How should we measure the digital economy?

By

Avinash Collis

B.Tech. Computer Science and Engineering
Indian Institute of Technology Delhi, 2011

M.S. Management
ESSEC Business School and University of Mannheim, 2014

S.M. Management Research
Massachusetts Institute of Technology, 2017

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2020

Signature of Author:____
Department of Management
March 3, 2020

Certified by: ____
Erik Brynjolfsson
Schussel Family Professor of Management Science
Thesis Supervisor

Accepted by: ____
Catherine Tucker
Sloan Distinguished Professor of Management
Professor, Marketing
Faculty Chair, MIT Sloan PhD Program

How should we measure the digital economy?

By

Avinash Collis

Submitted to the Department of Management on March 3, 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Management

ABSTRACT

Gross domestic product (GDP) measures production and is not meant to measure well-being. While many people nonetheless use GDP as a proxy for well-being, consumer surplus is a better measure of consumer well-being. This is increasingly true in the digital economy where many digital goods have zero price and as a result, the welfare gains from these goods are not reflected in GDP or productivity statistics. Chapter 1 proposes a way of directly measuring consumer's economic well-being using massive online choice experiments. It finds that digital goods generate a large amount of consumer surplus that is currently not captured in GDP. For example, the median Facebook user needed a compensation of around $48 to give it up for a month. Building up on these results, Chapter 2 extends the GDP framework to include welfare gains from new and free goods and construct a new metric called GDP-B, where B stands for benefits. It finds that including the welfare gains from Facebook would have added between 0.05 and 0.11 percentage points to GDP-B growth per year in the US. Chapter 3 proposes a way of measuring network effects on multi-sided platforms using choice experiments. It also models digital platforms allowing for heterogeneity in demand elasticity and network effects across users of different types. It then calibrates the model using an empirical application to Facebook and simulates six different taxation and regulatory policies. Chapter 4 looks at the impact of social media on subjective well-being and academic performance through a randomized controlled trial of University students. Chapter 5 summarizes the research agenda moving forward and concludes with a framework for measuring different aspects of well-being in the digital economy.

Thesis Supervisor: Erik Brynjolfsson

Title: Schussel Family Professor of Management Science, Professor of Information Technology, MIT Sloan School of Management

Acknowledgements

I could not have chosen a better place than MIT to do my PhD. I thank my advisor Erik Brynjolfsson for supporting me during my PhD journey. I could not have imagined the impact my dissertation research would have before I started working with Erik. He encouraged me to think big and aim to solve questions having direct impact on policy and society.

I thank my committee members Catherine Tucker and Sinan Aral for their support in my development as a researcher. My wonderful co-authors Felix Eggers, Seth Benzell and Kevin Fox made it fun to write papers.

My fellow PhD students and post docs in the department including Ananya Sen, Daniel Rock, Shan Huang, Sebastian Steffen, Xiang Hui and Dave Holtz were instrumental in keeping me sane and creating long lasting memories over the past 6 years. I made some dear friends across Sloan including Heather Yang, Brittany Bond, Michael Wahlen, Duanyi Yang, Madhav Kumar and Caroline Fry and they made the PhD process much more enjoyable. I thank Naomi Stephen, David Verrill and the rest of the team at the MIT Initiative on the Digital Economy for incredible support throughout my dissertation.

I thank my parents Jamuna and Sudhakar and my brother Anvesh for their unconditional love and setting me up for success since my childhood. Finally, I thank the love of my life Manuela Collis for everything. I dedicate this thesis to her.

Chapter 1 - Using Massive Online Choice Experiments to Measure Changes in Well-being

Abstract

GDP and derived metrics (e.g., productivity) have been central to understanding economic progress and well-being. In principle, the change in consumer surplus (compensating expenditure) provides a superior, and more direct, measure of the change in well-being, especially for digital goods, but in practice, it has been difficult to measure. We explore the potential of massive online choice experiments to measure consumers' willingness to accept compensation for losing access to various digital goods and thereby estimate the consumer surplus generated from these goods. We test the robustness of the approach and benchmark it against established methods, including incentive compatible choice experiments that require participants to give up Facebook for a certain period in exchange for compensation. The proposed choice experiments show convergent validity and are massively scalable. Our results indicate that digital goods have created large gains in well-being that are missed by conventional measures of GDP and productivity. By periodically querying a large, representative sample of goods and services, including those which are not priced in existing markets, changes in consumer surplus and other new measures of well-being derived from these online choice experiments have the potential for providing cost-effective supplements to existing national income and product accounts.

"If you don't know where you're going, you might not get there." -- Yogi Berra

1. Introduction

Digital technologies have transformed both the nature of production and the types of goods and services consumed in modern economies. Yet, our measurement framework for economic growth and well-being has not fundamentally changed since the 1930s. In principle, a more comprehensive approach is now feasible. By using massive online choice experiments to estimate changes in consumer surplus (compensating variation) we can supplement the traditional metrics based on Gross Domestic Product (GDP).

GDP measures the monetary value of the purchases of all final goods by households, businesses and government. It is the most widely used measure of economic activity and heavily influences policymakers in setting economic objectives and enacting interventions. GDP has been heralded as one of the greatest inventions of the 20th century by Paul Samuelson and William Nordhaus (Landefeld 2000). Economists and journalists routinely use GDP as if it were a welfare measure. Media articles regularly mention that the "economy grew by x%"[1] by measuring the growth in GDP and use this figure as a casual metric for the improvement in economic well-being. Similarly, economists widely use GDP per hour worked as a measure of productivity and infer links between productivity and improvement in living standards (OECD 2008).

Nonetheless, many economists consider GDP to be a significantly flawed measure of well-being, and several attempts have been made to design alternative measures (Stiglitz et al.

[1] E.g., "U.S. Economy Grew 1.4% in Fourth Quarter" (http://www.bloomberg.com/news/articles/2016-03-25/u-s-economy-grew-1-4-in-fourth-quarter-supported-by-consumers), "China's Economy Grew by 6.7% in First Quarter of 2016" (http://blogs.wsj.com/chinarealtime/2016/04/15/chinas-economy-grew-by-6-7-in-first-quarter-of-2016/)

2009). In fact, Simon Kuznets, the founding father of the system of national accounts that include GDP, explicitly warned against using it this way, writing, "The welfare of a nation can scarcely be inferred from a measurement of national income as defined [by the GDP.]" (Kuznets 1934).[2] Despite Kuznets' warning, growth in GDP is still the most widely used indicator of progress in our economic well-being.

For goods with a non-zero price, in theory it is often possible to infer welfare from national accounts including GDP measures (Hulten 1978; Jorgenson and Slesnick 2014), although in practice official estimates of welfare are not published. Research has looked at factors such as introduction of new goods, intangibles, quality adjustments and household production, in which GDP is biased away from welfare, and ways to correct these biases have been proposed. Feldstein (2017) provides an excellent survey and concludes that official measure significantly underestimate the true growth of GDP, personal income and productivity.

Using GDP as a welfare measure is especially problematic when prices are zero. This is the case in the emerging digital economy because most digital goods have nearly zero marginal cost and often a zero price. This makes it difficult to discern their contributions to welfare by looking at GDP calculations (Brynjolfsson and Saunders 2009; Brynjolfsson and McAfee 2014). For instance, although information goods have unquestionably become increasingly ubiquitous and important in our daily lives, the share of the information sector as a fraction of the total GDP (~ 4-5%) has not changed in the last 35 years (Figure 1). Moreover, in many sectors (e.g., music, media, encyclopedias) people substitute zero-price online services (e.g., Spotify, YouTube, Wikipedia) for goods with a positive price. As a result, the total revenue contributions of these sectors to GDP figures can fall even while consumers get access to better quality and more

[2] He underscored his views when accepting his Nobel Prize in 1971, saying that the conventional measures of national product (including GDP) omit various costs (e.g., pollution) and benefits (e.g., more leisure time) associated with technological innovations and predicted major changes in the way we measure the economy (Kuznets 1973).

variety of digital goods (Brynjolfsson and Saunders 2009). In other words, not only the magnitude, but even the sign of the change in well-being may be incorrectly inferred if decision makers rely solely on existing measures of GDP and productivity as a proxy for well-being.

[Insert Figure 1 here]

The societal benefits of technological advances are distinct from the expenditures on goods and services or profits to innovators. Nordhaus (2005) estimated that between 1948 and 2001 corporations were able to retain only 3.7% of the social returns from their technological advances while the remaining 96.3% of social returns went to consumers. Consumer surplus thus reflects most of the returns to improvements in technology.

Historically, the change in consumer surplus hasn't been widely used as a measure of change in well-being not because it is a poor measure of well-being, but because it is difficult to measure at scale. Estimating demand curves using traditional market data requires exogenous variations that shift the supply curve but not the demand curve, and it has not been practical to identify these variations for large bundles of goods.

However, with advances in digital technologies, it is now feasible to collect data about thousands of goods easily. Private sector organizations such as Microsoft, Amazon, Google and Facebook routinely conduct millions of online experiments to better understand consumer preferences and behavior. This scale of experimentation and inference would have been infeasible 20 years ago but is now routine at many organizations.

In this research, we propose a way of measuring changes in consumer surplus, not only for goods and services in the digital economy but also more broadly. Specifically, we implement a series of discrete choice experiments that measure consumers' willingness to accept payments in exchange for losing access to various goods. These experiments allow us to estimate the demand curves for these goods using data from thousands of consumers that are representative of the US population. We conclude that our approach is easily scalable and can be used to develop a system that tracks changes in consumer surplus of numerous goods and services in (near) real time via massive online choice experiments.

The paper proceeds as follows. In section 2, we illustrate the ways that GDP and consumer surplus change when prices change or new products are introduced, and the implications for welfare estimates. Section 3 describes the key methodologies we use to empirically assess consumer surplus. Section 4 presents results and sensitivity analyses of the proposed method. Section 5 applies the method to a broader set of goods. Section 6 concludes with a summary and discussion.

2. Background

2.1 GDP, consumer surplus and well-being

Perhaps no one has described the shortcomings of GDP[3] as a welfare measure as eloquently as Robert F. Kennedy:

> *Gross National Product counts air pollution and cigarette advertising, and ambulances to clear our highways of carnage. It counts special locks for our doors and the jails for the people who break them. It counts napalm and counts nuclear warheads and armored cars for the police to fight the riots in our cities...*
>
> *Yet the gross national product does not allow for the health of our children, the quality of their education or the joy of their play. It does not include the beauty of our poetry or the*

[3] Kennedy was technically discussing GNP, but his comments are equally applicable to GDP.

strength of our marriages, the intelligence of our public debate or the integrity of our public officials.

It measures neither our wit nor our courage, neither our wisdom nor our learning, neither our compassion nor our devotion to our country, it measures everything in short, except that which makes life worthwhile.[4]

Kennedy's poetic words contribute much to our understanding (if not to our GDP!). Subsequently, there have been a number of efforts to create a more comprehensive estimate of well-being. Since 2012, the United Nations Sustainable Development Solutions Network published an annual *World Happiness Report,* ranking countries based on measures of happiness (Helliwell et al. 2017). Jones and Klenow (2016) propose a measure that incorporates consumption, leisure, mortality and inequality to measure the economic well-being of a country. There is a growing stream of literature focusing on measuring subjective well-being and life satisfaction. However, while progress has been made (Krueger and Stone 2014), a survey of leading macroeconomists indicates that we are a long way off from reaching consensus on how to measure well-being so that they are reliable for policymaking (den Haan et al. 2017).

In this paper, we seek to stick more closely to a traditional microeconomic framework. In particular, we focus on the changes in consumer surplus generated by changes in consumption of digital goods and discuss ways in which our approach can be expanded to more goods and services. Brynjolfsson and Saunders (2009) paraphrase Robert Solow in noting that the influence of the information age is seen everywhere except in the GDP statistics. Almost all of us use more and more digital goods such as search engines, smartphones, social networking sites, and e-commerce platforms, but their revenues don't always reflect this increased use.

---

[4] Robert Kennedy speaking at University of Kansas in 1968 (Ref: http://www.jfklibrary.org/Research/Research-Aids/Ready-Reference/RFK-Speeches/Remarks-of-Robert-F-Kennedy-at-the-University-of-Kansas-March-18-1968.aspx).

One of the hypothesized explanations for productivity slowdown in the US since the past decade is that existing economic indicators (including GDP) do not properly measure the contributions of the latest wave to technological innovation, particularly digital goods and services (Brynjolfsson, Rock and Syverson 2017). Whereas average annual labor productivity growth was 2.8% per year over 1995-2004, it shrunk to 1.3% per year over 2005-2015. An optimistic interpretation is that recent productivity gains due to innovations in IT-related goods and services are not properly reflected in the current productivity measures (e.g., Brynjolfsson and McAfee 2014; Aeppel 2015; Hatzius 2015). However, recent literature (Byrne et al. 2016; Syverson 2016) has emphasized that while productivity mismeasurement may have been substantive in recent years, it was also likely substantive in the past, so its power to explain the productivity slowdown is limited.

Although motivated in part by this puzzle, our research focuses on the more fundamental issue that GDP, and thus productivity, is not a direct measure of well-being in the first place. Thus, whether or not GDP or productivity mismeasurement has grown is a distinct, albeit related, question from how well-being is changing. The gap between production (as measured by GDP) and well-being has been an issue since GDP was invented and, as we illustrate below, it is arguably an even bigger issue in the current digital era.

Consider the case of the music industry. Consumers shifted from buying physical units such as CDs, cassettes and vinyl records to downloading or streaming songs digitally through platforms such as iTunes, Pandora and Spotify. Digital goods have near-zero marginal cost and are hence priced much lower (often even at zero) than physical goods. Between 2004 and 2008, consumers listened to more music (units of music purchased increased from under 1 billion to over 1.5 billion without counting illegal downloads), but the recording industry's revenues declined by 40% (Brynjolfsson and Saunders 2009), and this trend has continued. Moreover, Waldfogel (2012) provides compelling evidence that the quality of music has likely increased since 1999. Therefore, although the financial contribution of music industry to GDP statistics has decreased, consumer well-being has presumably increased; consumers are listening to more and better music.

The relationships among GDP, consumer surplus and well-being can be understood by looking at three illustrative cases. First, consider a case that broadly describes many classic physical goods such as cars: consumer surplus is more or less proportional to firm revenue (Figure 2). Keeping the supply curve fixed, as more consumers enter the market, the size of the market increases, and the demand curve simply angles further to the right. In this case, both consumer surplus and quantity sold increase approximately proportionately.[5] The increased quantity sold shows up in GDP statistics as sales revenues increase, and hence both GDP and consumer welfare move in the same direction. At a given price, doubling the number of cars sold is likely to roughly double the revenues and contribution to GDP and consumer surplus. A similar logic applies to many services like haircuts, meals served or windows washed.

[Insert Figure 2 here]

A second case describes many purely digital goods such as email, messaging apps, Facebook and Google search, which have essentially zero marginal cost and are typically offered to the consumers for free. Although some digital goods may earn revenues from advertising, this is an intermediate good and does not contribute to GDP. As the value of these free goods increases, consumer surplus will also increase, but this change in well-being does not necessarily

[5] In the special case of horizontal supply curve and thus constant price, the effect is exactly proportional.

accrue to GDP (Figure 3). GDP may be completely unchanged due to this shift, even though consumers are better off.

[Insert Figure 3 here]

A third case illustrates the transitional situation faced by a number of sectors, in which physical goods and services are being substituted with digital goods and services. An apropos example of such a transition good is an encyclopedia. Since the 2000s, people have increasingly flocked to Wikipedia to get information about a wide variety of topics updated in real time by volunteers. In 2012, Encyclopedia Britannica, which had been one of the most popular encyclopedias, ceased printing books after 244 years (Pepitone 2012). Wikipedia has over 60 times as many articles as Britannica had, and its accuracy has been found to be on a par with Britannica (Giles 2005). Far more people use Wikipedia than ever used Britannica – demand and well-being have presumably increased substantially. But while the revenues from Britannica sales were counted in GDP statistics, Wikipedia has virtually no revenues and therefore doesn't contribute anything to GDP other than a few minimal costs for running servers and related activities and some voluntary contributions to cover these costs. Likewise, the transition from chemical to digital photography followed a similar arc. What's more, many people now have digital maps, streaming music, online newspapers, and other services available for no extra cost once they are able to access the Internet on mobile devices or home computers. For such transition goods, consumer surplus increases as free access spurs demand, but revenue decreases as prices become zero (Figure 4). Hence GDP and consumer welfare actually move in *opposite* directions.

