Measuring the value of free digital goods*

John Lourenze Poquiz†

September 22, 2022

Abstract

The goal of this research is to estimate and examine the value derived by households from the utilization of free digital goods. For this exercise, we estimate the gross value from the consumption of three forms of free digital goods: videoconferencing, personal email, and online news. As our measurement strategy, we employ the prices of "premium" or paid internet goods as proxy for the value from their free counterparts. We also use hedonic regression in order to extract the value of the 'free component' of these goods and untangle them from the value of the premium-exclusive components. Our estimates show that in 2020, the aggregate gross value derived by households from the consumption of the three digital services was between £6.1 billion and £22.7 billion. We also observe that the value derived by households from consuming these goods is growing much faster than aggregate household consumption. Our estimate show that in 2020, the initial year of the COVID pandemic, real household final consumption decline would have been 0.07 to 0.12 percentage points slower had the value of the three digital goods been incorporated in the estimates.

JEL-Classification: C13, C82, D60, E01, O47.

Keywords: Welfare measurement, digital economy, free goods, National Accounts, hedonic regression.

---

*The author would like to thank his supervisors Mary O'Mahony and Martin Weale to their substantial contributions to this paper. The author would also like to the participants of the King's Business School PhD Research Symposium 2021 and the Productivity Institute Brownbag Seminar Series for providing their comments.

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

1 Introduction

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

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

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

[1]The exceptions being the estimation of the value of ownership of dwellings, own-account production and government services, all of which are valued at cost.

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

to pay for the characteristics included in the specification.

Our estimates show that in 2020, the aggregate gross value derived by households from the consumption of the three forms of digital services was between £6.1 billion and £22.7 billion. This is around 1.1 percent of the UK's household final consumption. We also observe that the value derived by households from consuming these goods is growing much faster than aggregate household consumption. Our estimate show that in 2020, the initial year of the COVID pandemic, the real household final consumption decline would have been 0.07 to 0.12 percentage points slower had the value of the three digital goods been incorporated in the estimates. This tells us that the availability of free internet services was partially able to reduce welfare loss as a result of the lockdown.

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

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

The contingent valuation studies are aimed at identifying how much individuals value free digital goods by asking them how much are they willing to be compensated to give up the said good. These include the study by Corrigan et al. (2018), Brynjolfsson et al. (2019b,a), and Jamison and Wang (2021), where they employed randomized binary-choice experiments to generate demand curves for free digital goods. Nguyen and Coyle (2020) used an online survey to derive how much individuals value free digital goods by asking them much they are willing to be paid to give up the said good. The total cost approach, meanwhile, employs the cost of producing free goods as a representation of the value consumers derived from them. Nakamura et al. (2017) generated estimates on advertising-supported and marketing-supported internet media based on their cost of production. Similarly, Van Elp and Mushkudiani (2019) produced estimates of the aggregate value from the consumption of free digital goods for the Netherlands. One thing they all have in common is the goal of generating estimates of the aggregate value of free goods, as well as an augment measure of production (expanded GDP) that takes into account the imputed value of the said services.

While the total cost approach is consistent with how the National Accounts measure output from other non-market goods (i.e. government services and output from non-profit institution serving households), its main drawback stems from its limited ability in representing welfare gains from free digital goods. If the marginal cost of producing free digital goods is (close to) zero, then the value derived from an additional user may not be reflected in the total cost. Moreover, this approach be applied only to internet goods financed by either advertising or marketing marketing expenditures. As such, the value of goods made available through other business models, such as the "Freemium" model, where free versions of goods and services are offered in order to entice a portion of the market to avail of the paid version, cannot be estimated using this approach. On the other hand, even though contingent valuation studies are advantageous in terms of capturing the incremental level welfare received from an additional user of the service, the approach introduces inconsistencies with the accounting principles of the System of National Accounts. This approach values free digital goods based on the individual's willingness to accept (WTA). In contrast, the System of National Accounts (SNA) adheres to the willingness to pay (WTP) principle when valuing goods. Hanemann (1991) showed that WTA and WTP are not equal if there are no close substitutes. As such, estimates employing this approach may not be comparable with other aggregates compiled using the SNA framework such as GDP and household consumption.

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

For our second contribution, we provide a source of external validity for the estimates produced by other scholars working on this topic. We provide two levels of comparison. First, we compare the price estimates that we generated to valuation produced by some of the studies employing the contingent valuation approach. Second, we compare our estimated impact of free digital goods to GDP growth rates to the estimated impact from other studies. By doing this, we are able to provide insight on how the choice of methodology would have affect the estimates.

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

The outline of this paper is as follows. In section 2, we discuss and synthesize both the theoretical and empirical literature on the measurement of free internet goods. In section 3, we discuss the theoretical underpinnings of the the measurement strategy. In section 4, we detail our estimation strategy. In section 5, we show and describe our data. We then discuss our results and preliminary estimates in section 6. We end this paper this paper concluding remarks and our strategies moving forward.

2 Literature on Free Goods

In this section, we summarize and discuss the literature on free digital goods. We divide the literature in two broad categories: those focused on theory and those centered on empirics. For studies having both a theory component and an empirical component (such as papers of Bridgman et al. (2018) and Brynjolfsson et al. (2019a)), we separate the discussion of the theoretical and empirical aspect of their papers, accordingly. We end this section with a synthesis of the literature.

2.1 Theoretical Literature

A number of scholars have attempted to develop analytic models explaining how value can be derived from free goods. These models are important in constructing theory-consistent instruments that allow for the measurement of the goods in question.

Bridgman et al. (2018) developed a model that identifies returns from market wages as valid proxies for non-market leisure. His model aims to identify an appropriate means on how to impute the value of free leisure time.

In his framework, Bridgman et al. (2018) stated that households allocate their time to three activities: household work, H_t^h, market work, H_t^m, and leisure, H_t^l. The representative household's preference can be expressed as,

$$\sum_t \beta^t [u(C_t^h, C_t^m, C_t^l + l_t - v^m(H_t^m) - v^h(H_t^h)] \tag{1}$$

which implies that the household's period t utility is derived from the consumption of market goods, C_t^m, home consumption, C_t^h, market leisure, C_t^l, and non-market leisure, l_t. The household also receives disutilty from home work $v^h(H_t^h)$, as well as from market work $v^m(H_t^m)$, where it earns a wage w^m. Moreover, leisure is generated through a production technology $l_t = F^l(K_t^l, H_t^l)$ that combines leisure capital, K_t^l, and leisure time, H_t^l. Meanwhile, home consumption is produced using the technology $C_t^h = F^h(K_t^h, H_t^h + H_t^s)$, where home production capital, K_t^h, is combined with the household's time allocated to home work, H_t^h, and the labour hours of hired workers, H_t^s. Whenever the household hires a worker for home production, the workers is paid a wage W_t^s.

The representative household maximizes its utility subject to the budget constraint:

$$C_t^m + p_t^l C_t^l + \sum_{j \in \{m,h,l\}} X_t^j + B_{t+1} + W_t^s H_t^s = W_t^m H_t^m, R_t^m K_t^m + B_t(1 + R_t^b) \qquad (2)$$

where, p^l is the price of market leisure; $\sum_{j \in \{m,h,l\}} X_t^j$ represents the total investments outlays for capital used in home production j, market production m, and leisure ll; K_t^m is the capital stock for market production, B_t and B_{t+1} represent bonds for period t and $t+1$, respectively; and R_t^m and R_t^b and are the rates of return for market capital and bonds, respectively.

Bridgman et al. (2018) show that in equilibrium, $w^l = w^s - v^{h'(H_t^h)}/u_m(t)$. From this expression, we see that the imputed returns from leisure w^l is equal to the market wage for household work w^s, less the marginal disutility from labour. The household allocates time in such a way that the returns from leisure time and returns from household work are equal. This result implies that market wage can be used to impute for the value of non-market leisure.

In a separate paper, Goolsbee and Klenow (2006) shows that it is possible to impute the value derived from digital goods by using the time spent on the internet. They argue that as wages rise, people should spend less time on the internet. As such, they advocate the utilization of time-use data for the measurement of welfare gains from the digital goods.

Using index number theory, Brynjolfsson et al. (2019a, 2020) show that the money individuals are willing to receive to give up free digital goods can be used to generate estimates of the reservation price for these goods. They argue that it is possible to impute for household expenditures that takes into consideration the welfare gains from free digital goods, though willingness to accept function. Brynjolfsson et al. (2019a) also demonstrates that accounting for the business cost from advertising alone would not account for the benefits that advertising-supported products produce, since the choice in the level of cost are motivated by profit maximization considerations in a monopolistic competition setup.

In a paper presented to the 2021 International Association for Research in Income and Wealth (IARIW) General Conference, Schreyer (2021) argued that households are not only consumers of free digital service, rather, they are also producers of these services. He argued in his paper that free digital services are generated by households by means of own-account production. For instance, when an individual opens a social networking site such a Facebook for his own leisure, this individual is producing leisure services that he himself consumes. In his analytic framework, Schreyer (2021) incorporated the proposal of Brynjolfsson et al. (2019a, 2020) by expressing the individual's willingness to forgo free digital services as its cost of own-account production. He developed a cost index, which also incorporates the number of network users as an driver of quality change.

A set of models based on the Lancaster (1966) framework was developed by Hulten and Nakamura (2017) and was later expanded upon by Hulten and Nakamura (2019). They argue that households are not passive receivers of utility, rather they generate utility themselves through a production technology that transforms output into utility, which can be expressed as,

$$U(C_t) = U(\beta Q_t). \tag{3}$$

Based on the Lancaster (1966) model, households receive utility $U(*)$ from the consumption of a set of characteristics contained in the vector, C_t, which is functionally connected to a vector of output, Q_t, as shown in equation 3. In this framework, the set of parameters, β, defines the household's technology that transforms output into a bundle of characteristics that generates utility. In typical utility models, the value of $\beta = 1$.

Hulten and Nakamura (2017) argues that conventional growth accounting frameworks characterize productivity as *resource-saving* technical change that causes a shift in the production function. They argue that "information" allows households to increase utility without increasing output. They describe this as *output-saving* technical change. Hulten and Nakamura (2017) said that an alternative measure of GDP can be constructed by including the household's willingness to pay for the output-saving information.

Consider the conventional growth accounting framework where output per worker, Q_t/L_t at time t is expressed as a function of intangible capital per worker, R_t/L_t, tangible information and communication technology (ICT) capital per worker, E_t/L_t, non-ICT capital (which includes tangible and intangible capital) per worker, S_t/L_t, and an exogenous resource-saving technical change $e^{\lambda t}$:

$$Q_t/L_t = e^{\lambda t}(R_t/L_t)^{\alpha}(E_t/L_t)^{\delta}(S_t/L_t)^{\pi} \tag{4}$$

where α, δ, and π are the respective elasticities corresponding to each of the factors. Expressed in growth rates, equation 4 becomes:

$$q - l = \lambda + \alpha(r - l) + \delta(e - l) + \pi(s - l) \tag{5}$$

Hulten and Nakamura (2017) suggest that growth accounting should be extended to consider the utility function, rather than the production function. Moreover, they argue that the utility function should include a consumption technology:

$$U(C_t/N_t) = m(C_t/N_t) = me^{\theta t}[\rho_t(1-\sigma_t)(Q_t/L_t)] \tag{6}$$

where a representative household's utility $U(*)$ is derived from average consumption C_t/N_t. In this case, consumption C_t is the fraction $(1-\sigma_t)$ of output Q_t. In their model, it is assumed that the labour force L_t is connected to the population N_t, such that $L_t = \rho N_t$, where ρ is the participation rate. The availability of information allows consumption technology to shift at rate θ holding other factors constant. It also allows utility to increase over time without increasing Q_t. Expressing utility in terms of growth rates results into the expansion of equation 5:

$$u = \theta + (q-l) = \theta + \lambda + \alpha(r-l) + \delta(e-l) + \pi(s-l). \tag{7}$$

Equation 7 shows that growth in utility can be driven either through the growth of inputs r, e, and s, resource-saving technical change λ and output-saving technical change θ. The parameter θ represents the rate of which access to information is able to increase utility that is not reflected in measures of national output and conventional TFP estimates. This suggests that information results to welfare benefits that are not captured by aggregate output Q_t and its statistical analogue, GDP. As Hulten and Nakamura (2017) puts it, "GDP alone is not enough to assess the technological revolution."

Resource-saving technical change represented by productivity statistics can be estimated through the TFP residual approach because both sides of the growth accounting framework can be observed. The same cannot be said for output-saving innovations due to the fact that the utility function in equation 7 is unobserved. According to Hulten and Nakamura (2017), utility can be represented through the expenditure function, which is the minimum expenditure required to achieve a certain level of utility U^* given price price P and the level of information. The set of prices associated with the Hicks-neutral production function, $Q = e^{-\lambda t}f(L,K)$ (where L and K represents the labour and capital inputs, respectively), can be expressed as, $P = e^{-\lambda t}\phi(P_L, P_K)$. The expenditure function can be written as,

$$E(P, U^*) = e^{-\theta t}\xi(P, U^*) = e^{-\theta t}\xi(e^{-\lambda t}\phi(P_L, P_K), U^*). \tag{8}$$

Hulten and Nakamura (2017) state that the expenditure function $E(P, U^*)$ is expected to decline with an increase in information, since prices fall from improvements in efficiency. Equation 8 represents the dual function of equation 6. They argue that compensating variation V^C (and equivalent variation V^E), which measures the change in the expenditure function given a fixed set of prices P^*, would reflect the change in utility from an increase in information. Compensating variation can be expressed as,

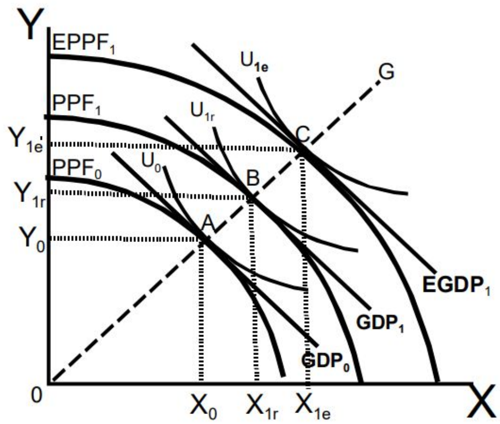
$$V^C = E(P^*, U_1) - E(P^*, U_0) \tag{9}$$

where $U_1 = e^{-\theta t}U_0$, implying that utility in the period 2 was driven by an increase in information θ. The expression in equation 9 thus represents a willingness-to-pay metric for a change in utility from the increase in information. A practical way of applying this concept is through the use of market prices to estimate compensating variation.

But in practical terms, how can one measure utility chances from the increase in information? 1 shows that market prices can be used to approximate compensating variation from information-induced utility change.

Consider the economy producing two goods in figure 1. Here, X_0 and Y_0 represent the output consumed (or produced) to reach the highest possible indifference curve prior to any innovation U_0. GDP_0, defined by the tangent line, is represented by $P_0^x X_0 + P_0^y Y_0$ and relative prices are denoted by the slope of this line P_x/P_y. Given the same set of inputs, production possibilities curve PPF_0 then shifts outward to PPF_1, allowing individuals to reach a higher level of indifference curve U_{1r}. At this point, $GDP_1 = P_0^x X_1 + P_0^y Y_1$ is also higher than GDP_0.

Figure 1



Notes: Figure sourced from Hulten and Nakamura (2019)

Output-saving technical change, on the other hand, would allow households to reach a higher indifference curve from U_{1r} to U_{1e} analogous to the effect of an outward shift the PPF from the origin. This occurs even as output (X_1 and Y_1), prices (P_x P_y) and GDP_1 remain unchanged. Hulten and Nakamura (2017) defines the new frontier as $EPPF_1$ the tangent line represented by $EGDP_1$. This form of innovation is associated with a pair of virtual outputs (X_{1e} and Y_{1e}) that are achieved by augmenting traditional outputs (X_1 and Y_1) through the process: $X_{1e} = (1+\theta)X_1$ and $Y_{1e} = 1 + \theta Y_1$. Here, θ represents the proportion of output saved from innovations. Hulten and Nakamura (2017) refer to this as utility-augmenting (output-saving) technical change, which is characterized by an increase in utility without increasing physical resources. According to these authors, this expansion can be captured by equivalent variation:

$$V = (P_0^x X_1^e + P_0^y Y_1^e) - (P_0^x X_0 + P_0^y Y_0) \tag{10}$$

where P_0^x and $P_0^y Y$ of good X_0 and good Y_0 at time 0, X_1^e and Y_1^e corresponds to the pair of virtual outputs following the output-saving innovation. Rearranging equation 10 we get[3]:

[3]Consider the expression:

$$(P_0^x X_1^e + P_0^y Y_1^e) - (P_0^x X_0 + P_0^y Y_0) = \theta(P_0^x X_0 + P_0^y Y_0)$$

$$= \theta(P_0^x X_0 + P_0^y Y_0)$$

$$V = \theta GDP_0 \tag{11}$$

where θ is the share of the output saved from the improved efficiency in the consumption of information. It also represents the output-saving technical innovation associated with a change in utility. Hulten and Nakamura (2017) devised welfare-based index called $EGDP$, which incorporates the willingness to pay for information resulting from the output-saving technical change:

$$EGDP = (P_0^x X_1^e + P_0^y Y_1^e) = V + P_0^x X_0 + P_0^y Y_0$$

$$= V + GDP_0 = (1+\theta)GDP_0 \tag{12}$$

The component V in equation 12 represents the individual's willingness to pay for information and is measured outside the National Accounts framework. It also captures the monetary effect in a $\theta-$shift in the utility function and operationally defines the new level of utility $U_1 e$. As such, in order to measure the degree of which the technological revolution has improved people's lives, it is critical to develop a measure of V.

