IARIW 2024

IARIW 2024

Thursday, August 22 – Friday, August 30

Measuring the Depreciation of Intangibles using Search Volume Data

John Lourenze S Poquiz

(King's College London)

Paper prepared for the 38th IARIW General Conference
August 26-30, 2024

Session 4A-1, Innovation: Benefits, Costs and Measurement and Conceptual Challenges I

Time: Wednesday, August 28, 2024 [14:00- 15:30 GMT]

Measuring the depreciation of intangibles using Search Volume Data*

John Lourenze Poquiz†

July 30, 2024

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 for non-crisis years, particularly from 1996 to early 2008, and again from 2011 to the end of 2016.

JEL-Classification: C80, E01, E22, O34, O47.

Keywords: Intangible Asset, capital measurement, depreciation, capital stock, Google Trends.

---

*The author would like to thank his supervisors Mary O'Mahony and Martin Weale for their substantial contributions to this paper. Additionally, he extends thanks to Rachel Soloveichik for her valuable contributions to the analysis of the results. The author would also like to thank the participants of the 2024 ESCoE Conference on Economic Measurement and the participants of the 2023 Productivity Institute Internal Conference, especially the paper's discussant Diane Coyle, for all the helpful inputs that went into the crafting of this version.

†King's Business School, King's College London, Bush House, 30 Aldwych, London WC2B 4BG. Email: john_lourenze.poquiz@kcl.ac.uk.

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 & 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 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 & 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 for the Information and Communication industry would be higher for non-crisis years, particularly from 1996 to early 2008, and again from 2011 to the end of 2016.

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; J. 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 & Prucha (1996) proposes methodologies for measuring depreciation rates for Research and Development. The work outlined by J. 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 & Inklaar, 2022; Roth, 2019). We examine the impact of changing the asset life on estimates of capital stocks and total factor productivity.

This paper 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 3. We proceed by elaborating our empirical methodology in section 4. We describe our data in section 5 and discuss our results in section 6. We present the impact on capital stock estimates and productivity figures in section 7. We end with some applications and concluding remarks and ways forward.

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 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). Con-

[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

ceptually, 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 & Sampayo, 2014; Görzig & 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 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 & Schankerman, 1984; Nakamura, 2010; De Rassenfosse & Jaffe, 2017; Li, 2014; Li & 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 & Jaffe, 2017), the estimation of a profit model (Li & Hall, 2020), the estimation of a production function (Hall, 2007; Huang & Diewert, 2011), amortization (Lev & 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 paper is to estimate the obsolescence of software and creative originals. Table 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 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.

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 & 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:

$$V_{k,t} = V_{k,0} e^{-\delta t} \tag{1}$$

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 & 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 & 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 & 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:

$$V_{k,t}\left(\sum_i g_{i,k,t}, \gamma_k, \mu_{k,t}\right) = \bar{g}_{k,t}^{\beta_k} \gamma_k \mu_{k,t} \tag{2}$$

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 1 and 2 and rearranging expression yields:

$$\bar{g}_{k,t}^{\beta_k} = \frac{V_{k,0} e^{-\delta t}}{\gamma_k \mu_{k,t}}. \tag{3}$$

For simplicity, we can take the log of equation 3:

$$log(\bar{g}_{k,t}) = \frac{1}{\beta_k}\left[log(V_{k,0}) - \delta t - log(\gamma_k) - log(\mu_{k,t})\right]. \tag{4}$$

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.

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:

$$log(\bar{G}_{k,l,t}) = \delta\tau_t + \Gamma_k + \theta_l + \varepsilon_{k,r,t} \quad (5)$$

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.

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 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 & 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 (W. Martin & Pham, 2020) and is widely used in the trade literature.

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.

5.1 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 & Ozsoyoglu, 2016) and the prediction of movie ratings (Hsu et al., 2014; Oghina et al., 2012; Gogineni & 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.

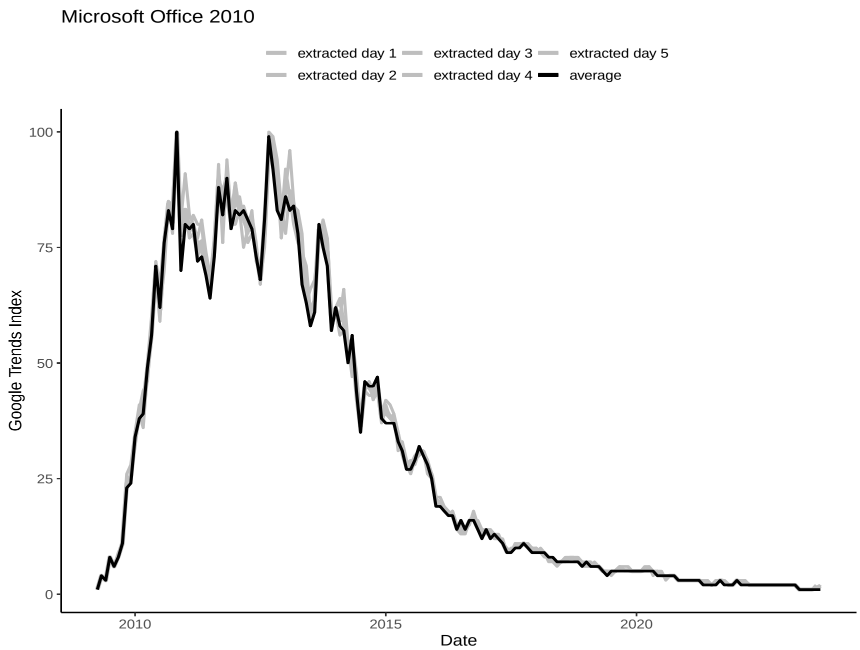
Users of Google Trends can filter the data in two ways: real-time and non-real-time. Within the real-time filtration, a random selection of queries from the preceding seven days is employed. Meanwhile, the non-real-time approach involves the use of a randomized subset of search queries extracted from 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 & Bhattacharjee, 2023; Ferrara & Simoni, 2022), as well as in predicting trends in tourism (Havranek & 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 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 & 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 & Shanbhogue, 2011; Carrière-Swallow & 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 & Shanbhogue, 2011), some 50 (McLaren & 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 & 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 stressing the Google IP address.

To illustrate, we show in figure 1 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.

5.2 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 C. For movies, we present some summary statistics from the data extracted from the IMDb website.

Table 2: 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

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 3: 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

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 4: 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

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: 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

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.

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 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.

6.1 Software

We present the results for software in table 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 6

	OLS (1)	PPML (2)	Negative Binomial (3)
δ	-0.006^{***} (0.0003)	-0.008^{***} (0.00001)	-0.009^{***} (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 7 that removing software titles with free versions, as well as those classified as open-source, does not make any changes to the estimates.

