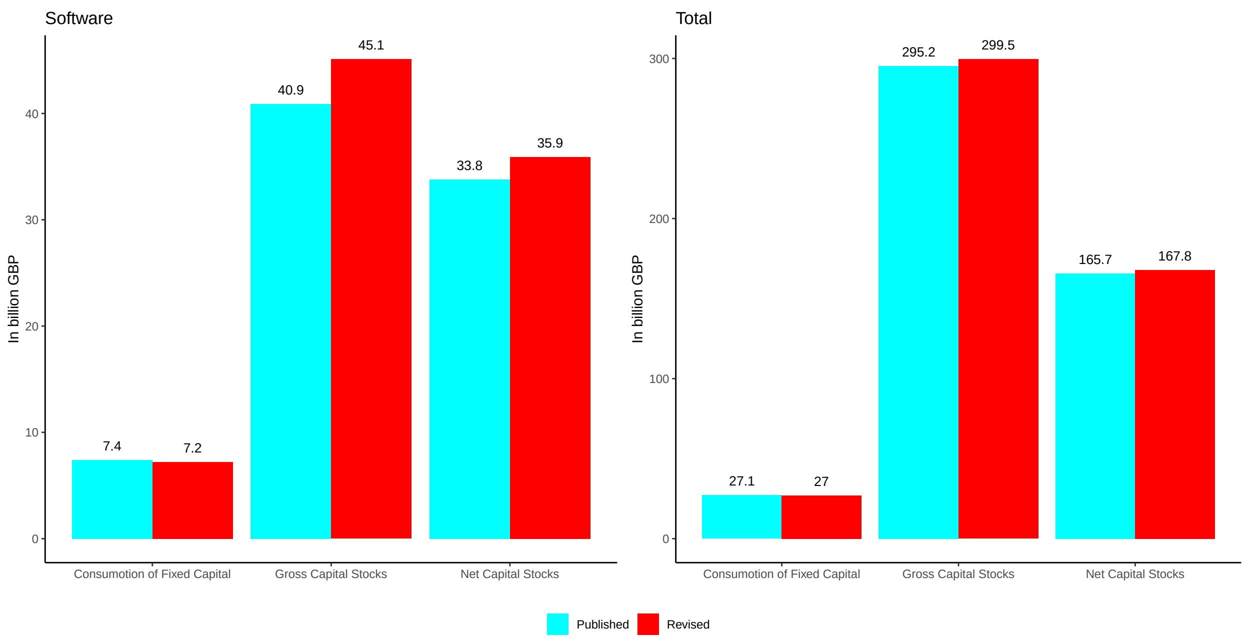
In this section, we examine how this methodology could impact some macroeconomic aggregates, such as consumption of fixed capital, gross capital stocks, and net capital stocks. We also estimate total factor productivity and examine how productivity statistics could change if we change the asset life assumed in the estimation of capital stocks.

In particular, we evaluate how these aggregates would change for the Information and Communication industry (Standard Industrial Classification Section J) of the UK. We compare the published levels with *revised* estimates that employ asset lives we calculated using our methodology.

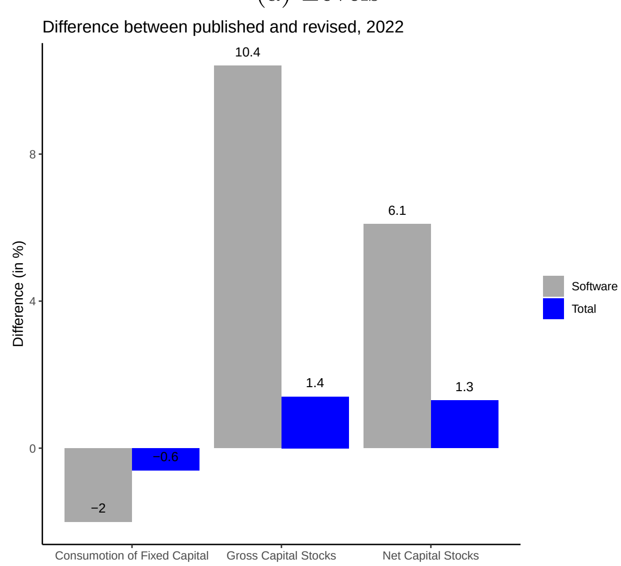
5.7.1 Impact on capital stocks for the UK's Information Communication industry

Due to time and resource constraints, we do not estimate a perpetual inventory model ourselves. Rather, we employ the codes provided by the Office for National Statistics on their websites[5]. The ONS provides the raw data, R codes, and the set of assumptions to replicate their estimates of various aggregates relating to capital at the 2-digit SIC industry level.

[5] https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/methodologies/capitalstocksuserguideuk last accessed: 15 February 2022



(a) Levels



(b) Percent difference

Figure 5.7.1: Estimates for Sector J, 2022

For this exercise, we changed the service life assumption for own-account software under SIC industries 62 (Computer programming, consultancy and related activities) and 63 (Information service activities)[6] from 4 years to 7 years, following our estimates for the original software. Since the ONS was not able to provide a breakdown of the type of software capital, we assumed that all software investment that goes to these industries are original software

[6]The ONS combines the two industries in their database. We are not able to separate the investments going to the two industries

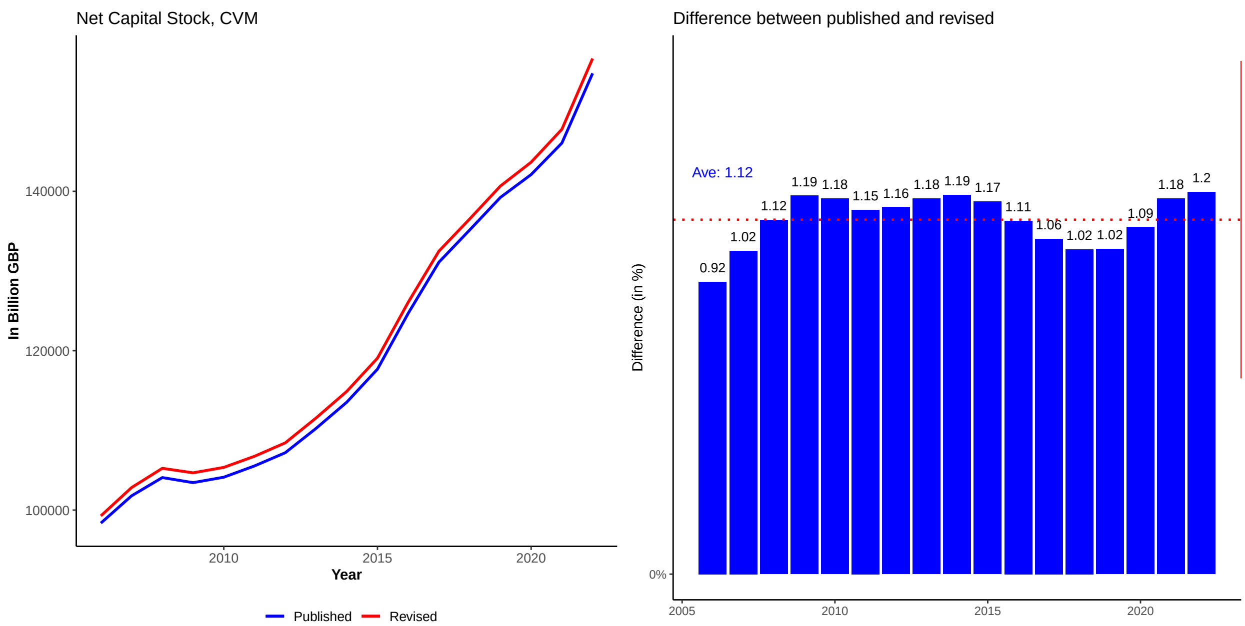
for reproduction.

Our results show that for 2022 (see figure 5.7.1), our estimates on the consumption of fixed capital for software are slightly lower than the published data. Our estimate of gross capital stock of software for sector J is 10.4 percent higher than the published data, while our estimates of net capital stock are 6.1 percent higher.

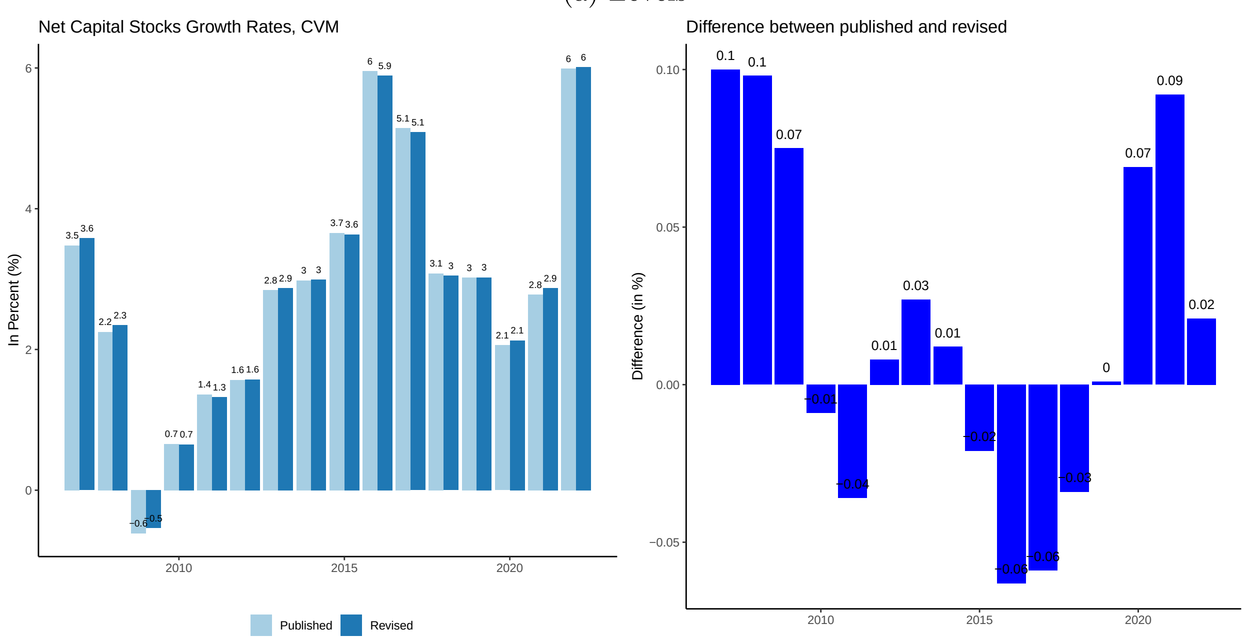
Not surprisingly, the difference is more modest for the total capital stocks. Our estimate of total gross capital stocks for sector J is only 1.4 percent higher than published, while the estimate for total net capital stocks is only 1.3 percent higher than what was published by the ONS.