In figure 1, the tangent lines corresponding to both GDP_1 and $EGDP_1$ are parallel, implying that their relative prices are equivalent. Moreover. the formulation of equivalent variation in equation 10 shows that only the price of market goods P_0^x and P_0^y are necessary for the estimation of V. This implies that the price of market goods can be used to provide estimates of welfare improvements from digital innovation.

---

where θ is any constant. We can show that θ represent the fraction of $(P_0^x X_1^e + P_0^y Y_1^e) - (P_0^x X_0 + P_0^y Y_0)$ to $(P_0^x X_0 + P_0^y Y_0)$,

$$\frac{(P_0^x X_1^e + P_0^y Y_1^e) - (P_0^x X_0 + P_0^y Y_0)}{(P_0^x X_0 + P_0^y Y_0)} = \theta$$

2.2 Empirical works on free digital goods

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

2.2.1 Contingent Valuation Approach

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

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

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

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

model with the respondent's decision to keep of give up the social media site as a outcome variable and the monetary value (in log scale) as the predictor variable.

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

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

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

Nguyen and Coyle (2020) conducted on online survey to estimate the WTA of online

90, 100, 1000).

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

goods in the United Kingdom. Employing a YouGov online panel, the researchers asked their participants how much they are willing to be compensated in order to give up a variety of goods, which includes both traditional goods and free goods. The goods that were identified in the survey include email, search engines, online banking, online maps, radio, TV, traditional and online news, streaming services, social media, among others. In terms of comprehensiveness, the study covers far more products than the study of Brynjolfsson et al. (2019a) and Jamison and Wang (2021). However, none of the participants were actually compensated for giving up the goods mentioned in the survey and random selection was not considered. Moreover, there were no randomization procedure involved in the data acquisition.

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

Nguyen and Coyle (2020) constructed demand curves from the percentage of their participants who are willing to give up the respective goods for the given compensation level. The authors did not, however, construct measures of aggregate monetary values of welfare from free goods.

Allcott et al. (2020) evaluated the effects of Facebook on welfare. Their study involved the conduct of a randomized control trial, employing 2,897 participants. Around 33 percent of the participants were paid $102[6] to deactivate Facebook for five weeks (the treatment group) while 67 percent were not paid any amount, but were asked to deactivate social media just the same. The participants were monitored for their social media activity throughout the experiment. An impact evaluation was conducted to examine the effect of Facebook deactivation on outcomes, namely 1) Substitute Time Use; 2) Social Interaction; 3) Substitute News Sources; 4) New Knowledge; 5) Political Engagement; 6) Political Polarization; and 7) Subjective Well Being. The analysis also tries to examine the post-experiment Facebook use and their opinions about Facebook. All these outcomes are self-reported. Indices were generated, taking into consideration a baseline survey and a

[6] The authors said that the amount $102 was chosen, since, based on their pilot data, a mass point of Willingness to Accept was at $100 would induce participants to deactivate Facebook.

post-experiment survey.

The results show that the treatment group spent less time online during the time they deactivated their accounts. They also spent more time watching television and spent more time with their families and friends. The treatment group also showed a lesser tendency to follow politics, following the experiment. Moreover, they were shown to follow President Donald Trump less, consume less news content, admit to having "less news knowledge" compared to the control group. In general, the treatment cohort reported that their political views are less polarized. In terms of subjective well-being, the experiment resulted to an overall positive impact on subjective well-being. Respondents for the treatment group were marginally happier and their life satisfaction score was higher. They also reported to being less depressed and less anxious after deactivation.

The authors did not estimate the monetary contribution of Facebook to the domestic economy, though they attempted to generate a back-of-the-envelop calculation of consumer surplus. They did so by multiplying the mean WTA from their experiment with the total number of Facebook users. They estimate that the consumer surplus from Facebook was around $31 billion. They also found the median WTA for the individuals in the control group actually decreased their valuation following deactivation.

Table 1 compares the WTA derived by the different studies employing the contingent valuation approach. It can be noticed the WTA values from the Nguyen and Coyle (2020) are substantially larger than those in the literature.

2.2.2 Total Cost Approach

The total cost approach employs the cost of producing free goods as to represent the value consumers derived from them. Many services that are freely available in the internet are financed either through advertising or marketing expense. For instance, Youtube is largely financed by non-tech firms advertising their products through the said medium. In the National Accounts, these expenditures are recorded as part of intermediate consumption. In the total cost approach, these expenditures would be recorded under Household Final Consumption Expenditure to reflect the welfare gained by households from consuming these services.

Soloveichik (2015) argued that the provision of free digital goods are products of a barter transaction between advertisers and internet users. She states that households are

Table 1: Comparison of WTA estimates for select digital goods

	Brynjolfsson et al. (2019b)	Corrigan et al.	Brynjolfsson et al. (2019a)		Nguyen and Coyle (2020)		Jamison and Wang
	(2017)	(2018)	US	Netherlands	May (Mean)	May (Median)	(2021)
Facebook	–	155.3	42.2	77.8	78	6.6-13	–
Instagram	–	–	–	5.1	36.9	0-0	–
LinkdIn	–	-	–	1.3	10.1	0-0	–
Twitter	–	–	–	¡1	20.9	0-0	–
Snapchat	–		–	1.6	16.6	0-0	–
Social Media	26.8	–	–	–	–	–	140.3
Whatsapp	–	–	–	384.1	103.8	13-32.5	–
Messenger	–	–	–	–	64.1	0.1-6.5	–
Instant Messengers	12.9	–	–	–	–	–	310.7
Maps	304	–	–	47.6	57.7	6.6-13	1,157.6
Skype	–	–		0.2	16.3	0-0	–
Zoom	–	–		–	–	–	44.9
Videoconferencing	–	–	–	–	–	–	337.5
Search engines	1,460.8	–	–	–	180.2	65-129.8	8,703.3
Personal email	701.2	–	–	–	192	129.9-324.5	2,095.7

Note: Table compares the estimated monthly WTA by Brynjolfsson et al. (2019b), Corrigan et al. (2018), Brynjolfsson et al. (2019a), Nguyen and Coyle (2020) and Jamison and Wang (2021). Figures are in USD. For comparability, WTA estimates by Brynjolfsson et al. (2019a) and Nguyen and Coyle (2020) were converted using their respective yearly average exchange rates during the period of data gathering. WTA estimates were acquired at different time periods. In particular, data by Brynjolfsson et al. (2019a) for the US (Facebook) were acquired in 2017; data from the Netherlands were acquired at different dates: January 2009 for Facebook, February 2004 for Whatsapp, October 2010 for Instagram, May 2003 for LindkIn, August 2003 for Skype, and March 2006 for Twitter. No inflation adjustments were applied to the data. Annual data from Brynjolfsson et al. (2019b) and Nguyen and Coyle (2020) were divided by 12 to generate an approximation of the monthly value.

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

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

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

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

2.2.3 Other studies

A number of scholars have attempted to estimate the aggregate value of welfare derived from free goods without resorting to the two approaches discussed above. These studies are more heterogeneous in terms of the approach they have taken to measure the value from free goods.

Bridgman et al. (2018) estimated the value from non-market leisure using the hourly returns from market leisure. Based on their estimates the value of non-market leisure was at $4.9 trillion in 2004, which grew to $7.8 trillion in 2015. They argue that the inability to account for free goods is likely to be the primary reason for the observed productivity slowdown in the past decade.

Syverson (2017) suggested the use of the GDP-GDI[7] as an indicator for the level of mismeasurement. He observed that over the past few decades, GDI has consistently outstripped GDP, which could indicate that workers are being paid for the production of goods with zero or substantially low market price. His analysis of the data showed that the GDP-GDI gap was, on average, $216.7 annually from 2013 to 2015.

The International Monetary Fund (2018) argued that the inability of National Accounts estimates to fully represent welfare improvements stems from the over estimation of price and inflation statistics. According to their report, this leads to a systematic underestimation of real growth. They suggest that improving the construction of price indices by into account quality change would ultimately result to the improvement of real growth estimates.

2.3 Synthesis of related works

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

The goal for most of the empirical studies is to generate a statistical aggregate rep-

[7]Gross Domestic Income

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

The main advantage of the total cost approach is that it requires little changes to the core accounting principles of the SNA. Measuring non-market output in terms of the total cost of producing them follows the practice for other non-market goods that are currently being recorded as part of GDP. This includes output from governments and non-profit institutions serving households (NPISH). However, they also suffer the same disadvantages, that is that they have a limited ability to reflect welfare changes[8] (see Bean (2016)).

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

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

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

[8]This is because output = input.

The valuation of goods and services in the SNA is based on the concept of an individual's WTP. Paragraph 3.119 of the 2008 SNA writes:

> Market prices for transactions are defined as amounts of money that willing buyers pay to acquire something from willing sellers; the exchanges are made between independent parties and on the basis of commercial considerations only, sometimes called "at arm's length." Thus, according to this strict definition, a market price refers only to the price for one specific exchange under the stated conditions. A second exchange of an identical unit, even under circumstances that are almost exactly the same, could result in a different market price.(see United Nations and others (2009))

Standard economic theory predicts that with small income effects, WTA and WTP should be equivalent or at least, close to each other (see Willig (1976) and Randall and Stoll (1980)). However, Hanemann (1991) showed in an analytic framework that the values are not equal, with WTA being greater than WTP, for goods with little to no close substitute. This was corroborated by various empirical studies (see the randomized control trial by Shogren et al. (1994), the literature review by Horowitz and McConnell (2002), and the meta-analysis by Tunçel and Hammitt (2014)). The WTA/WTP gap is also observed for digital goods such as social media, as demonstrated by the recent papers of Sunstein (2020) and Sindermann et al. (2022).

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

This is also reflected from the findings of Nguyen and Coyle (2020), where they also asked their respondents for their WTA for traditional goods. The authors found that (for May 2020) the mean WTA for paid goods like printed newspapers (£430), Cinema (£589), and Netfix (£1,373) are substantially higher than their market price, which is the valuation used in the SNA. As such, one can argue that aggregates generated from contingent valuation may not be truly consistent with GDP and other aggregates compiled using the same accounting framework. In a way, estimates from this approach may reflect the level of welfare individuals receive from from having access to these goods, but they are not necessarily comparable with estimates of the welfare value from the consumption of other goods as measured by National Accounts[9].

While it can be argued that the value of free digital goods can be recorded in a satellite account (as recommended by Schreyer (2021)), even satellite accounts attempt to preserve the core accounting principles of the SNA. For instance, the System of Environmental Economic Accounting (SEEA), the international standard for the compilation of satellite accounts for the environment, recommends the use valuation based on exchange (see paragraph 9.22 of United Nations (2014)). Other methods–such as direct survey and binary choice experiments–are not recommended for by the SEEA without validation or some form of adjustment (see paragraph 9.24 of United Nations (2014)). The reason for this restriction is to maintain the internal consistency of the account with the core national accounts estimates. If the goal is to measure the contribution of non-market output, such as ecosystem services, to total human activity, it is necessary that the accounting principle for non-market transactions should be consistent (or at leas similar) to the accounting principles applied to market transactions in order to ensure comparability.

What is missing from the literature is an estimation strategy that captures the welfare gains from the consumption of free digital goods, and is consistent with the accounting framework of the SNA. We address this gap in the literature by employing the price of premium versions of free goods as a source of valuation for their free counterparts. In the succeeding section, we discuss the theory why we believe that the price premium of services is a valid proxy for the price of free services.

3 Theoretical Underpinnings

We propose a theory on the valuation of "freemium" goods. We construct a static model employing the principles from the Lancaster (1966) framework. The model shows that value derived from free digital goods is equivalent to the price of its paid counterpart.

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

Van Elp and Mushkudiani (2019) explained that free goods are made available through six business models: 1) the advertising-financed model, 2) marketing-driven, 3) data-financed model, 4) ubiquity now, revenues later model, 5) charity and 6) freemium goods model. The framework we developed is specific towards the analysing the value of free goods delivered through the freemium model. In this business model, firms provide two versions of the digital good: a version that has no explicit cost upon consumption and one that consumers have to pay for. The paid versions are often called "premium" goods. Premium goods are often consumed as a bundle with other services (or features) that can exclusively be availed with the purchase of the premium version.

As an example, the videoconferencing Zoom offers both a free and a paid versions. The free version allows users to only access the videoconferencing feature of the service. However, the premium version allows users to access additional features such the creation of breakout rooms, live transcription, polling of participants, among others. Firms offering free services using the freemium model often do so in the hopes of attracting even a small percentage of users to their premium services in order to gain revenues.

The analytic model we developed is based on the Lancaster (1966) framework. His framework assumes that households are not passive receivers of utility. Rather, utility is generated by the households using a production technology. Lancaster (1966) also argued that households do not consume goods per se, but the characteristics embodied in these goods. For instance, an individual does not derive utility from the purchase of the car, rather utility is generated from the consumption of characteristics embedded in the car such as horsepower, millage, aesthetics, and so on.

In the model we developed, we consider the services from free goods, as well as premium-exclusive services, as individual characteristics consumed by household. Households are able to generate utility from the consumption of each characteristics. For simplicity, we shall refer to characteristics as "goods" in the following sections.

3.1 The Model

Consider an economy where a representative household j with demographic characteristics α consumes three types of goods. Good x is a form of digital good that can either be consumed with no explicit cost to the household or as part of a bundle with a set of digital goods $z_1, z_2, ..., z_n = \bar{z}$. The household's utility function can be expressed as:

$$u(x, y, z_1, z_2, ..., z_n, ; \alpha^j). \tag{13}$$

We assume that $u_x > 0$, $u_y > 0$, $u_{zi} > 0$, $u_{xx} < 0$, $u_{yy} < 0$, and $u_{z_i z_i} < 0$ which implies that that the household's utility is increasing and is strictly concave in consumption of goods x, y, and the set of goods $\bar{z}$.

The household's income w is exogenous. It uses its income to purchase all three goods. When the household consumes x as a bundle with $\bar{z}$, they would have to pay the amount $P^p(x, \bar{z})$. Alternatively, the household can also consume x free, allowing it to gain an implicit income $P^f(x)$ upon consumption. When purchased as a set, goods x and z_i are complements. The household cannot purchase any good z_i without acquiring good x. The household also consumes other goods y, which it pays for by the amount $P(y)$. $P(y)$ and $P^p(x, \bar{z})$ represents the expenditure functions for the consumption of goods y, x, and the set $\bar{z}$, respectively. The household's budget constrain is therefore constructed as,

$$P(y) + P^p(x, \bar{z}) \leq w + P^f(x). \tag{14}$$

In this model, expenditures on digital goods would allow the household to consume services that it could acquire free, x, coupled with services that it can only consume once they paid for it, $\bar{z}$. The household maximizes its utility choosing levels of x, y, and $\bar{z}$.

3.2 Valuing of free digital goods

The representative household maximization problem along with the first order conditions are detailed in appendix A. One result from the household's problem is the equivalence between the relative price of premium digital goods and the relative price of its free version. Consider the result,

$$\frac{P_x^f}{P_{z_i}} = \frac{P_x^p}{P_{z_i}} - \frac{u_x}{u_z} \tag{15}$$

where P_x^p and P_{z_i} are the marginal expenditures of the household for an extra unit of consumption of goods x and z_i, respectively. One can also think of them as the prices of goods x and z_i, respectively or the household's willingness to pay for each a unit of these goods. Meanwhile, P_x^f represents the additional implicit income received by household from the per-unit consumption of good x for free.

From equation 15, we see that the relative value of free goods P_x^f with respect to the value of premium-exclusive goods P_{z_i} is equal to the relative price of its paid version P_x^p (with respect to the price of value of premium-exclusive goods P_{z_i}) less the marginal rate of substitution between x and z. Since the two goods are complements, the MRS takes the value of either zero from the perspective of one good, infinity from the perspective on the other, or is undefined at the kink.

Since we know good z_i can only be acquired as a bundle with good x, then one can argue that households would always be in the horizontal part of the indifference curve. An increase in the consumption z_i would not increase the consumption of x. As such, the MRS is zero. We are left with,

$$\frac{P_x^f}{P_{z_i}} = \frac{P_x^p}{P_{z_i}}. \tag{16}$$

The expression in equation 16 shows that the relative value derived by households from the free version of good x is equivalent to the relative price of its premium version. From an empirical perspective, this suggests that one can can use a component of the price of 'premium' digital goods to represent the value of its free counterparts. For instance, the videoconferencing service Zoom provides both a free service and a premium service. The value gained by households from the use of its premium service is easy to derive, as it is partially reflected in the companies revenues. The same cannot be said from the value received by households from the use of the free version of its services.