Table 7: Results for PPML model removing free and open-source software

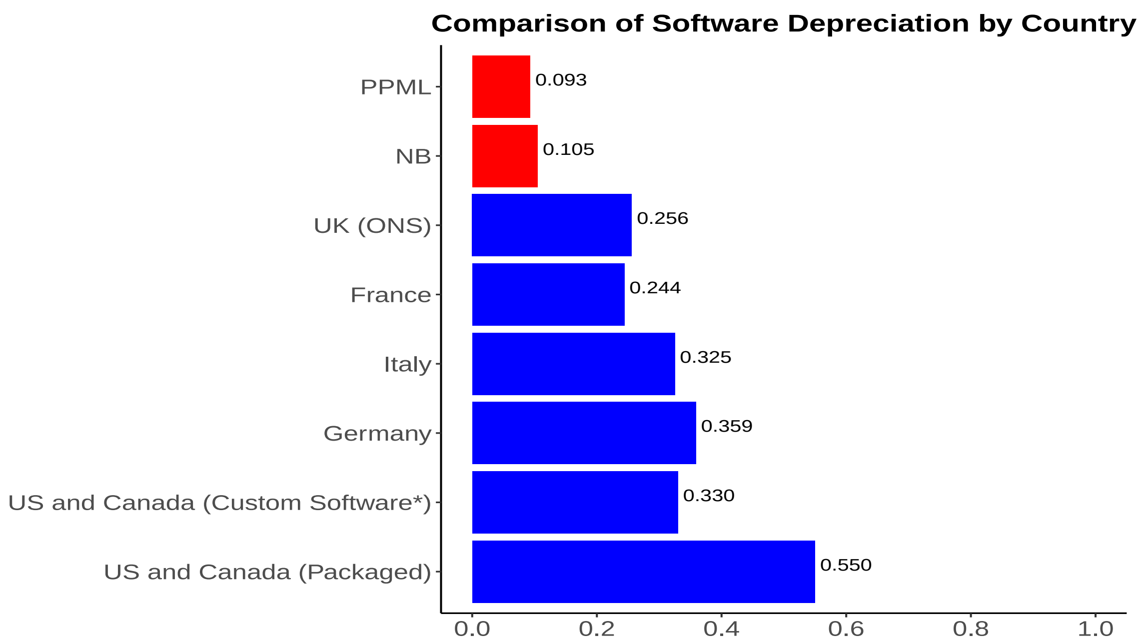
	All Software (1)	Removing those tagged "Free & Open Source" (2)	Removing those tagged Open Source only (3)
δ	-0.008^{***}	-0.008^{***}	-0.008^{***}
	(0.00001)	(0.00002)	(0.00001)
n	161,735	135,186	144,453
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

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 2). 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 2: 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 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 8: 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.

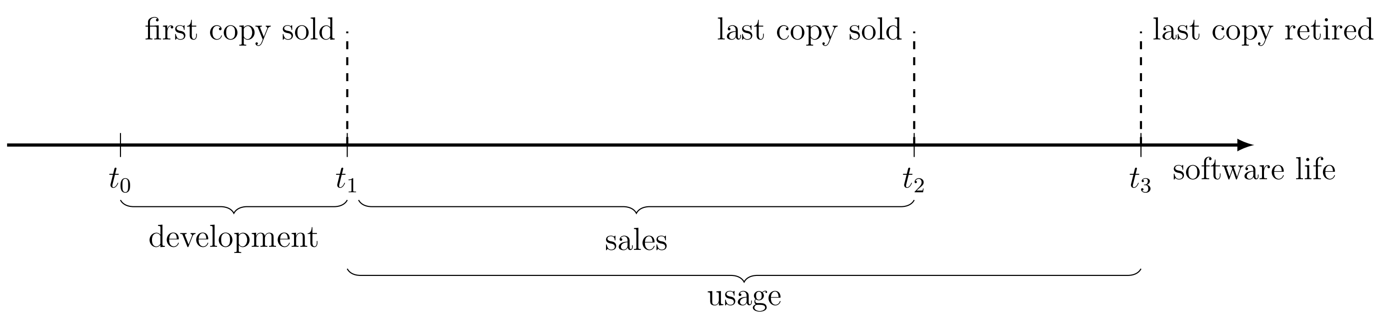
6.2 Original software

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 3 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 3



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 8. 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 9: 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 8. Our estimates for the asset life of original software in table 9 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.

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 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 10. We chose to present results for these countries because it would be easier to compare our estimates to official data later on.

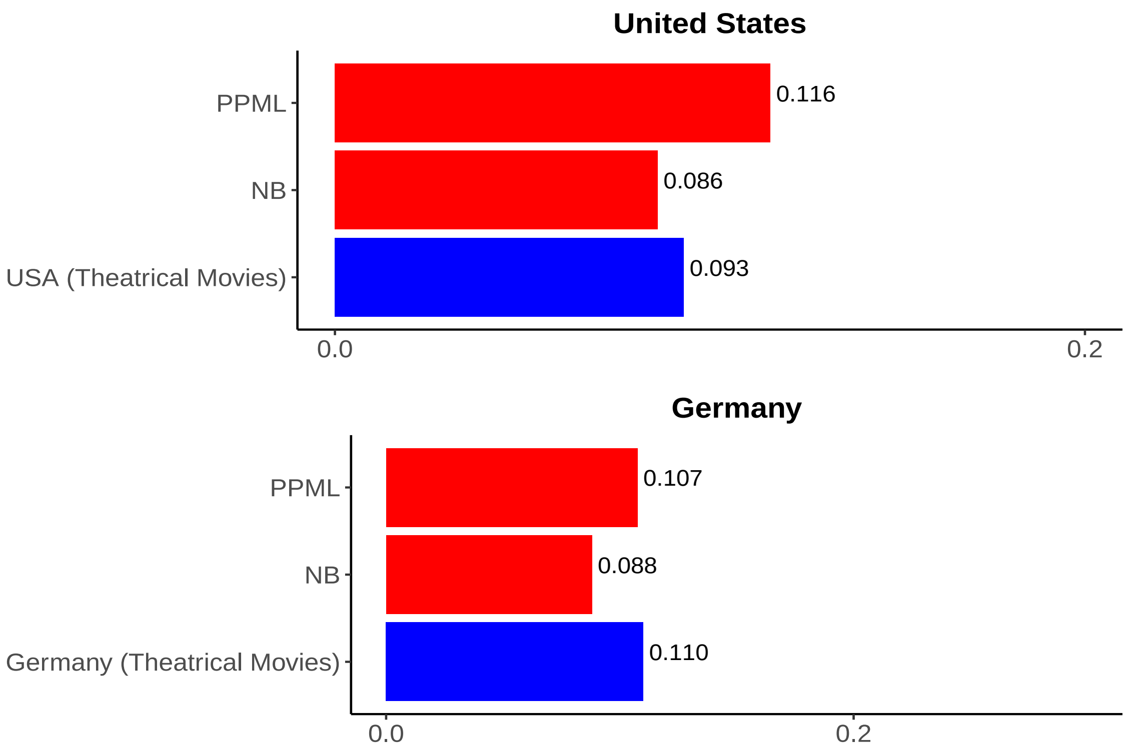
Table 10

	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 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 4: 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 10 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 1). We compare our estimates to the depreciation rates employed by

Table 11: 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

the US and Germany. We show the comparison in figure 4. Estimates for the UK, France, and Italy are shown in appendix table B.1.

For the US, our estimates are not materially different from those employed by BEA. Our estimated asset life for movies (see table 11) 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 12, 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 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.

Table 12: 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

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 & Schankerman, 1984; Nakamura, 2010; De Rassenfosse & Jaffe, 2017; Li, 2014; Li & Hall, 2020). There are limited avenues on how to test the relationship between GSV results and revenues that can be directly attributed to the asset. In the case of software, for instance, firms rarely break down revenue streams directly linked 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:

$$\bar{V}^B_{k,l,d,t} = \sum_{t=1}^{T} V^B_{k,l,d,t}. \tag{6}$$

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.

[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/

We estimate the model:

$$log(\bar{V}^B_{k,l,d,t}) = \tilde{\beta} \cdot log(\bar{G}_{k,l,d,t}) + \tilde{\delta}\tau_t + \Gamma_k + \theta_l + \omega_d + \tilde{\mu}_{k,r,d,t} \tag{7}$$

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*.