Between 2002 and 2006, our estimates of total net capital stocks (chain volume measure) for sector J are consistently higher than those published by the ONS, as illustrated in Figure 5.7.2. However, the disparity between our estimates and the published data varies over time, ranging from as low as 0.92 percent to as high as 1.2 percent.

The impact on growth rates also exhibits variability across different periods. Some years demonstrate faster growth in net capital stocks according to our calculations, while in others, our growth rates are slower than the published data. This could have implications for Total Factor Productivity (TFP) growth. Notably, from 2015 to 2018, our estimates indicate consistently slower growth rates in net capital stocks for sector J compared to the published data. This observation suggests that TFP growth during that period may be higher than initially estimates.



(a) Levels



(b) Growth rates

Figure 5.7.2: Estimates for Sector J, 2006 to 2022

Note: Figure describes the differences in net capital stocks (chain volume measure) for sector J calculated using the asset life from the GSV estimates against the published data on net capital stocks. The upper-left panels show the levels over time. Upper-right panels the differences between the levels (revised vs published) in percent. The lower-left panel shows the growth rates. The lower-right panel shows the difference in the growth rates revised vs published) in percentage points.

5.7.2 Impact on the productivity growth of the UK's Information and Communication industry

We estimate TFP following Bontadini et al. (2023) and Organisation for Economic Co-operation and Development (2021). We also tried to remain consistent with Office for National Statistics (2007). Details on the methodology are discussed in appendix 5.D.

There are some differences between the capital stocks that are part of the National Accounts of the ONS, and capital stocks employed in their calculation of TFP. The National Accounts assume a hyperbolic rate for all assets with he exception of R&D, which applies a geometric rate[7]. For capital stocks employed in productivity calculation, the ONS applies a geometric rate for all assets. Moreover, there are some technical differences, as well namely, 1) TFP statistics only cover the market sector, removing stocks employed by the government, and 2) the use of user cost (see appendix 5.D) to ensure that assets with higher depreciation rates are given higher weights than assets with longer services lives. The combination of these factors would cause differences between the intuition laid out in the previous section and the final results.

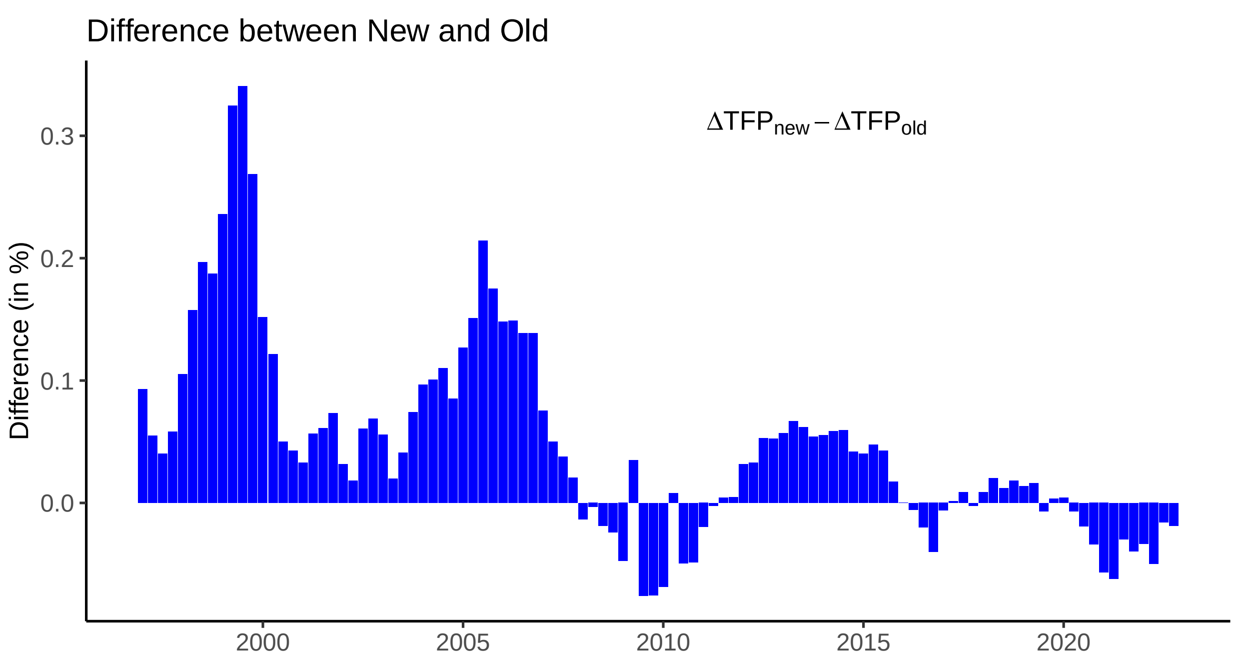


Figure 5.7.3: TFP difference for sector J, 1995 to 2022

Note: Figure shows the difference between the TFP calculated using the asset life of own-account software from estimates from the GSV methodology and TFP calculated using current assumptions on the own-account software's asset life (New - Old). Results are percentage points differences in TFP growth.

Figure 5.7.3 shows the difference between TFP using our estimated asset life ('New') and TFP growth calculated while maintaining the current assumptions of the ONS ('Old').

[7] https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/methodologies/capitalstocksuserguideuk

We find that TFP is largely underestimated from the years from 1995 up until the 2008 financial crisis for sector J. We also observe evidence of TFP growth being underestimated from 2011 to 2016. During the financial crisis and the years following the COVID recession and Brexit, we find that TFP is likely overestimated during those periods, the magnitude of overestimation is not large relative to the rate of underestimation in other periods.

Our calculations suggest that official TFP growth estimates are likely understated during the period leading to the 2008 financial crisis. Current assumptions on asset lives compressed the service life of software stocks when they should have been spread out over a longer period.

This discrepancy is likely attributable to the significant surge in software investments during the late 1990s and early 2000s. We notice that investments in intangibles overtook investments in tangible assets in the early 2000s for sector J (see Figure 5.B.2). Moreover, the share of own-account software investments to total intangibles for the sector also saw a substantial increase from 5.9 percent to 18.2 percent (see Figure 5.B.3). Capital stocks of the rising software were compressed to periods, coupled with a decline in tangible capital investment, which likely resulted in the underestimation that we observed in our results.

Perhaps we can interpret this as recent period software capital being more productive than those from the early 2000s. Clearly, more research is needed in this area. More importantly, given that these differences are systematic for certain periods, it is highly likely that this points to structural issues underlying the estimation of productivity and economic performance, particularly in sectors heavily reliant on intangible assets like software.

While our analysis is focused on TFP growth rates, we also find some systematic differences in the levels of the TFP index. In particular, we find that estimates of the TFP index, which employs the software asset life from our approach are consistently higher from 1995 to 2006. From 2007 onwards, we find that the TFP index from our estimates is consistently lower than TFP estimates using existing assumptions on the asset life (see table 5.D.3).

5.8 Conclusion

While various methodologies have been employed to measure intangible investments, persistent challenges remain in accurately estimating the depreciation of intangibles. We demonstrate that utilizing GSV can assist in this endeavor. Preliminary findings suggest a depreciation rate for software that diverges from estimates employed by statistical agencies, emphasizing the importance of refining measurement approaches to capture the true value dynamics of intangible assets within contemporary economic landscapes.

The advantage of this approach is its relatively easy implementation, requiring only the assumption that GSV directly correlates to changes in revenue streams associated with these assets. An assumption that we also tested and found support for in the study.

Additionally, the methodology holds the potential to extend its application to estimate the depreciation rates of other intangibles, such as TV series, songs, books, and music, as well as non-SNA intangibles, such as brands. Furthermore, the approach can incorporate additional dimensions into estimates, such as estimating the depreciation of intangibles for specific localities, and facilitate more frequent updates of asset lives required for capital stock measurement.

Appendix

5.A Capital Measurement

Consider the classical law of motion for capital:

$$K_{k,t} = K_{k,t-1}(1 - \delta_k) + I_{k,t} \tag{5.9}$$

where capital stock, $K_{k,t}$ at t t for asset k, is expressed as a function of current year investments $I_{k,t}$, and previous period capital stock $K_{k,t-1}$, discounted by a depreciation rate δ. This assumes a geometric depreciation pattern wherein the value of capital declines by a constant rate each period. Rearranging equation 5.9, we can arrive at an expression for CFC, which is typically defined as the difference between the present period capital stock and previous period capital stock:

$$CFC_{k,t} = K_{k,t-1} - K_{k,t} = \delta_k K_{k,t-1} + I_{k,t}. \tag{5.10}$$

While the above expression in equation 5.10 seems trivial, in practice the measurement exercise is often challenging. However, the value of aggregate capital stocks in the economy is typically not observed. Rather, statistical offices estimate capital stock using the perpetual inventory method approach. Consider a repeated substitution of equation 5.9 from the beginning period $t - 1$:

$$K_{k,t} = \sum_{i=0}^{\infty}(1 - \delta_k)^i + I_{k,t-(i+1)} \tag{5.11}$$

To implement equation 5.11, compilers need a historical investment time series that goes back to the initial investment period. This may not be possible in practice since the time series of investments for many countries are only available up to a specific period. However, it remains feasible to calculate capital stock if estimates of capital from the initial period of investment data are accessible, as described by Berlemann and Wesselhöft (2014):

$$K_{k,t} = (1-\delta_k)^{t-1}\bar{K}_k + \sum_{i=0}^{t-1}(1-\delta_k)^i I_{k,t-(i+1)}. \tag{5.12}$$

This requires three sets of information: historical investments data $I_{k,t}$ for asset k, initial capital stock when from the beginning to the investments time series, $\bar{K}_k$, and the depreciation rate δ_k. Investments are regularly recorded as gross fixed capital formation in the expenditure side of the National Accounts. The initial capital stocks are often estimated using information from a comprehensive set of financial statements or a country's economic census. Meanwhile, depreciation is often estimated by making assumptions on the asset's service life (how long the asset can contribute to production) and its retirement profile (when the asset is expected to be taken out of service).