From our results, it can be determined that the price paid for the premium version is a valid proxy for the value derived by households from the use of its free version. The challenge is to isolate the prices attributable only to the services present in the free versions. Econometric techniques such as hedonic regression can be used to isolate prices attributable to the premium-exclusive component and the price that is attributable to the service that is also available for free.

4 Estimation Strategy

We showed in the previous section the price of premium versions of digital goods can be employed as valid proxies for their free counterparts. This strategy is not new for non-market valuation. The SNA suggests the use of prices of products from similar markets as a source of valuation for non-market goods when prices cannot be observed. Paragraph 3.123 of the 2008 SNA states:

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

Compilers of National Accounts statistics often use market substitutes to impute the value of certain non-market goods, such as services from owner-occupied housing, barter transactions, extraction of groundwater, and agricultural products for own consumption, among others. Compilers of the Household Satellite Accounts also use this strategy to value household services such as childcare.

Subscribers to premium services would have access to the services provided by the free version with the addition of other features exclusive to the premium version. One can argue that the price of the premium versions pi would have two (2) components: a 'freely-available' p_f and a premium component p_p. If the relationship of the two components are additive, the price of premium services can be expressed as,

$$\underbrace{p_p}_{\text{price of premium service}} = \underbrace{p_f}_{\text{'freely-available' component}} + \underbrace{p_z}_{\text{premium-exclusive component}} . \tag{17}$$

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

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

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

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

4.1 Hedonic Regression

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

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

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

$$p_p = f(z_{i1}, ..., z_{i,n}, \varepsilon_i) \tag{18}$$

for a good with n characteristics. The marginal contribution of each characteristic can be estimated through a regression framework. In this study, we employ the logarithmic-linear (or semi-log) model[10]. In this exercise, we employ a modified time dummy variable model given by:

$$log(p_{i,j}^t) = \sum_{j=1}^{J} \sum_{t=1}^{T} (\delta_j \times \tau^t) + \sum_{k=1}^{K} \beta_k Z_{i,j} + \varepsilon_{i,j} \tag{19}$$

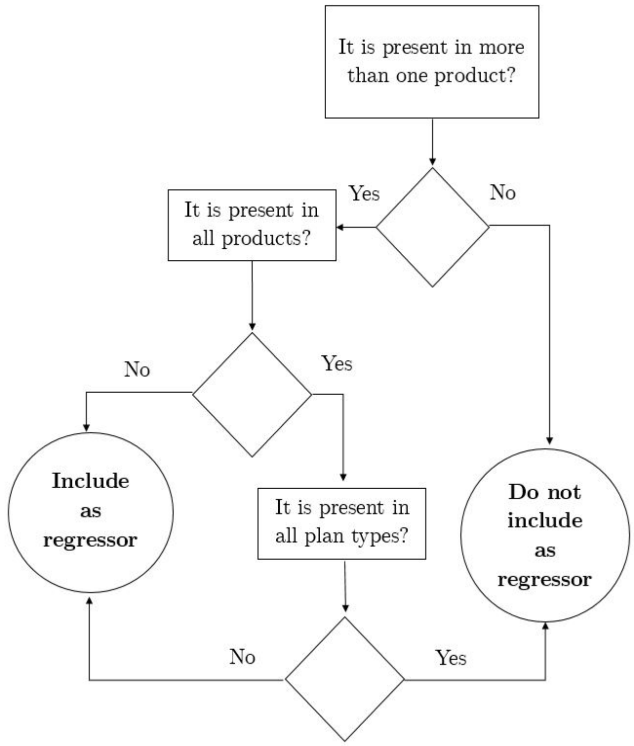
[10]An alternative is the linear specification where the levels of prices are used as the dependent variable. Diewert (2003) noted that it is more appropriate to employ the log-linear model for technological goods since it often mitigates the problem of heteroskedasticity as their prices tend to have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

where $log(p_{ij}^t)$ represents the the natural log of the prices at year t. The index i indicates the plan type (Standard, Pro, Business, etc) while the index j represents the service provider (Zoom, Cisco Webex, Microsoft Teams, etc). The list of service providers and their respective pages are listed in the appendix. These prices are regressed against a set of characteristics contained in matrix Z_{ij} and a set of service provider fixed effects δ_j. Details on the characteristics are described in section 5. In our specification, the term $(\delta_j \times \tau^t)$ represents the interaction term between the service provider dummies δ_j and year dummies τ^t. This ultimately generates separate intercept terms for each service provider for each year. We interpret each of the intercept terms as the quality-adjusted price for each service provider j for time t. The error term ε_{ij} is a assumed to be normally distributed with mean 0 and constant variance. For this paper, we follow the technical guidance of Aizcorbe (2014) and those of de Haan and Diewert (2013).

One of difficulties in implementing the hedonic approach is its sensitivity to the regression specification. It is critical that all characteristics that determines the price should be included as an explanatory variables, otherwise this would lead to omitted variable bias. Videoconferencing services advertise many features[11] that can be considered explanatory variables in the regression. In practical terms, however, we cannot include all features that appears in the website. In order to systematically identify the features that should be included as explanatory variables, we developed selection criteria that determine whether a feature should be included as a regressor or not (see figure 2).

[11]The terms "features" and "characteristics" would be used interchangeably in this research to denote characteristics in the hedonic regression framework.

Figure 2: Process for the selection of characteristics



Note: The figure shows the process in which premium-specific features are selected for the inclusion as an explanatory variable in the hedonic regression.

The criteria rely on two tests: 1) whether or not the inclusion of the characteristic would result in perfect multicollinearity, and 2) whether or not the inclusion of the characteristic would induce variation. If the characteristic is present in only one service provider, then it would be perfectly collinear with the service provider dummy δ_j. As such, the degree to which these characteristics have influence on the price should already be captured with the service provider fixed effects. These characteristics would, therefore, not be included as regressors. At the other extreme, if the characteristic is present in all service providers (and all of their plan types), this characteristic would obviously not induce variation.

In a typical regression framework[12], the intercept term a_0 represents the expected value of the dependent variable if the value of all explanatory variables are zero. In the context of this research, the said parameter represents the average (log) price of free goods if the

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

value all premium-exclusive characteristics are netted out. As such, $exp(a_0)$ would reflect the shadow price of the free version of digital goods. In the context of this research, since we allow each service provider to have a different intercept for each point in time, our hedonic regression equation would produce separate quality-adjusted price indices for each service provider. To impute for the price of free digital goods, we take the average of these quality-adjusted price indices for the specific year (see equation 20). For videoconferencing and email, we also include continuous variables as regressors that cannot be assumed to be zero. These are the number of participants in the case of videoconferencing and the mail storage in the case of email. We assume a certain value for these variables (z_1) in the prediction model and multiply them by their coefficient. Lastly, the expectation of the error term $E(exp(\varepsilon_{ij}))$ should be taken into consideration in the estimation of the price, otherwise, the estimates would be biased. The standard correction suggested by the literature (see Pakes (2003); Aizcorbe (2014); Erickson (2016)) the inclusion of the term $exp(0.5Var(\varepsilon_{ij}))$ for a log-linear model. The imputed price of free videoconferencing can be calculated by the expression:

$$\hat{p}_f^t = \left[exp\left(\frac{1}{J}\sum_{j=1}^{J}\delta_j^t\right) \times exp(\beta_1 log(z_1)) \right] \times exp(0.5Var(\varepsilon_{ij})) \tag{20}$$

In the area of official statistics, this approach has been adopted to generate quality-adjusted price indices for technological products by the US Bureau of Economic Analysis (see Groshen et al. (2017)) and the UK's Office of National Statistics (see Bean (2016)).

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

5 Data

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

We used the website Internet Archive (www.archive.org), a US-based digital library, to acquire data from years 2020 to 2017. The website allows for public access to past versions of websites, allowing us to acquire information on prices and characteristics from previous years. In order to minimize transfer errors, we scraped the data from their respective web pages using the Rvest package of R Studio. As much as possible, we tried to acquire data for the same week of the same month (second week of May) in our attempt to minimize the effect of seasonality. However, data for this particular month for some service providers are not available. One limitation of Internet Archive is that it only preserves data on a particular website once a user *archives* the website for a particular point in time (essentially taking a snapshot of the website for a point in time). If the website is not archived for that day, then its information would not be preserved in the portal. As such, for some providers, we were only able to use the data that is available for the month closest to May of that particular year. A description of the panel structure for the hedonic regressions is discussed in appendix B.

On average, the price of videoconferencing also appears to be increasing over time (see table 2). The average price of videoconferencing in 2017 was $25.5. This increased to $46.4 in 2021. However, the average number of participants each call can accommodate increased as well. In 2017, the average number of participants for videoconferencing services was at 72.5 participants. This increased to 183.2 participants in 2021. If we normalize the prize to the number of participants, prices actually declined slightly from $0.8 in 2017 to $0.4

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

Table 2: Descriptive statistics over time

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

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

in 2021. This could reflect improvements in technology, which we see in the trend of other information goods (see Roser and Ritchie (2013)).

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

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

6 Results

In this section, we describe the results of the hedonic regressions and estimates of the shadow price of videoconferencing, personal email, and online news. We describe the price estimates generated from by the hedonic approach and subject these estimates to a series of robustness checks.

6.1 Hedonic Regression Results

We estimate the hedonic regression model in equation 19 using ordinary least squares. We show the coefficient plots for the regressions in section. The full regression results can be viewed in the appendix . Each coefficient estimate for the hedonic regression represents the marginal contribution of each characteristic to the log price of videoconferencing. The exponential of each coefficients can also be interpreted as the willingness-to-pay for the said characteristic.

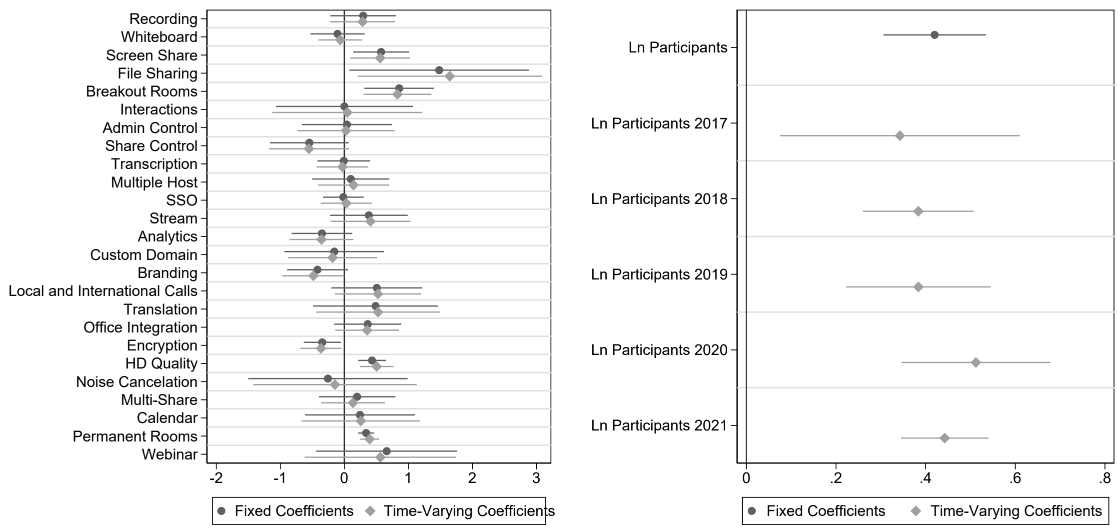
For videoconferencing (see figure 3), it can be noticed that not all coefficients are statistically significant. This suggests that the presence of some characteristics probably do not contribute substantially to the variations in prices across plan types and/or across service providers. We also observe the presence of negative coefficients that are statistically significant. If we were to interpret each coefficient as the marginal contribution of each characteristic to the price, it stands to reason that none of the variables should have a negative value for their coefficients. We offer two likely explanations for this. First, it is important to note that the coefficient estimates are partial elasticities and that we can only arrive at the marginal contribution of each characteristic by applying exponential transformations on the coefficients. In this case, the transformation would yield a positive value that is close to zero. Second, Erickson (2016) shows that it is possible for hedonic regressions to generate negative coefficient estimates if there are trade-offs between the characteristic with the negative coefficient and other characteristics in the regression. For instance, the trade-off between horsepower and mileage could result to negative coefficient estimates in a hedonic regression for automobiles. In the case of this exercise, only *Encryption* yielded a negative coefficient that is statistically significant. An examination of correlation between covariates (see appendix E) shows that the presence of encryption is negatively correlated with some of the statistically significant explanatory variables in the hedonic regression[14]. One can argue that the presence of these features makes it difficult to make calls more secure.

A major limitation of the panel hendonic regression is that it assumes that the marginal values of characteristics are fixed over time. It is possible that this assumption may not be true. From the descriptive statistics in table 2, we show that the average price per participants varies across years. We generate a second regression where in we interact the time dummies with the natural log of the maximum participants (z_1). This effectively generates a separate coefficient for the log of participants for each year.

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

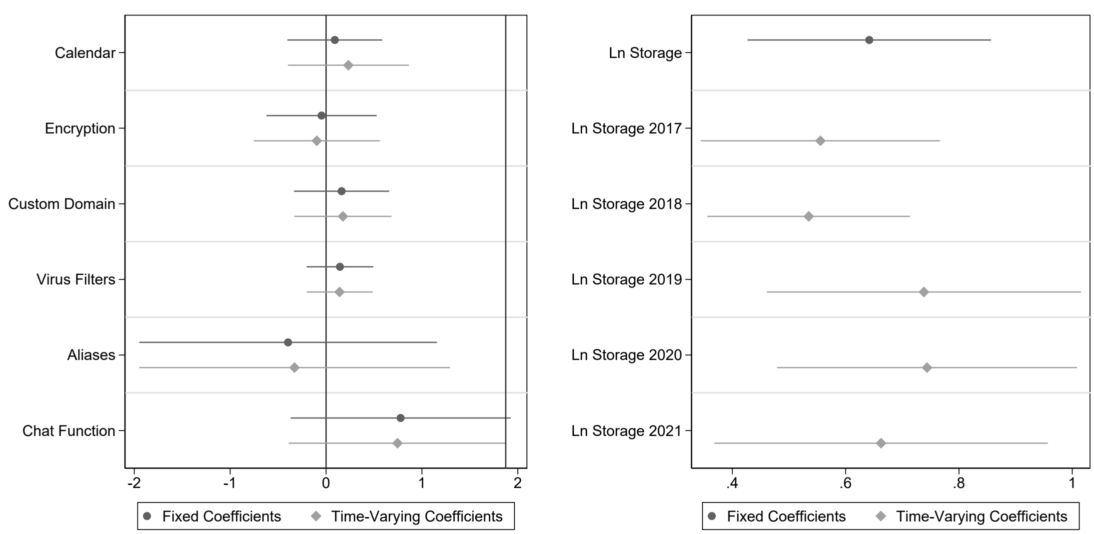
[14]These variables and their respective correlation coefficients with respect to Encryption are: File Sharing (-0.33), Breakout Rooms (-0.25), HD Quality (-0.13), and Log Participants (-0.11). See appendix for the correlation matrix.

Figure 3: Coefficient plot for videoconferencing



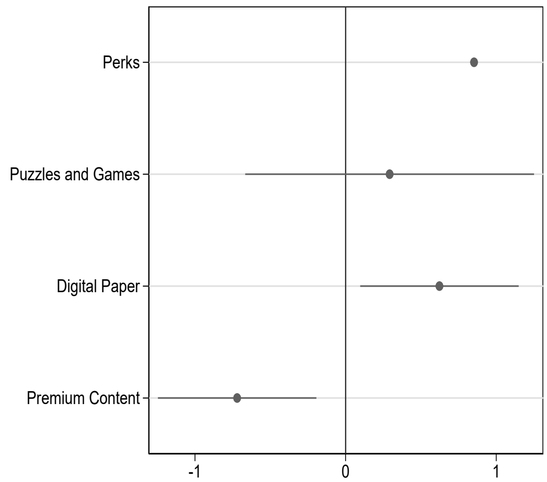
Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the the hedonic regression in equation 20. The coefficients represent column (3) of tables A.14 and A.15 in appendix D.

Figure 4: Coefficient plot for email



Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the the hedonic regression in equation 20. The coefficients represent column (2) of table A.16 in appendix D.

Figure 5: Coefficient plot for online news



Note: The figure shows coefficient plots and their corresponding 95 percent confidence interval generated by the the hedonic regression in equation 20. The coefficients represent column (2) of table A.17 in appendix D.

With the exception of the natural log of mail storage (in gigabytes), none of the characteristics included in the hedonic regression for email was statistically significant (see figure 4). The presence of chat functions generated a relatively large coefficient, however, its confidence interval still incorporated zero. As with videoconferencing, we interact the continuous variable (mail storage) with the time dummies in order to determine whether the coefficient would materially vary across time. Allowing the coefficient for mail storage to vary across time does not cause the estimates for the other coefficients to change. Moreover, the yearly estimates for the coefficients are not statistically different from the coefficients generated by the regression assuming that the parameter is fixed across time.