Table 13: 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 13. 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 8 incorporates lagged terms for GSV results, acknowledging the potential scenario wherein individuals frequently conduct internet searches for movies prior to viewing them:

$$log(\bar{V}^B_{k,l,d,t}) = \sum_{n=0}^{N} \left(\tilde{\beta}_{t+n} \cdot log(\bar{G}_{k,l,d,t-n})\right) + \tilde{\delta}\tau_t + \Gamma_k + \theta_l + \omega_d + \tilde{\mu}_{k,r,d,t} \tag{8}$$

Results shown in table 14 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. We also observe that as we increase the number of lags, the coefficient for the time trend gets smaller, implying lags contribute to the reduction of the omitted variable bias. In other words, incorporating lagged GSV results captures additional information that the time trend alone cannot, enhancing the model's explanatory power and reducing potential distortions caused by omitted variables. The relationship is also positive, suggesting that both variables move in the same direction. We believe that our results provide evidence of a direct and systematic relationship between Google Trends and movie revenues, which is a requirement in the framework that we discussed in section 3.

Table 14: 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

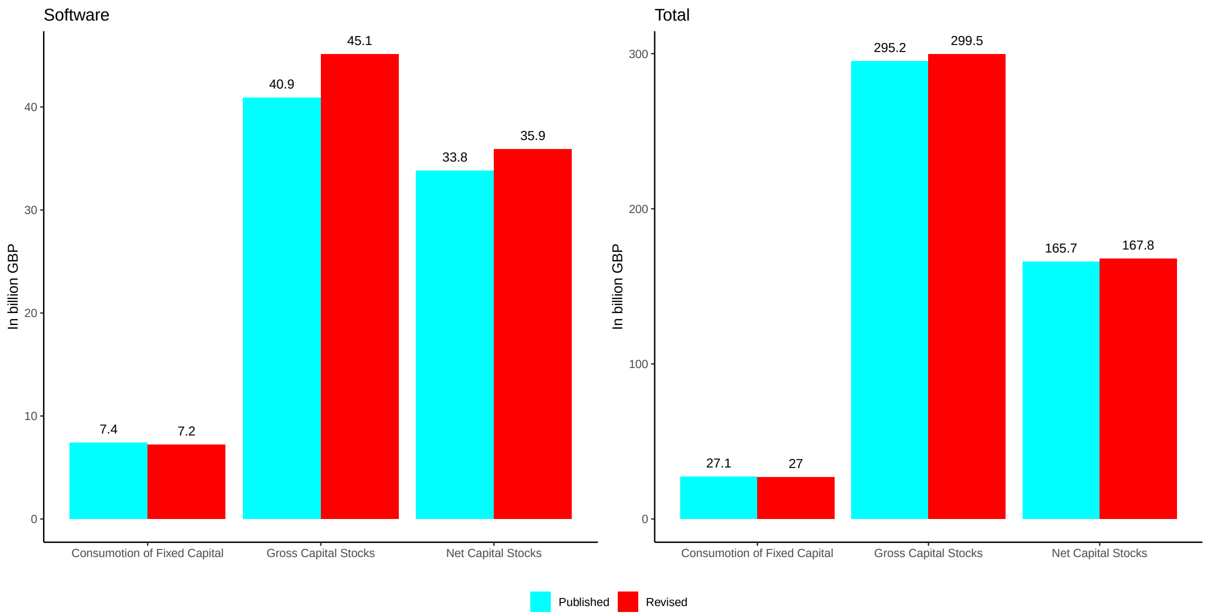
7 Impact on industry aggregates

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

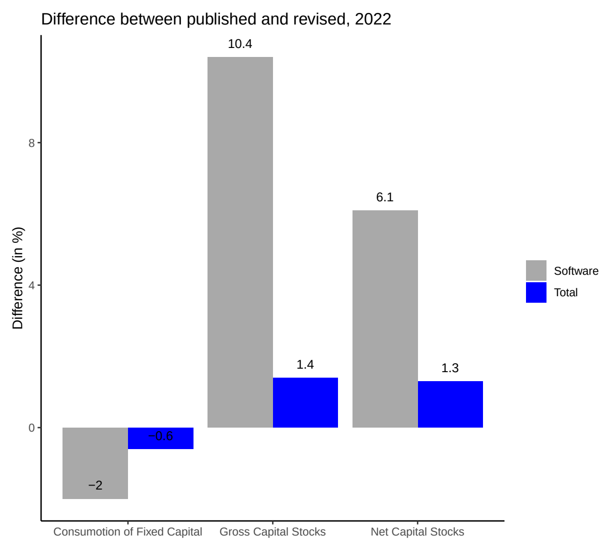
[5]https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/methodologies/capitalstocksuserguideuk last accessed: 15 February 2022

industry level.

Figure 5: Estimates for Sector J, 2022



(a) Levels



(b) Percent difference

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 for reproduction.

Our results show that for 2022 (see figure 5), 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.

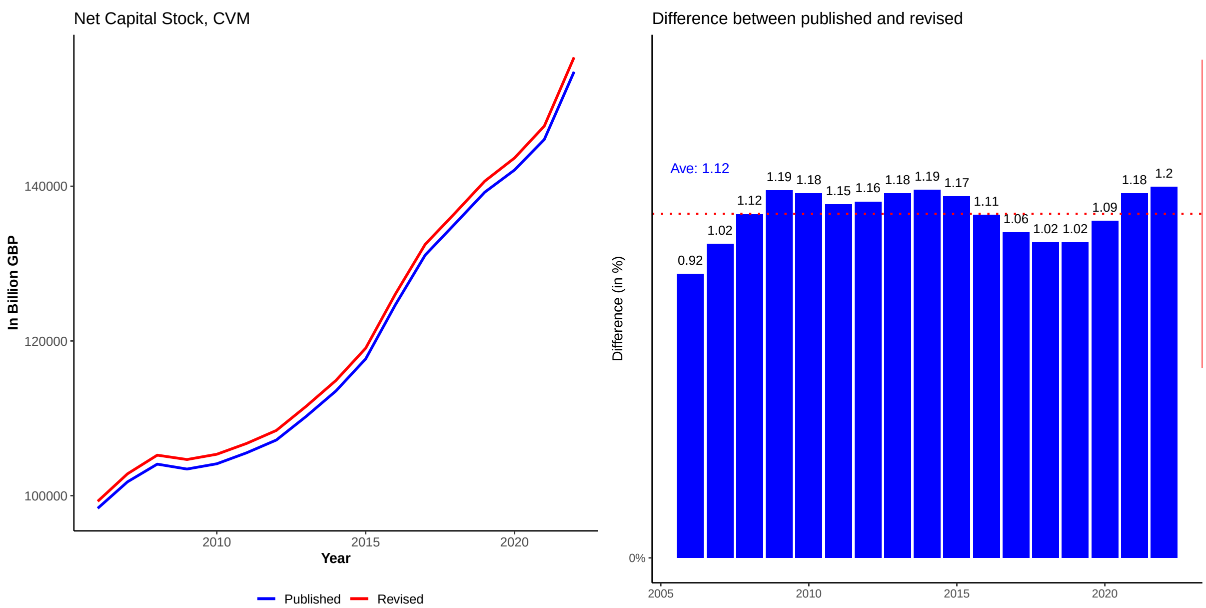
Not surprisingly, the difference is more modest for the total capital stocks. Our estimate of total gross capital stocks for sector J is only 1.4 percent higher than published, while the estimate for total net capital stocks is only 1.3 percent higher than what was published by the ONS.

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 6. 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.