5.B Additional results

Table 5.B.1

	OLS	PPML	NB
	(1)	(2)	(3)
δ	-0.109^{***}	-0.092^{***}	-0.065^{***}
	(0.002)	(0.0001)	(0.001)

Table 5.B.2: Results for Negative Binomial model removing free and open-source software

	All	No Free & Open Source	No Open Source Only
	(1)	(2)	(3)
δ	-0.009^{***}	-0.009^{***}	-0.009^{***}
	(0.0001)	(0.0001)	(0.0001)
n	161,735	135,186	144,453
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

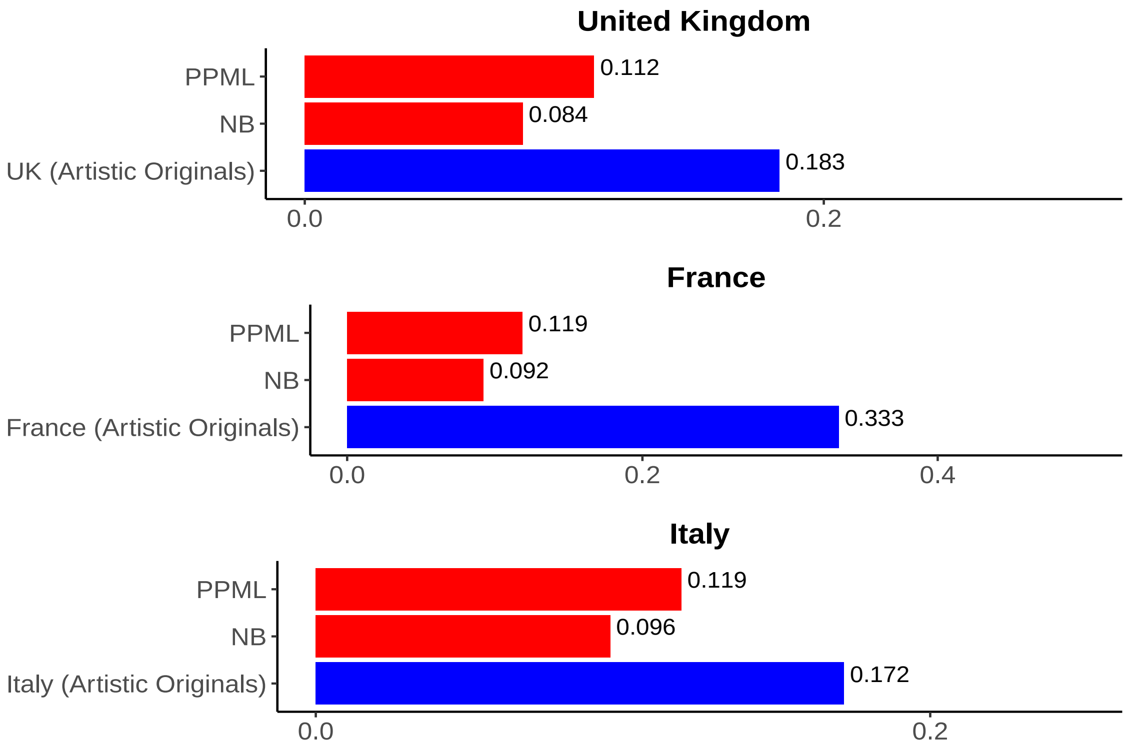


Figure 5.B.1: Comparing the annual depreciation rates estimates for movies with those from other countries

Note: The figure compares results from PPML and negative binomial (NB) model with the annual depreciation rates for theatrical movies employed by a select set of countries. The monthly estimates from table 5.6.5 were annualized assuming a geometric pattern. The depreciation rates employed by other countries were sourced from Pionnier et al. (2023).

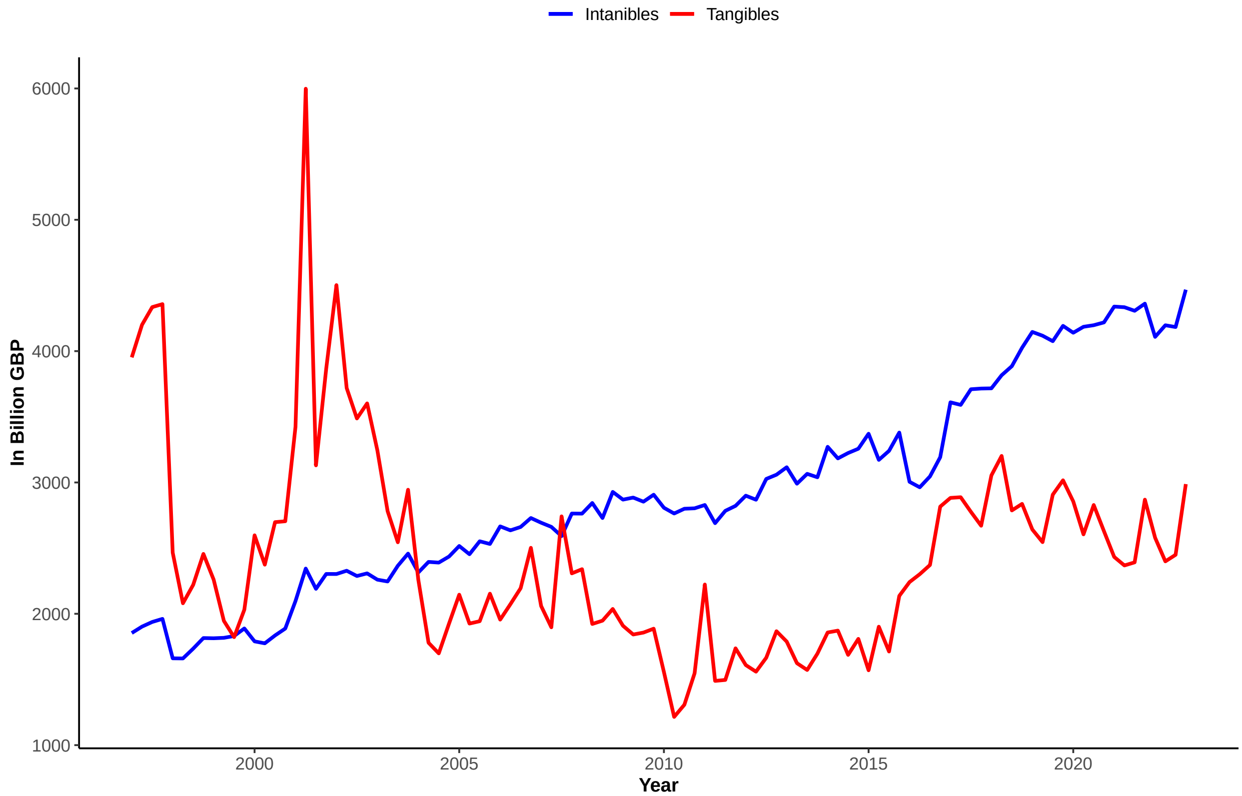


Figure 5.B.2: Investments for sector J from 1996 to 2022, current price estimates

Note: Figure shows the gross fixed capital formation for sector J at current price. In this figure, intangible investments only include intellectual property products (IPP) capitalized in the National Accounts, namely purchased and own-account software, literary and entertainment originals, mineral exploration, and research and development. The figure covers the period from 1996 to 2022.*Source of basic data:* Office for National Statistics, UK

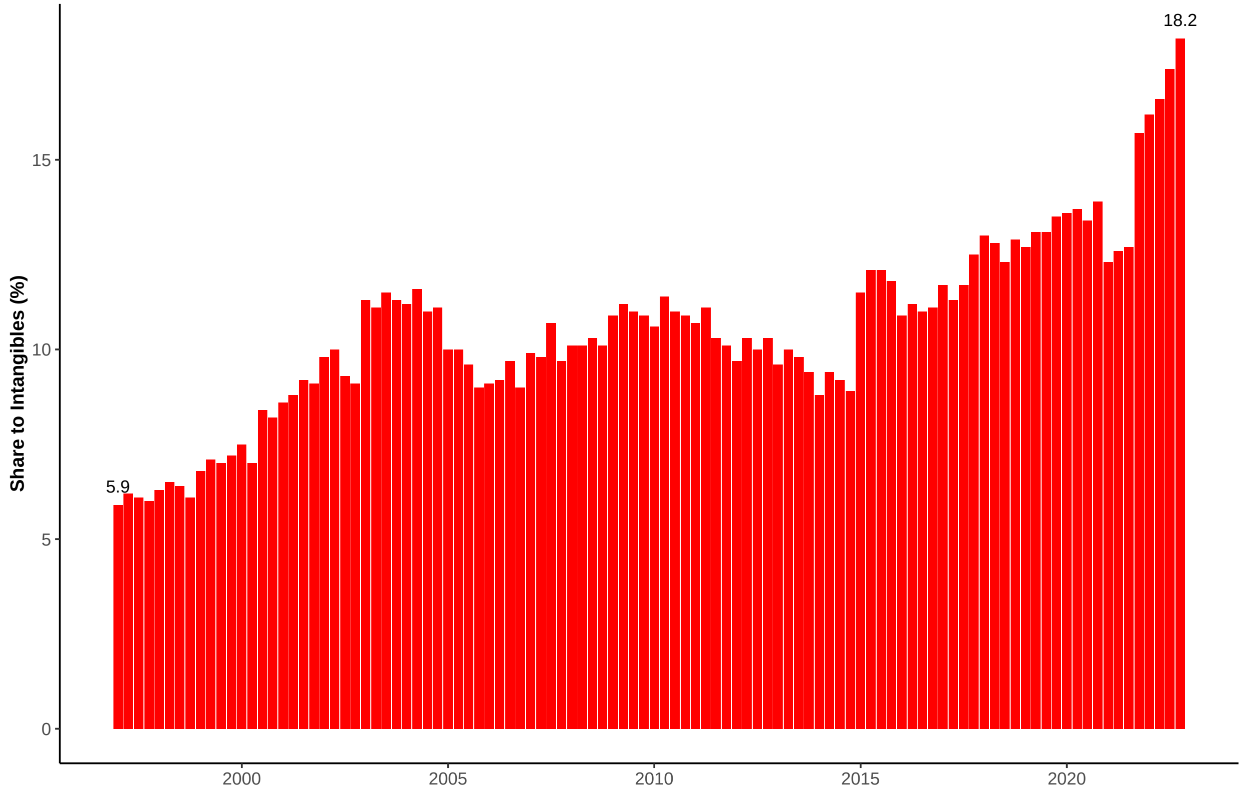


Figure 5.B.3: Share of own-account software to total intangibles for sector J from 1996 to 2022, current price estimates