[Insert Figure 4 here]

Figures 3 and 4 suggest that changes in consumer surplus are an important supplement to GDP as a measure of well-being for the current digital economy for either transition goods or purely digital goods. This is likely to become increasingly relevant as more and more goods transition from physical to digital in a variety of areas, including financial advising, customer service and law.

Total surplus can be thought of as the sum of consumer surplus and producer surplus.[6] While producer surplus cannot be inferred from consumer surplus, when it comes to technological advances, firms have typically been able to appropriate only a small fraction of the social returns (Nordhaus 2005). Accordingly, we can focus on consumer surplus. If the share of producer surplus contribution to the total social surplus remains relatively stable, then our results would have to be scaled up only slightly if one wanted to estimate total surplus. However, Furman and Orszag (2015) provide evidence that the top performing companies have been earning increasingly larger returns to capital. Therefore, measuring simply changes in consumer surplus might underestimate changes in total surplus more significantly if the producer surplus grows relative to the consumer surplus.

2.2 Prior work measuring consumer surplus from digital goods

Recently, there has been growing interest from researchers to estimate the changes in consumer surplus from digital goods. For instance, Greenstein and McDevitt (2011) estimate the additional consumer surplus created by broadband internet when consumers switched from dial-up to broadband. They estimate it to be between $4.8 and $6.7 billion per year in the US from 1999-2006. For 2015, this figure is estimated to be $55 billion (Syverson 2016). Although this

[6] More generally, there may also be externalities affecting neither consumers nor producers.

approach captures the welfare gains due to better internet access, it does not capture the increasing value of the digital information goods available online.

Another stream of literature has tried to measure the value of digital information goods by measuring the time spent using them. The underlying assumption behind these papers is that the value of free digital goods can be inferred from the time consumers spend on them. Using this approach, Goolsbee and Klenow (2006) estimate the effect of consumer gains from the internet for the median US resident to be $3000 per year as of the year 2005. Brynjolfsson and Oh (2012) extend this method to include substitutability between online and offline goods (e.g., TV). After accounting for this, they estimate the average annual change in consumer surplus of the internet to be about $25 billion per year between 2007 and 2011.

Nakamura and Soloveichik (2015) estimate the value of free media by computing the online advertising revenues generated by websites. Including ad revenues from free media in GDP increases real GDP growth by 0.019% according to their estimates, reflecting in part adjustments to price deflators. However, advertising is an intermediate good so advertising revenues do not contribute directly to GDP and do not track the value for consumers. More generally, advertising revenues are not proportional to consumer surplus. For example, in 2011 Google earned around $36 billion ad revenue (Miller 2012) while Varian (2011) estimated the consumer surplus of Google to be between $65-$150 billion. Spence and Owen (1977) argue that advertisers pay for numbers of views regardless of whether these views created low or high value for a consumer. For example, advertising revenues can be high for a program of broad interest (more views) but welfare need not be very high because consumers might only be marginally interested in that program. Conversely, for a niche program that is valued very highly by a small group of consumers, welfare will be high but advertising revenues will be low.

While these estimates of consumer surplus are based on available market data, our method uses choice experiments to elicit consumers' own valuation of goods. Specifically, we ask consumers to make a choice between keeping a digital good or taking a monetary equivalent compensation when foregoing it. This approach measures willingness-to-accept rather than willingness-to-pay money and experimentally varies the offered monetary values. It therefore addresses the limitation of market data in which the price of many digital goods is zero so that the market price in conjunction with demand does not reflect their consumer surplus value. Moreover, an experimental setting may be better able to isolate consumers' valuation of goods compared to market data that is typically confounded by many other variables; although, depending on the design of the experiment, it may come at the expense of being "hypothetical," i.e., inconsequential (Carson and Groves 2007) and therefore either noisy or biased, as we discuss and address below.

3. Methodology

3.1 Approaches to measuring consumer surplus

There are two general approaches for obtaining input data to measure consumer surplus: 1) market data and 2) choice experiments or survey techniques.

Approaches based on market data analyze longitudinal or cross-sectional variation in observed market prices for a good to derive demand curves and price elasticity (e.g., Cohen et al. 2016; Greenwood and Kopecky 2013). Similarly, hedonic pricing models try to decompose the overall value of a good into the value contribution of its characteristics by applying regression-type models to the cross-sectional covariation between observed market prices and characteristics of the goods (Williams 2008). However, both of these approaches require variance in the observed market prices and are therefore not directly applicable to goods that are provided for free. Alternatively, revealed preferences can be inferred if there is a proxy for market price, e.g., time spent using the digital goods (Goolsbee and Klenow 2006; Brynjolfsson and Oh 2012).

Choice experiments and survey techniques provide more flexibility because they do not require non-zero market prices or transactions to exist and they can be applied to contingent scenarios (leading to contingent valuation studies) (Bishop et al., 2017). One approach to determining stated preferences is to ask consumers directly about their maximum willingness-to-pay (WTP) in monetary terms. This question reveals a (potentially ratio-scaled) measure of a consumer's value of the good. However, this type of question can be less reliable because consumers are not used to formulating their own prices and because they may feel an incentive to hide their true preferences (Miller et al. 2011; Carson and Groves 2007).

The introduction of non-hypothetical, incentive compatible variants to elicit WTP in the form of auctions (e.g., Vickrey auctions, Vickrey 1961) or lotteries (e.g., BDM, Becker, DeGroot, and Marschak 1964; Wertenbroch and Skiera 2002) mitigates some of these disadvantages, but at the expense of being more complex and by introducing (artificial) competitive pressure in auctions (Carson, Groves, and List 2014; Völckner 2006). These incentive compatible direct question formats may thus be ill-suited to either digital goods, in which supply is not restricted, or to large-scale online choice experiments in which consumers need to easily comprehend and answer a preference-related question.

An alternative, indirect form of measuring preferences are discrete choice experiments (DCE) (Louviere, Hensher, and Swait 2000). DCEs ask consumers to choose between specific options and select the alternative that they value most. By experimental variation of the characteristics of the presented options (including prices) and applying logit or probit estimation models, it is then possible to estimate consumers' utility function for the characteristics, i.e., their valuation of features and sensitivity to price changes. DCEs have become a common synonym for choice-based conjoint experiments that typically involve about eight to 12 sequential choice tasks that present multiple alternatives, e.g., two to five alternatives, with variations on multiple attributes (Rao 2014). These DCEs have a long tradition in, among others, marketing (e.g., value of product features), transportation (e.g., valuation of travel time savings), contingent valuation (Carson et al. 2003), and are also applied to economic valuation contexts (e.g., Rosston, Savage, Waldman 2011). They are widely relied upon in the legal proceedings to estimate values of goods for the purposes of damages calculations (e.g., in the 2011-2014 Apple-Samsung lawsuit; see also McFadden 2014).

3.2 Proposed approach

We propose to measure consumer surplus of digital goods with DCEs. Instead of a conjoint-type experiment, we suggest a simpler implementation in which we only ask consumers to make a single choice among two options: Whether to keep access to a certain good or to forego the good in return for a specific amount of money. We only ask one question per consumer and vary the price points systematically between consumers. The procedure can therefore be termed single binary discrete choice (SBDC) experiment (Carson and Groves 2007; Carson, Groves, and List 2014). We deliberately elicit only limited information from each consumer, i.e., data that is nominal-scaled, with the benefit that this information can be captured faster and more reliably. Consumers only have to decide between two options instead of formulating and inputting a monetary figure themselves. Moreover, we can compensate for the lack in information at the individual level by using large-scale choice experiments and aggregating the responses from the overall sample in order to derive ratio-scaled demand data. Thus we use large (thousands of respondents), and potentially massive (hundreds of thousands or millions of respondents), sample sizes to overcome some of the limitations of earlier research relying on smaller samples. In some of the experiments, we enforce the consumers' choices, for instance be requiring them to give up Facebook for a given period before they get any payment. This makes their choices incentive-compatible: the rational thing to do is tell the truth when comparing alternatives options or being asked about valuations.

3.3 Utility theory and choice model

DCEs in general, including SBDC questions, are compatible with economic theory and can be used to estimate neoclassical Hicksian welfare measures (McFadden 1974, Carson and Czajkowski 2014). We will use utility theory and the random utility model to conceptualize the surplus that individual consumers obtain from consuming digital goods and the monetary value that they attach to them.

Specifically, we represent the utility that a consumer experiences from consuming a digital good g by $U(g)$. In our SBDC questions, utility is only affected by a change in the availability of the good with consumption quantities restricted to 1 and 0, i.e., a consumer can either use a good within a defined time period (g^1) or not (g^0). We abstract away from the intensity or duration of usage in this conceptual model but can account for it in our empirical application. We assume a constant market price of zero for the goods, which therefore does not have to be added to the utility function. We also do not need to explicitly consider the influence of other attributes, such as negative utility effects of advertising or limited privacy, because they are nested within g^1. These components can be easily added to the utility function when they are subject to experimental variation. We further assume that $U(g^1) \geq U(g^0)$, i.e., that consumers derive a non-negative utility of consuming the good (and would otherwise not use it). A measure of monetary value can then be estimated by introducing two Hicksian measures, either the compensating measure, *C*, or the equivalent measure, *E*, that have an effect on the consumer's income *y* (Carson and Czajkowski 2014), such that:

(3) $U(g^1, y - C^*) = U(g^0, y)$, or

(4) $U(g^1, y) = U(g^0, y + E^*)$,

with $C > 0$ and $E > 0$.

C^* is typically referred to as willingness-to-pay (WTP) for getting access to the good, while E^* can be seen as willingness-to-accept (WTA) to forego it.

While, theoretically, C^* should have the same magnitude as E^*, empirical studies show that typically $E^* > C^*$, e.g., due to an endowment effect (Hanemann 1991; Kahneman, Knetsch, and Thaler 1990; Kahneman, Knetsch, and Thaler 1991). It therefore becomes relevant to define the *status quo* of the valuation approach. When valuing the availability of free digital goods, it seems reasonable to focus on WTA and assume that $U(g^1, y)$ is the *status quo* because using the good requires no upfront investment $(y - C)$ from consumers.

When observing in the SBDC experiment that a consumer chooses to forego using a good for an offered amount *E* instead of keeping it, then we can assume that $U(g^0, y + E) > U(g^1, y)$, or $U(g^0, y + E) - U(g^1, y) > 0$. Therefore, only differential effects need to be considered between the choice options so that the overall income can be excluded and only the marginal effect of *E* needs to be considered. Without loss of generality, we can define the status quo utility as $U(g^1) =$

0. Consequently, a consumer will forego the good for amount E if $U(g^0, E)$ is positive, and will not if it is negative.

In order to estimate the equivalent measure E^*, we need estimates of how valuable consumers find using the good and how sensitive their choices are to differences in E. The random utility model is the standard framework to estimate the underlying utilities. It assumes that utility U consists of a systematic component V and a random component e that is inherent to consumer choice behavior and/or unobservable to the researcher (Manski 1977; Thurstone 1927), such that $U(g^0, E) = V(g^0, E) + e$. Typically, it is assumed that the systematic utility consists of part-worth utilities for each of the goods components, i.e., $V = b_0\, g^0 + b_1\, E$. The framework then allows us to express the observed choices as probabilities P within a binary logit model, i.e., the probability that a consumer chooses to forego the service (or, on an aggregate level, the share of consumers who are willing to accept E) is:

(5) $P(g_0, E) = exp(b_0\, g^0 + b_1\, E) / (1 + exp(b_0\, g^0 + b_1\, E))$

or 1 - $P(g_0, E)$, for keeping the service. The parameters can be estimated using closed-form maximum likelihood procedures. The median equivalent measure E^* is then the price that makes consumers indifferent between the two options so that $P(g_0, E^*) = 0.5$ or $b_0\, g^0 + b_1\, E = 0$, which leads to $E^* = - b_0\, g^0 / b_1$.

Here, we represent the utility function as linear in terms of monetary amounts. We will relax this assumption in the empirical application to handle non-linear terms and include further demographic variables to account for consumer differences.

4. Consumer Surplus of Facebook

We use Facebook as a useful case in order to measure the consumer surplus with SBDC choice experiments. We benchmark the approach against a BDM lottery and explore its robustness in sensitivity analyses. In section 5, we apply the proposed SBDC approach to a broader list of goods and present an additional benchmarking study using best-worst scaling.

4.1 Incentive-compatible Single Binary Discrete Choice experiment

In order to avoid any bias that may affect consumer choices when the options are purely hypothetical choices, we applied the SBDC experiment in a non-hypothetical, incentive compatible procedure to measure the consumer surplus of Facebook. We asked consumers if they would prefer to 1) keep access to the Facebook or 2) give up Facebook for one month[7] in return for a payment of $\$E$. We varied $\$E$ across twelve[8] price points (E = 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000). To make the SBDC question consequential for the consumer, we informed them that we will randomly pick one consumer out of every 200 respondents[9] and fulfill that person's selection. Specifically, we told respondents that if they choose "Keep access to Facebook" nothing will change for them, however, they will also not receive any money. If they choose "Give up Facebook and get paid $\$E$," we promised them the money in cash provided that they do not access Facebook for one month. We further informed them of our procedure for

[7] We initially restrict the time frame to one month in order to keep the incentive compatibility procedure manageable. We address the sensitivity of the valuation depending on the time frame in the sensitivity analysis.
[8] In a follow-up study, we included additional price points, i.e., \$0.01, \$5, \$200, \$500 and found consistent results.
[9] Carson, Groves, and List (2014) show that stochastically binding procedures (here: one out of every 200 respondents) do not significantly affect the results compared to deterministically binding procedures. We can confirm this result for our Facebook study, in which we also tested a condition in which one out of every 50 respondents was selected (E was kept at \$50 in this condition). We did not find significant differences in the choice behavior when varying the chances to win ($p = 0.236$).

remotely monitoring their Facebook online status and the requirement to provide their email address (see Figure A.1 in the appendix for the exact question wording and monitoring process).

We recruited consumers for this study from a professional panel provider with 2.9 million active panelists and member of several survey research organizations, including CASRO, ESOMAR, and MRA (Peanut Labs 2015). We invited respondents in June/July 2016 and 2017 to be able to measure annual changes. We targeted consumers who were 18 years or older and lived in the US. We further asked consumers to select all online services they had used in the last twelve months from a list of 14 options, including a non-existent online service. Consumers had to select Facebook in order to qualify for the survey; if they (also) selected the nonexistent service that we included in the survey, they were disqualified. We set quotas for gender, age, and US regions to match US census data (File and Ryan 2014) and applied post-stratification for education and household income.

Consumers who accessed the survey were randomly allocated to one of the tested price points. We sampled the highest and lowest price points twice as often in order to obtain more reliable estimates for the endpoints of the demand function. We received 2885 complete responses (n_{2016} = 1497, n_{2017} = 1388).