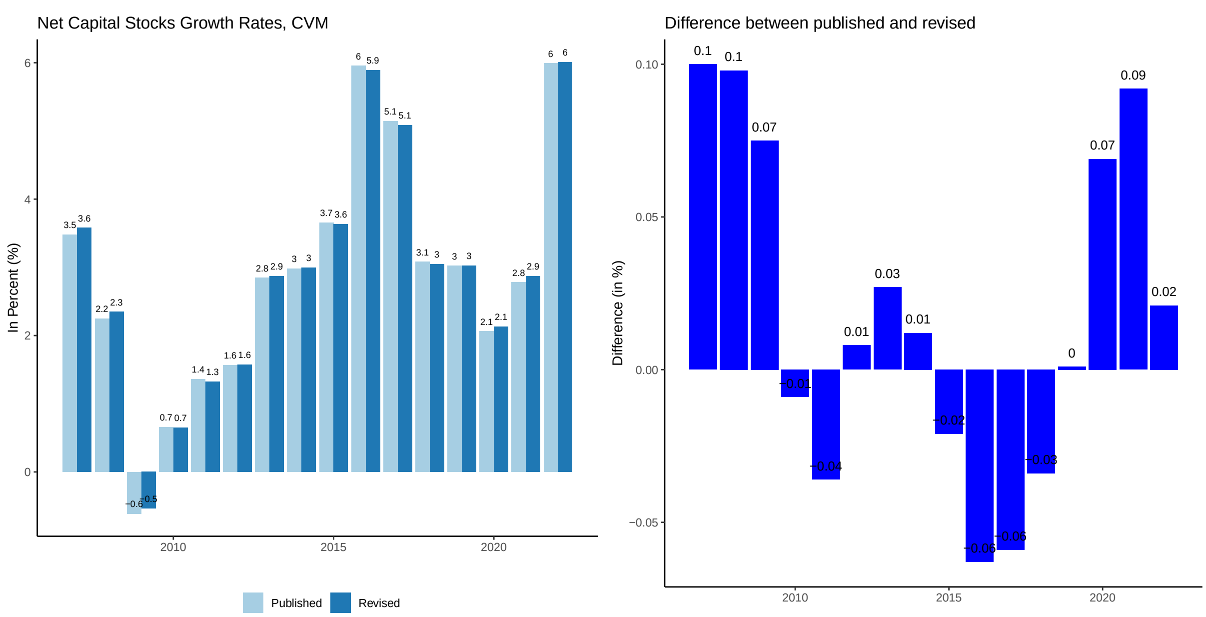
The impact on growth rates also exhibits variability across different periods. Some years demonstrate faster growth in net capital stocks according to our calculations, while in others, our growth rates are slower than the published data. This could have implications for Total Factor Productivity (TFP) 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.

[6]The ONS combines the two industries in their database. We are not able to separate the investments going to the two industries

Figure 6: Estimates for Sector J, 2006 to 2022



(a) Levels



(b) Growth rates

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 that difference in the growth rates revised vs published) in percentage points.

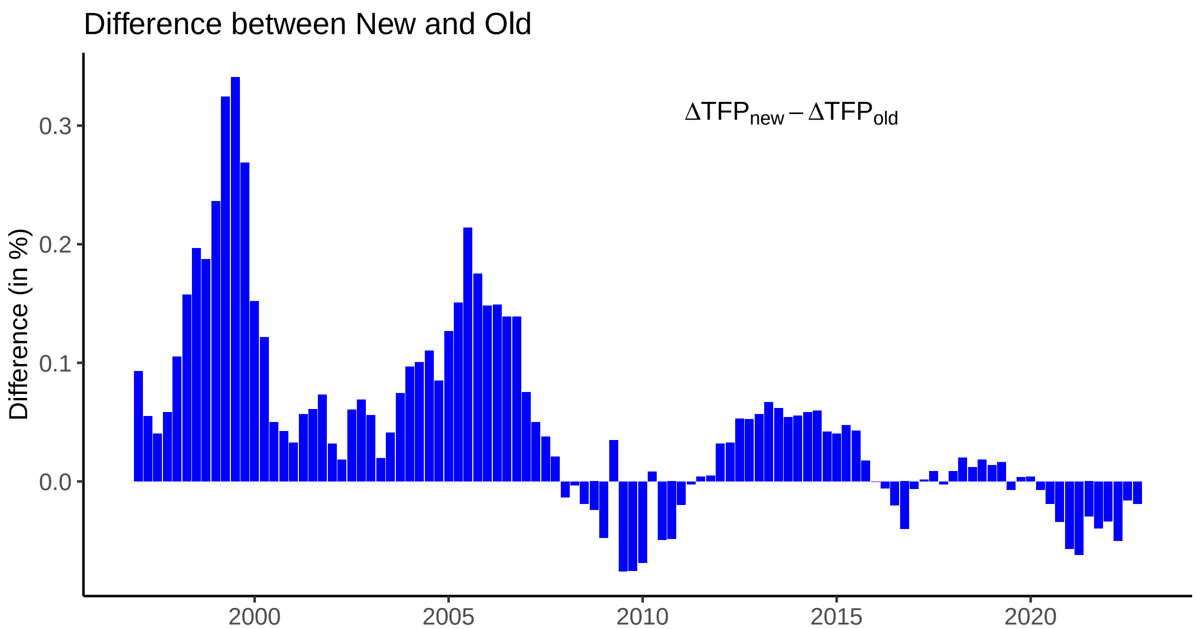
7.1 Impact on the productivity growth of the UK's Information and Communication industry

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 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 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.

[7]https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/methodologies/capitalstocksuserguideuk

Figure 7: 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 7 shows the difference between TFP using our estimated asset life ('New') and TFP growth calculated while maintaining the current assumptions of the ONS ('Old').

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 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 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 D.11).

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.

References

Ahouraian, M. (2021, 9 16). *The effects of a shrinking window for theatrical release exclusivity.* Retrieved from https://www.forbes.com/sites/forbesbusinesscouncil/2021/09/16/the-effects-of-a-shrinking-window-for-theatrical-release-exclusivity/

Awano, G., Franklin, M., Haskel, J., & Kastrinaki, Z. (2010). Investing in innovation: Findings from the UK investment in intangible assets survey. *NESTA Index Report (July)*.

Ballester, M., Garcia-Ayuso, M., & 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., & Sebastia-Barriel, M. (2014). The UK productivity puzzle. *Bank of England Quarterly Bulletin*, Q2.

Berlemann, M., & 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.

Bontadini, F., Corrado, C., Haskel, J., Iommi, M., & 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*.

Calderón, J. B. S., Rassier, D. G., et al. (2022). *Valuing the US data economy using machine learning and online job postings* (Tech. Rep.). Bureau of Economic Analysis.

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289–298.

Cervellin, G., Comelli, I., & 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.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, *88*, 2–9.

Chute, J. W., McCulla, S. H., & 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., & Iommi, M. (2014). Intangibles and industry productivity growth: Evidence from the EU.

Corrado, C., Haskel, J., Jona-Lasinio, C., & Iommi, M. (2022). Intangible capital and modern economies. *Journal of Economic Perspectives*, *36*(3), 3–28.

Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and us economic growth. *Review of income and wealth*, *55*(3), 661–685.

Del Rio, F., & Sampayo, A. (2014). Obsolescence and productivity. *Portuguese Economic Journal*, *13*, 195–216.

De Rassenfosse, G., & Jaffe, A. B. (2017). *Econometric evidence on the R&D depreciation rate* (Tech. Rep.). National Bureau of Economic Research.

Diewert, W. E., Wykoff, F. C., et al. (2006). Depreciation, deterioration and obsolescence when there is embodied or disembodied technical change. *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*.

Effenberger, M., Kronbichler, A., Shin, J. I., Mayer, G., Tilg, H., & Perco, P. (2020). Association of the covid-19 pandemic with internet search volumes: a Google Trendstm analysis. *International Journal of Infectious Diseases*, *95*, 192–197.

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google Trends. *Economic Inquiry*, *60*(2), 694–705.

Fernald, J. G., & Inklaar, R. (2022). The UK productivity "puzzle" in an international comparative perspective..

Ferrara, L., & Simoni, A. (2022). When are google data useful to nowcast GDP? an approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, 1–15.

Field, S., Franklin, M., et al. (2012). *Results from the second survey of investment in intangible assets, 2010.* Retrieved from https://webarchive.nationalarchives.gov.uk/ukgwa/20160105194600/http:/www.ons.gov.uk/ons/rel/icp/results-from-the-second-survey-of-investment-in-intangible-assets/2010/art-results-from-the-second-survey-of-investment-in-intangible-assets.html

Gogineni, S., & Pimpalshende, A. (2020). Predicting imdb movie rating using deep learning. In *2020 5th international conference on communication and electronics systems (icces)* (pp. 1139–1144).