Note: Figure shows the share of own-account software investments to total gross fixed capital formation on itangibles for sector J at current price. In this figure, intangible investments only include intellectual property products (IPP) capitalized in the National Accounts, namely purchased and own-account software, literary and entertainment originals, mineral exploration, and research and development. The figure covers the period from 1996 to 2022.*Source of basic data:* Office for National Statistics, UK

5.C Classifying software into categories

To classify software into categories, we exploit the categories tag at the bottom of the Wikipedia page of each software title. See example below:

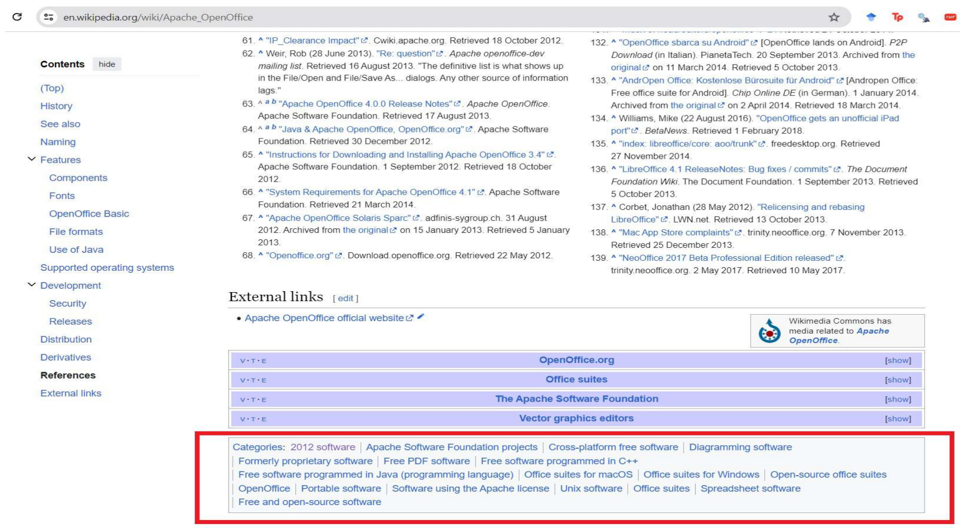


Figure 5.C.1: Example: Wikipedia page for Open Office

We scrape all of the tags and attach them to their corresponding software title. Each title would often have more than one tag. We were able to scrape 2,119 tags. Each of these tags corresponds to one or more software titles. For instance, the tag “Office suites” was attached to both Open Office, Microsoft Office Versions, and other office software. We show in figure 5.C.2 as word cloud for the top 1,000 most common tags.

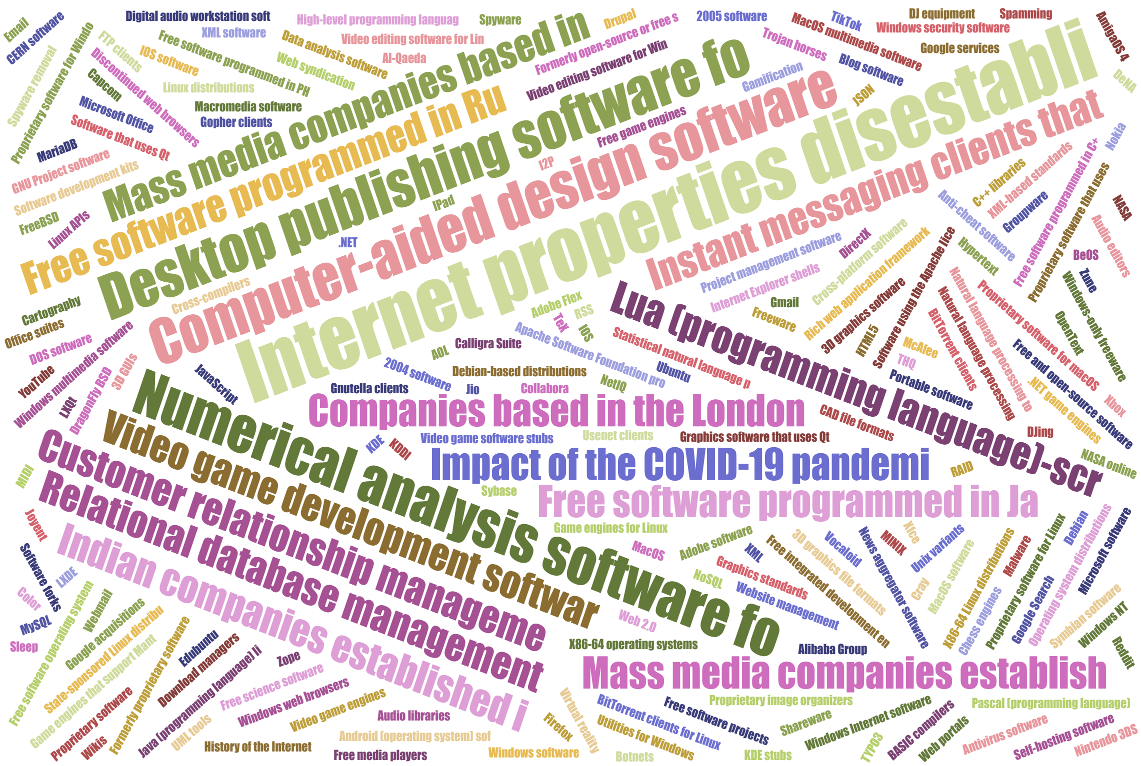
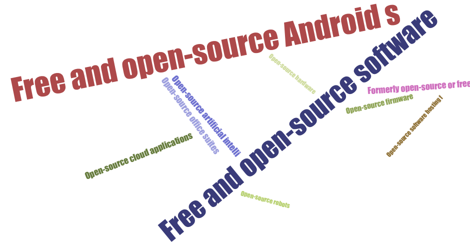


Figure 5.C.2: Word cloud of the category tags from software Wikipedia entries

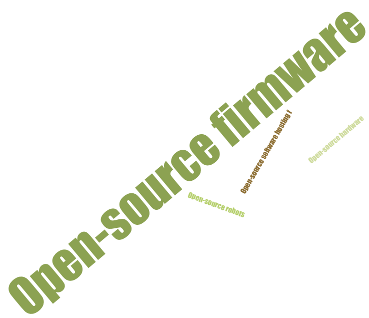
Our interest is to see whether the estimates would change if we remove “free-to-use” and “open-source” software. As such, we generated broad categories by identifying tags with the words “free” and “open source” and lumping them into a single broad category. We also created a third category that identifies tags containing the word “open source” but does not contain the word “free”. We show the word cloud for these tags in figure 5.C.3.

To extend our descriptive analysis, we also generated other broad categories using the same method. We identified the operating system with which these software are compatible and the type of software they can be classified into. We show the word cloud for the tags identified for each category in figures 5.C.4 and 5.C.5, respectively.

(a) Free software



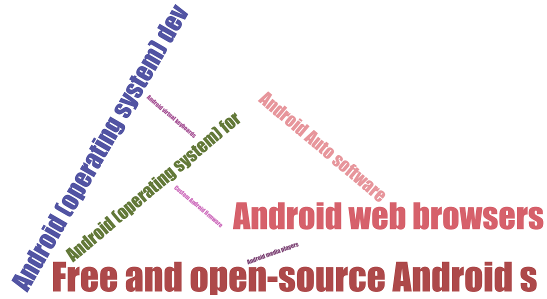
(b) Open-source software



(c) Strictly open-source software

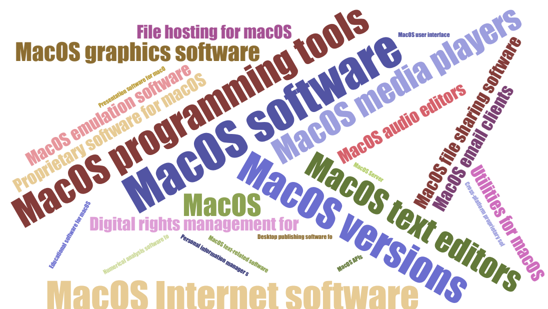
Figure 5.C.3: Free and open-source software