Figure 5 plots the estimated WTA demand curves, separated for 2016 and 2017.[10]

[Insert Figure 5 here]

In order to measure the WTA and quantify the annual change, we estimated a binary logit model that accounts for the magnitude of E (here, $\log(E)$ provided a better fit to the data), year (dummy variable), and whether the samples in the different years differ in sensitivity towards E. Table 1 shows the estimation results. The intercept represents the share of consumers in 2016

[10] In order to be consistent to normal practice for representing demand curves, the plots show the shares of consumers who prefer to keep using Facebook instead of being willing to accept the money. That is, we plotted the data in a way that makes it easier to see the negative effect of price.

who prefer to keep Facebook at E = \$1 (i.e., log(1) = 0). This share is estimated to be exp(1.2)/(1 + exp(1.2)) = 76.9%. This share is slightly larger in 2017 (p = 0.166) with exp(1.2 + 0.29)/(1 + exp(1.2 + 0.29)) = 81.6% but the difference is not statistically significant. In 2016, the sample's utility decreased by -0.309 with every one-unit increase in log(E), implying a median WTA_{2016} = \$48.49 per month. In other words, 50% of the Facebook users in our sample would give up all access to Facebook for one month if we paid them about \$50 or more. The Facebook users in 2017 appear to be more sensitive to differences in E (p = 0.049). A one-unit increase in log(E) results in a utility decrease of -0.309 - 0.101 = -0.410. As a consequence, consumers in 2017 were willing to accept a lower amount to give up Facebook, i.e., median WTA_{2017} = \$37.76 per month. Since the sample consists solely of Facebook users, a surplus measure also needs to consider the overall number of consumers who use Facebook. However, the share of Facebook users in the US increased from 2016 to 2017 by just 2.6%[11], which does not offset the negative tendency in median WTA.

We used bootstrapping to calculate 95% confidence intervals for the median WTA values, i.e., CI_{2016} = [\$32.04, \$72.24], CI_{2017} = [\$27.19, \$51.97]. The range of the confidence intervals illustrates the limitation of the approach in being less precise, given the current sample size. Although the median WTA values suggest a substantial drop in value, the confidence intervals are very broad, so we can't reasonably rule out that this is simply due to chance. We address the effect on precision by using larger sample sizes in the sensitivity analyses below.

[Insert Table 1 here]

We added usage and demographic variables to further understand differences in consumer value. The estimation results can be found in Table 2. The usage of Facebook per

[11] https://www.statista.com/statistics/408971/number-of-us-facebook-users/

week (self-reported, measured on a 5-point scale from “less than 1 hour” to “more than 14 hours”) is a significant predictor for the value of Facebook ($p = 0.006$). The more time a consumer spends on Facebook, the more likely they are to keep their access.[12] Similarly, the more friends someone has on Facebook (self-reported, measured on a 6-point scale from “less than 50” to “more than 1000”) the more compensation they require to leave Facebook ($p = 0.024$). In terms of activities on Facebook (measured on a 6-point scale ranging from “never” to “several times a day,”) consumers perceive significantly more value in Facebook the more they post status updates or share pictures and videos ($p = 0.010$), the more they like and comment ($p = 0.018$), and play games ($p = 0.025$). Watching videos is marginally significant ($p = 0.080$), while using the messenger and chat is associated with no additional value ($p = 0.100$). Consistently, we find significant substitution effects due other social media services, i.e., Instagram ($p = 0.025$), and video platforms, i.e., YouTube ($p = 0.003$). Thus, consumers who also use Instagram or YouTube are more likely to give up Facebook. Services that are not related to activities that provide value on Facebook show no significant substitution effects (e.g., Wikipedia, $p = 0.601$).

In terms of socio-demographics, we find significant effects for gender and age of the respondent, as well as household income. Specifically, we see that female respondents are more likely to keep Facebook than male users ($p = 0.011$). The same holds for older consumers ($p < 0.001$). The effects for household income are less consistent. Households with an income between 100K and 150K perceive significantly less value in Facebook ($p = 0.019$), while higher income households value Facebook more ($p = 0.008$). The effect is also significantly positive for consumers who preferred not to disclose their income ($p = 0.004$). Education and US region are not significant (not shown in Table 2).

---

[12] This confirms the assumptions made in estimating consumer surplus from consumer time allocation (Brynjolfsson and Oh, 2012)

[Insert Table 2 here]

To summarize, the SBDC experiment leads to plausible demand functions and plausible effects of usage and demographic variables. The results indicate that Facebook provides substantial value to consumers who would require a median compensation of about $40-$50 per month for leaving this service. We find no evidence that this valuation increased from 2016 to 2017; if anything, it may have declined somewhat. However, given the nature of choice data, the estimated median WTA values are limited in terms of precision compared to directly elicited values, which we will use as a benchmark method in the next section.

4.2 Benchmark method: BDM lottery

As a benchmark to check the convergent validity of the SBDC approach, we applied an incentive compatible BDM lottery procedure (Becker, DeGroot, and Marschak 1964) in order to elicit direct, numeric responses from consumers about their WTA. Specifically, we asked consumers about the minimum amount of money they would request in order to give up Facebook for one month. In order to achieve incentive compatibility, we informed respondents that the amount will serve as their bid in a lottery. The BDM lottery process instructs that, after the survey, a random price will be drawn from a uniform distribution of values. If the random price is higher than the bid, the respondent will be paid the random price when giving up Facebook for one month. If the random price is lower than the bid, the respondent will receive no money but can keep the access to Facebook. Thus, the rational, utility-maximizing strategy for the respondent is to bid exactly their true value for Facebook.

We conducted the BDM lottery in the lab of a European university, parallel to an incentive compatible SBDC experiment. The lab setting allowed us to explain the BDM procedure in detail and ensure that the respondents understood the pay-off mechanism. In total, 139 students took part in the lottery. We compare this sample to a sample of respondents that took part in the incentive compatible SBDC experiment of the same lab (n = 356). The SBDC procedure was identical to the Peanut Labs study but used monetary offers in Euros (€).

Figure 6 shows the estimated demand functions that result from both approaches. The SBDC derived function is closely aligned to the BDM demand function. The observed shares correlate strongly (correl. = 0.891). Fitting a regression model to the observed shares (R^2 = 0.755) shows that the BDM approach estimates a larger intercept than the SBDC approach (p = 0.013), i.e., more respondents are willing to keep Facebook even at low monetary values. This is plausible because BDM gives respondents more control over their bids and few respondents submitted low monetary values, while the SBDC approach follows a take-it-or-leave-it mechanism with exogenous monetary offers. More importantly, however, there is no significant statistical difference (p = 0.278) in the estimated price sensitivity for the two approaches. While this result gives us confidence in our estimates from the SBDC experiment, we explore its robustness in further sensitivity analyses.

[Insert Figure 6 here]

4.3 Sensitivity analyses

We assess the robustness of the SBDC approach regarding its sensitivity to a hypothetical bias, random responses, sample size, and the analyzed time frame.

4.3.1 Hypothetical bias

In order to measure the hypothetical bias, we applied a hypothetical scenario parallel to the incentive compatible SBDC experiments in section 4.1. Specifically, we conducted the same surveys as in the incentive compatible scenarios with Peanut Labs in June/July 2016 and 2017 but without informing consumers that their answers were consequential. We allocated respondents randomly to the incentive compatible (IC) and non-incentive-compatible (NIC) scenarios. In addition to the 2885 respondents in the IC studies, we tested 2878 consumers under NIC conditions ($n_{2016,NIC} = 1500$, $n_{2017,NIC} = 1378$).

For illustration, we detail the results for the 2016 study first. Figure 7 compares the observed shares between IC and NIC groups. For very low prices, i.e., a price of $1, the IC and NIC condition produce almost identical shares, which is reasonable. For higher prices, the disparities increase, leading to consistently higher shares in willingness to keep Facebook in the IC condition. The estimation of the binary logit model confirms that the IC consumers do not differ in the intercept ($p = 0.905$) but they react significantly less sensitive towards differences in *E* ($p = 0.002$, see Table 3). Consequently, the IC consumers are less attracted by the monetary offers and require a significantly higher amount in order to give up Facebook ($WTA_{IC,2016}$ = $48.49, $CI_{IC,2016}$ = [$32.04, $72.24]). Consumers in the NIC setting are satisfied with lower amounts, i.e., $WTA_{NIC,2016}$ = $13.80 per month (95% $CI_{NIC,2016}$ = [$9.80, $19.19]). Consequently, the hypothetical WTA is understated in this research context and needs to be calibrated by a factor of 3.5.

[Insert Figure 7 here]

[Insert Table 3 here]

The results for the 2017 study are consistent. In this case, the median WTA in the NIC condition is $9.18 (95% $CI_{NIC,2017}$ = [$6.07, $13.70]), compared to $37.76 in the IC scenario

($CI_{IC,2017}$ = [$27.19, $51.97].), which leads to a calibration factor of 4.1 (see appendix, Table A.1 for the full estimation model that accounts for year and group membership).

Our results suggest that the hypothetical bias can be substantial. More importantly, however, our primary interest is not the absolute amount of consumer surplus for Facebook but annual changes in value. In this case, the incentive compatible study would estimate a loss in value of 20% from 2016 to 2017, while the hypothetical study calculates a loss of 32%. Despite the hypothetical bias, the annual changes move in the same direction and are more closely aligned than the absolute valuations.

4.3.3 Effect of random answers

Random answers increase the error variance in choice model estimations. The error variance, in turn, has a negative effect on the precision, i.e., scale of the estimates *S* in logit choice models (Hauser, Eggers, and Selove 2016). Specifically, the scale *S* is inversely proportional to the error variance. The scale *S* cannot be separately identified, such that it is incorporated in the “raw” estimated utilities *b*:

$V = (S * b_0)\, g^0 + (S * b_1)\, E.$

Lower scaled estimates (more error), i.e., estimates with lower magnitude, cause the logit function to become more linear. Higher scaled estimates (less error) lead to a stepwise function that allows us to predict decisions and identify the median WTA more precisely (see Figure 8).

[Insert Figure 8 here]

The effect is demonstrated empirically in Table 4. The table shows the result of a modified bootstrapping procedure in which 1,000 subsamples were drawn from the 2016 IC

Facebook sample for illustration.[13] In each subsample, we replaced *R* randomly selected original responses with the same amount of *random* answers and re-estimated the logit model. The results show that more random noise in the answers decreases the scale of the estimates. The scale *S* is proportional to the relative share of non-random answers. Having more random answers than original responses ($R = 800$) causes the magnitude of the estimates to be less than half the size of the original estimation without additional random answers ($R = 0$). However, the median WTA (averaged across the 1,000 subsamples) as well as the absolute standard error of the estimates remain largely unaffected. Surplus measures that consider the overall demand function by integrating the demand function, here in the interval from $1 to $1000, are biased by random answers. We therefore only report WTA measures in our analyses.

[Insert Table 4 here]

The simulated results illustrate that the possibility of random answers cannot account for the observed hypothetical bias. Interestingly, when we compare how many of the observed choices can be predicted correctly based on the estimated model, we find a better fit for the NIC group (hit $\text{rate}_{\text{NIC}} = 69.9\%$) than for the IC group (hit $\text{rate}_{\text{IC}} = 62.1\%$). This suggests that consumers in the IC group faced a decision that was more difficult to make, likely because their choices were consequential. It is important to note that the misclassified choices are not necessarily due to purely random responses. These cases can also be explained by heterogeneity among consumers that is not accounted for in the estimation models, either with respect to their valuation of Facebook or regarding their general price sensitivity (or both).

4.3.4 Effect of sample size

---

[13] We obtain similar results for the NIC group and for the 2017 samples.

Next to random noise, the precision of the WTA estimates also depends on the sample size. To analyze the magnitude of the effect, we used bootstrapping with varying subsample sizes to observe the effect on standard errors and confidence intervals for the WTA estimate. Each subsample of a given size was again randomly drawn 1000 times from the original sample (IC group in 2016). As expected, Table 5 demonstrates that the standard errors of the estimates are reduced by the square-root of 2 when doubling the sample size (in this case the scale of the estimates remains largely unaffected). This general pattern also holds for the standard error of the WTA estimate. However, since WTA is a ratio of two stochastic variables, this generalization is approximate. The results show how the 95% confidence interval narrows when increasing the sample size. There is uncertainty in the measure even with a sample size of 1500. For the case of Facebook, a 95% confidence interval of ±$10 can be achieved with a sample of 6000 consumers. This result highlights the need for more massive sample sizes to measure consumer surplus precisely.

[Insert Table 5 here]

4.3.2 Effect of the analyzed time frame

In the previous incentive compatible studies, we used one month instead of one year as the time frame that respondents should forego Facebook. This raises the question to what extent consumers are sensitive to the time frame. To address this question, we conducted SBDC experiments in an incentive compatible setting in which, in addition to prices E, we varied the time frame across three periods, i.e., T = 1 week, 2 weeks, 1 month. We recruited another sample from Peanut Labs in 2017 using the same criteria as in the previous studies, but we did not screen out respondents who do not use Facebook (assuming that these respondents would accept any low monetary compensation; empirical valuations are therefore lower than in the previous study). A total of 1499 respondents were available for the analysis.

Table 6 shows the estimation results. As expected, the time frame has a significant positive effect on the probability to keep Facebook. Accordingly, the median WTAs for the different time frames are \$3.92 for one week, \$10.53 for two weeks, and \$17.61 for one month. Interestingly, these values and coefficient estimates suggest that the effect of time is not necessarily linear. As with any good, the value of Facebook depends on the context.

[Insert Table 6 here]

In order to get a better overview of the effect of time, we sampled 5021 additional respondents in a hypothetical setting using Google Consumer Surveys (see section 5). We allocated these respondents randomly to one of ten conditions that differ in the time frame: $T = 1$ hour, 1 day, 1 week, 2 weeks, 3 weeks, 4 weeks, 1 month, 2 months, 3 months, 6 months, 1 year (operationalized in the estimation model in terms of number of days). We kept E constant at \$50 in this study. Figure 9 shows the observed shares of respondents who prefer to keep Facebook at the different time frames and the predicted time function according to the binary logit model (using $\log(T)$ and $\log(T)^2$ as predictors, see Table A.2 in the appendix). It confirms a positive effect of time with increasing marginal effects. Accordingly, consumers are more likely to keep Facebook the longer the time frame, and this effect is reinforced with increasing duration. We use a time frame of one year in the large-scale studies we present next.

[Insert Figure 9 here]

5. Large-scale Studies to Measure Consumer Surplus

5.1 Google Consumer Surveys: Single Binary Discrete Choices

For the implementation of our large-scale studies, we use Google Consumer Surveys (GCS) as our primary platform. GCS allows us to run short one-question surveys inexpensively and quickly and is therefore well suited for our SBDC experiments. A number of online publishers (including news and arts/ entertainment sites) participate in GCS and host these choice experiments on their site as a gateway to access premium content (Stephens-Davidowitz and Varian 2015). Users must answer the survey in order to unlock premium content (Figure 10). Survey creators pay per response, part of which goes to the publisher for hosting it. In addition to the responses, some demographic characteristics of the respondents such as region, age, gender and income are also provided, which are inferred from IP address, location, browsing history (provided by Google's DoubleClick cookies which are also used to serve ads) and census data. Prior research has found that GCS results are very similar to those obtained from other surveys conducted by professional organizations such as Pew (Stephens-Davidowitz and Varian 2015).[14]

[Insert Figure 10 here]

We identified the most widely used apps and websites on various devices and combined them into the following eight product categories: Email, Search Engines, Maps, E-commerce, Video, Music, Social Media, and Instant Messaging. We ran SBDC surveys for each of these categories in June/July 2016 and 2017. In these studies, we asked consumers to consider giving

[14] To confirm that there is no selection bias, we compared the NIC group from the Peanut Labs sample (see section 4.3.1) to a GCS sample (n = 1451). Because Google Surveys do not screen respondents if they are Facebook users or not, unlike in the Peanut Labs study, we matched the NIC group by accounting for the share of non-Facebook users. A binary logit model confirms that there are no significant differences between both samples, either in terms of their intercept (p = 0.991) or their sensitivity towards E (p = 0.474). See appendix for details (Table A.3, Figure A.2).

up access to these categories for one year. As compensation, we offered one of 6 to 15 price levels for each product category and gathered around 500 responses per price level per year. If the median WTA was outside the range of our initial set of price levels, we increased the number of price levels in the following year in order to accommodate higher prices (for Search Engines, Email, Maps).