Goodridge, P., Haskel, J., & Wallis, G. (2013a). Can intangible investment explain the UK productivity puzzle? *National Institute Economic Review*, *224*, R48–R58.

Goodridge, P., Haskel, J., & Wallis, G. (2013b). Film, television & radio, books, music and art: estimating UK investment in artistic originals. *Measuring the Contribution of Knowledge*, 67.

Goodridge, P., Haskel, J., & Wallis, G. (2018). Accounting for the UK productivity puzzle: a decomposition and predictions. *Economica*, *85*(339), 581–605.

Görzig, B., & Gornig, M. (2015). *The assessment of depreciation in the case of intangible assets* (Tech. Rep.). SPINTAN Working Paper Series No.

Hall, B. H. (2007). *Measuring the returns to R&D: The depreciation problem.* National Bureau of Economic Research Cambridge, Mass., USA.

Hamulka, J., Jeruszka-Bielak, M., Górnicka, M., Drywień, M. E., & 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.

Harish, B., Kumar, K., & Darshan, H. (2019). Sentiment analysis on imdb movie reviews using hybrid feature extraction method.

Haskel, J., Corrado, C., Jona-Lasinio, C., & Iommi, M. (2018). Intangible investment in the EU and US before and since the great recession and its contribution to productivity growth.

Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics*, *27*(1), 129–148.

Hsu, P.-Y., Shen, Y.-H., & 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* (pp. 444–453).

Huang, N., & 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.

Kohns, D., & Bhattacharjee, A. (2023). Nowcasting growth using Google Trends data: A bayesian structural time series model. *International Journal of Forecasting*, *39*(3), 1384–1412.

Lev, B., & 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.* Carnegie Mellon University.

Li, W. C., & Hall, B. H. (2020). Depreciation of business R&D capital. *Review of Income and Wealth*, *66*(1), 161–180.

Martin, J. (2019). Measuring the other half: new measures of intangible investment from the ONS. *National Institute Economic Review*, *249*, R17–R29.

Martin, W., & 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., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*(2011), Q2.

Nadiri, M. I., & 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. I. (2010). Intangible assets and national income accounting. *Review of Income and Wealth*, *56*, S135–S155.

OECD. (2009). Measuring capital, Second Edition. *OECD Publishing,*. Retrieved from https://doi.org/10.1787/9789264068476-en

OECD. (2010). OECD Handbook on Deriving Capital Measures of Intellectual property products,. *OECD Publishing,*. Retrieved from https://doi.org/10.1787/9789264079205-en

Office for National Statistics. (2007). *The ons productivity handbook: a statistical overview and guide.* Palgrave Macmillan.

Oghina, A., Breuss, M., Tsagkias, M., & 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* (pp. 503–507).

Organisation for Economic Co-operation and Development. (2021). *OECD productivity statistics methodological notes.* Palgrave Macmillan.

O'Mahony, M., & 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., & 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* (pp. 73–88). University of Chicago Press.

Pionnier, P.-A., Zinni, M. B., & Baret, K. (2023). Sensitivity of capital and mfp measurement to asset depreciation patterns and initial capital stock estimates.

Riley, R., Bondibene, C. R., & Young, G. (2015). The UK productivity puzzle 2008-13: evidence from british businesses.

Riley, R., Rincon-Aznar, A., & Samek, L. (2018). Below the aggregate: a sectoral account of the UK productivity puzzle. *ESCoE Discussion Papers*, *6*.

Rincon-Aznar, A., Riley, R., & Young, G. (2017). Academic review of asset lives in the UK. *Discussion Papers*, *474*.

Rogers, S. (2016). *What is Google Trends data — and what does it mean?* Published in Medium. Retrieved from https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8

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.

Shaukat, Z., Zulfiqar, A. A., Xiao, C., Azeem, M., & Mahmood, T. (2020). Sentiment analysis on imdb using lexicon and neural networks. *SN Applied Sciences*, *2*, 1–10.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, *88*(4), 641–658.

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., 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).

Topal, K., & 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)* (pp. 1170–1176).

Van Criekingen, K., Bloch, C., & Eklund, C. (2022). Measuring intangible assets—a review of the state of the art. *Journal of Economic Surveys*, *36*(5), 1539–1558.

Warusawitharana, M. (2015). Research and development, profits, and firm value: A structural estimation. *Quantitative Economics*, *6*(2), 531–565.

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., ... 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.

Appendix

A Capital Measurement

Consider the classical law of motion for capital:

$$K_{k,t} = K_{k,t-1}(1-\delta_k) + I_{k,t} \tag{A.1}$$

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 A.1, 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:

$$CFC_{k,t} = K_{k,t-1} - K_{k,t} = \delta_k K_{k,t-1} + I_{k,t}. \tag{A.2}$$

While the above expression in equation A.2 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 A.1 from the beginning period $t-1$:

$$K_{k,t} = \sum_{i=0}^{\infty}(1-\delta_k)^i + I_{k,t-(i+1)} \tag{A.3}$$

To implement equation A.3, 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 & Wesselhöft (2014):

$$K_{k,t} = (1-\delta_k)^{t-1}\bar{K_k} + \sum_{i=0}^{t-1}(1-\delta_k)^i I_{k,t-(i+1)}. \tag{A.4}$$

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).

B Additional results

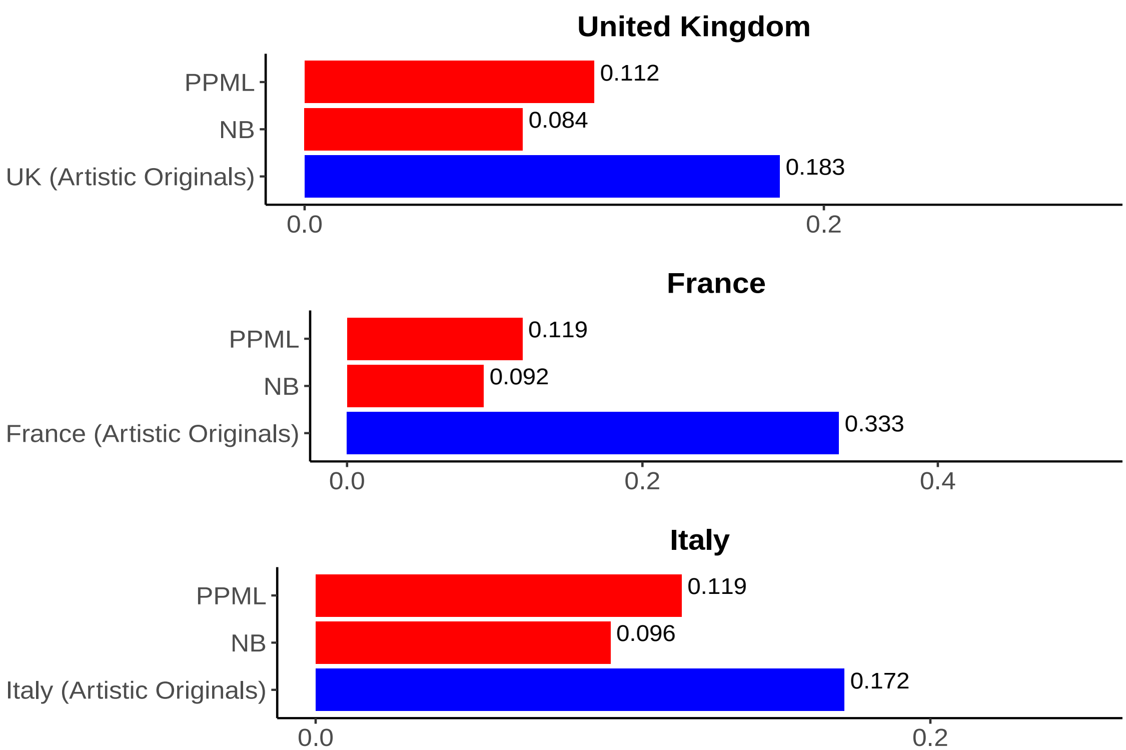
Table B.1

	OLS (1)	PPML (2)	NB (3)
δ	-0.109^{***} (0.002)	-0.092^{***} (0.0001)	-0.065^{***} (0.001)