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

6.2 Shadow price of free digital goods

We impute the price of free digital goods using equation 20. We used the confidence interval for the coefficient of the number of participants $[\beta_1^{\hat{U}}, \beta_1^L]$ to generate our upper and lower bound estimates of the price[15]. For online news, however, this is not possible because the regression for online does not incorporate any continuous variable. As such, we only take the average of the upper and lower bound estimates of the quality-adjusted price indices, δ_j for email in order to generate interval estimates of its shadow price.

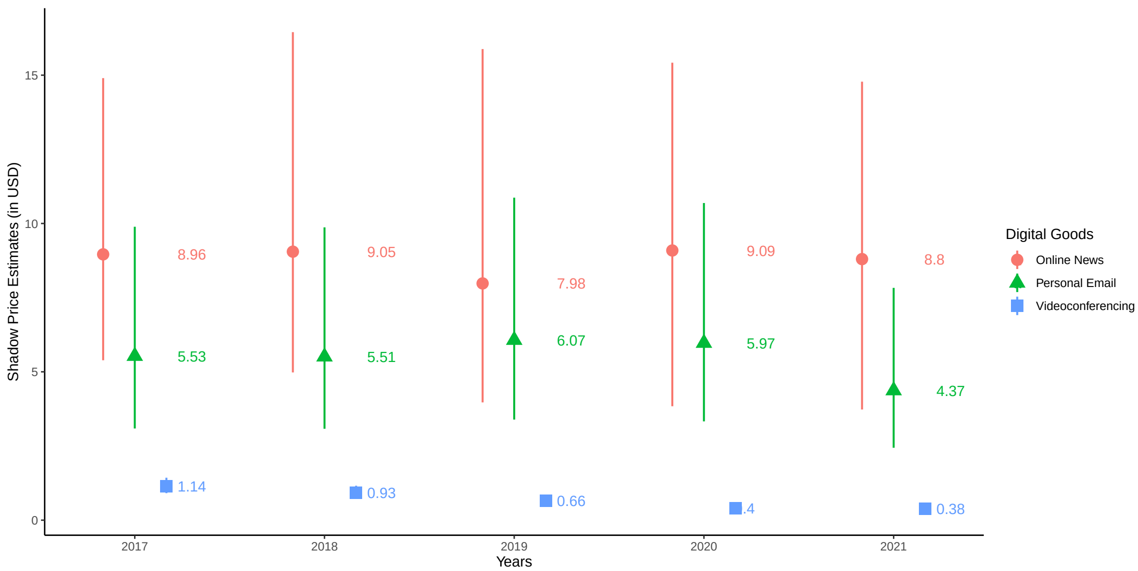
The specification in equation 20 requires us to assume a value for the continuous variable, z_1. For videoconferencing, this is the number of participants. The number of participants for free videoconferencing services is different for each provider. The top three messenging apps that offer free videoconferencing features are Whatsapp, Facebook Messenger, and We Chat[16]. The maximum number of participants for both Whatsapp and Facebook Messenger is 8 while the maximum for We Chat is 9 participants. The maximum participants for other popular applications offering free videoconferencing services also vary: for Viber, it is 20 participants, for Discord, 25, and for Telegram the maximum is 1,000. To be on the conservative side, we adopt the assumption of 8 participants for the generation of our price estimates. This is consistent with the maximum number of participants for Whatsapp and Facebook Messenger, the two largest provider of free videoconferencing as of writing.

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

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

[16]Statista Research Department (2021b)

Figure 6: Imputed price of videoconferencing, personal email, and online news



Note: The figure shows and shadow price estimates for videoconferencing, personal email, and online news, generated by the prediction model in equation 20. The price estimates, upper and lower bound estimates can be viewed in appendix G.

We present our estimates for the shadow price of videoconferencing, personal email, and online news in figure 6 (see appendix G for the table showing the interval estimates of the price levels for each year). Of the three forms of digital goods covered in this study, our price estimate for videoconferencing was lowest. Our estimates show that the shadow price for videoconferencing was approximately \$0.40 in 2020, lower compared to \$1.14 in 2017. This decline is consistent with the decline in the price per participants that we observe in table 2. For personal email, we estimate a shadow price of about \$6.0 in 2020, slight higher than the 2017 estimate, which is at \$5.5. Based on 95 percent prediction intervals, however, estimates across years for personal email are not statistically different from one another. As such, we cannot make any conclusive conjecture on the price trend for email. The largest price estimate we generated across years is for online news. Our estimates show that the shadow price of online news is approximately \$9.1. As with email, we observe no apparent trend for the price movement of online news across time.

The validity of our imputed price estimates is contingent on the validity of our hedonic regression model. We assume that the regression parameters should approximate *true* WTP for each characteristic. To test this, we generated a predicted price of premium versions of the digital goods, i using the prediction model in equation 21. Here, we include all characteristics in the regression model to generate a estimated price $\hat{p}_i^t$. If our predicted price is approximates the observed price, then it should be fair to argue that our model is valid.

$$\hat{p}_i^t = \left[exp\left(\frac{1}{J}\sum_{j=1}^{J}\delta_j^t\right) \times exp\left(\sum_{k=1}^{K}\beta_k log(z_k)\right)\right] \times exp(0.5Var(\varepsilon_{ij})) \tag{21}$$

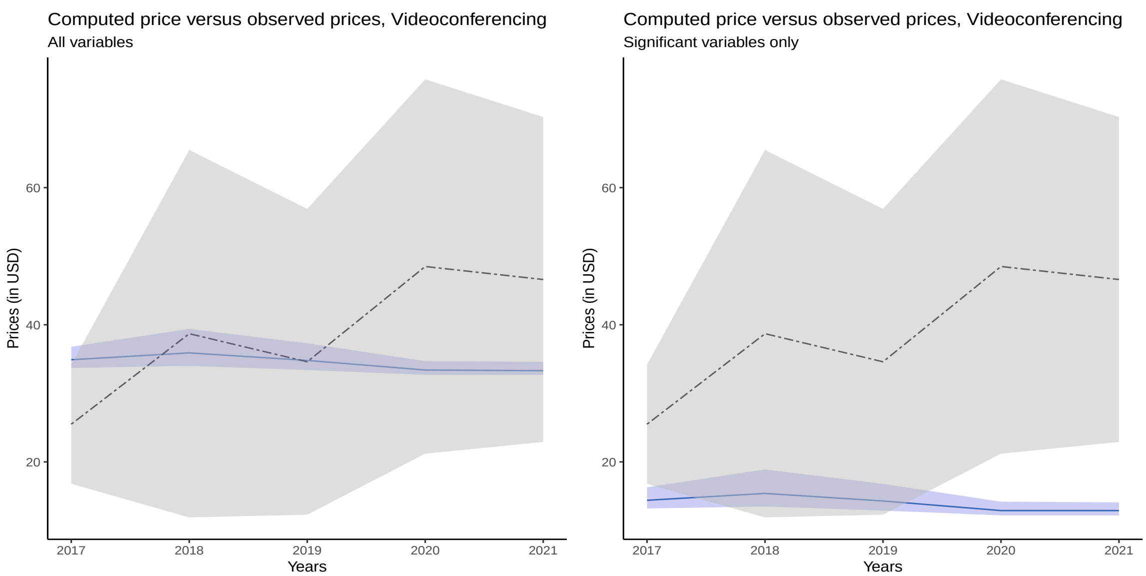
In figure 7 we compare the estimated premium price of videoconferencing (blue line) with the average price of paid videoconferencing from our data (dashed line) with its corresponding prediction interval[17]. We observe that the predicted price is within interval estimates of the average price for all years. However, the predicted price does not exhibit the same upward trend apparent in the mean price data. Moreover, we also observe that the prediction interval for our estimated price (blue shaded region) is smaller compared to the interval of the mean price. These discrepancies probably arose because the hedonic regression model we employed assumes that the marginal contributions to price of all premium-exclusive characteristics are fixed across time. By keeping the coefficients fixed over time, the model does not capture some of the time variation we see in the data. Allowing every characteristics to vary over time by interacting them with time dummies would likely capture this dynamic. Since the inclusion of additional variables increases the standard error, the said action would also likely to broaden the prediction interval. However, we choose not to do this for two reasons. First, the price estimates are still within the intervals of the observed data, implying that our estimates are reasonable. Second, have large intervals maybe a good thing when the goal is to generate unbiased estimates of parameters (for example, when examining relationships). But for the purposes of estimation, this is not the case. Large intervals are often not useful when generating prediction, as it reflects a large degree of uncertainty for the estimates. Estimates with large intervals are often not useful for policy purposes as well.

An alternative way to estimate the price of digital goods using hedonic regression is by including only variables that are statistically significant (at $\alpha = 0.05$). One can argue that characteristics that are not statistically significant do not contribute materially to the

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

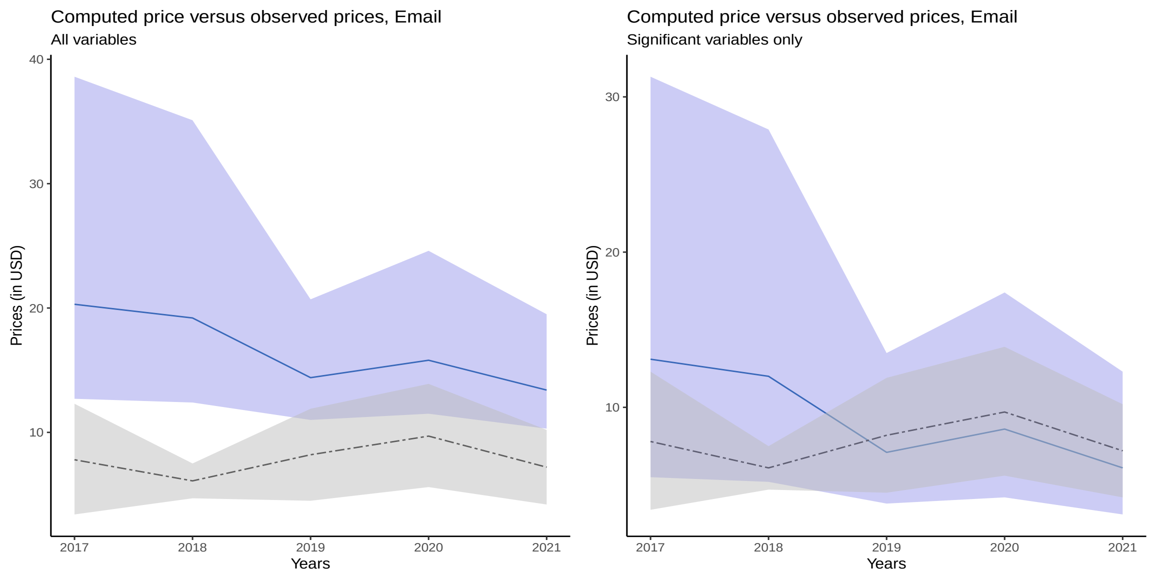
price of the good and could be ignored. Dropping all the variables that are not statistically significant would cause the estimated price to drop substantially for videoconferencing (see right panel of figure 7). Interestingly, this is not the case for email. Dropping variables that are not statistically significant causes the estimates to better align the observed data (see figure 8). This is likely to be because most of the coefficients for email were not significant in the first place.

Figure 7: Estimated price versus observed price for videoconferencing



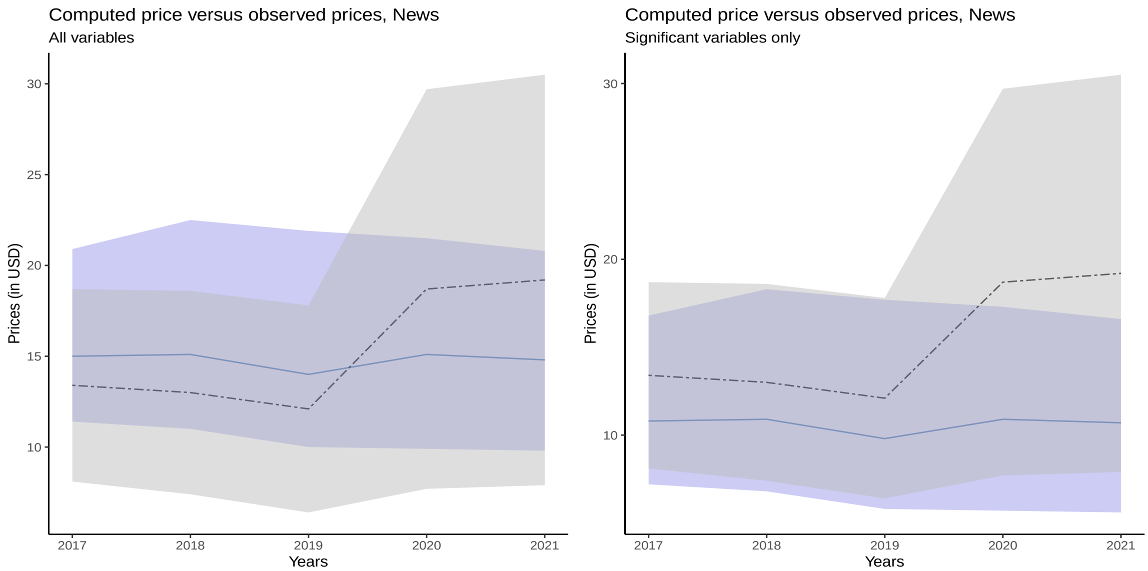
Note: The figure shows predicted monthly price of paid videoconferencing (blue line), its corresponding prediction intervals (blue area), the average monthly price of videoconferencing (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 21. The left panel shows all the price estimates which employs all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 where incorporated in the prediction model. All figures are in USD.

Figure 8: Estimated price versus observed price for email



Note: The figure shows predicted monthly price of paid email (blue line), its corresponding prediction intervals (blue area), the average monthly price of email (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 21. The left panel shows all the price estimates which employs all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 where incorporated in the prediction model. All figures are in USD.

Figure 9: Estimated price versus observed price for videoconferencing



Note: The figure shows predicted monthly price of paid online news (blue line), its corresponding prediction intervals (blue area), the average monthly price of online news (dashed line), and its corresponding confidence interval (grey area). The estimated price was generated by the prediction model in equation 21. The left panel shows all the price estimates which employs all characteristics in the prediction model while the right panel shows the price estimates where only characteristics that are significant at alpha = 0.05 where incorporated in the prediction model. All figures are in USD.

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

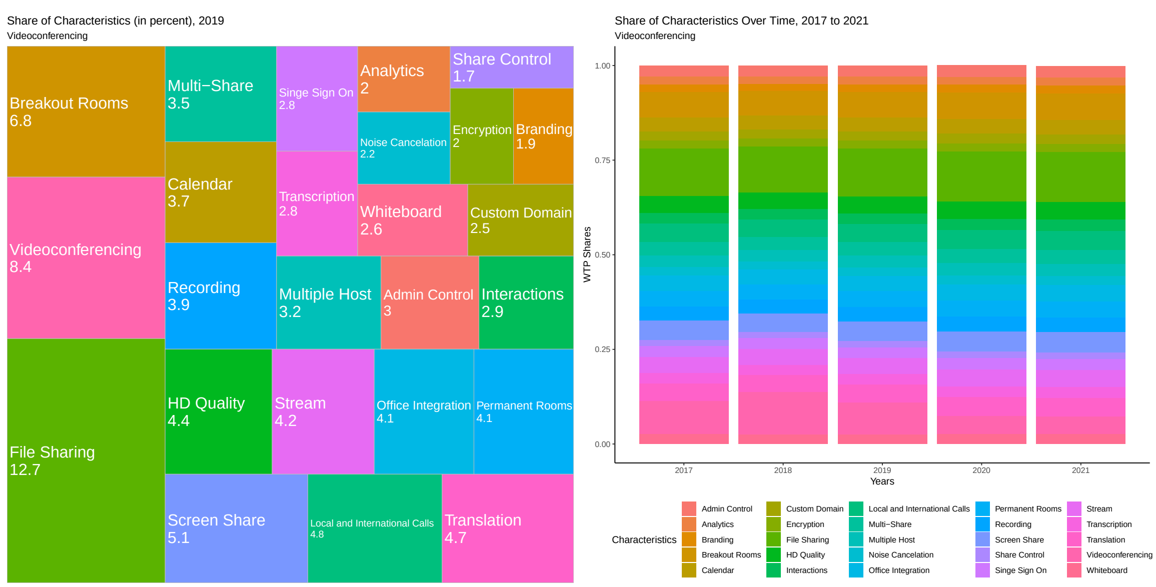
For online news, both the predicted price and the observed mean price are within the 95 percent prediction intervals of one another. One thing to note is that the price increase for the predicted price of online news from 2019 to 2020 is less pronounced compared with the jump in prices seen in the data.