The observed shares and estimated demand curves are shown in Figure 11. The demand curves appear plausible and are consistent across time (solid lines represent 2016, dashed lines 2017). The annual changes suggest an increase in the valuation for these categories, albeit not statistically significant. This notion is confirmed when inspecting the median annual WTA values per year in Table 7. As in the Facebook study, the range of the confidence intervals is large, meaning that the significance of the changes cannot be estimated reliably.

According to the median WTA estimates for 2017, Search Engines ($17,530) is the most valued category of digital goods followed by Email ($8,414) and digital Maps ($3,648). One possible reason that these values are high relative to the other goods in our analysis may be the lack of effective substitutes for search engines, email or digital maps compared to the other categories in our sample. Because most consumers do not directly pay for these services, almost all of the WTA for these goods contributes towards consumer surplus. What’s more, for many people, these services are essential to their jobs, making them reluctant to give up these goods.

Video streaming services (e.g., YouTube, Netflix) are valued by consumers with a median WTA of $1,173 per year. Some consumers do pay for some of these services. However, these amounts are of the order of $10-$20 per month, or $120-$240 per year (for those who pay). Our measure suggests that the surplus the median consumers receive from these goods is a 5-10 multiple of what they actually pay (and which is visible in national accounts). The remaining categories for which we estimated the median WTA are (in descending order) E-Commerce ($842), Social Media ($322), Music ($168), and Instant Messaging ($155).

These estimates are potentially biased downwards due to lack of incentive compatibility in these studies. Nevertheless, the sum of these estimates suggests there is a significant amount of consumer surplus from digital goods and a positive tendency over time.

[Insert Figure 11 here]

[Insert Table 7 here]

The available demographic variables (gender, age, income and urban density) reported by Google were added to an extended model to determine effects for different consumer segments. These extended logit models are reported in Table A.4 in the appendix. These results reveal a number of patterns that are interesting and may have implications for research and business. For instance: The value of search engines increases by age and income and is higher for female consumers. Similar effects of age and gender can be observed for the email category. In this case, consumers in urban areas and with a median income of $50K to $75K also perceive a higher value. For maps, the effect of age on WTA follows an inverse U-shape. Middle-aged consumers of 35-44 years value maps most. Income has a positive effect on the valuation of maps. A similar inverse U-shaped effect between age and valuation can be seen for e-commerce. In this case, the maximum value is experienced by 55-64 year-old consumers. In addition, female consumers perceive a higher value from online shopping. Age has a negative effect for the video and music categories. Whereas this pattern is consistent across all age groups for videos, the negative pattern only starts at an age of 45 years or older for music. Male consumers value videos more. The music category is preferred in urban areas. Female users value social media more. The same holds for instant messaging. In this category, the youngest age group (18-24 years) perceives the highest value. Older consumers perceive significantly less value. Our approach opens the door to testing a variety of hypotheses and uncovering most such patterns relatively easily.

Our approach can be used for digital and non-digital goods alike. As an example, we also ran SBDC surveys to estimate the WTA to give up the option of eating breakfast cereal[15] for one year. Figure 12 plots the WTA demand curve for breakfast cereal. We estimate the median WTA to give up breakfast cereal to be $44.27 in the US in 2017 (95% CI_{2017}: [$37.19; $52.47]).[16] This estimate is almost identical to the results from 2016 (95% CI_{2016}: [$37.98; $49.74]). Examining non-digital goods can help us calibrate the relative importance of some of the digital goods we examine. We therefore also incorporate non-digital goods in the benchmark study using best-worst scaling.

[Insert Figure 12 here]

5.2 Benchmark method: Best-worst scaling

As a benchmark to GCS, we conducted additional choice experiments based on the best-worst scaling (BWS) approach (Flynn et al. 2007; Marley and Louviere 2005). Best-worst scaling asks consumers to repeatedly select the best and worst options from sets of alternatives. Collecting more information, both within the choice set and across sequential choice sets, for each consumer makes this approach more efficient compared to the SBDC approach, which elicits only one decision. Moreover, consumers are required to make a tradeoff when deciding

[15] Economists have studied this industry using a variety of approaches. See e.g., Hausman (1996), Schmalensee (1978), Nevo (2001), and others. Hausman (1996) estimates the consumer surplus due to entry of a new cereal brand (Apple-Cinnamon Cheerios) to be $0.3136 per person per year.

[16] This figure is in addition to the price paid by consumers for buying breakfast cereal.

which goods they perceive as most and least valuable. This may mitigate or even eliminate the systematic hypothetical bias, at least with respect to the ordinal ranking of the choices.

We used nineteen digital goods, six non-digital goods and nine price points ranging from $1 to $20,000, that consumers compared three at a time. Because we examined the value of foregoing access to specific services or amenities for one year, the price options were also expressed as losses (foregoing a specific amount of salary for one year) in order to be comparable, e.g., "earning $10,000 less for 1 year." The price sensitivity we are measuring is therefore closer to WTP than WTA.

We presented three options within each choice set for each individual so that respondents created a full ranking of the three options in a set by indicating the best and worst options. Figure 13 shows an example of such a choice set. Respondents answered 10 or 11 sets[17] in order to be exposed to each good. We randomized the allocation of goods and prices to choice sets across respondents.

[Insert Figure 13 here]

We recruited consumers for this online study via Peanut Labs in 2017. We targeted consumers that were 18 years or older and lived in the US. Consumers who did not fulfill these criteria were screened out. We controlled quotas for gender, age, and US regions to match US census data (File and Ryan 2014). In total, 503 respondents completed the study.

We estimated utility parameters using a multinomial logit model. We considered both best and worst choices in the same model by interpreting utilities from best choices as the negative of worst choices. The estimation leads to interval-scaled utility scores that represent the disutility of not having access to the goods (or earning less income) for one year, which are

[17] We used two subsamples that differed in the number of goods and number of choice sets in order to accommodate different goods and price points. One subsample (n = 204) evaluated 30 options in 10 choice sets; the other subsample (n = 299) 33 options in 11 sets.

depicted in Figure 14 (see also Table A.5 in the appendix). We have set the lowest ranked service for the US, WhatsApp, as a reference category so that utilities are expressed relative to WhatsApp. The ranking of the goods is consistent to the SBDC experiments for the eight most widely used categories using GCS, with only one exception: online shopping is valued more than maps and video streaming in the best-worst scaling approach, while we find it to be valued less in the GCS surveys. When comparing the utilities of the services to the utility scores of the price levels we find, as expected, consistently lower implied WTP values than WTA estimates according to the GCS survey. Estimating a demand function and interpolating WTP shows very strong correlation among BWS and SBDC valuations (Correl. = 0.911). Overall, comparing the results of both approaches indicates convergent validity.

[Insert Figure 14 here]

6. Discussion

With advances in information technologies, we can now gather data at a large scale in close to real time. Initiatives such as MIT's Billion Prices project[18] and Adobe's Digital Price Index[19] are collecting price data from online retailers in real time to compute price and inflation indices. We explore the potential to reinvent and supplement the measurement of economic well-being by taking advantage of the ease of gathering data in the digital era. The end goal of this research agenda is to design a scalable method of measuring changes in consumer surplus induced by technological advancements. We explore a potential way of measuring changes in consumer surplus through SBDC experiments. Our method is highly scalable and relatively

[18] http://bpp.mit.edu/

[19] https://blogs.adobe.com/digitalmarketing/analytics/introducing-digital-economy-project/

inexpensive. Therefore, it can be run at very frequent, regular intervals to track changes in consumer surplus. As argued previously, this measure can be an important complementary indicator of consumer well-being for the digital economy.

In a series of online experiments, we show that the SBDC approach leads to plausible demand functions that are consistent with other validated approaches. We find that free digital goods provide substantial value to consumers even if they don't contribute substantially to GDP and may even displace products that do contribute to GDP. We further find that our approach can detect consumers' sensitivity towards different time frames, e.g., whether consumers use (or not use) the goods for one week, one month, or one year. We find that time has a positive effect on the probability to keep a service with increasing marginal returns.[20]

In order to address the limitation of the bias in answering hypothetical questions of the proposed approach, we have compared consumers' valuation of Facebook in an incentive compatible and a hypothetical setting. We confirm that a hypothetical bias exists, such that valuations of Facebook in the hypothetical scenarios tend to be significantly underestimated. The magnitude of the bias and potential correction factors need to be analyzed further in future studies. However, the differences between hypothetical and incentive compatible approaches are much less severe when analyzing annual *changes* in valuations, rather than levels.

A major limitation of our study remains the lack of precision in our estimates. While the BEA is able to measure GDP very precisely (e.g. US GDP was reported as $19,736.5 billion in

---

[20] Some consumers seem to be willing to undergo "digital detox" for a short duration by giving up internet or individual services like Facebook either through self-control or by installing software which blocks particular sites. This might explain consumers' weaker sensitivity towards short term abstinence and raises interesting questions about neoclassical economic models of rational choice, self-control and the nature of utility functions. Economics continues to evolve to take account mental biases that deviate from traditional notions of rationality, e.g., Kahneman et al. 1990, Kahneman 2011, Thaler 2015.

the fourth quarter of 2017[21]), we are only able to provide a relatively coarse estimate of changes in consumer surplus, even in our large-scale studies. Future work should use larger sample sizes to narrow the confidence interval of the WTA estimates.

Although the median WTA is robust to random noise in the data, the overall demand functions are not: small numbers of extreme valuations can have undue influence. In contrast, focusing only on the median valuations, while much more robust to noise, limits the application of the SBDC approach to those goods that are used by at least 50% of the consumers. Thus, research can benefit from reporting other key percentiles, e.g., the valuation for people at the 90th percentile, or other benchmarks, when comparing goods to each other. Before being able to derive surplus measures along the overall demand curve, we need further evidence to confirm that the error variance in the data remains consistent over time and therefore cancels out when calculating annual changes.

Another limitation of our study is that it is biased towards people using the internet. The massive variants of our choice experiments are only accessible online, therefore people not using the internet at all are excluded. Pew estimates that about 15% of Americans don't use the internet.[22] Accordingly, our results must be interpreted as relevant to this audience, but not necessarily others.

That said, our approach is at least attempting to directly measure a concept that we know is not correctly measured by other official data. In short, we believe it is better to be approximately correct than precisely wrong.

---

[21] https://www.bea.gov/newsreleases/national/gdp/gdpnewsrelease.htm (Accessed March 2018)

[22] http://www.pewresearch.org/fact-tank/2015/07/28/15-of-americans-dont-use-the-internet-who-are-they/. Of course, to the extent that the unmeasured consumer surplus dynamics are occuring in digital goods, it may be safe to surmise that those who are not on the internet are probably not using many digital goods and have negligible effects on such surplus.

References

Aeppel, T. (2015). "Silicon Valley Doesn't Believe U.S. Productivity Is Down," Wall Street Journal, July 16, www.wsj.com/articles/silicon-valley-doesnt-believe-u-s-productivity-is-down1437100700?tesla=y&cb=logged0.28855257923714817.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. The Quarterly Journal of Economics, 118(1), 73-106.

Bates, W. (2009). Gross national happiness. Asian-Pacific Economic Literature, 23(2), 1-16.

Becker, G.M., DeGroot, M.H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. Behavioral Science, 9, 226–232.

Bishop, R. C., Boyle, K. J., Carson, R. T., Chapman, D., Hanemann, W. M., Kanninen, B., ... & Paterson, R. (2017). Putting a value on injuries to natural assets: The BP oil spill. Science, 356(6335), 253-254.

Bishop, R.C. and Heberlein, T.A. (1979). Measuring values of extramarket goods: Are indirect measures biased? American Journal of Agricultural Economics, 61(5), 926-930.

Brynjolfsson, E. and Oh, J. (2012). The attention economy: measuring the value of free digital services on the Internet. Proceedings of the 33rd International Conference on Information Systems, Orlando, December 2012.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Brynjolfsson, E., Rock, D, and Syverson, C. (2017). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. NBER Working Paper No. 24001. Issued in November 2017

Brynjolfsson, E., & Saunders, A. (2009). What the GDP gets wrong. (Why managers should care). MIT Sloan Management Review, 51(1), 95.

Byrne, D. M., Fernald, J. G., & Reinsdorf, M. B. (2016). Does the United States have a productivity slowdown or a measurement problem?. Brookings Papers on Economic Activity, 2016(1), 109-182.

Carson, R. and Czajkowski, M. (2014). The discrete choice experiment approach to environmental contingent valuation, in: Hess, S. and Daly, A. eds., 2014. Handbook of Choice Modelling. Edward Elgar Publishing.

Carson, R.T. (2012). Contingent valuation: A practical alternative when prices aren't available. The Journal of Economic Perspectives, 26(4), 27-42.

Carson, R.T., Groves, T. (2007). Incentive and informational properties of preference questions, Environmental and Resource Economics, 37(1), 181-210.

Carson, R.T., Groves, T., List, J.A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. Journal of the Association of Environmental and Resource Economists, 1(1/2), 71-207.

Carson, R.T., Mitchell, R.C., Hanemann, M., Kopp, R.J., Presser, S. and Ruud, P.A. (2003). Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. Environmental and Resource Economics, 25(3), 257-286.

Cohen, P., Hahn, R., Hall, J., Levitt, S., & Metcalfe, R. (2016). Using Big Data to Estimate Consumer Surplus: The Case of Uber (No. w22627). National Bureau of Economic Research.

Diamond, P.A. and Hausman, J.A. (1994). Contingent valuation: Is some number better than no number? The Journal of Economic Perspectives, 8(4), 45-64.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. Journal of Marketing Research, 44(2), 214-223.

Ding, M., Grewal, R. and Liechty, J. (2005). Incentive-aligned conjoint analysis. Journal of marketing research, 42(1), 67-82.

den Haan, W., Ellison, M. Ilzetzki, E., McMahon, M., and R Reis (2017) “Happiness and wellbeing as objectives of macroeconomic policy: Views of economists” https://voxeu.org/article/views-happiness-and-wellbeing-objectives-macroeconomic-policy

Feldstein, M. (2017). Underestimating the Real Growth of GDP, Personal Income and Productivity, Journal of Economic Perspectives, 31(2), 145-164

File, T. and Ryan, C. (2014). Computer and Internet Use in the United States: 2013, U.S. Census Bureau. (Accessed at: http://www.census.gov/history/pdf/2013computeruse.pdf)
Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: what it can do for health care research and how to do it. Journal of Health Economics, 26(1), 171-189.

Furman, J., & Orszag, P. (2015). A firm-level perspective on the role of rents in the rise in inequality. Presentation at “A Just Society” Centennial Event in Honor of Joseph Stiglitz Columbia University.

Giles, J. (2005). Internet encyclopaedias go head to head: Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries. Nature 438(7070), 900–1

Goolsbee, A., & 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.

Greenstein, S. and McDevitt, R.C., 2011. The broadband bonus: Estimating broadband Internet's economic value. Telecommunications Policy, 35(7), 617-632.

Greenwood, J., & Kopecky, K. A. (2013). Measuring the welfare gain from personal computers. Economic Inquiry, 51(1), 336-347.

Haab, T.C., Interis, M.G., Petrolia, D.R. and Whitehead, J.C. (2013). From hopeless to curious? Thoughts on Hausman's “dubious to hopeless” critique of contingent valuation. Applied Economic Perspectives and Policy, p.ppt029.

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

Hatzius, J. (2015). “Productivity Paradox 2.0.” Top of Mind, Issue 39, p. 6–7. Goldman Sachs. http://www.goldmansachs.com/our-thinking/pages/macroeconomic-insights-folder/the-productivity-paradox/report.pdf.

Hauser, J.R., Eggers, F., and Selove, M. (2016), The Strategic Implications of Precision in Conjoint Analysis, MIT Sloan School of Management working paper.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In The Economics of New Goods (pp. 207-248). University of Chicago Press.

Hausman, J., (2012.) Contingent valuation: from dubious to hopeless. The Journal of Economic Perspectives, 26(4), 43-56.

Helliwell, J., Layard, R., & Sachs, J. (2017). World Happiness Report 2017, New York: Sustainable Development Solutions Network.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. The Review of Economic Studies, 45(3), 511-518.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. The American Economic Review, 106(9), 2426-2457.