Table B.2: Results for Negative Binomial model removing free and open-source software

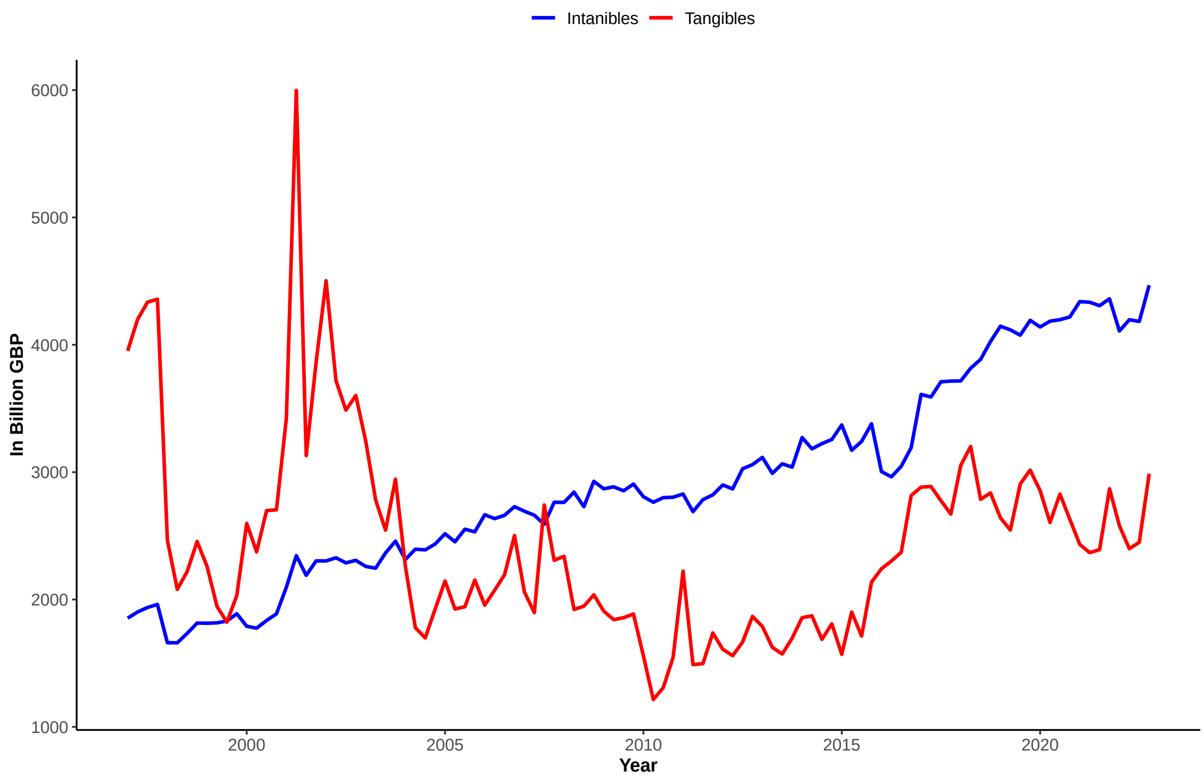
	All (1)	No Free & Open Source (2)	No Open Source Only (3)
δ	-0.009^{***} (0.0001)	-0.009^{***} (0.0001)	-0.009^{***} (0.0001)
n	161,735	135,186	144,453
Keyword FE	Yes	Yes	Yes
Release Date FE	Yes	Yes	Yes

Figure B.1: Comparing the annual depreciation rates estimates for movies with those from other countries



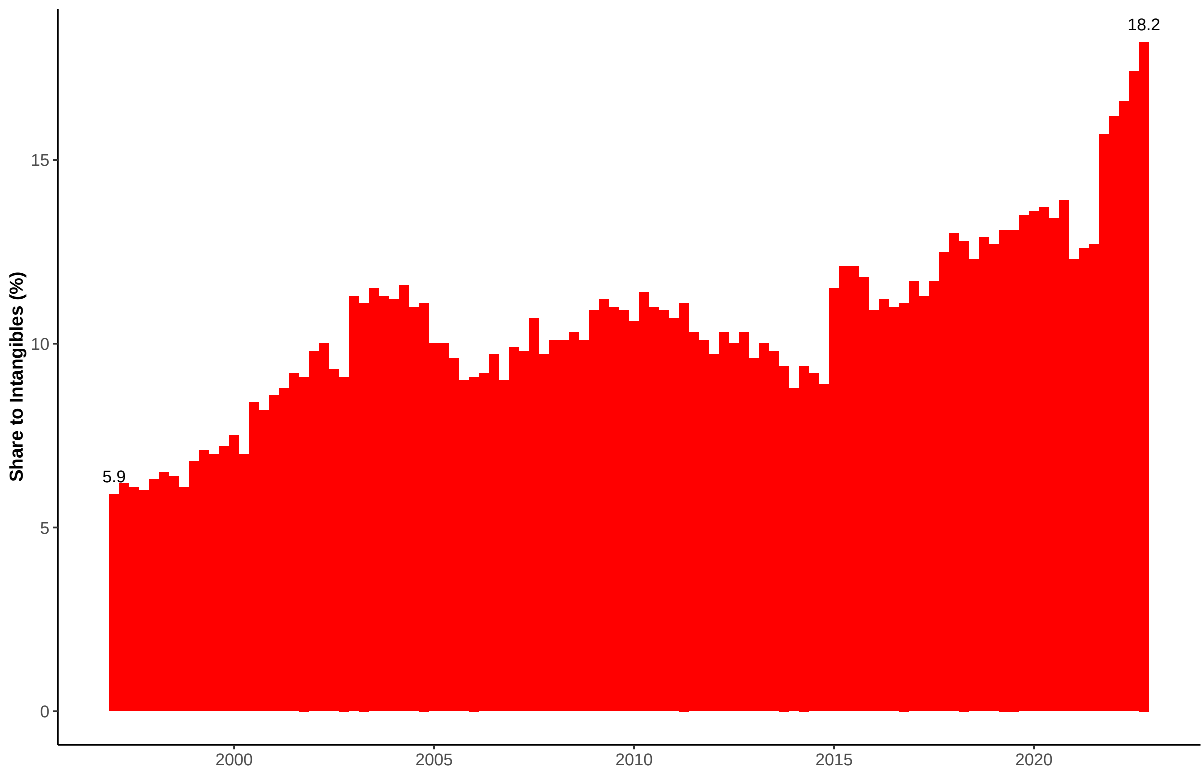
Note: The figure compares results from PPML and negative binomial (NB) model with the annual depreciation rates for theatrical movies employed by a select set of countries. The monthly estimates from table 10 were annualized assuming a geometric pattern. The depreciation rates employed by other countries were sourced from Pionnier et al. (2023).