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

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

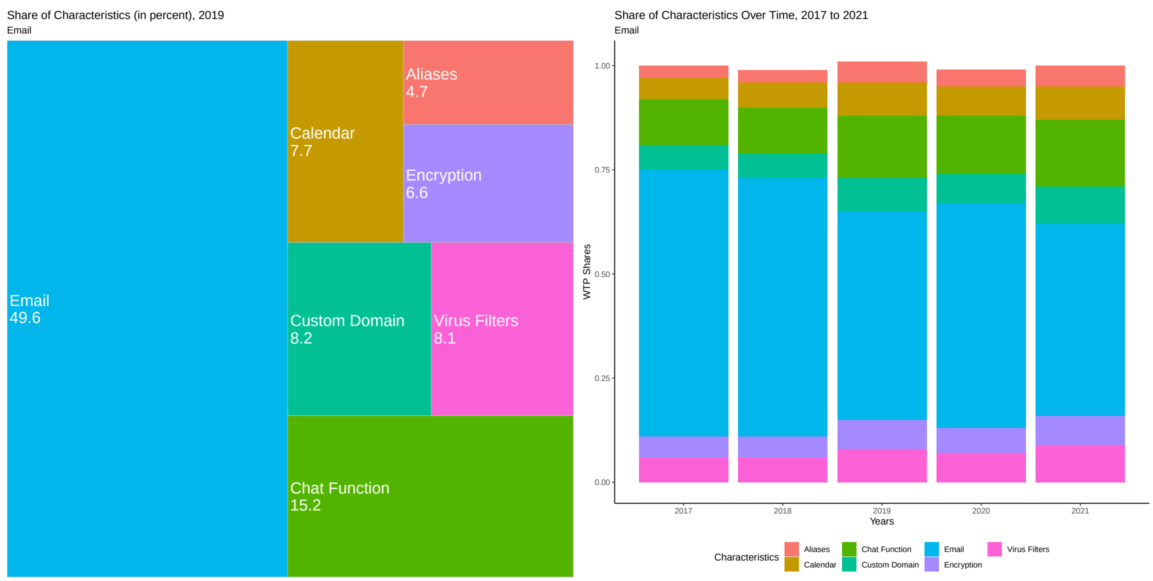
$$s_k^t = \frac{exp(\beta_k log(z_k))}{\hat{p}_i^t} \times 100. \tag{22}$$

Figure 10: Percentage share of characteristics to the predicted price, videoconferencing



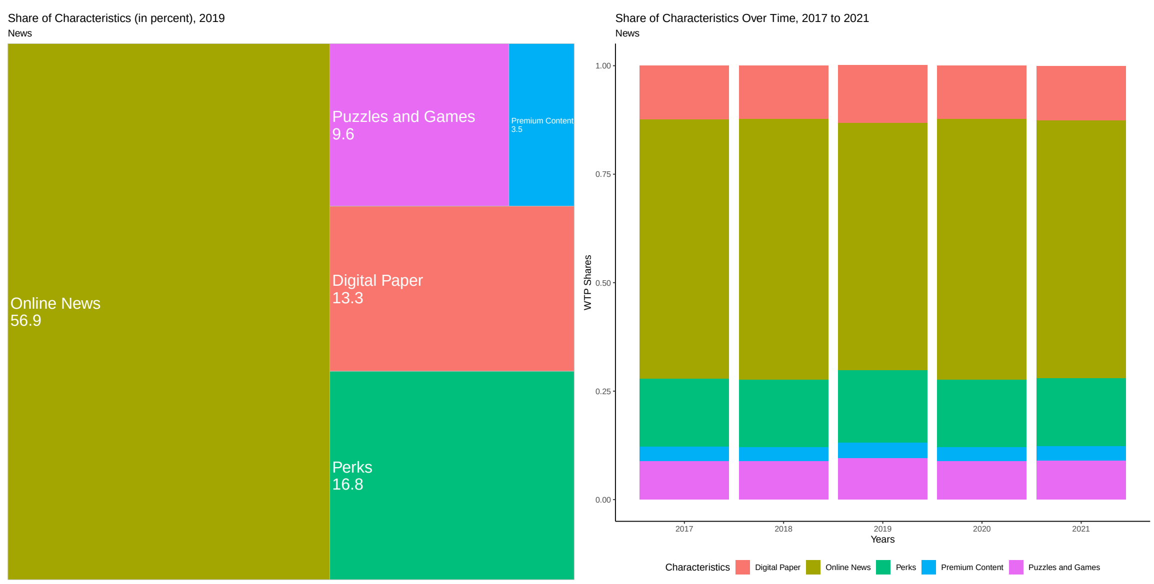
Note: The figure in the left shows the percentage share of each characteristic to the predicted price for videoconferencing in 2019, as computed in equation 22. The figure on the right shows the share of each characteristic of videoconferencing from 2017 to 2021.

Figure 11: Percentage share of characteristics to the predicted price, email



Note: The figure in the left shows the percentage share of each characteristic to the predicted price for email in 2019, as computed in equation 22. The figure on the right shows the share of each characteristic of email from 2017 to 2021.

Figure 12: Percentage share of characteristics to the predicted price, online news



Note: The figure in the left shows the percentage share of each characteristic to the predicted price for online news in 2019, as computed in equation 22. The figure on the right shows the share of each characteristic of online news from 2017 to 2021.

We report that that percentage share of the estimated WTP for all characteristics in figures 10, 11, and 12. For both email and online news, the estimated share of the "free component" accounts for about half of the predicted price. For videoconferencing, it accounts for the second largest share to its predicted price. This is consistent with our prior. The shares also appear to be consistent over time.

6.3 Robustness Check

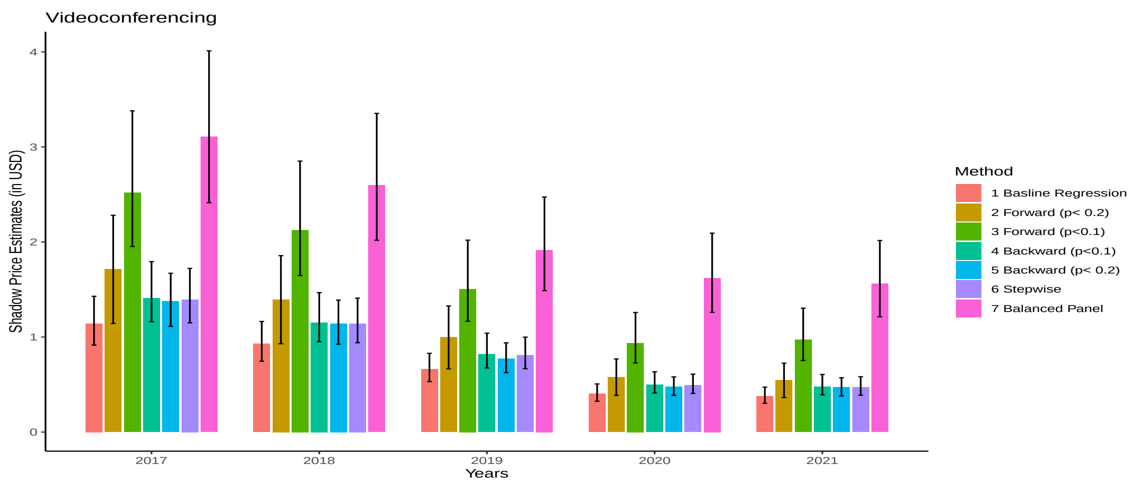
Price estimates from hedonic regressions can be sensitive to characteristics included in the specification. It is possible that the inclusion or exclusion of any of explanatory variable in the regression can result in substantial changes in the estimates. In this section, we examine the degree of which our price estimates would change given varying specifications.

The criteria for the selection of explanatory variables in figure 2 were aimed at maximizing the number of characteristics we can incorporate in the regressions given the information set published by the service providers in their websites. As mentioned earlier, the goal of this approach is to minimize omitted variable bias. We wanted to assume with confidence that all steps were taken in order to incorporate all *observable* characteristics in the baseline specification.

One of the pitfalls of this approach is that it could result in over fitting. The coefficient estimates for some of the characteristics in regression results displayed in the results tables are not statistically significant. A more parsimonious model could be a better fit for our purposes.

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

Figure 13: Robustness check, videoconferencing



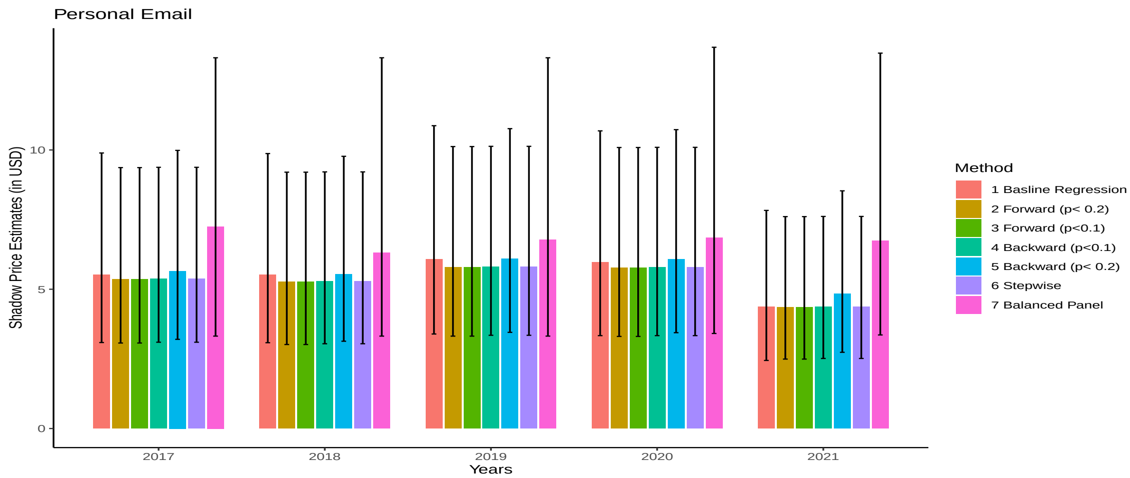
Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for videoconferencing. The figure also the interval estimates for the price of videoconferencing if service providers that are not present in all years are dropped from the data set. Figures can be views in table A.30.

For videoconferencing, the shadow price estimates from the forward, backward and stepwise selection models are shown in figure 13 (table A.30). Estimates from the forward selection are slightly higher than the baseline regression. Estimates from both the backward and stepwise selection models were closer to those generated from the baseline hedonic regression. Only the estimates from the restrictive 0.1 p-value threshold of the forward selection were outside the interval estimates from the baseline regression. Even so, the deviation is arguably not that substantial.

Lastly, we earlier noted that the panel data set we used for our regressions is not balanced. As shown in table A.1, we do not have data for all service providers for all of the years covered in the panel. This is either because the service provider had not started operating in those years, or that data simply cannot be acquired. To test how much attrition affects our estimates, we run the hedonic regression is equation 19, dropping the all service providers with incomplete data. The results are shown in table A.30 column (7).

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

Figure 14: Robustness check, personal email

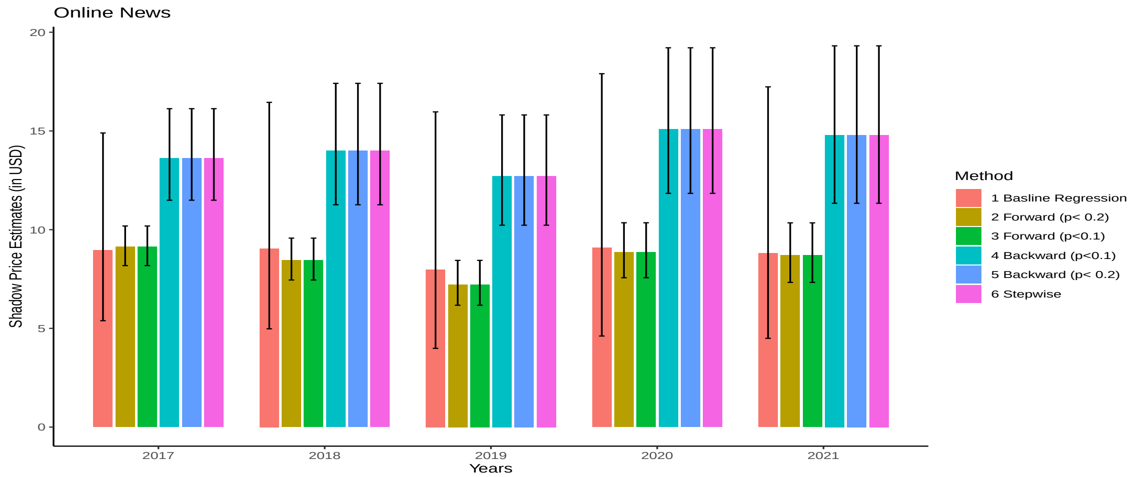


Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for personal email. The figure also the interval estimates for the price of personal email if service providers that are not present in all years are dropped from the data set. Figures can be views in table A.31.

The results are more robust for email. From figure 14 (table A.31), we can observe that the shadow price estimates from the baseline specification are noticeably similar to the estimates generated from the forward, backward and stepwise selection models. Moreover, the intervals for all estimates overlap, implying that there is no statistical difference between the estimates for all six models.

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

Figure 15: Robustness check, online news



Note: The figure shows point (bar) and interval for the baseline, forward, backward and stepwise estimation for online news. Figures can be views in table A.32

For online news, shadow price estimates from the forward, backward, and stepwise estimation are within the prediction interval of the baseline specification (see figure 15 and table A.32). Estimates from the forward estimation tends to be lower and estimates from the backward estimation tend to be higher.

6.4 Comparison with other studies

We compare our imputed prices of videoconferencing, personal email, and online news to the WTA estimates by other authors, namely Brynjolfsson et al. (2019a), Nguyen and Coyle (2020), and Jamison and Wang (2021). For videoconferencing, the comparison can be viewed from figure 16 (table A.33).

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

The mean WTA estimates from the online survey of Nguyen and Coyle (2020) for Skype, Messenger, and Whatsapp are higher that those generated by the hedonic regression. Their median bands are substantially lower compared to the mean estimates. For Messenger the mean estimate is 243 folds higher than the median, while for Whatsapp, the mean estimate higher by 261 folds greater[18] This suggests that many individuals reported extreme value in their response to the survey, causing the mean estimates to be high.

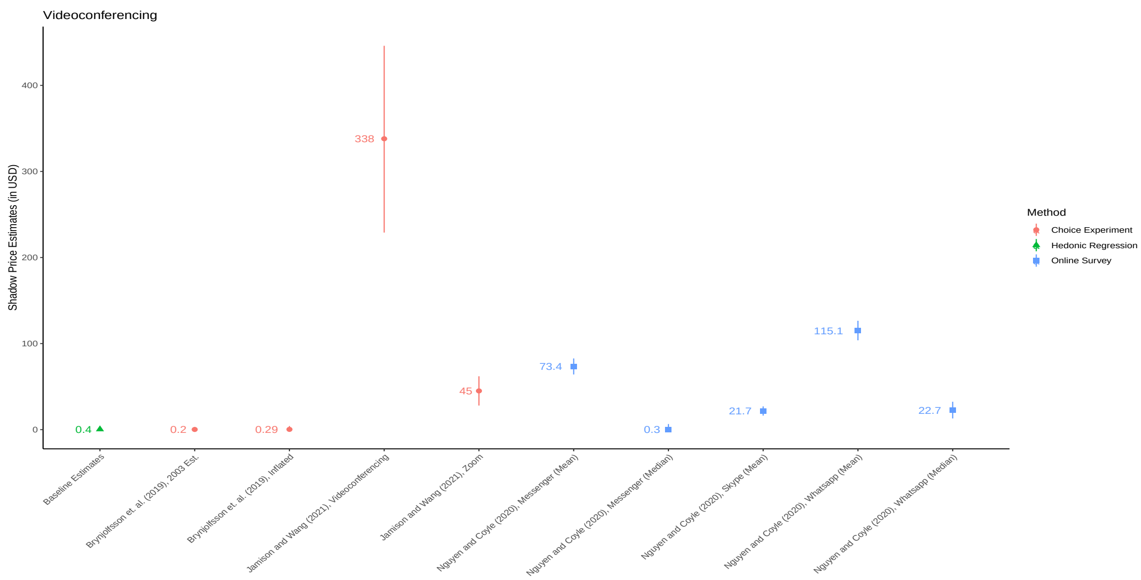
The estimates by Jamison and Wang (2021) for videoconferencing is also considerably higher than the esimtates from the hedonic regression. It is interesting to note that for Jamison and Wang (2021), their median WTA for videoconferencing is 7.5 fold greater than those of Zoom, which is one of the most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

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

The same pattern can be observed for both personal email and online news. We compare the estimates to those from Nguyen and Coyle (2020) and Jamison and Wang (2021) in figure 17 (table A.34). For both goods, estimates from the hedonic imputation is substantially lower than the WTA estimates from the contingent valuation studies. We offer two explanations for this observation.

[18]The median band for Skype by Nguyen and Coyle (2020), which we assume meant that it is less than zero. Therefore, we did not include it in the comparison.