Jorgenson, D. W., & Slesnick, D. T. (2014). Measuring social welfare in the US national accounts. In Measuring Economic Sustainability and Progress (pp. 43-88). University of Chicago Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. Journal of Political Economy, 98(6), 1325-1348.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. The Journal of Economic Perspectives, 5(1), 193-206.

Kahneman, D. (2011). Thinking, Fast and Slow. Macmillan.

Krueger, A. B., & Stone, A. A. (2014). Progress in measuring subjective well-being. Science, 346(6205), 42-43.

Kuznets, S. (1934). “National Income, 1929-1932” 73rd Congress, 2nd Session, Senate Document no. 124.

Kuznets, S. (1973). Modern economic growth: findings and reflections. The American Economic Review, 63(3), 247-258.

Landefeld, J. S. (2000). GDP: One of the great inventions of the 20th century. Survey of Current Business, 80(1), 6-14.

List, J.A. and Gallet, C.A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20(3), 241-254.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000). Stated Choice Methods: Analysis and Applications. Cambridge University Press.

Manski, C.F. (1977). The structure of random utility models. Theory and Decision, 8(3), 229-254.

Marley, A. A. J. and Louviere, J. J. (2005). Some probabilistic models of Best, Worst, and Best-Worst choices. Journal of Mathematical Psychology. 49, 464-480.

McFadden, D.L. (1974). Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), Frontiers in Econometrics, New York: Academic Press.

McFadden, D.L. (2014). In the Matter of Determination of Rates and Terms for Digital Performance in Sound Recordings and Ephemeral Recordings (WEB IV), Before the Copyright Royalty Board Library of Congress, Washington DC, Docket No. 14-CRB-0001-WR, October 6.

Miller, K.M., Hofstetter, R., Krohmer, H. and Zhang, Z.J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. Journal of Marketing Research, 48(1), 172-184.

Miller, M. (2012). How Google made $37.9 billion in 2011. Search Engine Watch. (Accessed at: https://searchenginewatch.com/sew/news/2140712/google-usd379-billion-2011)

Morwitz, V.G., Steckel, J.H. and Gupta, A. (2007). When do purchase intentions predict sales? International Journal of Forecasting, 23(3), 347-364.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation, Environmental and Resource Economics, 30(3), 313-325.

Nakamura LI, and Soloveichik R. (2015). Valuing ‘free’ media across countries in GDP. Federal Reserve Board of Philadelphia Working Paper No. 15-25.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica, 69(2), 307-342.

Nordhaus, W. D. (2005). Schumpeterian Profits and the Alchemist Fallacy.Yale Economic Applications and Policy Discussion Paper, (6).

OECD (2008). OECD compendium of productivity indicators. (Accessed at: http://www.oecd.org/std/productivity-stats/40605524.pdf)

Peanut Labs (2015). Peanut Labs Panel Book. (Accessed at: http://news.peanutlabs.com/rs/peanutlabs/images/PL-PanelBook_2015.pdf)

Pepitone, J. (2012). Encyclopedia Britannica to stop printing books. CNN. (Accessed at: http://money.cnn.com/2012/03/13/technology/encyclopedia-britannica-books)

Plott, C.R. and Zeiler, K. (2005). The willingness to pay–willingness to accept gap, The American Economic Review, 95(3), 530-545.

Rao, V.R. (2014). Applied Conjoint Analysis (p. 56). New York, NY: Springer.

Rosston, G., Savage, S. and Waldman, D. (2011). Household demand for broadband internet service. Communications of the ACM, 54(2), 29-31.

Schmalensee, R. (1978). Entry deterrence in the ready-to-eat breakfast cereal industry. The Bell Journal of Economics, 9(2), 305-327.

Spence, M., B. Owen. 1977. Television programming, monopolistic competition, and welfare. Quarterly Journal of Economics, 91(1). 103–126.

Stephens-Davidowitz, S., H. Varian. (2015). A Hands-on Guide to Google Data. (Accessed at: http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf)

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). Report by the commission on the measurement of economic performance and social progress. Paris: Commission on the Measurement of Economic Performance and Social Progress.

Syverson, C. (2016). Challenges to Mismeasurement Explanations for the US Productivity Slowdown (No. w21974). National Bureau of Economic Research.

Thaler, R. H. (2015). Misbehaving: The Making of Behavioral Economics. WW Norton & Company.

Thurstone, L.L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273.

Varian H. (2011, March 29). The economic value of Google. Presentation, San Francisco.

Varian, H. (Sep 2016). A microeconomist looks at productivity: A view from the valley. Presentation at Brookings (Accessed at: https://www.brookings.edu/wp-content/uploads/2016/08/varian.pdf)

Vickrey, W. (1961). Counterspeculation, auctions and competitive sealed tenders. Journal of Finance, 16, 8–37.

Völckner, F. (2006). An empirical comparison of methods for measuring consumers' willingness to pay. Marketing Letters, 17(2), 137-149.

Waldfogel, J. (2012). Copyright protection, technological change, and the quality of new products: Evidence from recorded music since Napster. Journal of Law and Economics, 55(4), 715-740.

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. Journal of Marketing Research, 39, 228–241.

Whitehead, J.C. (2002). Incentive incompatibility and starting-point bias in iterative valuation questions. Land Economics, 78(2), 285-297.

Williams, B. (2008). "A hedonic model for Internet access service in the Consumer Price Index." Monthly Labor Review, July, 33-48.

Wlömert, N. and Eggers, F. (2016). Predicting new service adoption with conjoint analysis: external validity of BDM-based incentive-aligned and dual-response choice designs. Marketing Letters, 27(1), 195-210.

Tables and Figures

Table 1: Estimation results of binary logit model comparing consumers' valuation of Facebook in 2016 and 2017

	beta	Std. Error	z	p
(Intercept)	1.200	0.125	9.624	<0.001
log(E)	-0.309	0.030	-10.327	<0.001
Year_2017	0.290	0.209	1.385	0.166
Year_2017*log(E)	-0.101	0.051	-1.966	0.049

Table 2: Facebook value diagnostic

	beta	Std. Error	z	p
(Intercept)	0.321	0.254	1.261	0.207
log(E)	-0.346	0.032	-10.801	<0.001
Year_2017	0.306	0.220	1.392	0.164
Year_2017*log(E)	-0.105	0.054	-1.940	0.052
Facebook usage per week (scale)	0.117	0.043	2.740	0.006
Facebook number of friends (scale)	0.074	0.033	2.257	0.024
Facebook activity: Posting status updates or sharing pictures and videos (scale)	0.095	0.037	2.577	0.010
Facebook activity: Liking and commenting (scale)	0.093	0.039	2.363	0.018
Facebook activity: Playing games (scale)	0.054	0.024	2.234	0.025
Facebook activity: Using the messenger or chat (scale)	0.053	0.032	1.643	0.100
Facebook activity: Watching videos (scale)	0.066	0.037	1.748	0.080
Instagram user	-0.225	0.100	-2.245	0.025
Skype user	-0.067	0.092	-0.733	0.464
Google maps user	-0.076	0.107	-0.712	0.477
Google search user	-0.188	0.127	-1.482	0.138
YouTube user	-0.420	0.141	-2.983	0.003
Wikipedia user	0.049	0.096	0.510	0.610
Gender female (reference)	(0.000)			
Gender male	-0.220	0.086	-2.546	0.011
Age 18-24 (reference level)	(0.000)			
Age 25-34	-0.012	0.152	-0.079	0.937
Age 35-44	0.245	0.151	1.620	0.105
Age 45-54	0.367	0.155	2.371	0.018
Age 55-64	0.590	0.161	3.669	<0.001

Age 65+	0.936	0.176	5.335	<0.001
Income less than 25K (reference)	(0.000)			
Income 25K to 50K	0.081	0.140	0.578	0.563
Income 50K to 100K	-0.030	0.131	-0.229	0.819
Income 100K to 150K	-0.370	0.157	-2.355	0.019
Income 150K or more	0.441	0.165	2.671	0.008
Income "prefer not to answer"	0.784	0.273	2.873	0.004

Table 3: Estimation results of binary logit model comparing IC and NIC scenarios (2016 study)

	beta	**Std. Error**	**z**	**p**
(Intercept)	1.178	0.135	8.726	<0.001
log(E)	-0.449	0.034	-13.147	<0.001
IC	0.022	0.184	0.119	0.905
IC*log(E)	0.140	0.045	3.076	0.002

Table 4: Effects of random answers on estimate errors

Random sample R	Non-random sample	**Mean intercept**	**Mean beta log (E)**	Std. error Intercept	Std. error beta log(E)	WTA	Surplus	**Scale S**
800	700	**0.517**	**-0.135**	0.139	0.033	$46.29	$430.53	**0.431**
400	1100	**0.846**	**-0.218**	0.149	0.035	$48.52	$390.61	**0.700**
200	1300	**1.020**	**-0.262**	0.151	0.037	$49.06	$371.32	**0.844**
100	1400	**1.122**	**-0.289**	0.157	0.038	$48.91	$359.69	**0.929**
0	1500	**1.206**	**-0.311**	0.163	0.039	$48.18	$349.72	**(1.000)**

Table 5: Effects of sample size on estimate errors

Sample size	Mean intercept	Mean beta log (E)	**Std. error Intercept**	**Std. error beta log(E)**	WTA	**95% CI lower**	**95% CI upper**
200	1.242	-0.319	**0.462**	**0.110**	$49.65	**$13.13**	**$187.73**
400	1.227	-0.316	**0.324**	**0.077**	$48.72	**$21.16**	**$112.28**
800	1.214	-0.311	**0.226**	**0.053**	$49.30	**$27.83**	**$87.27**
1500	1.206	-0.311	**0.163**	**0.039**	$48.18	**$31.69**	**$73.26**

Table 6: Estimation results for the marginal effect of time (IC study)

	beta	Std. Error	z	p
(Intercept)	0.324	0.126	2.572	0.010
log(E)	-0.237	0.024	-10.009	<0.001
Time 1 week (reference)	(0.000)			
Time 2 weeks	0.235	0.135	1.734	0.083
Time 1 month	0.357	0.133	2.688	0.007

Table 7: Median WTA estimates for eight digital goods categories

Category	WTA/year 2016	WTA/year 2017	95% CI 2016		95% CI 2017		n
			lower	upper	lower	upper	
All Search Engines	$14,760	$17,530	$11,211	$19,332	$13,947	$22,080	8,074
All Email	$6,139	$8,414	$4,844	$7,898	$6,886	$10,218	9,102
All Maps	$2,693	$3,648	$1,897	$3,930	$2,687	$5,051	7,515
All Video	$991	$1,173	$813	$1,203	$940	$1,490	11,092
All E-Commerce	$634	$842	$540	$751	$700	$1,020	11,051
All Social Media	$205	$322	$156	$272	$240	$432	6,023
All Messaging	$135	$155	$98	$186	$114	$210	6,076
All Music	$140	$168	$112	$173	$129	$217	6,007

Figure 1: Share of information sector's contribution to GDP (Source: BEA)

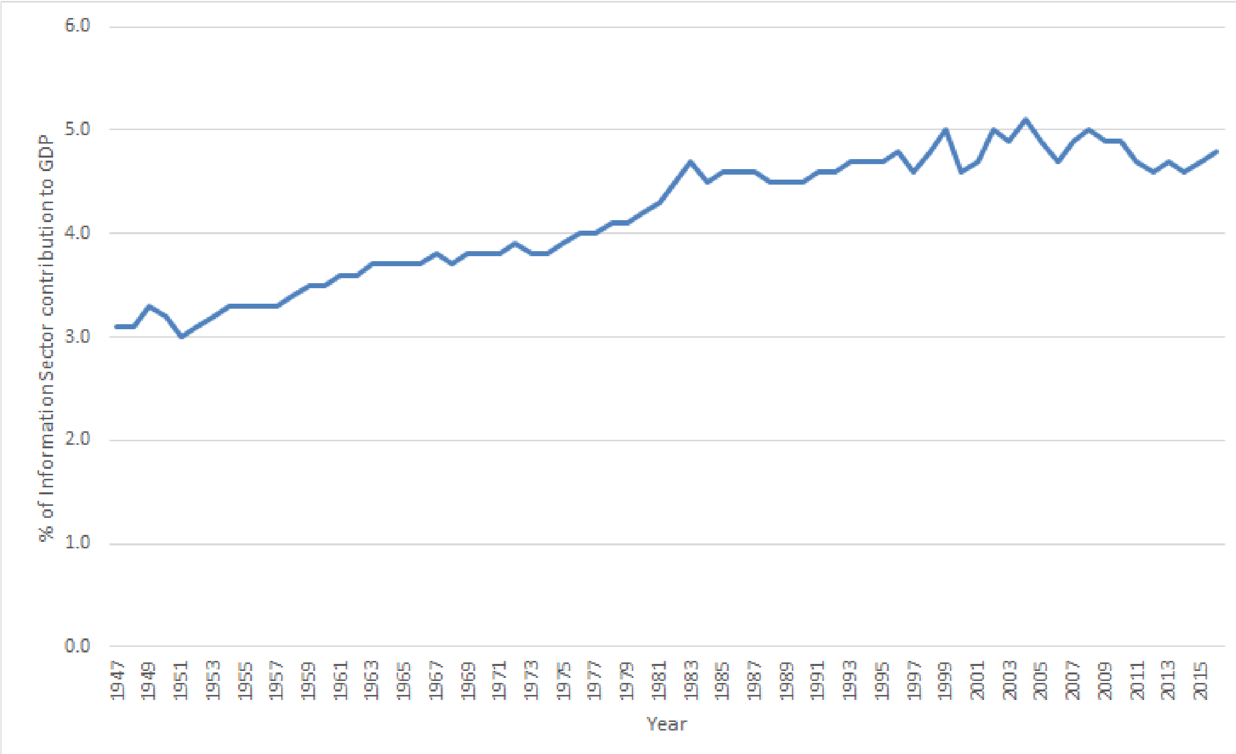


Figure 2: Consumer surplus and revenue for classic goods such as cars

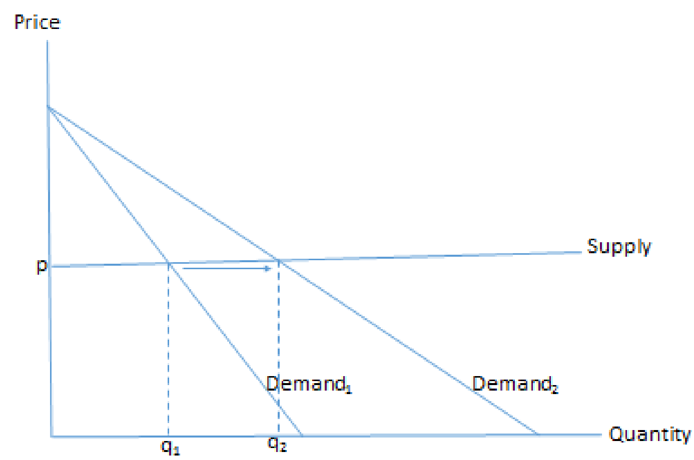


Figure 3: Consumer surplus and revenue for purely digital goods

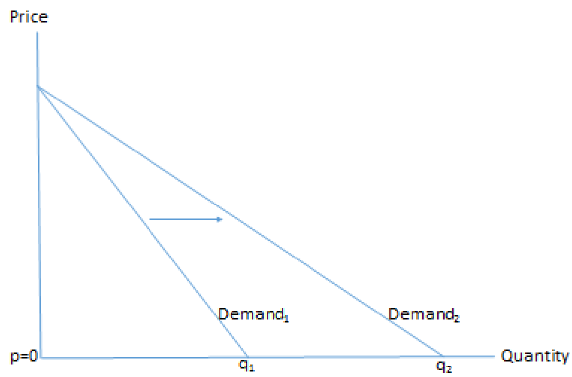


Figure 4: Consumer surplus and revenue for transition goods such as encyclopedias