(a) Windows

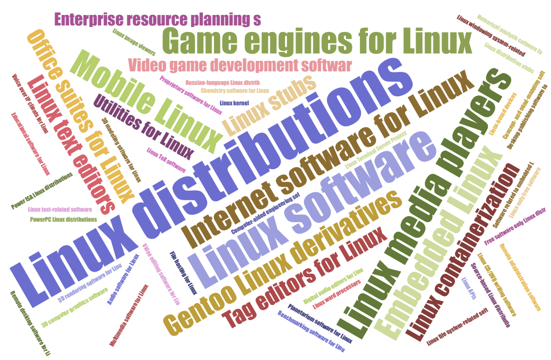


(b) Android

(c) iOS

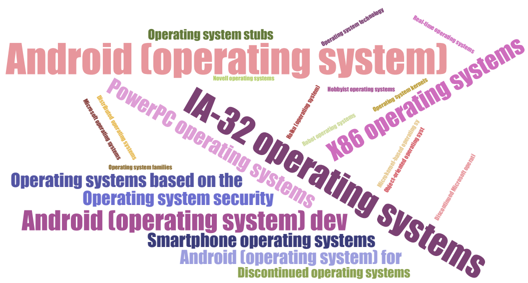


(d) MacOS



(e) Linux

Figure 5.C.4: By operating system

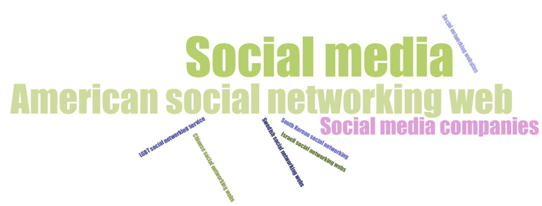


(a) Operating system

(b) Media

(c) Office

(d) Web browser



(e) Social media and social network

Figure 5.C.5: By type

5.D Constructing TFP estimates for sector J

We estimate TFP following Bontadini et al. (2023) and Organisation for Economic Co-operation and Development (2021). We also tried to remain to be consistent with Office for National Statistics (2007). The first step was to calculate user cost:

$$u_{i,j,t} = r_t p_{i,j,t} + \delta_i p_{i,j,t} + (p_{i,j,t} - p_{i,j,t-1}) \tag{5.13}$$

where $u_{i,t}$ is the user cost for asset i, for industry j, at time t, is the rate of return, δ_i is the depreciation rate, $p_{i,t}$ is the investment price, and r_t is the rate of return. We assumed a geometric pattern for the depreciation rate for all assets. To maintain consistency with the official methodology, we also used the internal rate of return, which we calculate by taking the ratio of Gross Operating Surplus to GDP[8]. We sourced all our data from the ONS website[9]. We proceeded by calculating thing net capital stock K_t:

$$K_{j,t} = \sum_i k_{j,i,t} \times w_{j,t} \tag{5.14}$$

where the weights $w_{i,t}$ is derived from:

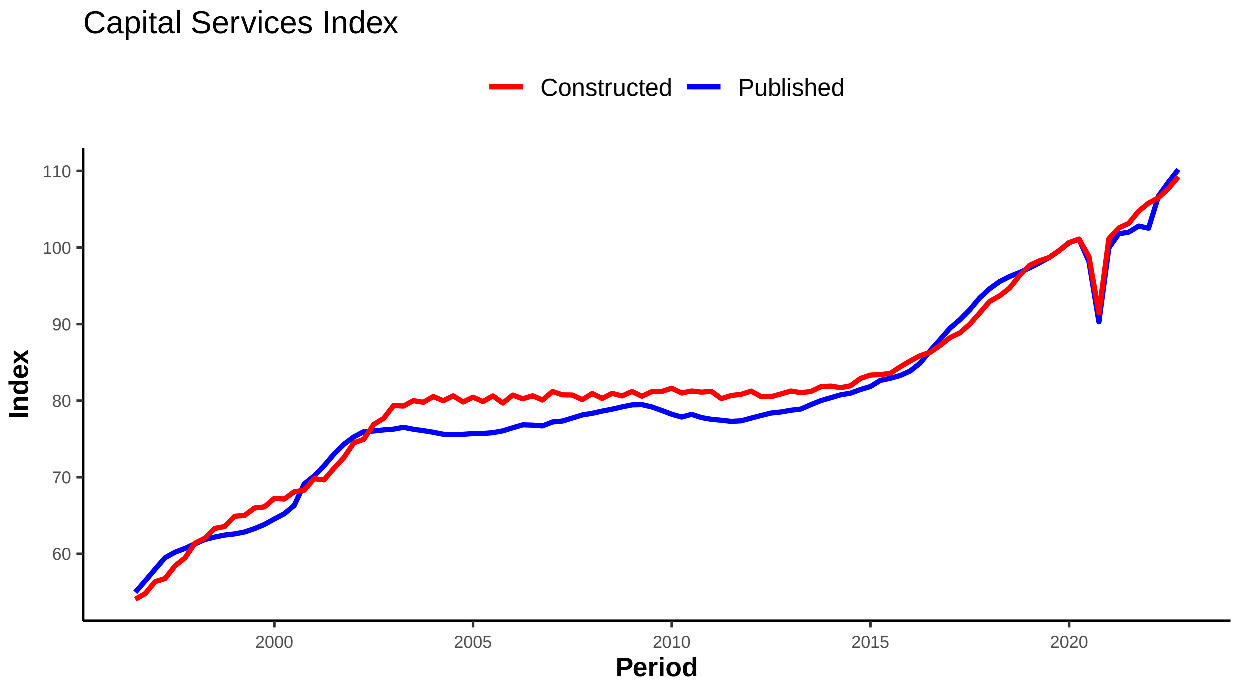
$$w_{j,t} = \frac{k_{j,i,t} \times u_{i,j,t}}{\sum_i k_{j,i,t} \times u_{i,j,t}} \tag{5.15}$$

Our estimates of the capital services index do not align with the capital services index published by the ONS. One of the reasons is likely due to the tax adjustment that the ONS applies when computing for user cost, as well as the difference in deflating computers. Note that at this point, we have not made any changes yet to the assumptions on the service life for any asset.

[8]This is also the recommendation of Bontadini et al. (2023).

[9]https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/methodologies/capitalstocksuserguideuk

Figure 5.D.1: Constructed versus published Capital Services Index



Note: The figure shows the capital service index we estimated (red line) and the capital services index employed by the ONS for their TFP calculations (blue line).

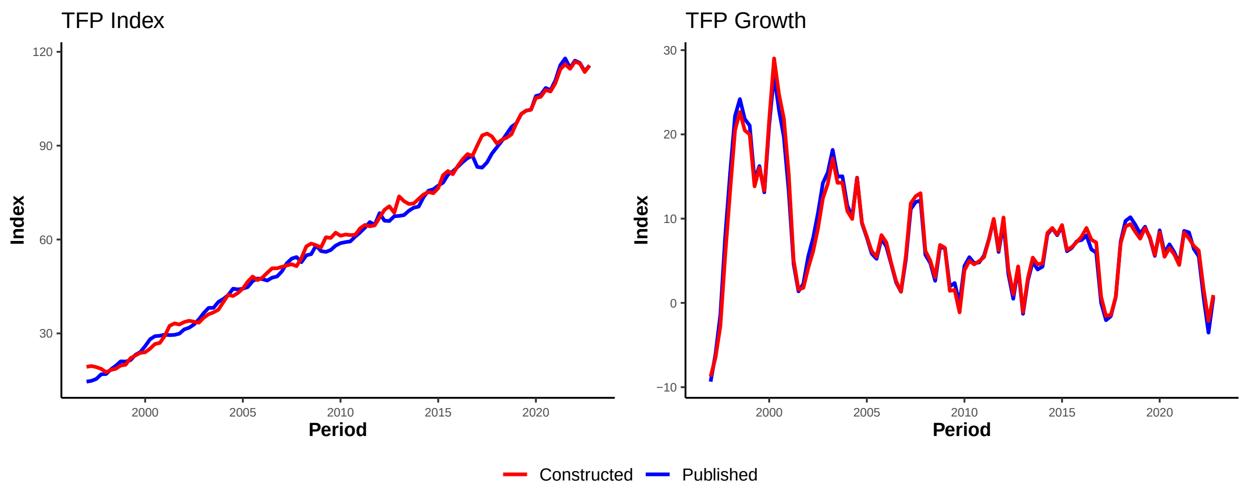
We proceed by calculating TFP using the standard growth accounting formula.

$$\Delta log(TFP_t) = \Delta log(Y_t) - v_t^l \Delta log(L_t) - v_t^k \Delta log(K_t) \tag{5.16}$$

where we express TFP_t as the difference between changes in the output Y_t and changes in Labor inputs L_t, and capital inputs K_t. The terms v_t^l and v_t^k are labor and capital weights.

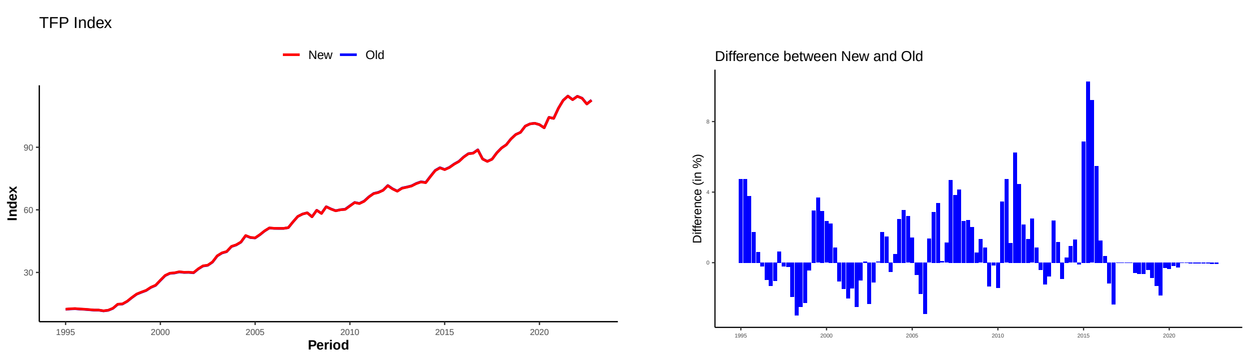
Despite the difference in our constructed capital services index to those published by the ONS, we see little discrepancies between the published TFP index and those that we constructed using equation 5.16 (see figure 5.D.2). We do not find any systematic difference between our constructed index and the TFP index and growth rates published by the ONS.

Figure 5.D.2: Constructed versus published TFP



Note: The figure shows TFP we estimated (red line) and the TFP published by the ONS (blue line).

Following the steps earlier, we constructed a third TFP index for sector J where we changed the assumption on the asset life of own-account software for industries 62 and 63, using the asset life we estimated using GSV results. We also made changes to the user cost of own-account software by employing the new set of depreciation rates in our calculations. We compare the indices in figure 5.D.3.