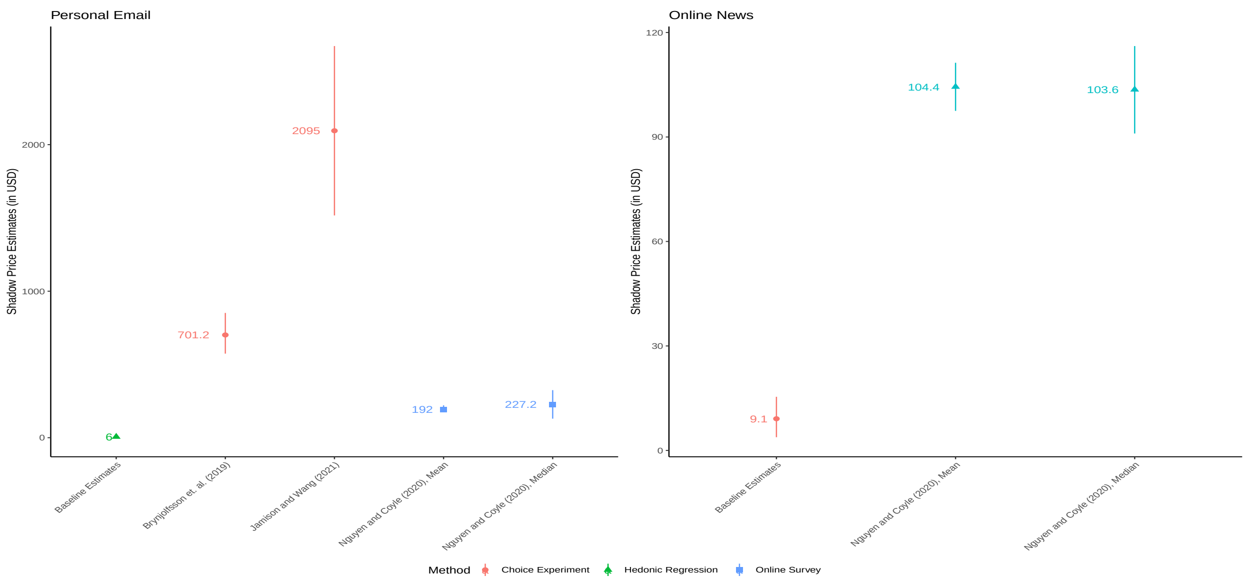
Figure 16: Comparison of WTA values with the price imputations for videoconferencing



Note: The figure compares the WTA estimates from Brynjolfsson et al. (2019a), Nguyen and Coyle (2020), and Jamison and Wang (2021) with the price estimates from the hedonic regression. Estimates by Brynjolfsson et al. (2019a) in column (2) were inflated to 2020 prices using Dutch CPI inflation from 2017 to 2020. Figures can be viewed from table A.33.

It is possible that respondents from the consumer surplus studies are not able to internalize the available substitutes. For instance, when they are asked how much they are willing to be paid to give up online news, they are not not thinking that they can purchase printed news as a substitute for online news when they make their choice. As such, the individual's willingness to pay for the consumption of news service in that case would not necessarily equate with how much they are willing to be compensated for giving up access to the service entirely. Furthermore, we see from the Nguyen and Coyle (2020) study that these discrepancies extends to traditional goods where their WTA are substantially larger than their market equivalent.

Figure 17: Comparison of WTA values with the price imputations for email and online news



Note: The figure compares the WTA estimates from Nguyen and Coyle (2020) and Jamison and Wang (2021) with the price estimates from the hedonic regression. Figures can be viewed from A.34.

7 Gross value of free digital goods

To estimate the aggregate willingness to pay for free digital services, we multiply its imputed price from equation 20 by a volume measure. The total monetary value of free goods V^t can be expressed as,

$$V^t = \sum_{f=1}^{F} \hat{p}_f q_f^t \tag{23}$$

where p_f is the shadow price[19] of free digital good f and q_f^t is a measure of its volume (or quantity). The expression V^t would represents that aggregate value derived by individuals from the consumption of free internet goods and could be part of household final consumption. This begs the question, what is the most appropriate measure of volume for our purposes?

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

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

Figure 18

TABLE 5: INTERNET ACTIVITIES, 2007 TO 2020

Within the last 3 months %

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

Base: Adults (aged 16+) in Great Britain.

: Data not available.

Source: Office for National Statistics

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

For our application, the only feasible course of action is to adopt the second case since the only information we have on prices is based on subscriptions. The task of acquiring reliable data on the number of subscribers to free goods is not straight forward. This type of information is not readily available from any source that we know of at this point. As such, we estimate the number of individuals who have access to videoconferencing and video calls using the ONS' Internet Access Survey and population statistics. In particular, we employ table 5 of the said survey (see figure 18). We multiply the proportion of adults with access to certain internet activities–which in our case, "Making voice and video calls", "Sending and receiving emails", and "Reading online news"–by the estimated number of individuals 18 years old and above based on the ONS' population projection data set. We arrive at the gross value of free goods by multiplying our estimated number of subscribers for each activity to their respective implied prices.

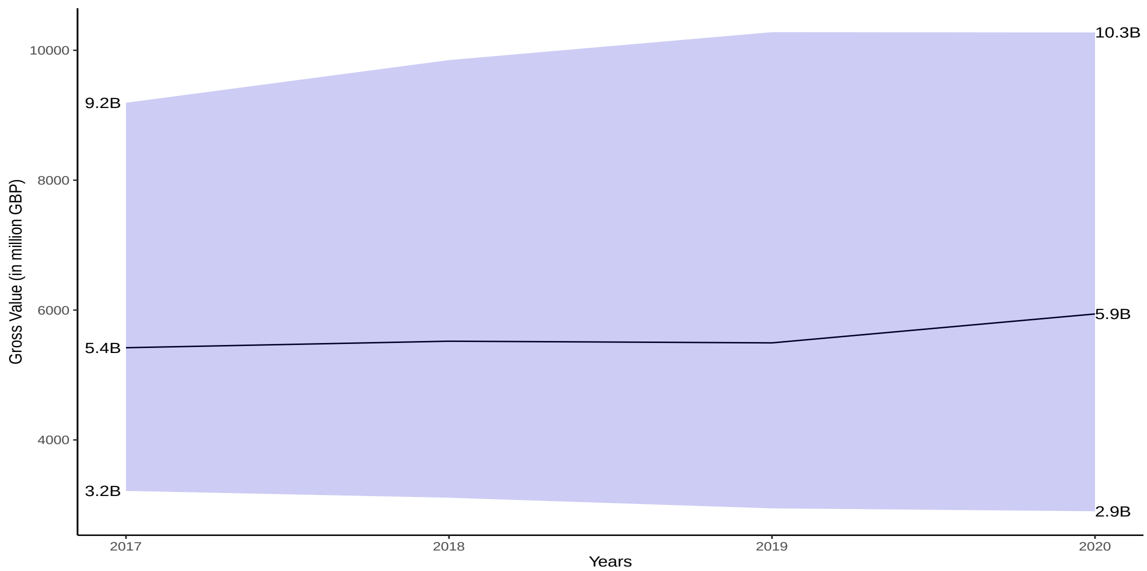
It is important to note that there would probably be double counting in the estimates, when we aggregate them with HFCE and/or GDP levels. The volume measure we used q_f^t includes *both* free and and paying users of the good. In order to appropriately aggregate these estimates with official statistics, it is important either to identify the number of free users or to net out the value derived by paying users. A counter argument to this is that individuals who subscribe to paid services often subscribe to their free counterpart as well. For instance, people who read the news through the paid version of Telegraph also read the news from free sources such as the BBC or CNN. Therefore, while double counting may be a problem, to which it affects our estimates may not be as severe.

7.1 Estimates of the gross value of free digital goods

We interpret our estimates as measures of the gross value of free digital goods. As such, we consider our estimates as part of the consumption side of GDP rather than the production side. The current price estimates of the gross value of digital goods are shown in figure 19. The initial figures that we generated were in USD. In order to be comparable with the UK's National Accounts data, we convert the estimates to GBP. We apply only one exchange rate (which is the average exchange rate from 2017 to 2020), in order to avoid having foreign exchange fluctuations affect our results.

Based on our estimates, the point estimate for the gross value of these goods is £5.9 billion, higher by 9.2 percent than the £5.4 billion in 2017. The data shows that the gross value of free digital goods makes up less than 1 percent of Household Final Consumption Expenditures (around 0.482 percent of HFCE in 2020) and GDP (around 0.280 percent of GDP in 2020). The interval shows that the the gross value of free digital goods is between £2.9 billion to £10.3 billion in 2020.

Figure 19: Gross Value of free digital goods



Note: the interval estimate of the aggregate gross value for the three digital goods, videoconferencing, personal email, and online news, as well as the estimates for household final consumption expenditures (HFCE) and gross domestic product by the ONS. All estimates are in million £. Figures can be viewd from table A.35 of appendix J.

We also generate constant price estimates (see appendix J) by deflating the nominal figures with an implicit Laspeyres-type price index. We chose 2018 as the base year in order to be consistent with the National Accounts estimates of the ONS. We add the constant price estimates to the chain volume measure estimates of the UK's HFCE and GDP to generate "expanded HFCE" and "expanded GDP" measures that includes the consumption of the three digital products. We show the growth rates in table 3.

For both 2018 and 2019, the gross value of free digital goods has grown substantially faster than both aggregate household consumption and GDP. Our estimates show that with the inclusion of three digital goods, the decline in HFCE would have been slower by 0.03 to 0.1 percentage points in 2020. The decline in GDP for 2020, meanwhile, would be 0.02 to 0.06 percentage points slower with the inclusion of the value of digital goods.

Table 3: Growth rates of digital goods and household consumption

	2017-2018	2018-2019	2019-2020
HFCE	1.03	1.14	-10.01
GDP	1.25	1.45	-9.83
Digital goods			
Point Estimate	2.33	2.83	3.23
Lower	2.26	2.98	3.12
Upper	2.37	2.72	3.32
HFCE + digital goods			
Point Estimate	1.03	1.14	-9.95
Lower	1.03	1.14	-9.98
Upper	1.04	1.15	-9.91
GDP + digital goods			
Point Estimate	1.25	1.45	-9.8
Lower	1.25	1.45	-9.81
Upper	1.25	1.45	-9.77

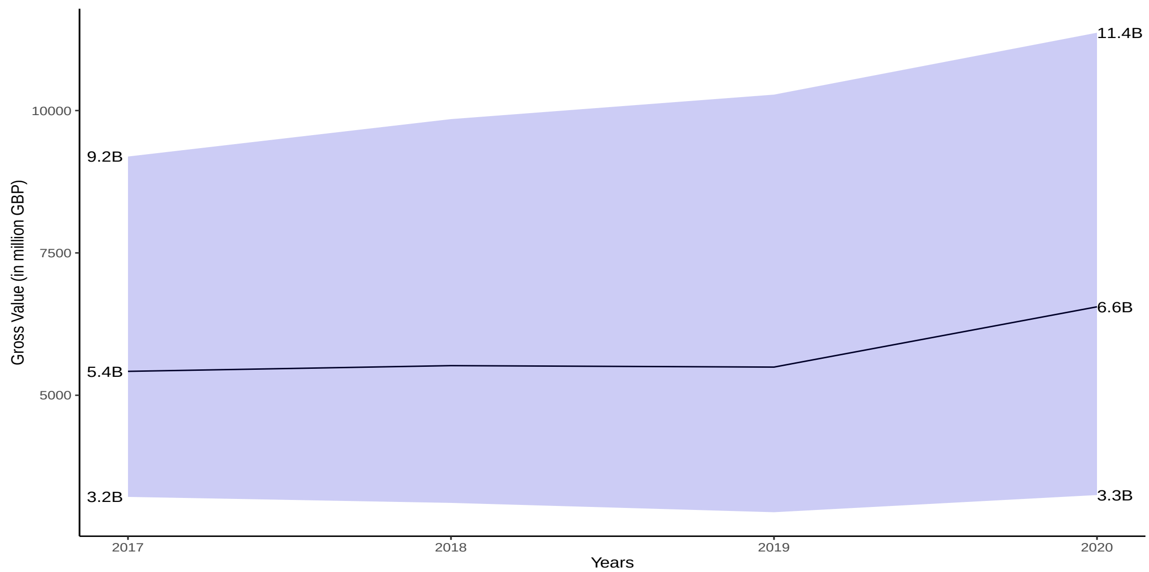
Note: The table shows the the growth rates of the household final consumption expenditure and gross domestic product chain volume measure estimates estimates of the ONS, constant price estimates of the gross value of digital goods, HFCE + digital goods, and GDP + digital goods. Figures are in percent.

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

7.2 Effect of the pandemic on the gross value of free digital goods

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

Figure 20: Gross value of digital goods adjusted for Ofcom data, at current prices



Note: the interval estimate of the aggregate gross value (at current prices) for the three digital goods, videoconferencing, personal email, and online news, after estimates in 2020 is adjusted using Ofcom data. All estimates are in million £. Figures can be viewd from table A.37 of appendix J.

Our estimates show that the point estimate for the gross value of free digital goods was at £6.6 billion in 2020, higher by 22.2 percent compared to the 2017 figures. The interval shows that the the gross value of free digital goods is between £3.3 billion to £11.4 billion in 2020.

We present the growth rates of the gross value of digital goods, calculated using volume measures from the Ofcom report in table 5. In contrast to the earlier estimates, the gross value from free digital goods during the 2020 grew at double digits. The impact to HFCE is more substantial compared to earlier. Household consumption decline was slower by 0.06 to 0.18 percentage points. The estimates also show that the inclusion of free digital goods to GDP would slow its decline by 0.03 to 0.11 percentage points.

Table 4: Growth rates of digital goods and household consumption using Ofcom volume indicators

	2017-2018	2018-2019	2019-2020
HFCE	1.03	1.14	-10.01
GDP	1.25	1.45	-9.83
Digital goods			
Point	2.33	2.83	13.53
Lower	2.26	2.98	13.54
Upper	2.37	2.72	13.51
HFCE + digital goods			
Point	1.03	1.14	-9.91
Lower	1.03	1.14	-9.95
Upper	1.04	1.15	-9.83
GDP + digital goods			
Point	1.25	1.45	-9.77
Lower	1.25	1.45	-9.80
Upper	1.25	1.45	-9.72

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

7.3 Accounting for multiple provider usage

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

In reality, people often use multiple service providers for the same purpose. For instance, it is common that a person who uses Whatsapp for video calls would also engage services of other videoconferencing providers such as Facebook Messenger or WeChat. One can argue that the utility received by individuals from the use of one service provider is separate from the utility it derives from another provider[20]. In the case of market goods, if a person is subscribed to both Netflix and Disney Plus, subscription to the two services would be counted separately in GDP and HFCE.

We generate a separate set of estimates, which accounts for the use of multiple providers. Ideally, the best way to achieve this is by employing the number of users for each service provider. Unfortunately, precise data on the number of users are not readily available.

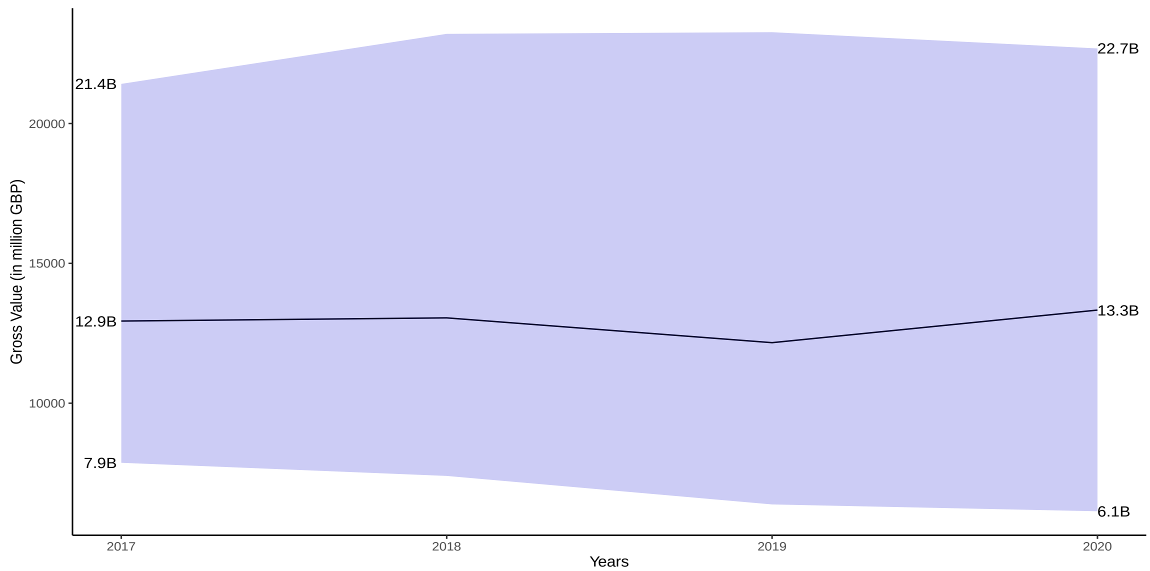
The top two providers of videoconferencing service (in terms of user share) in the UK are Whatsapp and Facebook Messenger (see Statista Research Department (2021b)). The top two most downloaded videoconferencing applications in the UK are Whatsapp and Telegram (see Statista Research Department (2021a)).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

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

To arrive at the estimates of gross value, we multiply our indicators for the number of subscribers for free videoconferencing and online news to the price estimates we generated in section 6. Unfortunately, we are not able to find any data on the number of users for Gmail, Outlook, or Yahoo mail (the top three providers of free email services in the UK). As such, we maintain our earlier estimates for email.

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

Figure 21: Gross value of digital goods accounting for multiple service provider usage, at current prices



Note: The table shows the interval estimate of the aggregate gross value (at current prices) for the three digital goods, videoconferencing, personal email, and online news, accounting for multiple service provider use. Figures can be viewd from table A.39 of appendix J.