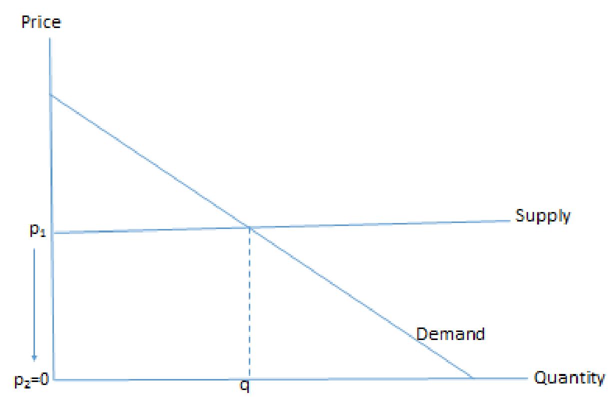


Figure 5: WTA demand curves for Facebook in 2016 and 2017

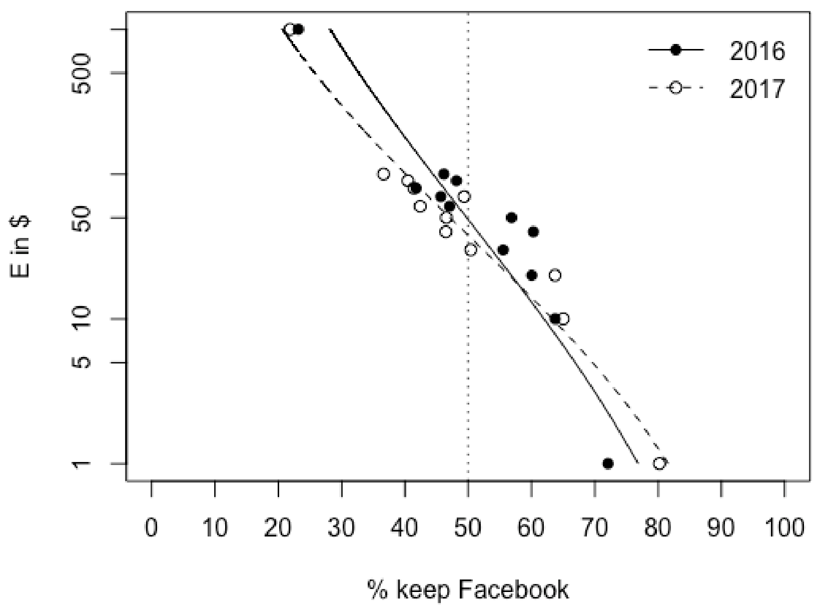


Figure 6: Comparison of demand curves estimated by BDM lottery and SBDC survey

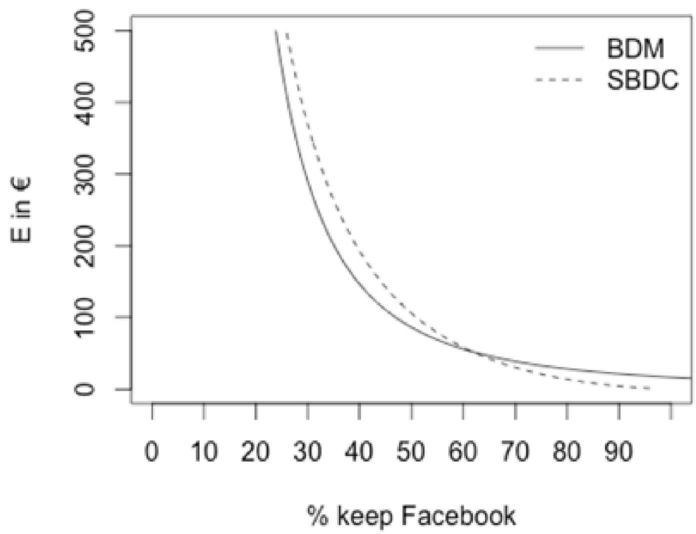


Figure 7: Assessment of hypothetical bias for Facebook

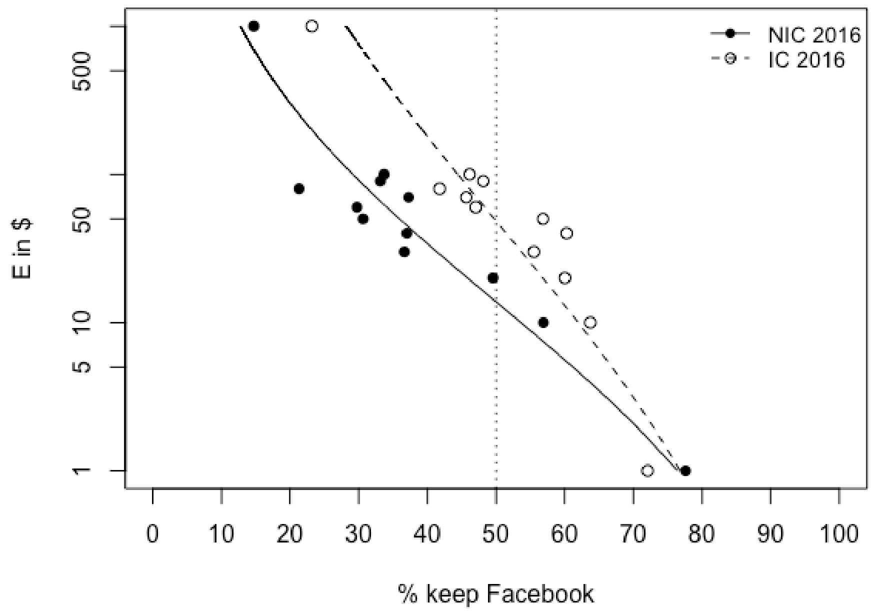


Figure 8: Effects of scale of the estimates on logit function

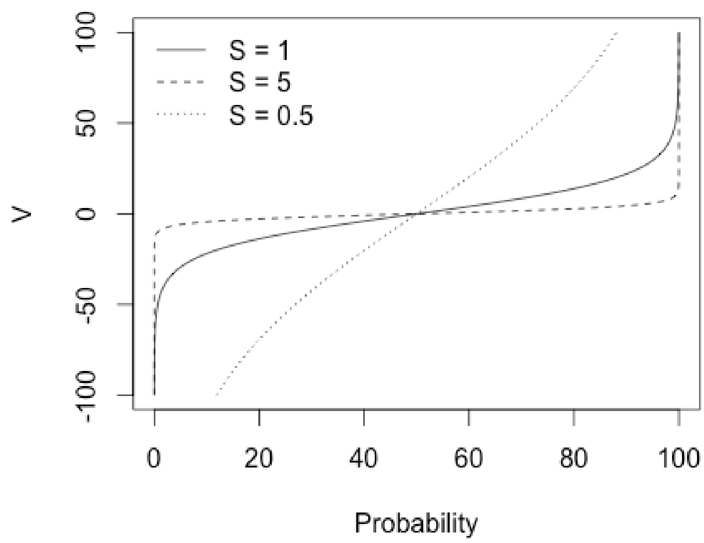


Figure 9: Effects of required abstinence time on the probability to keep Facebook

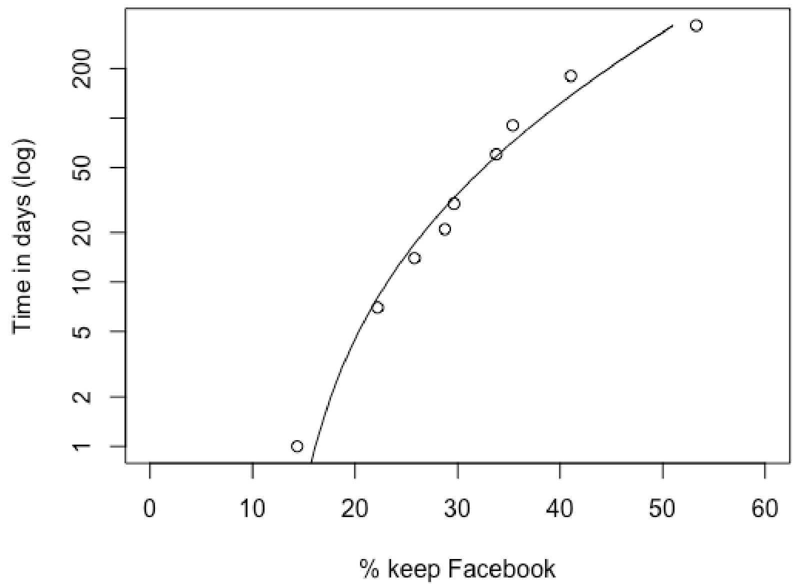


Figure 10: Example of Google Consumer Surveys

Figure 11: WTA demand curves comparing 2016 (solid line) and 2017 (dashed line) for the most widely used categories of digital goods

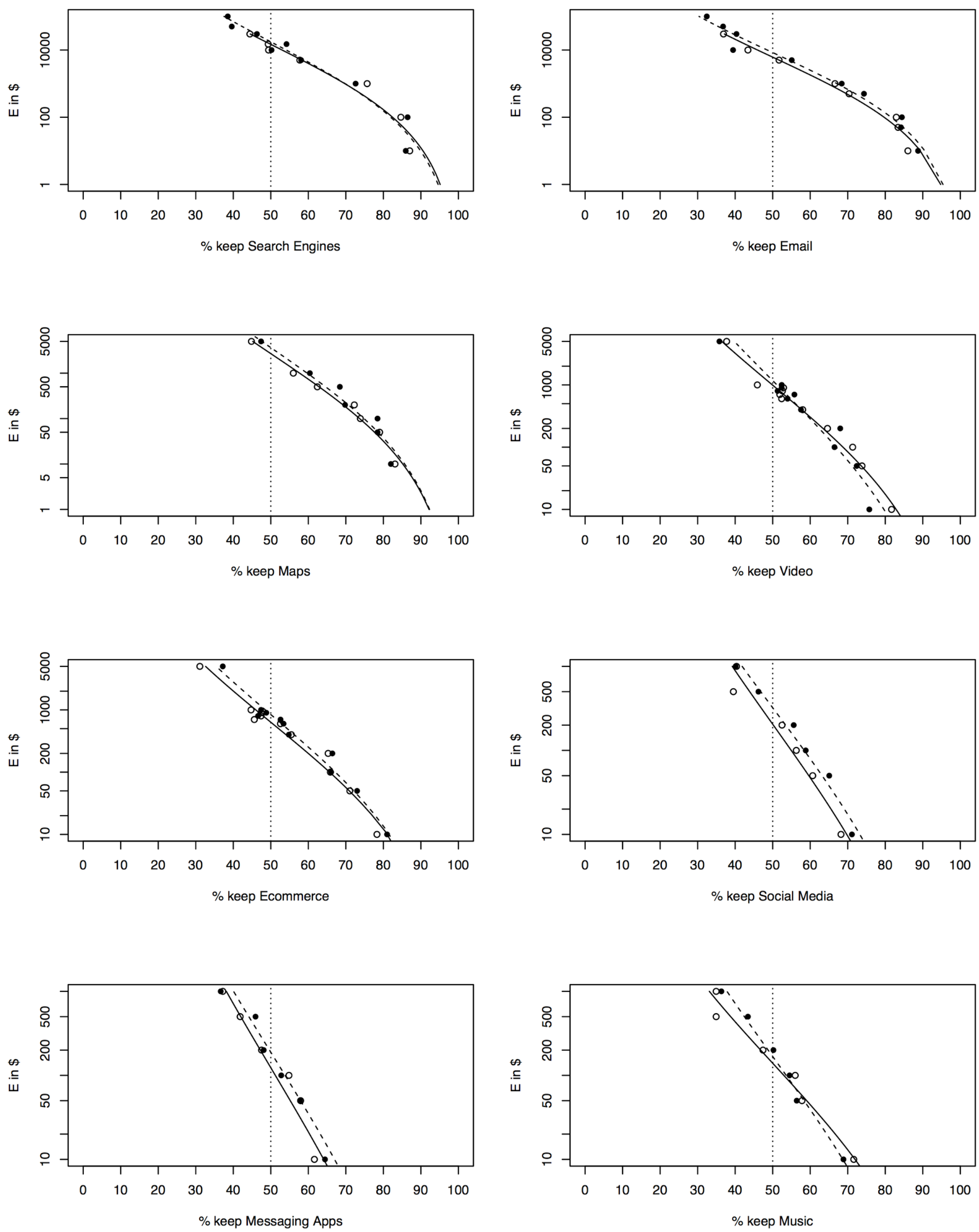


Figure 12: WTA demand curves for breakfast cereal

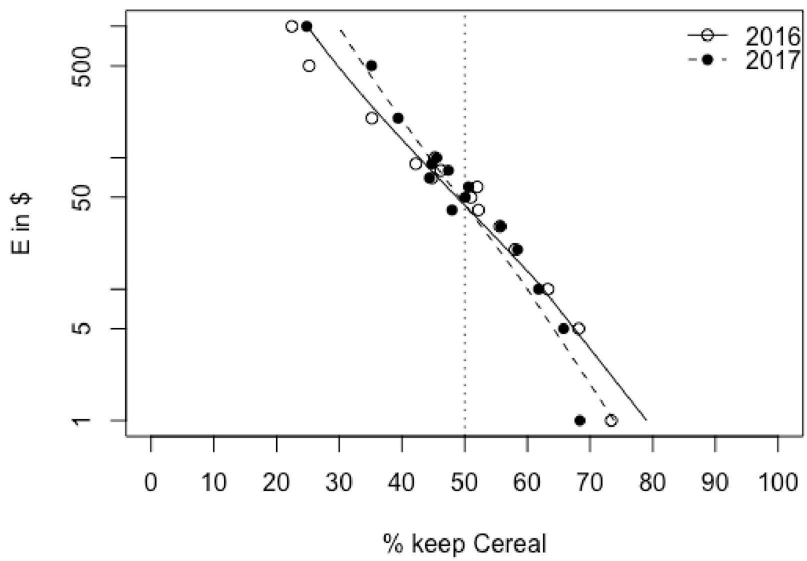
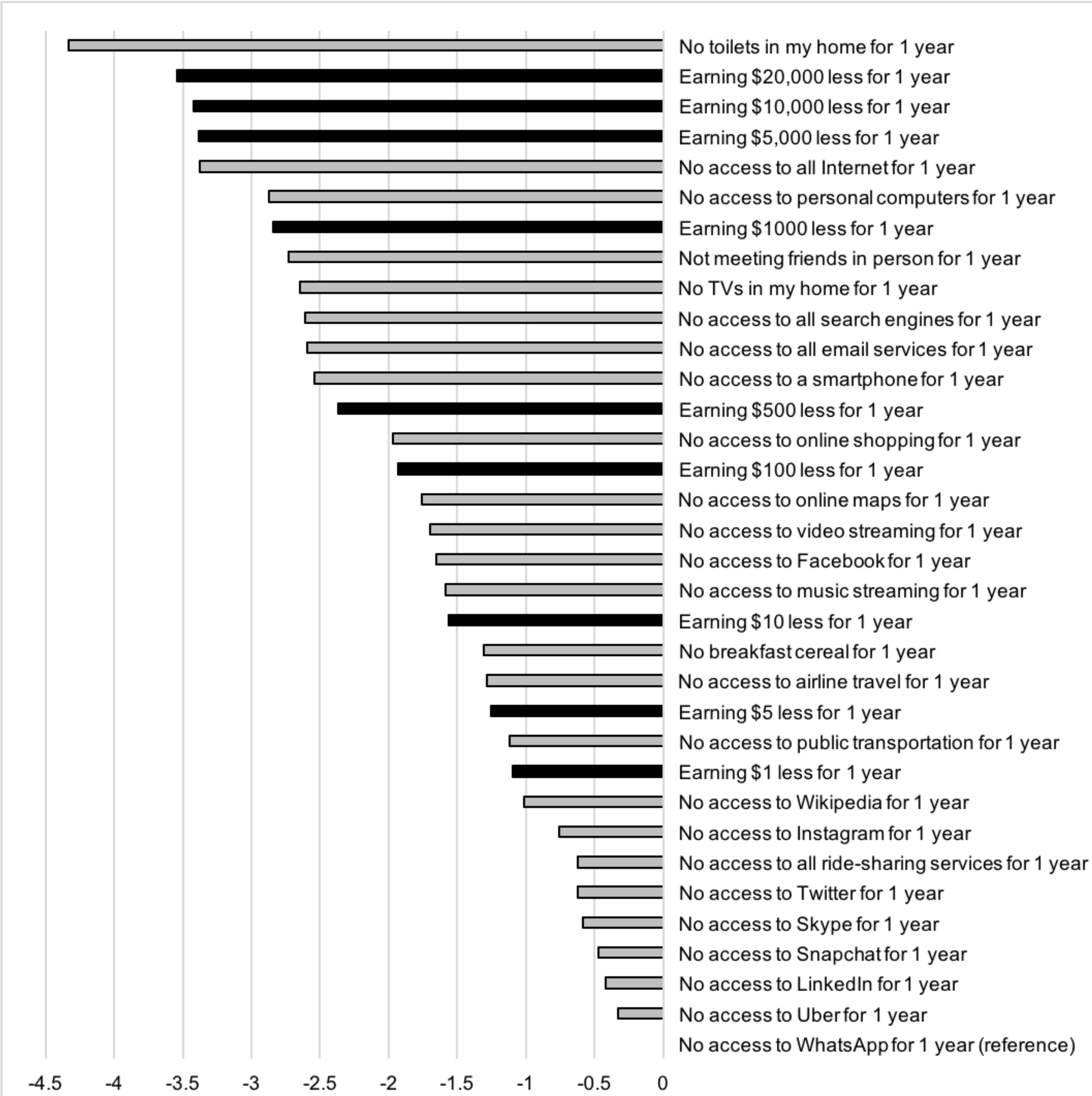


Figure 13: Example of a best-worst scaling survey question

Please assume that you would have to give up access to some services or amenities for 1 year. Please consider the options below. Which of these options do you find worst and best?