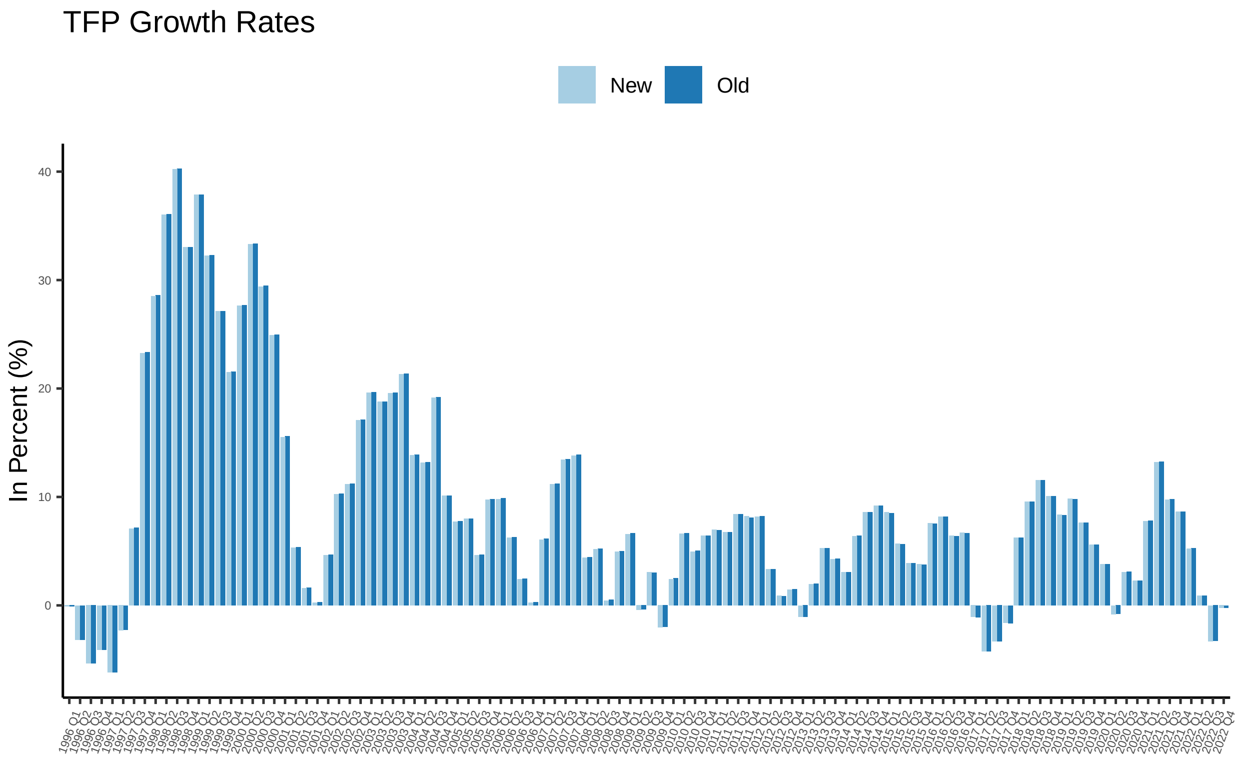


(a) Levels

(b) Difference

Figure 5.D.3: TFP Index

Figure 5.D.4: TFP growth rates



CHAPTER 6: Discussion of Key Findings and Concluding Remarks

In this set of essays, we provide some estimation strategies that could be employed to improve the measurement of the modern economy. The goal is to provide a set of methodologies that can be used to compile macroeconomic aggregates and provide a better understanding of the modern economy. Most of the methodologies are limited in scale and scope. Despite the limitations, we were able to draw some interesting insights from these exercises.

In the first essay (Chapter 2), we show analytically that applying the price of paid platforms to the valuation of their free counterpart allows for a surplus that distinguishes the value of viewership from the independent value of digital services. This addresses some of the limitations of the barter model in terms of capturing welfare.

While the second essay (Chapter 3) is limited in scope, covering only three categories of digital services, we find that the contribution of these services to household final consumption in the UK is substantial. Our estimates show that the aggregate value of free digital services is comparable to other expenditure items in the National Accounts such as private health and communication services. This speaks about how important these services have become in our daily lives.

Considering that this study does not even cover other widely-used services such as online maps, search engines, instant messaging, and social media, it is not difficult to imagine that a comprehensive set of estimates covering all free services would provide an even more striking picture of their significance in the contemporary economy. As society becomes more dependent on free digital services—not just for leisure but for communication, navigation, and daily transactions—policymakers need to be informed about the degree to which consumers are vulnerable to shocks in these services.

Often, providers of free digital products deliver their services to consumers with no oversight nor are they accountable to their users. They can unilaterally impose prices on services that are currently free (consider Zoom for education) or discontinue certain services without warning. Without knowing how much consumers value these services overall, how should governments respond appropriately to such actions? How can they justify their response without credible data backing them?

On a related note, we also provide estimates households receive from digital piracy (Chapter 4). The IPO report shows that more than a quarter of the British population is engaged in digital piracy. We find that by including the value of digital piracy in household consumption, the *measured* growth in the consumption of information goods will slow down. This implies that our official figures overestimate the growth in welfare to some extent.

Central to this discussion is the notion that households have historically engaged with movie and music services, albeit through piracy. The popularity of low-cost streaming platforms and increasingly stringent measures against piracy has prompted a shift, with households now more likely to pay for these services. Consequently, the consumption of such products is now reflected in GDP figures. However, their previous consumption of services through piracy was not.

Lastly, in our last essay (Chapter 5), we show that big data could be used to estimate the depreciation rate of intangibles. We show that for assets where current estimates are based on empirical data (such as theatrical movies in the US), the decline in Google Trends results provides a good approximation of the depreciation rates of these assets.

The methodology that we propose has the potential to provide compiling organizations such as statistical agencies and multilateral institutions the flexibility to measure capital stocks with more details and accuracy. For instance, the approach allows statistical agencies to update the asset lives they employ for TFP more regularly, and allow these agencies to be more responsive to the changing landscape of modern society.

Direction for future research

Future work in this area could also examine the contribution of free digital services to the production of market goods and services. While we find this as a very important area of research, there is little work in this direction. As we discuss in appendix 2.A.2, a possible measurement strategy would be to estimate a production function, which incorporates free digital products as part of inputs. We imagine that time-use surveys could provide indicators of how much time workers spend using free digital services in their work routine.

How much time do office workers spend using Google Docs or Google Sheets? How much time do researchers spend on Google Scholar? Answering these questions could provide an indication of how much industries rely on free digital services as inputs to their production process. The linkages between free digital service providers and producers of market goods are very relevant in terms of describing how the internet is shaping human society. It answers the question, to what extent are businesses vulnerable to shocks in the provision of these services?

Moreover, our valuation method only covers services where paid alternatives can be found. Future research could focus on measuring the value of free digital products with no paid counterparts. Taking off from Chapter 2, the key is to find indicators that reflect the marginal benefits of users.

A possible way forward is to consider the use of surveys and experiments, otherwise known as contingent valuation. Recent studies employing this approach aim to elicit their respondents' WTA for abstaining from the use of these services. We discussed in Chapter 2, that WTA estimates would likely capture the value of the activity enhanced by the digital product and not the value of the product itself. A possible way to address this problem is to design surveys (or experiments) that elicit the users' WTP instead of WTA. The National Accounts value goods and services in terms of exchange value. If researchers are able to capture the median WTP for digital services such as search engines and online maps, this might not necessarily reflect the exchange value required for consistency with the SNA. However, this could be a starting point for the development of simulated exchange values, which have been applied to the valuation of ecosystem services (Atkinson and Obst, 2017). We provide some discussion on this in appendix 2.A.2.

Future research could also focus on developing a comprehensive satellite account that includes the value of free digital services. The estimates provided in Chapters 3 and 4 represent the gross value from the consumption of these services and would be reflected on the demand side of the accounts. A thorough satellite account should also consider the production and income sides, necessitating the distinction between value-added and intermediate consumption. Advertising spending or revenue from data sales offers valuable insights into how much advertisers and third parties value viewership and user data, respectively. As discussed in Chapter 2, this information could be used to account for the value of household viewership, ultimately considering it as the household's intermediate consumption.

The estimates derived from our methodology in Chapters 3 and 4 are analogous to gross output on the production side. Therefore, once the value of viewership is accounted for, it would be possible to populate table 2.2 in Chapter 2, which describes the production, income, and consumption aspects of free digital services provision. This could extend the digital economy satellite account proposed by Ahmad and Ribarsky (2018) or be used to create a stand-alone satellite account that highlights the value derived by households and firms from the consumption of free digital products.

Lastly, our approach to measuring the depreciation of intangibles provides enough flexibility for various applications. It is possible to apply this methodology to other categories of intangibles such as music, TV series, books, and perhaps brands. Aside from movies, it has always been challenging to compile detailed and consistent time series data on revenues from other artistic originals (Goodridge et al., 2013b; Soloveichik et al., 2013a,b; Soloveichik, 2014).

The methodology also allows for the estimation of asset lives for sub-categories of assets. For instance, it could be possible to have separate depreciation rates for word processors, operating systems, statistical software, and others. It would be interesting to examine how the application of specific depreciation rates could impact macroeconomic aggregates. At the moment, statistical agencies assume a single average asset life for all software, regardless of categories. If there is heterogeneity in the asset lives of each category, a change in the composition of software investments could have an impact on measured growth. Future research could explore this avenue. Given the economy is becoming increasingly dependent on intangibles, it is critical that we have an accurate estimate to guide policymakers for sound evidence-based decision-making.

****

At the end of the day, GDP is a human construct. It measures what we want it to measure. But one important thing to think about is that the official statistics are compiled to help policymakers make informed decisions. In this set of essays, we argued that due to the changing nature of the economy, traditional ways of compiling macroeconomic aggregates make it difficult to provide a complete picture of contemporary society. We believe that it is critical that we consider better ways of reflecting new economic realities. By incorporating the value of free digital services, the impact of digital piracy, and improving how we measure intangibles, we can offer a more comprehensive and accurate representation of economic activity. This should ultimately lead to better policy decisions that reflect the true dynamics of the modern economy.

References

Abramsky, J. and Drew, S. (2014). Changes to national accounts: inclusion of illegal drugs and prostitution in the UK national accounts. *UK Office for National Statistics Working Paper*.

Ahmad, N. and Ribarsky, J. (2018). Towards a framework for measuring the digital economy. In *16th Conference of the International Association of Official Statisticians*, pages 1–33.

Ahouraian, M. (2021). The effects of a shrinking window for theatrical release exclusivity.

Aizcorbe, A. M. (2014). *A practical guide to price index and hedonic techniques*. Practical Econometrics.