Accounting for multiple service provider use, the estimates for the gross value of digital goods would more than double compared to the baseline figures. Our estimates show that the gross value of digital goods is around £13.3 billion in 2020. The interval estimates show that the the gross value of free digital goods is between £6.1 billion to £22.7 billion in 2020 (see figure 21).

Table 5: Growth rates of digital goods and household consumption using Ofcom volume indicators

	2017-2018	2018-2019	2019-2020
HFCE	1.03	1.14	-10.01
GDP	1.25	1.45	-9.83
Digital goods			
Point Estimate	1.51	2.25	1.90
Lower	1.62	2.14	2.24
Upper	1.43	2.32	1.66
HFCE + digital goods			
Point Estimate	1.03	1.15	-9.89
Lower	1.03	1.14	-9.94
Upper	1.03	1.16	-9.80
GDP + digital goods			
Point Estimate	1.25	1.45	-9.76
Lower	1.25	1.45	-9.79
Upper	1.25	1.45	-9.71

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

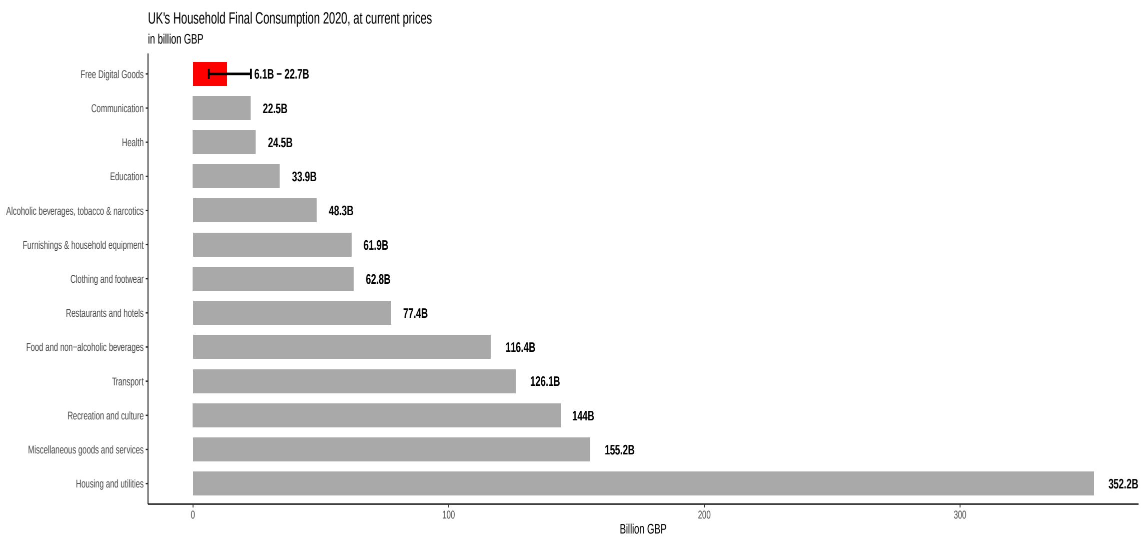
Not surprisingly, the percentage points impact on GDP growth rates is also larger. Our estimates show that the impact on real HFCE decline in 2020 was between 0.06 to 0.18 percentage points. For GDP, the impact to real GDP decline in 2020 was between 0.03 to 0.11 percentage points.

7.4 Discussion

Even accounting for multiple a service provider use, the estimates that our methodology generated are small relative to the UK economy. Based on our results, free digital goods account for 0.5 to 1.8 percent of the UK's HFCE in 2020 and 0.3 to 1.1 percent of the UK's GDP in the same year.

The figures that we generated, however, are likely conservative estimates of the true value of free digital goods for two reasons. First, we are unable to account for multiple service provider use for email. It is possible that many internet users hold multiple accounts from different free email providers. Second, we only accounted for the users of the top three videoconferencing providers. Due to data constraints, our estimates do not include users of Facetime, WeChat, Skype, and even Zoom. Both of these reasons likely to caused our estimates to have a downward bias.

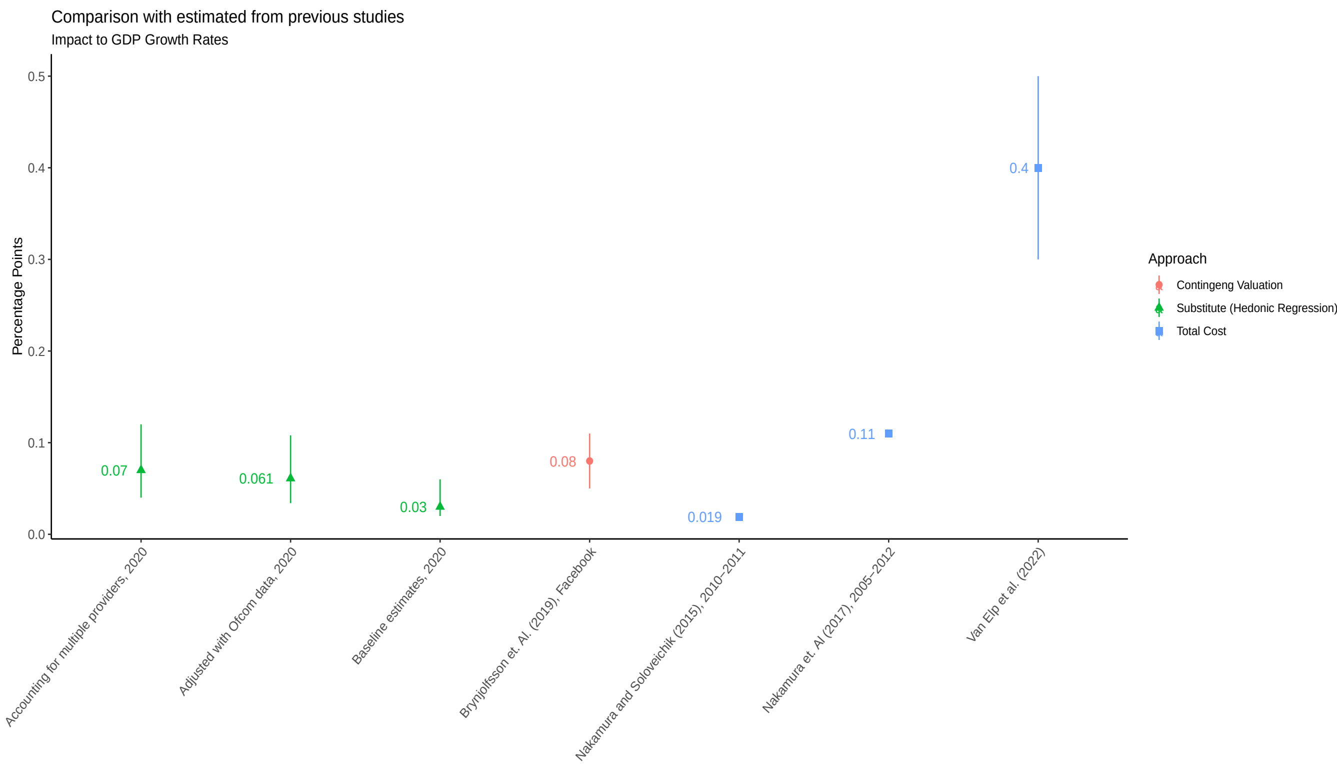
Figure 22: Comparison with other expenditure items



Note: The figure compares the current price estimates of gross value of free digital goods (accounting for multiple provider use) in table A.39 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.

Despite this, we argue that our estimates for the value of free digital goods is economically significant. Our estimates shows that the gross value of free digital goods was between £6.1 billion to £22.7 billion in 2020. For context, the lower limit of our estimates is already 30 percent of the total final consumption expenditures for communications, which is at £22.6 billion (see figure 22). Meanwhile, the upper limit exceeds the value of the same expenditure item.

Figure 23: Comparing estimated impact to GDP with estimates from other studies



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

We also compare our estimates to the findings of other authors (see figure 23). In particular, we compare the impact to GDP growth rates to the estimated impact by Nakamura and Soloveichik (2015), Nakamura et al. (2017), Brynjolfsson et al. (2019a), and Van Elp et al. (2022). It is important to note that these estimates cover the different time periods and that the four studies are concerned with the impact of free digital goods to the US economy (in the case of Nakamura and Soloveichik (2015), Nakamura et al. (2017), Brynjolfsson et al. (2019a)) and the Dutch economy (in the case of Van Elp et al. (2022)), as opposed to the UK economy, which the focus of this paper. However, we believe that this comparison would still provide valuable insight on how our approach differs as compared to others.

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

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

8 Conclusion and way forward

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

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

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

Our goal is to apply the methodology to other forms of online free services such as games, video and music streaming, dating apps, pornography, among others. The same principles can also be extended to generate estimates for the gross value of illegal illegal streaming (digital piracy). The national accounts does not discriminate between legal and illegal activities as a source of value and welfare for the aggregate ecooknomy. One can argue that many individuals gain some level of welfare from the illegal streaming and downloading of movies. Therefore, we recommend further studies on the application of this method on producing estimates of the value derived from free digital goods.

Since our approach employs the price of premium services to derive the value of their free counterpart, the method effectively limits our application to digital goods with paid versions. It is possible that in the subject of measuring the value of free goods, multiple approaches are needed to generate a complete picture. Another thing to note is that digital services that operate under the Freemium model often finance the operations of the free versions through cross-subsidies. As such, the shadow price estimates that we generate could form part of the value of cross-subsidies from paid digital good to its free version.

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

References

Aizcorbe, A. M. (2014). *A practical guide to price index and hedonic techniques*. Practical Econometrics.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3):629–76.

Bean, C. R. (2016). *Independent review of UK economic statistics*. HM Treasury.

Bridgman, B. et al. (2018). Is productivity on vacation? the impact of the digital economy on the value of leisure. *Macroeconomic Dynamics*, pages 1–22.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., and Fox, K. J. (2019a). Gdp-b: Accounting for the value of new and free goods in the digital economy. Technical report, National Bureau of Economic Research.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., and Fox, K. J. (2020). Measuring the impact of free goods on real household consumption. In *AEA Papers and Proceedings*, volume 110, pages 25–30.

Brynjolfsson, E., Collis, A., and Eggers, F. (2019b). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, 116(15):7250–7255.

Corrigan, J. R., Alhabash, S., Rousu, M., and Cash, S. B. (2018). How much is social media worth? estimating the value of facebook by paying users to stop using it. *PloS one*, 13(12):e0207101.

Coyle, D. (2015). *GDP: a brief but affectionate history-revised and expanded edition*. Princeton University Press.

Coyle, D. (2019). Do-it-yourself digital: the production boundary, the productivity puzzle and economic welfare. *Economica*, 86(344):750–774.

de Haan, J. and Diewert, E. (2013). Hedonic Regression Methods. *Handbook on Residential Property Price Indices*, pages 49–64.

Diewert, W. E. (2003). Hedonic regressions: A review of some unresolved issues. In *7th Meeting of the Ottawa Group, Paris, May*, volume 29.

Dynan, K. and Sheiner, L. (2018). Gdp as a measure of economic well-being. *Working Paper*, 43.

Erickson, T. (2016). *On" incorrect" Signs in Hedonic Regression.* US Department of Labor, US Bureau of Labor Statistics, Office of Prices and . . . .

Goolsbee, A. and Klenow, P. J. (2006). Valuing consumer products by the time spent using them: An application to the internet. *American Economic Review*, 96(2):108–113.

Groshen, E. L., Moyer, B. C., Aizcorbe, A. M., Bradley, R., and Friedman, D. M. (2017). How government statistics adjust for potential biases from quality change and new goods in an age of digital technologies: A view from the trenches. *Journal of Economic Perspectives*, 31(2):187–210.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: how much can they differ? *The American Economic Review*, 81(3):635–647.

Heys, R., Martin, J., Mkandawire, W., et al. (2019). Gdp and welfare: A spectrum of opportunity. Technical report, Economic Statistics Centre of Excellence (ESCoE).

Heys, R. and Taylor, C. (2021). Valuing free digital platforms in a national accounting framework.

Horowitz, J. K. and McConnell, K. E. (2002). A review of wta/wtp studies. *Journal of environmental economics and Management*, 44(3):426–447.

Hulten, C. and Nakamura, L. (2017). Accounting for growth in the age of the internet: The importance of output-saving technical change. Technical report, National Bureau of Economic Research.

Hulten, C. and Nakamura, L. (2019). Expanded gdp for welfare measurement in the 21st century. In *Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.

International Monetary Fund (2018). Measuring the digital economy. Technical report, IMF Policy Papers.

Jamison, M. A. and Wang, P. (2021). Valuation of digital goods during the coronavirus outbreak in the united states. *Telecommunications Policy*, 45(5):102126.

Jigsaw Research (2021). News consumption in the uk: 2020. Technical report, Office of Communications.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Nakamura, L., Samuels, J., and Soloveichik, R. H. (2017). Measuring the'free'digital economy within the gdp and productivity accounts.

Nakamura, L. I. and Soloveichik, R. H. (2015). Valuing 'free' media across countries in gdp.

Nguyen, D. and Coyle, D. (2020). Free goods and economic welfare. Technical report, Economic Statistics Centre of Excellence Discussion Papers.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to pc's. *American Economic Review*, 93(5):1578–1596.

Randall, A. and Stoll, J. R. (1980). Consumer's surplus in commodity space. *The American Economic Review*, 70(3):449–455.

Roser, M. and Ritchie, H. (2013). Technological progress. *Our World in Data*.

Schreyer, P. (2021). Accounting for free digital services and household production - an application to facebook.

Shogren, J. F., Shin, S. Y., Hayes, D. J., and Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, pages 255–270.

Sindermann, C., Yang, H., Yang, S., Elhai, J. D., and Montag, C. (2022). Willingness to accept (wta), willingness to pay (wtp), and the wta/wtp disparity in chinese social media platforms: Descriptive statistics and associations with personality and social media use. *Acta psychologica*, 223:103462.

Soloveichik, R. (2015). Valuing 'free'entertainment in gdp: an experimental approach. In *AEA Meeting in Boston, Boston, MA*.

Statista Research Department (2021a). Leading communication apps in the google play store in great britain (gb) in october 2021, by number of downloads.

Statista Research Department (2021b). Most popular global mobile messenger apps as of july 2021, based on number of monthly active users.

Statista Research Department (2021c). Whatsapp users in the united kingdom 2017-2025.

Statista Research Department (2022). Number of facebook messenger users in the united kingdom from september 2018 to december 2021.

Sunstein, C. R. (2020). Valuing facebook. *Behavioural Public Policy*, 4(3):370–381.

Syverson, C. (2017). Challenges to mismeasurement explanations for the us productivity slowdown. *Journal of Economic Perspectives*, 31(2):165–86.

Triplett, J. (2006). Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes. *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes.*

Tunçel, T. and Hammitt, J. K. (2014). A new meta-analysis on the wtp/wta disparity. *Journal of Environmental Economics and Management*, 68(1):175–187.

United Nations (2014). System of environmental-economic accounting 2012. Technical report, The World Bank Group.

United Nations and others (2009). *System of National Accounts 2008*. New York: United Nations.

Van Elp, M., Kuijpers, N., and Mushkudiani, N. (2022). Free services in the netherlands.

Van Elp, M. and Mushkudiani, N. (2019). Free services. Technical report, Statistics Netherlands.

Willig, R. D. (1976). Consumer's surplus without apology. *The American Economic Review*, 66(4):589–597.

Appendix

A Market Equilibrium

In section, we describe the household's problem and the first order conditions. The household maximizes its utility choosing levels of consumption for goods available free x, the set of premium-exclusive goods $z_1, z_2, ..., z_n = \bar{z}$, and other goods y. The households pays the amount $P^p(x, \bar{z})$ to acquire good x in a bundle with set $\bar{z}$ and pays $P(y)$ to consume good y. It gains an income $P^f(x)$ when it chooses to consume good x for free. The household also receives an exogenous income w.

$$\max_{x,y,z_1,z_2,...,z_n} u(x, y, z_1, z_2, ..., z_n, ; \alpha^j)$$

subject to its budget constraint,

$$P(y) + P^p(x, \bar{z}) \leq w + P^f(x).$$

Since utility is increasing in the consumption of all goods, a utility-maximizing household would use up all its income to acquire x, y and $\bar{z}$. The first order conditions are as follows.

- $U_y/P_y = U_{z_i}/P_{z_i}$
- $U_x/[P^p_x - P^f_x] = U_y/P_y$
- $U_x/[P^p_x - P^f_x] = U_{z_i}/P_{z_i}$

B Panel Structure

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

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

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

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

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

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

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

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

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

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

C Detailed description of the data

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

A.1 Videoconferencing

The standard descriptive statistics of the monthly price for each provider of videoconferencing services are shown in table A.4. From the descriptive statistics, it can be notice that the range between the minimum and maximum prices is large [$1.0 to £750]. The standard deviation is relatively large as well, which in this case is $88.4 (more than double the average monthly price). The average price for the pooled data set is at $40.7. The likely reason why this is so is because of the presence of services that are dedicated and optimized for webinars and large online conferences.