	Option 1	Option 2	Option 3
	No breakfast cereal for 1 year	No access to online shopping for 1 year	Earning $500 less for 1 year
Worst option:			
Best option:			

Figure 14: (Dis-)Utility according to best-worst scaling



Appendix

Table A.1: Full estimation model for Facebook study

	beta	Std. Error	z	p
(Intercept)	1.178	0.135	8.726	<0.001
log(E)	-0.449	0.034	-13.147	<0.001
IC	0.022	0.184	0.119	0.905
IC*log(E)	0.140	0.045	3.076	0.002
Year_2017	-0.097	0.208	-0.465	0.642
Year_2017*log(E)	-0.039	0.054	-0.721	0.471
Year_2017*IC	0.386	0.295	1.310	0.190
IC*Year_2017*log(E)	-0.062	0.074	-0.838	0.402

Table A.2: Effect of experimentally varied time frame

	beta	Std. Error	z	p
(Intercept)	-1.650	0.060	-27.550	<0.001
log(T)	0.137	0.021	6.419	<0.001
log(T)^2	0.025	0.005	5.520	<0.001

Table A.3: Estimation results of binary logit model comparing Peanut Labs (non-incentive compatible group) and GCS

	beta	**Std. Error**	**z**	**p**
(Intercept)	0.579	0.114	5.091	<0.001
log(E)	-0.374	0.029	-12.686	<0.001
GCS	0.002	0.168	0.011	0.991
GCS*log(E)	0.031	0.043	0.715	0.474

Table A.4: Estimated logistic functions for eight widely used categories of digital goods

	E-commerce	Email	Maps	Messaging	Music	Search	Social	Video
(Intercept)	2.108***	2.49***	2.12***	1.555***	1.669***	2.587***	1.827***	2.626***
log(E)	-0.351***	-0.342***	-0.316***	-0.234***	-0.362***	-0.313***	-0.282***	-0.345***
Year_2017	-0.013	0.033	0.028	0.054	-0.309.	-0.135	0.195	-0.279.
Year_2017*log(E)	0.014	-0.002	0.009	-0.003	0.084*	0.016	-0.012	0.057*
Age 18-24 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Age 25-34	0.113	0.012	0.125	-0.257**	-0.022	-0.008	-0.175.	-0.092
Age 35-44	0.295***	0.096	0.339**	-0.2.	0.025	0.171.	0.001	-0.181*
Age 45-54	0.359***	0.472***	0.309**	-0.254*	-0.174	0.159	0.096	-0.301***
Age 55-64	0.401***	0.684***	0.255*	-0.295**	-0.314**	0.382***	-0.119	-0.588***
Age 65+	0.282**	1.089***	0.053	-0.338**	-0.552***	0.518***	-0.078	-0.555***
Age Unknown	-0.035	0.195	-0.108	0.078	0.308	0.248	0.013	0.096
Gender Female (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gender Male	-0.099*	-0.117*	-0.099.	-0.355***	-0.023	-0.204***	-0.486***	-0.03
Gender Unknown	0.192	0.095	0.14	-0.669***	-0.582**	-0.209	-0.55*	-0.509***
Income \$0-\$24.999 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Income \$25,000-\$49,999	-0.073	0.101	0.092	0.105	0.051	0.293**	-0.038	-0.074

Income $50,000-$74,999	-0.025	0.297**	0.29**	-0.058	0.02	0.348***	-0.039	0.019
Income $75,000-$99,999	-0.018	0.143	0.405**	0.131	-0.083	0.479***	-0.134	0.004
Income $100,000-$149,999	-0.012	0.036	0.992***	-0.101	0.357	0.435*	0.051	0.046
Income $150,000+	0.344	0.026	0.843.	-0.503	-0.17	0.888.	0.373	-0.183
Income “Prefer not to say”	0.046	-0.562*	-0.068	-0.018	-0.023	0.399.	-0.148	0.034
Income Unknown	-0.154	0.149	-0.093	-0.354	0.094	0.146	-0.517.	-0.276
Urban Density Rural (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Urban Density Suburban	0.064	0.101	0.073	-0.051	0.197*	-0.007	0.112	0.072
Urban Density Urban	0.024	0.218**	0.141.	-0.029	0.37***	0.137.	0.015	0.094
Urban Density Unknown	-0.038	-0.067	0.227	0.269	0.17	0.343*	0.128	-0.273.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . < 0.1

Table A.5: Best-worst scaling estimation results

Good	Utility	Str. Error	WTP implied from demand function
No toilets in my home for 1 year	-4.331	0.139	$346,345.39
Earning $20,000 less for 1 year	-3.540	0.144	$18,079.67
Earning $10,000 less for 1 year	-3.424	0.123	$11,729.62
Earning $5,000 less for 1 year	-3.382	0.161	$10,023.70
No access to all Internet for 1 year	-3.373	0.123	$9,694.25
No access to personal computers for 1 year	-2.870	0.134	$1,482.92
Earning $1000 less for 1 year	-2.839	0.117	$1,323.45
Not meeting friends in person for 1 year	-2.725	0.116	$866.24
No TVs in my home for 1 year	-2.647	0.116	$645.66
No access to all search engines for 1 year	-2.610	0.115	$563.80
No access to all email services for 1 year	-2.592	0.115	$525.43
No access to a smartphone for 1 year	-2.542	0.115	$437.16
Earning $500 less for 1 year	-2.371	0.114	$230.40
No access to online shopping for 1 year	-1.967	0.113	$51.13
Earning $100 less for 1 year	-1.933	0.113	$45.03
No access to online maps for 1 year	-1.756	0.113	$23.24
No access to video streaming for 1 year	-1.695	0.112	$18.56
No access to Facebook for 1 year	-1.654	0.112	$15.91
No access to music streaming for 1 year	-1.587	0.112	$12.36
Earning $10 less for 1 year	-1.565	0.112	$11.41
No breakfast cereal for 1 year	-1.307	0.113	$4.36

No access to airline travel for 1 year	-1.287	0.112	$4.04
Earning $5 less for 1 year	-1.254	0.127	$3.58
No access to public transportation for 1 year	-1.120	0.113	$2.17
Earning $1 less for 1 year	-1.097	0.128	$1.99
No access to Wikipedia for 1 year	-1.016	0.112	$1.47
No access to Instagram for 1 year	-0.754	0.114	$0.55
No access to all ride-sharing services for 1 year	-0.621	0.115	$0.34
No access to Twitter for 1 year	-0.621	0.114	$0.34
No access to Skype for 1 year	-0.586	0.114	$0.30
No access to Snapchat for 1 year	-0.474	0.116	$0.19
No access to LinkedIn for 1 year	-0.415	0.115	$0.16
No access to Uber for 1 year	-0.326	0.117	$0.11
No access to WhatsApp for 1 year (reference)	0.000		$0.03

Figure A.1: Example of Incentive Compatible (IC) Questionnaire for Facebook SBDC question (for E = $80)

Would you prefer to keep access to Facebook or go without access to Facebook for 1 month and get paid $80?

We want to reward you for thinking carefully about this question. Therefore, we will randomly pick 1 out of every 200 respondents and her/his selection will be fulfilled:

- If you choose "Keep access to Facebook" you can keep using Facebook as before. However, you will not receive the $80 in cash.
- If you choose "Give up Facebook and get paid $80" you will receive the $80 in cash, provided that you do not access Facebook for 1 month. Facebook collects the date and time when you have last used your account. Given your permission, we can access this time with an app (e.g., see this link for an example). In order to get your permission, we will contact you via email. You can revoke this permission at any time.

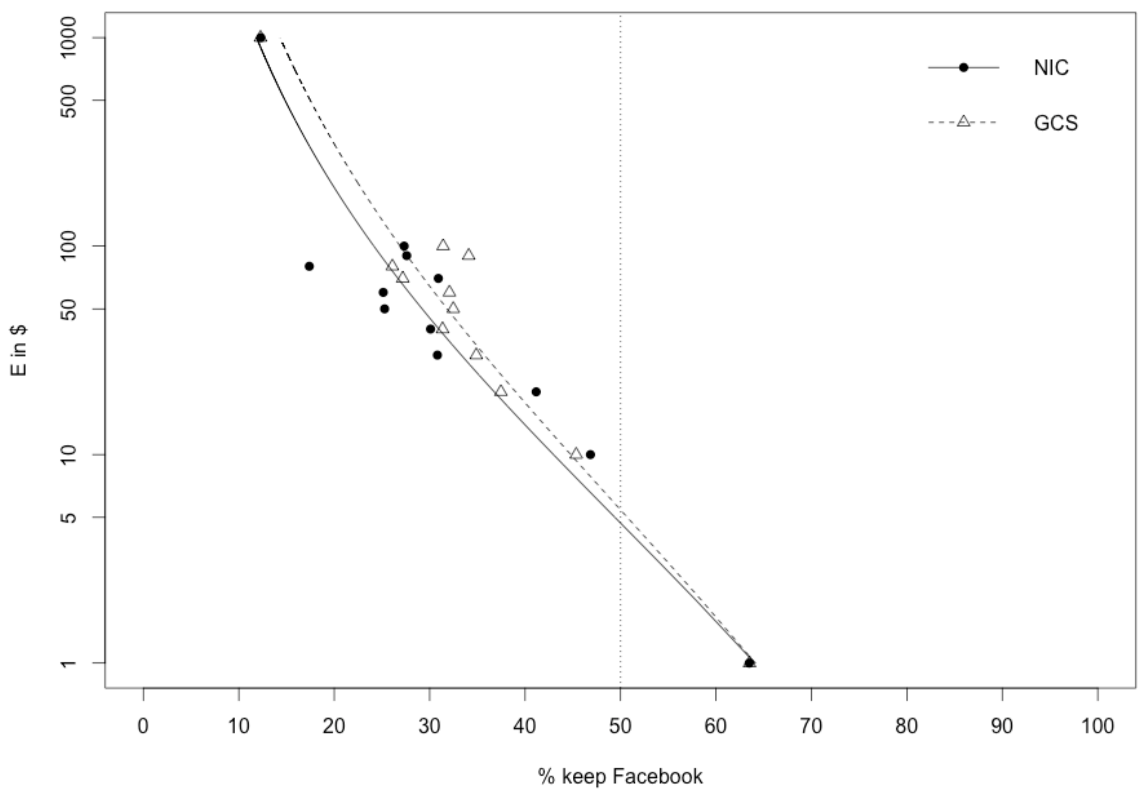
Therefore, it is in your best interest to think carefully about how valuable you find Facebook.

What is your decision: Would you prefer to keep access to Facebook or go without access to Facebook for 1 month and get paid $80?

- ○ Keep access to Facebook
- ○ Give up Facebook and get paid $80

Proceed to next page

Figure A.2: Assessment of selection bias



Chapter 2 - GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy

Abstract

The welfare contributions of the digital economy, characterized by the proliferation of new and free goods, are not well-measured in our current national accounts. We derive explicit terms for the welfare contributions of these goods and introduce a new metric, GDP-B which quantifies their benefits, rather than costs. We apply this framework to several empirical examples including Facebook and smartphone cameras and estimate their valuations through incentive-compatible choice experiments. For example, including the welfare gains from Facebook would have added between 0.05 and 0.11 percentage points to GDP-B growth per year in the US.

"The welfare of a nation can scarcely be inferred from a measure of [GDP]."
– Simon Kuznets, 1934

1. Introduction

We develop a new framework for measuring welfare change and real GDP growth in the presence of new and free goods.[1] The increased proliferation of such goods is a key characteristic of the digital economy. New, sometimes very specialized, goods appear with increasing rapidity,[2] and free goods (such as information and entertainment services) are increasingly available at zero price, reflecting the very low marginal costs of digital replication and distribution. Even when free goods have an implicit price,[3] this price is not usually observed so a price of zero is applied. Thus, the positive quantities of these goods that are consumed have a measured price of zero and measured value of zero in the conventional national accounts. Hence, they are not reflected in standard statistical agency reports for GDP or related metrics like productivity, which are typically defined in terms of GDP. Furthermore, despite GDP's widespread use as a proxy for welfare, it is not the correct metric for this purpose, at least as conventionally measured.

Our framework provides a means by which to understand the welfare contributions from these goods and the potential mismeasurement that arises from not fully accounting for them. We use this framework to derive an explicit term that is the marginal value of a new good on welfare change, providing a means for estimating welfare change mismeasurement if the good is omitted from statistical

---

[1] Throughout this paper, we use the word "goods" to refer to goods and services collectively.
[2] Goolsbee and Klenow (2018, Table 3), using Adobe Analytics data on online transactions for millions of products across many different categories, find that roughly half the sales volume online for 2014-2017 is for products that did not exist in the previous year.
[3] See Nakamura, Samuels and Soloveichik (2016) and Brynjolfsson and Oh (2012) for examples of how to think about the valuation of "free" media.

agency collections. This can shed light on the debate regarding the potential of the digital economy to generate productivity, economic growth and welfare gains, and the paradox implicit in the gap between technological advances and low (measured) productivity.[4] In particular, if measurement is lacking, through methodological challenges, statistical agency budgets or data availability, then we are severely hampered in our ability to understand the impact of new technologies, goods on the economy, and consequently the prospects for future productivity, economic growth and welfare improvements.[5]

A problem in assessing the full impact of the introduction of a new good on real GDP growth is that we would really need national statistical offices to recalculate their estimates of real GDP including the new goods with, for example, estimated Hicksian reservation prices for the period before they are sold in positive quantities; the reservation price of a good is the price which would induce a utility maximizing potential purchaser of the product to demand zero units of it.[6] However, we are able to use our framework to derive a close approximation to the addition to real GDP growth that would be required to account for the welfare gains from the introduction of a new good, without having to recalculate the official GDP numbers published by national statistical offices.

---

[4] Among others, see, for example, Gordon (2016) and Cowen (2011) giving a pessimistic view and Sichel (2016), Mokyr, Vickers and Ziebarth (2015) and Brynjolfsson and McAfee (2011, 2014) giving a more optimistic view.