Atkinson, G. and Obst, C. (2017). Prices for ecosystem accounting. *World Bank, Wealth Accounting and Valuation of Ecosystem Services*, 38.

Awano, G., Franklin, M., Haskel, J., and Kastrinaki, Z. (2010). Investing in innovation: Findings from the UK investment in intangible assets survey. *NESTA Index Report (July)*.

Ballester, M., Garcia-Ayuso, M., and Livnat, J. (2003). The economic value of the R&D intangible asset. *European Accounting Review*, 12(4):605–633.

Barnett, A., Batten, S., Chiu, A., Franklin, J., and Sebastia-Barriel, M. (2014). The UK productivity puzzle. *Bank of England Quarterly Bulletin*, page Q2.

Bean, C. R. (2016). *Independent review of UK economic statistics*. HM Treasury.

Belleflamme, P. and Peitz, M. (2021). *The economics of platforms*. Cambridge University Press.

Berlemann, M. and Wesselhöft, J.-E. (2014). Estimating aggregate capital stocks using the perpetual inventory method: A survey of previous implementations and new empirical evidence for 103 countries. *Review of Economics*, 65(1):1–34.

Besen, S. M. and Kirby, S. N. (1989). Private copying, appropriability, and optimal copying royalties. *The Journal of Law and Economics*, 32(2, Part 1):255–280.

Blackburn, D. (2007). The heterogeneous effects of copying: the case of recorded music. *Harvard Working Papers*.

Blackburn, D., Eisenach, J. A., and Harrison, D. (2019). Impacts of digital video piracy on the US economy. *NERA Economic Consulting, the Chamber's Global Innovation Policy Center.*

Blades, D. (2011). Estimating value added of illegal production in the Western Balkans. *Review of Income and Wealth*, 57(1):183–195.

Bontadini, F., Corrado, C., Haskel, J., Iommi, M., and Jona-Lasinio, C. (2023). Euklems & intanprod: industry productivity accounts with intangibles. *Sources of growth and productivity trends: methods and main measurement challenges, Luiss Lab of European Economics, Rome.*

Boyle, R. (2006). *Measuring public sector productivity: lessons from international experience*, volume 35. Institute of Public Administration.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., and Fox, K. J. (2019a). GDP-B: Accounting for the value of new and free goods in the digital economy. In *NBER Working Paper Series*. National Bureau of Economic Research.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., and Fox, K. J. (2020). Measuring the impact of free goods on real household consumption. In *AEA Papers and Proceedings*, volume 110, pages 25–30.

Brynjolfsson, E., Collis, A., and Eggers, F. (2019b). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, 116(15):7250–7255.

Calderón, J. B. S., Rassier, D. G., et al. (2022). Valuing the us data economy using machine learning and online job postings. Technical report, Bureau of Economic Analysis.

Caparrós, A., Oviedo, J. L., Álvarez, A., and Campos, P. (2017). Simulated exchange values and ecosystem accounting: Theory and application to free access recreation. *Ecological Economics*, 139:140–149.

Carrière-Swallow, Y. and Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, 32(4):289–298.

Cervellin, G., Comelli, I., and Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? insights from different clinical settings. *Journal of Epidemiology and global health*, 7(3):185–189.

Chantler-Hicks, L. (2022). Londoners face £700 million local service cuts unless government steps in.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, 88:2–9.

Chute, J. W., McCulla, S. H., and Smith, S. (2018). Preview of the 2018 comprehensive update of the national income and product accounts. *Survey of Current Business*, 98(4).

Corrado, C., Haskel, J., Jona-Lasinio, C., and Iommi, M. (2014). Intangibles and industry productivity growth: Evidence from the eu.

Corrado, C., Haskel, J., Jona-Lasinio, C., and Iommi, M. (2022). Intangible capital and modern economies. *Journal of Economic Perspectives*, 36(3):3–28.

Corrado, C., Hulten, C., and Sichel, D. (2009). Intangible capital and us economic growth. *Review of income and wealth*, 55(3):661–685.

Corrigan, J. R., Alhabash, S., Rousu, M., and Cash, S. B. (2018). How much is social media worth? estimating the value of facebook by paying users to stop using it. *PloS one*, 13(12):e0207101.

Coyle, D. (2015). *GDP: a brief but affectionate history-revised and expanded edition*. Princeton University Press.

Coyle, D. (2019). Do-it-yourself digital: the production boundary, the productivity puzzle and economic welfare. *Economica*, 86(344):750–774.

Coyle, D. and Nguyen, D. (2023). Free digital products and aggregate economic measurement. *Economie et Statistique/Economics and Statistics*, (539):27–50.

de Haan, J. and Diewert, E. (2013). Hedonic Regression Methods. *Handbook on Residential Property Price Indices*, pages 49–64.

De Rassenfosse, G. and Jaffe, A. B. (2017). Econometric evidence on the R&D depreciation rate. Technical report, National Bureau of Economic Research.

De Vany, A. S. and Walls, W. D. (2007). Estimating the effects of movie piracy on box-office revenue. *Review of Industrial Organization*, 30(4):291–301.

Dejean, S. (2009). What can we learn from empirical studies about piracy? *CESifo Economic Studies*, 55(2):326–352.

Del Rio, F. and Sampayo, A. (2014). Obsolescence and productivity. *Portuguese Economic Journal*, 13:195–216.

Diewert, W. E. (2003). Hedonic regressions: A review of some unresolved issues. In *7th Meeting of the Ottawa Group, Paris, May*, volume 29.

Diewert, W. E., Wykoff, F. C., et al. (2006). Depreciation, deterioration and obsolescence when there is embodied or disembodied technical change. *forthcoming in Price and Productivity Measurement, WE Diewert, BM Balk, D. Fixler, KJ Fox and AO Nakamura (eds.), Canada: Trafford Press. http://www. econ. ubc. ca/diewert/dp0602. pdf.*

Digitalization Task Team (2022). Guidance note on recording and valuing "free"' digital products in an SNA satellite account. Technical report, Inter-secretariat Working Group on National Accounts, United Nations.

Dragusin, L. (2015). Macro-statistical challenges: measuring illegal activities in Romania. *Competitiveness of Agro-Food and Environmental Economy*, page 449.

Dunleavy, P. (2017). Public sector productivity: Measurement challenges, performance information and prospects for improvement. *OECD Journal on Budgeting*, 17(1):1–28.

Dynan, K. and Sheiner, L. (2018). GDP as a measure of economic well-being. *Working Paper*, 43.

Effenberger, M., Kronbichler, A., Shin, J. I., Mayer, G., Tilg, H., and Perco, P. (2020). Association of the COVID-19 pandemic with internet search volumes: a Google Trends tm analysis. *International Journal of Infectious Diseases*, 95:192–197.

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., and Sax, C. (2022). Obtaining consistent time series from Google Trends. *Economic Inquiry*, 60(2):694–705.

Erickson, T. (2016). *On" incorrect" Signs in Hedonic Regression*. US Department of Labor, US Bureau of Labor Statistics, Office of Prices and ....

Fernald, J. G. and Inklaar, R. (2022). The UK productivity" puzzle" in an international comparative perspective. Federal Reserve Bank of San Francisco.

Ferrara, L. and Simoni, A. (2022). When are google data useful to nowcast GDP? an approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, pages 1–15.

Field, S., Franklin, M., et al. (2012). Results from the second survey of investment in intangible assets, 2010.

Gogineni, S. and Pimpalshende, A. (2020). Predicting IMDB movie rating using deep learning. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 1139–1144. IEEE.

Goodridge, P., Haskel, J., and Wallis, G. (2013a). Can intangible investment explain the UK productivity puzzle? *National Institute Economic Review*, 224:R48–R58.

Goodridge, P., Haskel, J., and Wallis, G. (2013b). Film, television & radio, books, music and art: estimating UK investment in artistic originals. *Measuring the Contribution of Knowledge*, page 67.

Goodridge, P., Haskel, J., and Wallis, G. (2018). Accounting for the UK productivity puzzle: a decomposition and predictions. *Economica*, 85(339):581–605.

Görzig, B. and Gornig, M. (2015). The assessment of depreciation in the case of intangible assets. Technical report, SPINTAN Working Paper Series No.

Groshen, E. L., Moyer, B. C., Aizcorbe, A. M., Bradley, R., and Friedman, D. M. (2017). How government statistics adjust for potential biases from quality change and new goods in an age of digital technologies: A view from the trenches. *Journal of Economic Perspectives*, 31(2):187–210.

Hall, B. H. (2007). Measuring the returns to R&D: The depreciation problem.

Hamulka, J., Jeruszka-Bielak, M., Górnicka, M., Drywień, M. E., and Zielinska-Pukos, M. A. (2020). Dietary supplements during covid-19 outbreak. results of Google Trends analysis supported by plifecovid-19 online studies. *Nutrients*, 13(1):54.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: how much can they differ? *The American Economic Review*, 81(3):635–647.

Harish, B., Kumar, K., and Darshan, H. (2019). Sentiment analysis on IMDB movie reviews using hybrid feature extraction method.

Haskel, J., Corrado, C., Jona-Lasinio, C., and Iommi, M. (2018). Intangible investment in the EU and US before and since the great recession and its contribution to productivity growth.

Havranek, T. and Zeynalov, A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics*, 27(1):129–148.

Heys, R., Martin, J., Mkandawire, W., et al. (2019). GDP and welfare: A spectrum of opportunity. Technical report, Economic Statistics Centre of Excellence (ESCoE).

Heys, R. and Taylor, C. (2021). Valuing free digital platforms in a national accounting framework.

Hsu, P.-Y., Shen, Y.-H., and Xie, X.-A. (2014). Predicting movies user ratings with IMDB attributes. In *Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9*, pages 444–453. Springer.

Huang, N. and Diewert, E. (2011). Estimation of R&D depreciation rates: a suggested methodology and preliminary application. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(2):387–412.

Hulten, C. and Nakamura, L. (2017). Accounting for growth in the age of the internet: The importance of output-saving technical change. In *NBER Working Paper Series, organization=National Bureau of Economic Research*.

Jamison, M. A. and Wang, P. (2021). Valuation of digital goods during the coronavirus outbreak in the united states. *Telecommunications Policy*, 45(5):102126.