Figure 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 B.3: Share of own-account software to total intangibles for sector J from 1996 to 2022, current price estimates

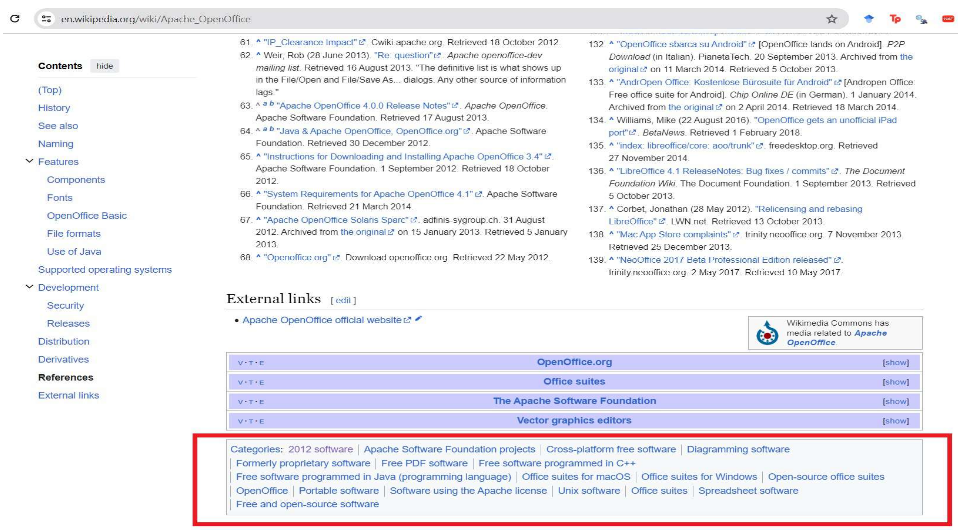


Note: Figure shows the share of own-account software investments to total gross fixed capital formation on itangibles for sector J at current price. In this figure, intangible investments only include intellectual property products (IPP) capitalized in the National Accounts, namely purchased and own-account 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

C Classifying software into categories

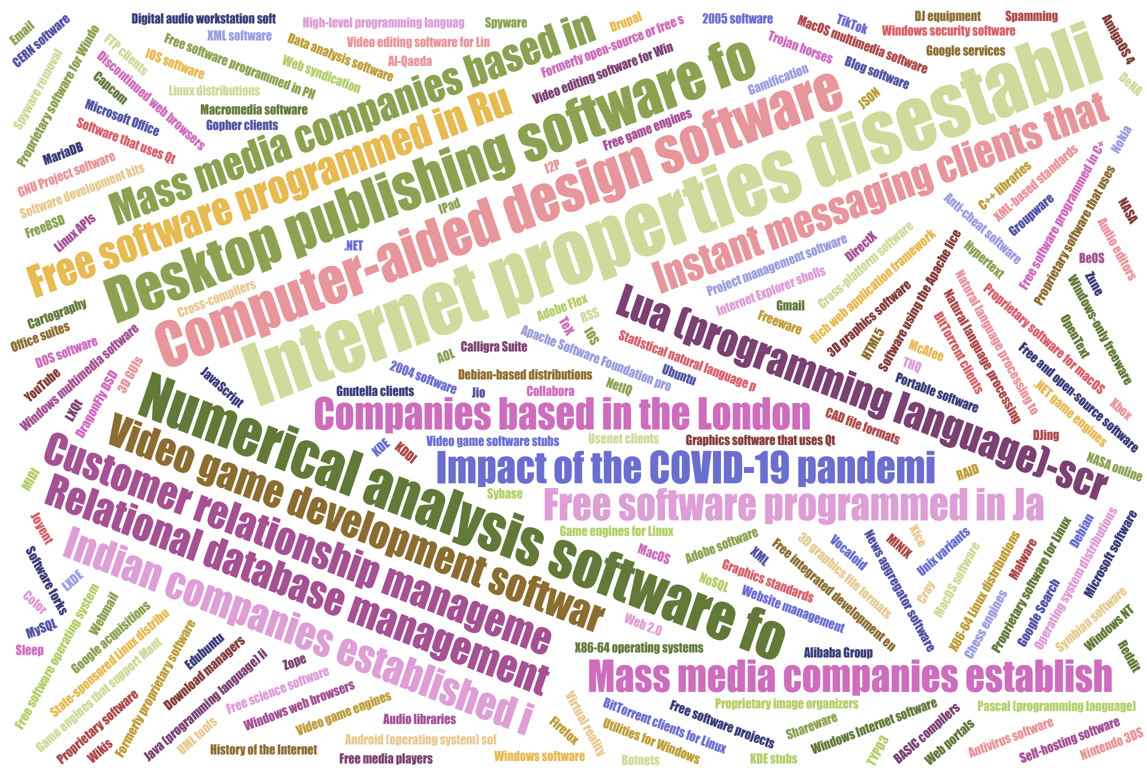
To classify software into categories, we exploit the categories tag at the bottom of the Wikipedia page of each software title. See example below:

Figure C.4: 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 C.5 as word cloud for the top 1,000 most common tags.

Figure C.5: Word cloud of the category tags from software Wikipedia entries

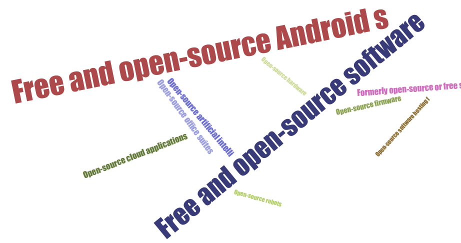


Our interest is to see whether the estimates would change if we remove “free-to-use” and “open-source” software. As such, we generated broad categories by identifying tags with the words “free” and “open source” and lumping them into a single broad category. We also created a third category that identifies tags containing the word “open source” but does not contain the word “free”. We show the word cloud for these tags in figure C.6.

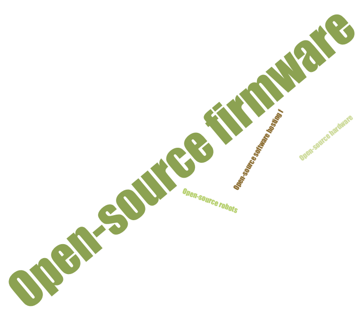
To extend our descriptive analysis, we also generated other broad categories using the same method. We identified the operating system with which these software are compatible and the type of software they can be classified into. We show the word cloud for the tags identified for each category in figures C.7 and C.8, respectively.

Figure C.6: Free and open-source software

(a) Free software



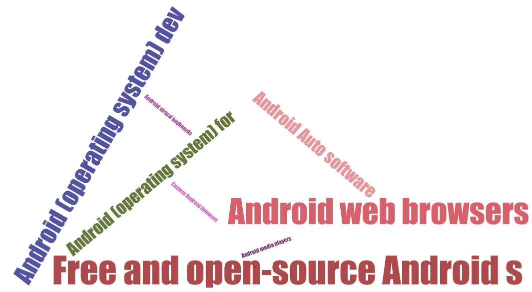
(b) Open-source software



(c) Strictly open-source software

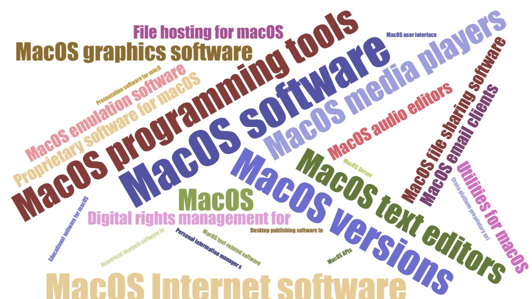
Figure C.7: By operating system

(a) Windows

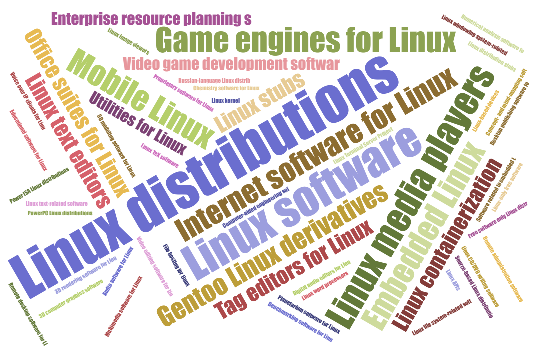


(b) Android

(c) iOS

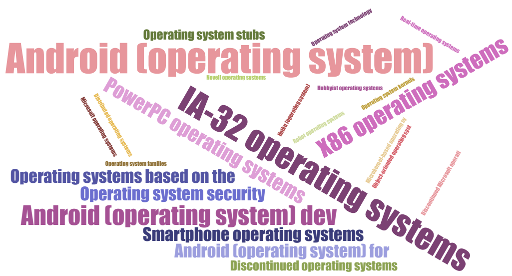


(d) MacOS



(e) Linux

Figure C.8: By type

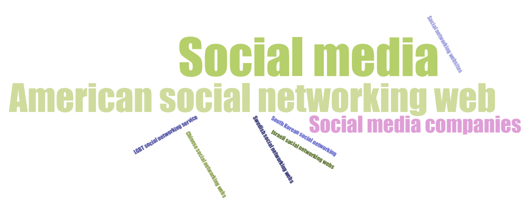


(a) Operating system

(b) Media

(c) Office

(d) Web browser



(e) Social media and social network

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:

$$u_{i,t} = r_t p_{i,t} + \delta_i p_{i,t} + (p_{i,t} - p_{i,t-1}) \tag{D.1}$$

where $u_{i,t}$ is the user cost for asset i 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:

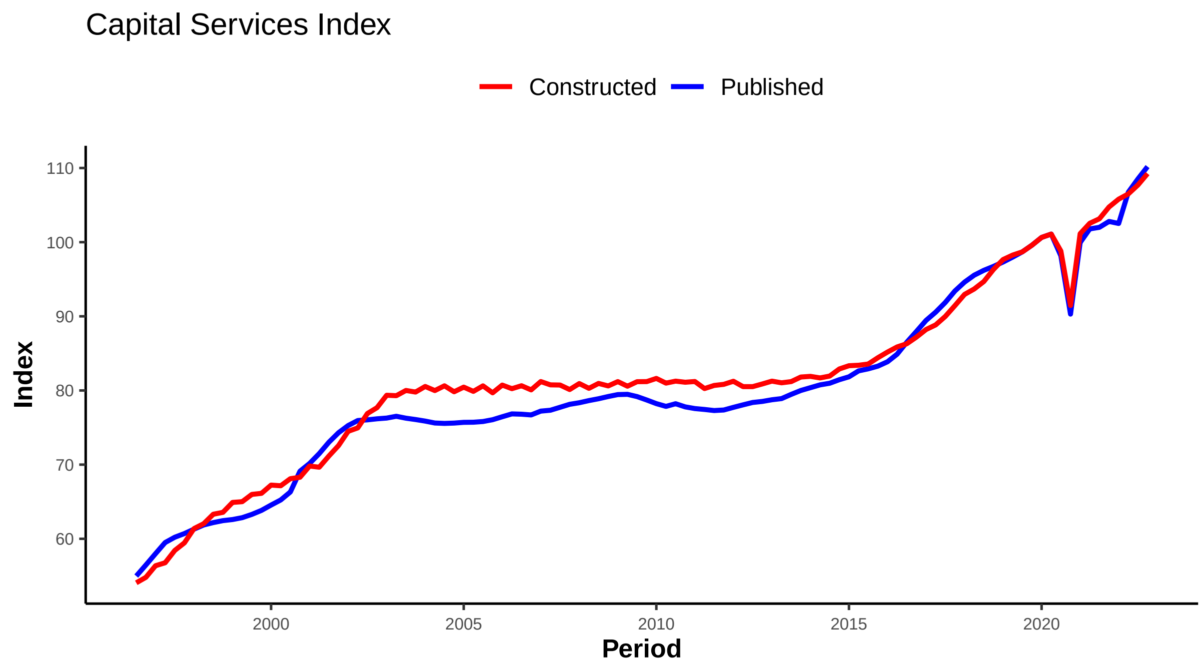
$$K_{i,t} = k_{i,t} \times \frac{k_{i,t} \times u_{i,t}}{\sum_i^I k_{i,t} \times u_{i,t}} \tag{D.2}$$

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 D.9: 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).

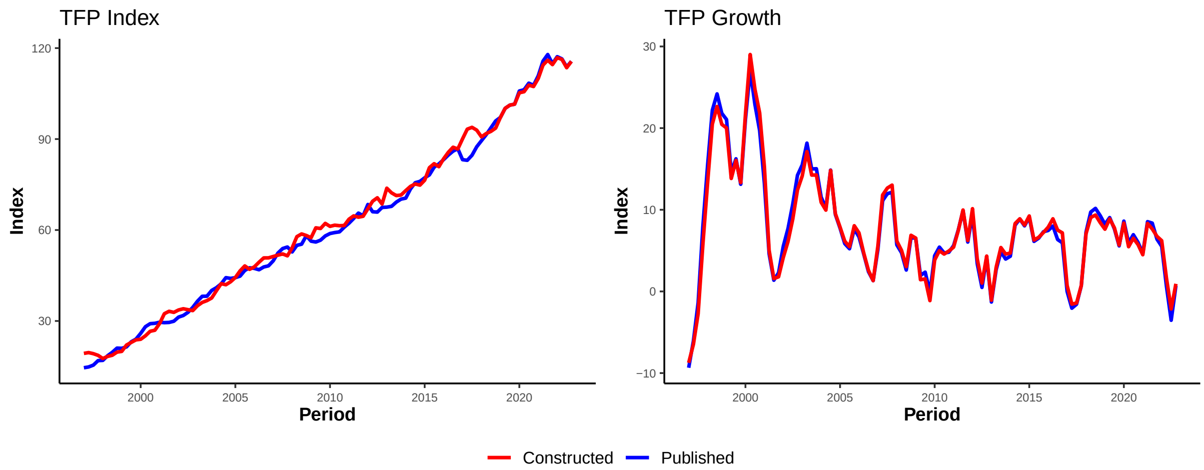
We proceed by calculating TFP using the standard growth accounting formula.

$$\Delta log(TFP_t) = \Delta log(Y_t) - v_t^l \Delta log(L_t) - v_t^k \Delta log(K_t) \tag{D.3}$$

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 D.3 (see figure D.10). We do not find any systematic difference between our constructed index and the TFP index and growth rates published by the ONS.

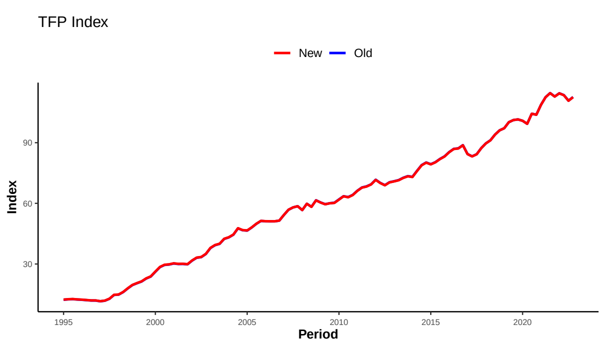
Figure D.10: Constructed versus published TFP



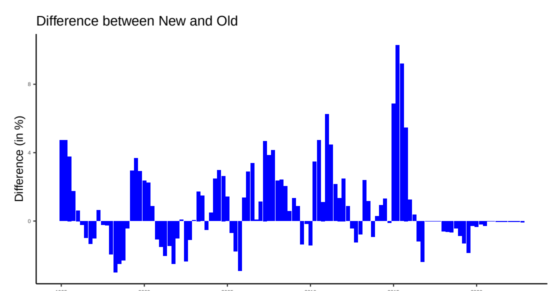
Note: The figure shows TFP we estimated (red line) and the TFP published by the ONS (blue line).

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 D.11.

Figure D.11: TFP Index



(a) Levels



(b) Difference

Figure D.12: TFP growth rates