Videoconferencing service providers cater to two types of customers 1) those that require a venue for online meetings and 2) those needing to reach a broader audience (with 100 or more participants). Zoom, one of the most popular videoconferencing service providers at the time of the writing of this manuscript, was able to cater to both types of customers. However, some service providers opted to specialize and cater to the second type of customers[21]. These services are often priced higher than those that are targeted for smaller online meetings.

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

Table A.4: Summary statistics of monthly subscription price by videoconferencing provider

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

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

In table A.5, the sample is divided into two segments: plan types that focuses on participants requiring large audiences or "webinar-focused" plans, and those that do not. We identified 21 webinar-focused plan types in our data sets. As anticipated, the monthly average price is higher for the services that are focused on webinars compared to those that are not. Plan types that are not focused on webinars have a mean monthly price of \$14.6, substantially lower than the average for the pooled data set. Due to this source of heterogeneity, we control for whether the plan type is webinar-focused in the hedonic regressions.

Table A.5: Comparison between webinar and non-webinar providers

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

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

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

Table A.7 presents the number of observations possessing each of the respective characteristic in the data set, as well as the percent share of the observation with said characteristics. As mentioned earlier (with the exception of the number of participants), each characteristic would be represented by a dummy variable with a value of 1 if the characteristic is present in the particular plan, 0 otherwise. Information on the characteristics of these services were acquired through the scraping of the service providers' websites. One limitation of this approach is that our information on characteristics are dependent of whether the service provider were able to accurately reflect the features of their services on their websites. In order to address this, we subscribed to the trial versions of these services in order to validate the presence (or absence) of the said characteristics for each of the providers.

Table A.6: Summary statistics of videoconferencing data over time

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

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

Table A.7 shows the number of plan types possessing each of the respective characteristics in Panel A and their proportion to the total number of plan types in Panel B. In terms of their count and share to the total number of participants, videoconferencing services in recent years possess more characteristics than past years. This may be a manifestation of how technology is improving these services. Moreover, the pandemic might have contributed to this forcing some service providers to offer more services because the demand for videoconferencing substantially rose during the lockdowns.

Table A.7: Frequency and share of premium-exclusive features from videoconferencing across time

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

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

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

A.2 Email

The standard set of descriptive statistics was also generated for the price data on personal email (see table A.8). Similar to videoconferencing, the range for the pooled data set is noticeably large [$0.8 to $57.0]. The mean of the pooled data set is $7.8 while the standard deviation is at $10.1.

Table A.8: Summary statistics of monthly subscription price by email provider

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

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

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

Table A.9: Summary statistics of email data over time

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

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

Table A.10: Frequency and share of premium-exclusive features from email services across time

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

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

A.3 Online News

The standard descriptive statistics for the price of online news is shown in table A.11. For the pooled data set, the prices of online news subscription ranges between \$3.1 to \$67.0. The average price of news subscription is at \$15.4. It can be noticed that the average price of business and financial news providers such as the Wall Street Journal, Bloomberg, and the Financial Times are generally higher than those of the other providers. Because of this, we include a dummy variable for business-focused news providers.

Table A.11: Summary statistics of monthly subscription price by online news provider

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

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

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

Table A.12: Summary statistics of online news prices over time

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

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

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

Table A.13: Frequency and share of premium-exclusive features from online news services across time

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

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

D Hedonic regression coefficients

Table A.14: Hedonic regression results for videoconferencing

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

Note: Table shows the results of the hedonic regression in equation 19. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.15: Hedonic regression results the coefficient for the number of participants varying over time

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

Note: Table shows the results of the hedonic regression in equation 19. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16: Hedonic regression results for personal email

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

Note: Table shows the results of the hedonic regression in equation 19. Columns (1) shows the results of the classical time dummy variable model. Columns (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Hedonic regression results for online news

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

Note: Table shows the results of the hedonic regression in equation 19. Columns (1) shows the results of the classical time dummy variable model. Columns (2) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Correlation among characteristics

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

Given the nature of hedonic regressions, one would expect that all dummy variable representing the presence of a characteristics should be positive. After all, the presence of an additional feature to any good or service can only contribute positively to its price. However, this is not what we observe from the results of the hedonic regression in table A.14. The coefficient estimate for one of the variables (share control) was negative and significant at 10 percent level. Moreover, the variable representing the presence if call encryption was also negative and is significant at 5 percent level. This is in contrast with the intuition of interpreting coefficients for hedonic regressions. An individual cannot possibly have a negative value for their willingness to pay.

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

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

Table A.18: Correlation matrix of characteristics for videoconferencing

(1)

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

	Share Control	Transcription	Multiple Host	SSO	Stream	Analytics	Custom Domain	Branding	Local and International Calls	Translation	Office Integration	HD Quality	Noise Cancelation	Multi-Share	Calendar
Encryption															
Share Control	1														
Log Participants	0.154*														
Recording	0.282***														
Whiteboard	0.353***														
Screen Share	0.144*														
File Sharing	0.0512														
Breakout Rooms	0.351***														
Interactions	0.260***														
Virtual Background	0.248***														
Admin Control	0.344***														
Transcription	0.312***	1													
Multiple Host	0.211***	0.283***	1												
SSO	0.0995	0.0487	0.345***	1											
Stream	0.245***	0.507***	0.157*	-0.0422	1										
Analytics	0.252***	0.534***	0.0644	0.157*	0.262***	1									
Custom Domain	0.0779	0.186**	0.298***	0.200**	-0.207**	0.0121	1								
Branding	0.283***	0.624***	0.0675	-0.0656	0.376***	0.335***	0.238***	1							
Local and International Calls	0.254***	0.593***	0.133*	-0.0484	0.247***	0.556***	0.00166	0.513***	1						
Translation	0.301***	0.636***	-0.111	-0.197**	0.752***	0.469***	-0.165*	0.573***	0.464***	1					
Office Integration	-0.0950	-0.149*	-0.0870	0.382***	-0.115	0.197**	-0.101	-0.305***	-0.0215	-0.172**	1				
HD Quality	0.542***	0.195**	0.314***	0.122	0.384***	0.293***	-0.158*	0.0591	0.325***	0.283***	0.00858	1			
Noise Cancelation	-0.0550	-0.0693	-0.0280	-0.0470	-0.0585	0.0885	-0.0367	-0.0768	-0.0907	-0.0440	-0.0380	-0.0892	1		
Multi-Share	0.360***	0.114	0.494***	0.250***	-0.0936	0.231***	0.268***	0.112	0.0611	-0.106	-0.125	0.186**	0.304***	1	
Calendar	-0.102	-0.338***	0.151*	0.224***	-0.133*	-0.341***	0.0249	-0.174**	-0.326***	-0.391***	0.0490	-0.228***	0.0969	0.0185	1
Permanent Rooms	0.0207	-0.254***	-0.00856	0.220***	-0.210***	-0.209**	-0.173**	-0.237***	-0.295***	-0.190**	-0.00749	-0.0820	-0.0458	-0.00123	0.370***

t statistics in parentheses

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

Table A.19: Correlation matrix of characteristics for personal email

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

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

Table A.20: Correlation matrix of characteristics for online news

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

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

F Quality-Adjusted Price Indices

In a typical cross-section hedonic regression, the intercept term would represent the quality-adjusted price index[22] of the good being analyzed (see de Haan and Diewert (2013)). An alternative to the cross-section hedonic regression is the time dummy variable model, which allows for the intercept term to vary over time. The model can be written as,

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

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

In this paper, however, we emphasized a modified time dummy variable model, which allows for the intercept term to vary for different service provider for different points in time. The specification for the model is given by equation 19. We present the quality adjusted price indices for both specifications in tables A.21, A.22, and A.23.

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

Table A.21: Coefficient estimates of the service provider dummies for videoconferencing

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

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

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

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

Note: Table shows the coefficient estimates for the year fixed effects and service provider dummies equation 19. Columns (1) and (2) shows the results of the classical time dummy variable model. Columns (3) and (4) shows the hedonic regressions that include the interaction term between service provider fixed effects and time dummies. Coefficients estimates for the service provider fixed effects are not displayed. Standard errors in parentheses are clustered around service providers. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.22: Coefficient estimates of the service provider dummies for personal email

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

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

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

Table A.23: Coefficient estimates of the service provider dummies for online news

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

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

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

G Price Estimates

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

Table A.24: Imputed price estimates for videoconferencing

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

Note: The table shows the the price estimates for each year. Panel A estimates the prices using equation 20. Panel B presents an alternative model, where the variable for the log of participants was interacted with the time dummies. All estimates are in USD.

Table A.25: Imputed price estimates for personal email

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

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

Table A.26: Imputed price estimates for online news

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

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

A.2 Time Dummy Variable Model

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

$$\hat{p}^t = exp(\tau^t) \times exp(\beta_1 log(z_1)) \times exp(0.5Var(\varepsilon_{ij})). \tag{A.2}$$

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

Table A.27: Price estimates using the time dummy variable model for videoconferencing

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

Note: The table shows the the price estimates for each year using the time dummy variable model in equation A.2 All estimates are in USD.

Table A.28: Price estimates using the time dummy variable model for personal email

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

Note: The table shows the the price estimates for each year using the time dummy variable model in equation A.2 All estimates are in USD.

Table A.29: Price estimates using the time dummy variable model for online news

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

Note: The table shows the the price estimates for each year using the time dummy variable model in equation A.2 All estimates are in USD.

H Robustness check

Table A.30: Robustness Check of Price Estimates for Videoconferencing

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

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

Table A.31: Robustness Check of Price Estimates for Personal Email

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

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

Table A.32: Robustness Check of Price Estimates for Online News

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

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

I Comparison with other studies estimates

Table A.33: Comparison of WTA values with the price imputations for videoconferencing services

	Hedonic Regression	Brynjolfsson et al. (2019a)		Nguyen and Coyle (2020)					Jamison and Wang (2021)	
	May 2020 (1)	August 2003 (2)	Inflated to 2020 (3)	Skype Mean (4)	Messenger Mean (5)	Messenger Median (6)	Whatsapp Mean (7)	Whatsapp Median	Video Conferencing	Zoom
Point	0.41	0.2	0.29	21.7	73.4	0.3	115.1	22.7	337.5	44.9
Lower	0.33	0.01	0.02	16.3	64.1	0.1	103.8	13	228.9	28
Upper	0.52	2.92	4.18	27	82.7	6.5	126.5	32.4	446.2	61.9

Note: The table compares the WTA estimates from Brynjolfsson et al. (2019a), Nguyen and Coyle (2020), and Jamison and Wang (2021) with the price estimates from the hedonic regression. Estimates from Nguyen and Coyle (2020) were based on their May 2020 data collection. Estimates by Brynjolfsson et al. (2019a) in column (2) were inflated to 2020 prices using Dutch CPI inflation from 2017 to 2020. All estimates are in USD.

Table A.34: Comparison of WTA values with the price imputations for email and online news

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

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

J Gross Value of free digital goods, levels

A.1 Baseline estimates

Table A.35: Gross value of digital goods and Household Final Consumption Expenditures, at current prices

	2017	2018	2019	2020
Point Estimate	5,420	5,520	5,495	5,940
Lower	3,215	3,110	2,946	2,902
Upper	9,193	9,850	10,280	10,275
HFCE	1,272,638	1,318,800	1,350,697	1,225,209
GDP	2,068,757	2,141,792	2,218,439	2,112,039

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

Table A.36: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2018)

	2017	2018	2019	2020
Point Estimate	5,394.6	5,520.2	5,676.5	5,860.1
Lower	3,041.0	3,109.8	3,202.6	3,302.6
Upper	9,621.2	9,849.6	10,117.8	10,453.9
HFCE	1,305,386	1,318,800	1,333,795	1,200,329
GDP	2,115,293	2,141,792	2,172,810	1,959,220

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

A.2 Adjusted using Ofcom data

Table A.37: Gross value of digital goods adjusted for Ofcom data, at current prices

	2017	2018	2019	2020
Point Estimate	5,420	5,520	5,495	6,554
Lower	3,215	3,110	2,946	3,248
Upper	9,193	9,850	10,280	11,367

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

Table A.38: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2018)

	2017	2018	2019	2020
Point Estimate	5,394.6	5,520.2	5,676.5	6,444.7
Lower	3,041.0	3,109.8	3,202.6	3,636.3
Upper	9,621.2	9,849.6	10,117.8	11,484.5

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

A.3 Accounting for multiple provider use

Table A.39: Gross value of digital goods and Household Final Consumption Expenditures, at current prices

	2017	2018	2019	2020
Point Estimate	12,937	13,051	12,164	13,329
Lower	7,869	7,399	6,380	6,136
Upper	21,417	23,206	23,266	22,684

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

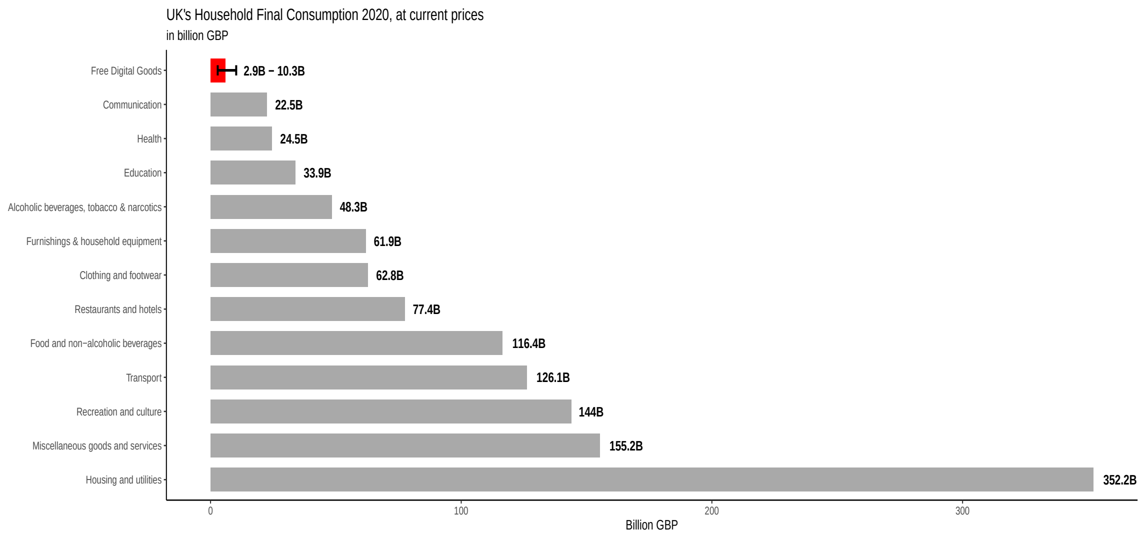
Table A.40: Gross value of digital goods and Household Final Consumption Expenditures, at constant prices (100=2018)

	2017	2018	2019	2020
Point Estimate	12,856.2	13,050.6	13,343.8	13,597.1
Lower	7,281.5	7,399.4	7,557.7	7,726.7
Upper	22,877.9	23,206	23,745.5	24,138.9

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

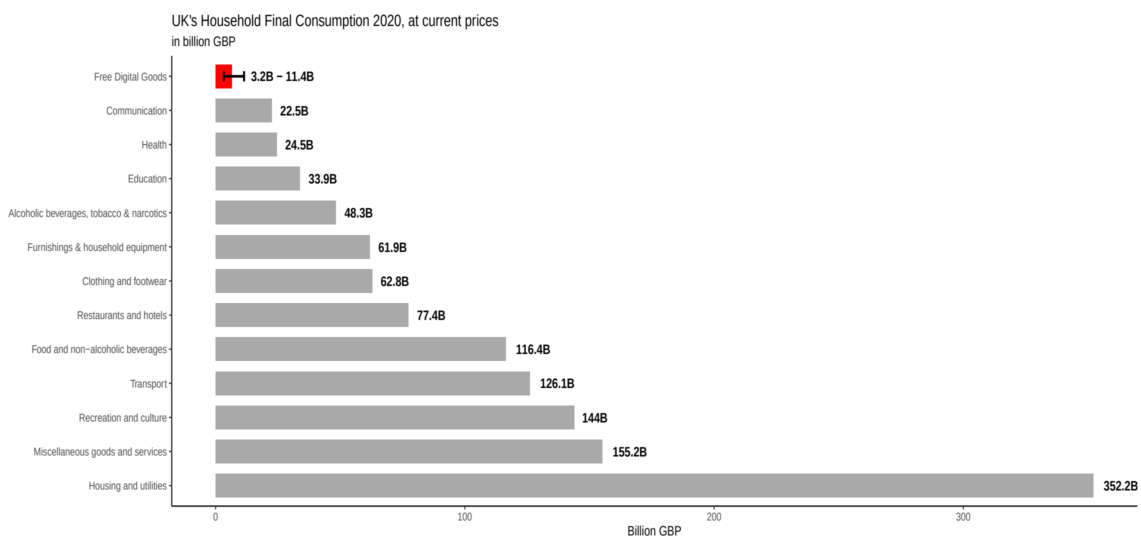
K Comparison with other expenditure items

Figure A.1



Note: The figure compares the current price estimates of gross value of free digital goods in table A.35 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.

Figure A.2



Note: The figure compares the current price estimates of gross value of free digital goods (adjusted using Ofcom data) in table A.39 with other expenditure items under UK's HFCE for 2020. HFCE data is sourced from the ONS.