Jigsaw Research (2021). News consumption in the UK: 2020. Technical report, Office of Communications.

Johnson, W. R. (1985). The economics of copying. *Journal of Political Economy*, 93(1):158–174.

Kazemier, B., Bruil, A., Steeg, A. v. d., and Rensman, M. (2013). The contribution of illegal activities to national income in the netherlands. *Public Finance Review*, 41(5):544–577.

Kohns, D. and Bhattacharjee, A. (2023). Nowcasting growth using Google Trends data: A bayesian structural time series model. *International Journal of Forecasting*, 39(3):1384–1412.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Leontief, W. et al. (1974). Environmental repercussions and the economic structure: an input-output approach: a reply. *The Review of Economics and Statistics*, 56(1):109–110.

Lev, B. and Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. *Journal of accounting and economics*, 21(1):107–138.

Li, W. C. (2014). Organizational capital, R&D assets, and offshore outsourcing.

Li, W. C. and Hall, B. H. (2020). Depreciation of business R&D capital. *Review of Income and Wealth*, 66(1):161–180.

Liebowitz, S. (2008). Testing file-sharing's impact on music album sales in cities. research note: School of Management, University of Texas at Dallas. *Management Science*.

Liebowitz, S. J. (1985). Copying and indirect appropriability: Photocopying of journals. *Journal of Political Economy*, 93(5):945–957.

Martin, J. (2019). Measuring the other half: new measures of intangible investment from the ONS. *National Institute Economic Review*, 249:R17–R29.

Martin, J. and Riley, R. (2024). Productivity measurement: Reassessing the production function from micro to macro. *Journal of Economic Surveys*.

Martin, W. and Pham, C. S. (2020). Estimating the gravity model when zero trade flows are frequent and economically determined. *Applied Economics*, 52(26):2766–2779.

McLaren, N. and Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, (2011):Q2.

Michel, N. J. (2006). The impact of digital file sharing on the music industry: An empirical analysis. *Topics in Economic Analysis & Policy*, 6(1).

Nadiri, M. I. and Prucha, I. R. (1996). Estimation of the depreciation rate of physical and R&D capital in the us total manufacturing sector. *Economic Inquiry*, 34(1):43–56.

Nakamura, L., Samuels, J., and Soloveichik, R. H. (2017). Measuring the'free'digital economy within the GDP and productivity accounts.

Nakamura, L. I. (2010). Intangible assets and national income accounting. *Review of Income and Wealth*, 56:S135–S155.

Nakamura, L. I. and Soloveichik, R. H. (2015). Valuing 'free' media across countries in GDP.

Nordhaus, W. D. (2006). Principles of national accounting for nonmarket accounts. In *A new architecture for the US national accounts*, pages 143–160. University of Chicago Press.

Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy*, 115(1):1–42.

OECD (2009). Measuring capital, Second Edition. *OECD Publishing,.*

OECD (2010). OECD Handbook on Deriving Capital Measures of Intellectual property products,. *OECD Publishing,.*

Office for National Statistics (2007). The ons productivity handbook: a statistical overview and guide.

Office for National Statistics (2018). Value of nature implicit in property prices – hedonic pricing method (hpm) methodology note.

Office for National Statistics (2019a). Consumer prices indices technical manual, 2019.

Office for National Statistics (2019b). Valuing green spaces in urban areas: a hedonic price approach using machine learning techniques.

Oghina, A., Breuss, M., Tsagkias, M., and De Rijke, M. (2012). Predicting IMDB movie ratings using social media. In *Advances in Information Retrieval: 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings 34*, pages 503–507. Springer.

Ordover, J. A. and Willig, R. D. (1978). On the optimal provision of journals qua sometimes shared goods. *The American Economic Review*, 68(3):324–338.

Organisation for Economic Co-operation and Development (2021). OECD productivity statistics methodological notes.

O'Mahony, M. and Weale, M. (2021). Depreciation and net capital services: how much do intangibles contribute to economic growth? In *IARIW-ESCoE Conference on Intangible Assets*.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to pc's. *American Economic Review*, 93(5):1578–1596.

Pakes, A. and Schankerman, M. (1984). The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&D, patents, and productivity*, pages 73–88. University of Chicago Press.

Peitz, M. and Waelbroeck, P. (2004). The effect of internet piracy on cd sales: Cross-section evidence. *Journal of Political Economy.*

Pionnier, P.-A., Zinni, M. B., and Baret, K. (2023). Sensitivity of capital and mfp measurement to asset depreciation patterns and initial capital stock estimates.

Quintais, J. and Poort, J. (2018). Global online piracy study. Technical report, University of Amsterdam - Institute for Information Law.

Randall, A. and Stoll, J. R. (1980). Consumer's surplus in commodity space. *The American Economic Review*, 70(3):449–455.

Riley, R., Bondibene, C. R., and Young, G. (2015). The UK productivity puzzle 2008-13: evidence from british businesses.

Riley, R., Rincon-Aznar, A., and Samek, L. (2018). Below the aggregate: a sectoral account of the UK productivity puzzle. *ESCoE Discussion Papers*, 6.

Rincon-Aznar, A., Riley, R., and Young, G. (2017). Academic review of asset lives in the UK. *Discussion Papers*, 474.

Rob, R. and Waldfogel, J. (2006). Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. *The Journal of Law and Economics*, 49(1):29–62.

Rogers, S. (2016). What is Google Trends data — and what does it mean? Published in Medium.

Roser, M. and Ritchie, H. (2013). Technological progress. *Our World in Data*.

Roth, F. (2019). Intangible capital and labour productivity growth: a review of the literature.

Schreyer, P. (2004). Capital stocks, capital services and multi-factor productivity measures. *OECD Economic Studies*, 2003(2):163–184.

Schreyer, P. (2021). Accounting for free digital services and household production - an application to facebook. In *36th International Association for Research in Income and Wealth (IARIW)) Virtual General Conference*. International Association for Research in Income and Wealth.

Seung-Hyun, H. (2005). Measuring the effect of digital technology on the sales of copyrighted goods: Evidence from napster. Technical report, Working Paper, University of Illinois at Urbana-Champaign.

Shaukat, Z., Zulfiqar, A. A., Xiao, C., Azeem, M., and Mahmood, T. (2020). Sentiment analysis on IMDB using lexicon and neural networks. *SN Applied Sciences*, 2:1–10.

Shogren, J. F., Shin, S. Y., Hayes, D. J., and Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, pages 255–270.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Simpson, H. (2009). Productivity in public services. *Journal of Economic Surveys*, 23(2):250–276.

Soloveichik, R. (2010). Artistic originals as a capital asset. *American Economic Review*, 100(2):110–114.

Soloveichik, R. (2014). Books as capital assets. *Journal of scholarly publishing*, 45(2):101–127.

Soloveichik, R. (2015). Valuing 'free'entertainment in GDP: an experimental approach. In *AEA Meeting in Boston, Boston, MA*.

Soloveichik, R. (2019). Including illegal activity in the us national economic accounts. Technical report, BEA Working Paper 2019-4, https://www. bea. gov/system/files/papers/WP2019 ....

Soloveichik, R. et al. (2013a). *Long-Lived Television Programs as Capital Assets*. Bureau of Economic Analysis.

Soloveichik, R. et al. (2013b). *Music originals as capital assets*. Bureau of Economic Analysis.

Soloveichik, R. et al. (2013c). *Theatrical Movies as Capital Assets*. Bureau of Economic Analysis (BEA).

Statista Research Department (2021a). Leading communication apps in the Google Play Store in Great Britain (GB) in october 2021, by number of downloads.

Statista Research Department (2021b). Most popular global mobile messenger apps as of july 2021, based on number of monthly active users.

Statista Research Department (2021c). Whatsapp users in the united kingdom 2017-2025.

Statista Research Department (2022). Number of facebook messenger users in the united kingdom from september 2018 to december 2021.

Sturgess, B. et al. (2018). Measuring illegal activities in the national accounts. *World Economics*, 19(2):1–12.

Syverson, C. (2017). Challenges to mismeasurement explanations for the US productivity slowdown. *Journal of Economic Perspectives*, 31(2):165–86.

Takeyama, L. N. (1994). The welfare implications of unauthorized reproduction of intellectual property in the presence of demand network externalities. *The Journal of Industrial Economics*, pages 155–166.

Topal, K. and Ozsoyoglu, G. (2016). Movie review analysis: Emotion analysis of IMDB movie reviews. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1170–1176. IEEE.

Triplett, J. (2006). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes. *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*.

United Nations (2012). System of environmental-economic accounting 2012. Technical report, The World Bank Group.

United Nations (2014). System of environmental: Ecosystem accounting. Technical report, The World Bank Group.

United Nations and others (2009). *System of National Accounts 2008*. New York: United Nations.

Van Criekingen, K., Bloch, C., and Eklund, C. (2022). Measuring intangible assets—a review of the state of the art. *Journal of Economic Surveys*, 36(5):1539–1558.

Van Elp, M., Kuijpers, N., and Mushkudiani, N. (2022). Free services in the netherlands. In *37th IARI General Conference*, Luxembourg. International Association for Research in Income and Wealth.

Van Elp, M. and Mushkudiani, N. (2019). Free services. Technical report, Statistics Netherlands.

Warusawitharana, M. (2015). Research and development, profits, and firm value: A structural estimation. *Quantitative Economics*, 6(2):531–565.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *American Economic Review*, 100(4):1642–1672.

Willig, R. D. (1976). Consumer's surplus without apology. *The American Economic Review*, 66(4):589–597.

Woloszko, N. (2021). Tracking GDP using Google Trends and machine learning: A new OECD model. *Central Banking*, 12:12.

Zattoni, F., Gül, M., Soligo, M., Morlacco, A., Motterle, G., Collavino, J., Barneschi, A. C., Moschini, M., and Moro, F. D. (2021). The impact of covid-19 pandemic on pornography habits: a global analysis of Google Trends. *International journal of impotence research*, 33(8):824–831.

Zentner, A. (2005). File sharing and international sales of copyrighted music: An empirical analysis with a panel of countries. *The BE Journal of Economic Analysis & Policy*, 5(1).

Zentner, A. (2006). Measuring the effect of file sharing on music purchases. *The Journal of Law and Economics*, 49(1):63–90.