[5] Among others, see, for example, Feldstein (2017), Groshen *et al*. (2017), Hulten and Nakamura (2017), Syverson (2017), Ahmad and Schreyer (2016), Byrne, Fernald and Reinsdorf (2016), Brynjolfsson and Saunders (2009), Brynjolfsson and Oh (2012), Brynjolfsson, Rock and Syverson (2017), Greenstein and McDevitt (2011), Brynjolfsson, Eggers and Gannamaneni (2018) and Brynjolfsson, Collis and Eggers (2019).

[6] See Hicks (1940), Diewert (1980), Hausman (1981, 1996), Feenstra (1994), Diewert, Fox and Schreyer (2018), and Diewert and Feenstra (2017).

Free goods are addressed through generalizing the standard microeconomic model of household cost minimization. It is then possible to re-work our welfare change and real GDP growth adjustment terms to allow for there to be free goods. Our new metric is labelled *GDP-B,* as it captures the *benefits* associated with new and free goods and thus goes "beyond GDP".[7] In addition, our calculations of GDP-B make it easy to calculate a corresponding productivity metric, *Productivity-B* which simply uses GDP-B as its numerator.

We provide several empirical examples of free digital goods where we quantify these welfare and GDP growth adjustment terms. Specifically, we draw on the work of Brynjolfsson, Collis and Eggers (2019) who developed an approach to directly estimate consumer welfare by running massive online choice experiments. They explored both hypothetical and incentive compatible choice experiments to estimate willingness to accept values for giving up access to a good. While hypothetical choice experiments might suffer from hypothetical bias, incentive aligned choice experiments, which make participants' choices consequential, have been shown to be externally valid (Ding, Grewal and Liechty 2005; Ding 2007; Carson, Groves and List 2014; Bishop et. al. 2017). We therefore constructed incentive compatible discrete choice experiments to estimate the potential impact on welfare growth by Facebook, a free social networking service which had rapid diffusion and quickly accumulated many diverse users. We ran our experiments on a representative sample of the US internet population recruited through an online survey panel. We use the results to provide estimates of the adjustments to

[7] See e.g. Jones and Klenow (2016), Coyle and Mitra-Kahn (2017), Corrado *et al.* (2017) and Jorgenson (2018). Some national statistical offices are considering producing a spectrum of expanded GDP measures. Heys (2018) presented options being considered by the UK Office of National Statistics, which include incorporating welfare adjustments for private and publically provided free goods. Our approach in this paper provides a way of doing this.

welfare change and real GDP-B growth from Facebook's launch in 2004 through 2017.

In a laboratory setting in the Netherlands, we also ran incentive compatible choice experiments to estimate the consumer welfare created by several other popular digital goods, including Instagram, Snapchat, Skype, Digital Maps, LinkedIn, Twitter as well as Facebook. Although we did not have a representative sample of the population in the laboratory, our results are indicative of the approximate size of the adjustment term to real GDP-B growth which would need to be added to account for the welfare gain from these digital goods.

We also show the need for properly adjusting for quality changes in calculating GDP-B growth so that welfare changes are properly inferred. This issue is particularly acute for smartphones which have substituted (to varying degrees) a panoply of other devices including cameras, GPS, landline phones, gaming consoles, ebook readers, personal computers, video and audio players, maps/atlases, alarm clocks, calculators and sound recorders,[8] as well as numerous new capabilities that previously were unavailable at any price like real-time traffic and various types of social networking and messaging applications. What is more, new features are added frequently and quality of existing features changes rapidly. In fact, application developers conduct thousands of A/B tests every day and tweak features to improve user experience. Groshen *et al*. (2017) discuss how the US Bureau of Labor Statistics (BLS) adjusts for quality changes using hedonic methods. However, they mention that this approach is ruled out for smartphones since the set of relevant characteristics for the hedonic models constantly keep on

[8] See https://www.huffingtonpost.com/steve-cichon/radio-shack-ad_b_4612973.html (accessed Feb 10, 2019) and also Hal Varian's presentation at Brookings (https://www.brookings.edu/wp-content/uploads/2016/08/varian.pdf, accessed March 19, 2019).

changing. While there has been a subsequent development in that the US BLS commenced some hedonic quality adjustments for smartphones in January 2018,[9] such explicit hedonic quality adjustment is still very limited internationally, with the UK ONS being a standout early adopter of this approach for smartphones, commencing in 2011 (see Wells and Restieaux (2014)).

Hence, to advance understanding of the consumer benefits from quality change, we conduct an incentive compatible BDM lottery study (Becker, DeGroot, and Marschak 1964) in a university laboratory in the Netherlands to elicit consumers' valuations for smartphone cameras. We find that there is a large difference between the contribution of smartphone cameras towards conventionally-measured GDP and the welfare generated by these cameras for consumers as reflected in GDP-B. As a result, not accounting for quality adjustments in smartphones leads to a significant underestimate of GDP-B growth.

The rest of the paper is organised as follows. The next section sets out some preliminary definitions that will be used in the subsequent sections. Section 3 looks at the problem of new goods, and shows how the impact of new goods on welfare change and real GDP growth can be estimated to a high degree of approximation. Section 4 extends this framework to the case of free goods and introduces our preferred measure, GDP-B. Section 5 provides the empirical examples of Facebook and other popular free digital goods to quantify adjustments to welfare change and GDP-B growth for not accounting for these goods. Section 6 presents results from the smartphone camera laboratory study to highlight potential biases due to not performing quality adjustments. Section 7 concludes with a summary and some implications.

---

[9] See "Measuring Price Change in the CPI: Telephone hardware, calculators, and other consumer information items", available at https://www.bls.gov/cpi/factsheets/telephone-hardware.htm.

2. Preliminaries

We assume that a consumer has a utility function, $f(q)$, which is continuous, quasiconcave and increasing in the components of the nonnegative quantity vector $q \geq 0_N$. For each strictly positive price vector $p >> 0_N$ and each utility level u in the range of f, we can define the dual cost function C as follows:

$$(1)\ C(u,p) \equiv \min_q \{p \cdot q\ ;\ f(q) \geq u\}.$$

We are given the price and quantity data, (p^t, q^t) for periods $t = 0, 1$. We assume that the consumer minimizes the cost of achieving the utility level $u^t \equiv f(q^t)$ for $t = 0, 1$ so observed expenditure in each period is equal to the minimum cost of achieving the given utility level in each period; i.e., we have

$$(2)\ p^t \cdot q^t = C(f(q^t),p^t) \text{ for } t = 0, 1.$$

Valid measures of utility change over the two periods under consideration are the following Hicksian *equivalent and compensating variations* (Hicks, 1942):[10]

$$(3)\ Q_E(q^0, q^1, p^0) \equiv C(f(q^1), p^0) - C(f(q^0), p^0)\ ;$$
$$(4)\ Q_C(q^0, q^1, p^1) \equiv C(f(q^1), p^1) - C(f(q^0), p^1)\ .$$

[10] These are Hick's original definitions of equivalent and compensating variations. Hicks (1946, 331-332) appears to provide an alternative definition of the equivalent variation as $C(f(q^1), p^1) - C(f(q^1), p^0)$ and the compensating variation as $C(f(q^0), p^1) - C(f(q^0), p^0)$. The existence of alternative definitions has caused significant confusion in the literature; see Diewert (1992, p. 567, footnote 10).

The above variations are special cases of the following Samuelson (1974) family of quantity variations: for $p >> 0_N$, define:[11]

(5) $Q_S(q^0, q^1, p) \equiv C(f(q^1), p) - C(f(q^0), p)$.

Hence there is an entire family of cardinal measures of utility change defined by (5), with one measure for each reference price vector p.

The variations defined by (3) and (4) are not observable (since $C(f(q^1), p^0)$ and $C(f(q^0), p^1)$ are not observable) but the following Laspeyres and Paasche variations, V_L and V_P, are observable:

(6) $V_L(p^0, p^1, q^0, q^1) \equiv p^0 \cdot (q^1 - q^0)$;
(7) $V_P(p^0, p^1, q^0, q^1) \equiv p^1 \cdot (q^1 - q^0)$.

Note that V_L and V_P are difference counterparts to the Laspeyres and Paasche quantity indexes, $Q_L = p^0 \cdot q^1 / p^0 \cdot q^0$ and $Q_P = p^1 \cdot q^1 / p^1 \cdot q^0$, respectively. Hicks (1942) showed that V_L approximates Q_E and V_P approximates Q_C to the accuracy of a first order Taylor series approximation; see also Diewert and Mizobuchi (2009; 345-346). The observable Bennet (1920) variation or indicator of quantity change V_B is defined as the arithmetic average of the Laspeyres and Paasche variations in (6) and (7):

$$
\begin{aligned}
(8)\ V_B(p^0, p^1, q^0, q^1) &\equiv \tfrac{1}{2}(p^0 + p^1) \cdot (q^1 - q^0) \\
&= p^0 \cdot (q^1 - q^0) + \tfrac{1}{2}(p^1 - p^0) \cdot (q^1 - q^0)
\end{aligned}
$$

[11] These measures of overall quantity change are difference counterparts to the family of Allen (1949) quantity indexes in normal ratio index number theory. The Allen quantity index for reference price vector p is defined as the ratio $C(f(q^1), p)/C(f(q^0), p)$.

$$= V_L\left(p^0,p^1,q^0,q^1\right) + \frac{1}{2}\sum_{n=1}^{N}\left(p_n^{\,1} - p_n^{\,0}\right)\left(q_n^{\,1} - q_n^{\,0}\right).$$

Thus the Bennet variation is equal to the Laspeyres variation $V_L(p^0, p^1, q^0, q^1)$ plus a sum of *N* Harberger (1971) consumer surplus triangles of the form $(1/2)(p_n^{\,1} - p_n^{\,0})(q_n^{\,1} - q_n^{\,0})$.

An alternative decomposition of the Bennet variation is the following one:

$$\text{(9) } V_B(p^0,p^1,q^0,q^1) \equiv \tfrac{1}{2}(p^0 + p^1)\cdot(q^1 - q^0)$$
$$= p^1\cdot(q^1 - q^0) - \tfrac{1}{2}(p^1 - p^0)\cdot(q^1 - q^0)$$
$$= V_P\left(p^0,p^1,q^0,q^1\right) - \frac{1}{2}\sum_{n=1}^{N}\left(p_n^{\,1} - p_n^{\,0}\right)\left(q_n^{\,1} - q_n^{\,0}\right).$$

Thus the Bennet variation is also equal to the Paasche variation $V_P(p^0, p^1, q^0, q^1)$ minus a sum of *N* Harberger consumer surplus triangles of the form $(1/2)(p_n^{\,1} - p_n^{\,0})(q_n^{\,1} - q_n^{\,0})$.

It is possible to relate the observable Bennet variation to a theoretically valid Samuelson variation of the form defined by (5). However, in order to do this, we need to assume a specific functional form for the consumer's cost function, C(u, p). If the cost function has a flexible,[12] translation-homothetic normalized quadratic functional form, then Proposition 1 in Diewert and Mizobuchi (2009; 353) relates the observable Bennet variation, $V_B(p^0, p^1, q^0, q^1)$ defined by (8) or (9) to the unobservable equivalent and compensating variations defined by (3) and (4); i.e., we have the following exact equality:

[12] Diewert (1974) defined a flexible functional form as one that provides a second order approximation to a twice continuously differentiable function at a point.

$$(10)\ V_B(p^0, p^1, q^0, q^1) = ½Q_E(q^0, q^1, p^0) + ½Q_C(q^0, q^1, p^1).$$

That is, with certain assumptions on the functional form for the consumer's cost function (and using normalized price vectors), the observable Bennet variation can be shown to be *exactly equal* to the arithmetic average of the unobservable equivalent and compensating variations.[13] Hence, there is a strong justification from an economic perspective for using the Bennet quantity variation. Also, it has a strong justification from an axiomatic perspective (Diewert, 2005).

Finally, we can note that value change can be decomposed into Bennet quantity and price variations, as follows:

$$(11)\ p^1\cdot q^1 - p^0\cdot q^0 = V_B(p^0, p^1, q^0, q^1) + I_B(p^0, p^1, q^0, q^1),$$

where $V_B(p^0, p^1, q^0, q^1) \equiv ½(p^0 + p^1)\cdot(q^1 - q^0)$ and $I_B(p^0, p^1, q^0, q^1) \equiv ½(q^0 + q^1)\cdot(p^1 - p^0)$. Equation (11) can thus provide a decomposition into quantity and price components for any value change, including a change in nominal GDP.

3. The New Goods Problem

[13] Normalized prices are needed for this result to be true: "If there is a great deal of general inflation between periods 0 and 1, then the compensating variation will be much larger than the equivalent variation simply due to this general inflation, and an average of these two variations will be difficult to interpret due to the change in the scale of prices. To eliminate the effects of general inflation between the two periods being compared, it will be useful to scale the prices in each period by a fixed basket price index of the form $\alpha\cdot P$, where $\alpha \equiv [\alpha_1, \ldots, \alpha_N] > 0_N$ is a nonnegative, nonzero vector of price weights." Diewert and Mizobuchi (2009, 352-353). They recommend choosing α so that a fixed-base Laspeyres price index is used to deflate nominal prices (footnote 34, page 368).

We can now apply the above results to measure the benefits of the introduction of a new good to a consumer who cannot purchase the good in period 0 but can purchase it in period 1. First, we have to make an additional assumption. We assume that there is a shadow or reservation price for the new good in period 0 that will cause the consumer to consume 0 units of the new good in period 0. This type of assumption dates back to Hicks (1940; 114).[14]

Let the new good be indexed by the subscript 0 and let the *N* dimensional vectors of period t prices and quantities for the continuing goods be denoted by p^t and q^t for t = 0,1. The period 1 quantity of good 0 purchased during period 1 is also observed and is denoted by q_0^1. The period 0 reservation price for good 0 is not observed but we make some sort of estimate for it, denoted as $p_0^{0*} > 0$. The period 0 quantity is observed and is equal to 0; i.e., $q_0^0 = 0$. Thus the price and quantity data (for the *N+1* goods) for period 0 is represented by the *1+N* dimensional vectors (p_0^{0*}, p^0) and $(0, q^0)$ and the price and quantity data for period 1 is represented by the *1+N* dimensional vectors (p_0^1,p^1) and (q_0^1, q^1). We adapt our first expression for the Bennet variation, (8), to accommodate the new good. We find that our new Bennet variation is equal to the following expression:

$$
\begin{aligned}
(12)\ V_B([p_0^{0*}, p^0], [p_0^1, p^1], [0, q^0], [q_0^1, q^1])
&= \tfrac{1}{2}(p^0 + p^1)\cdot(q^1 - q^0) + \tfrac{1}{2}(p_0^{0*} + p_0^1)(q_0^1 - 0) \\
&= p^0\cdot(q^1 - q^0) + \tfrac{1}{2}(p^1 - p^0)\cdot(q^1 - q^0) + p_0^1(q_0^1 - 0) - \tfrac{1}{2}(p_0^1 - p_0^{0*})\,(q_0^1 - 0) \\
&= p^0\cdot(q^1 - q^0) + \tfrac{1}{2}(p^1 - p^0)\cdot(q^1 - q^0) + p_0^1 q_0^1 - \tfrac{1}{2}(p_0^1 - p_0^{0*})q_0^1.
\end{aligned}
$$

[14] There is now quite a literature on this topic and for alternative approximate welfare gain estimates; see Hausman (1981) (1996), Feenstra (1994) and Diewert and Feenstra (2017), and the references in these publications. Diewert has been applying the above Hicksian reservation analysis in the ratio context (i.e., in the context of the true cost of living index) for a long time; see Diewert (1980; 498-505), (1987; 378) (1998; 51-54). A weakness in these theories is the difficulty in determining the appropriate reservation prices.

From the last equation on the right hand side of (12), we see that the first term, $p^0 \cdot (q^1 - q^0)$ is simply the change in consumption valued at the real prices of period 0, a Laspeyres variation as in (6); the second term, $\frac{1}{2}(p^1 - p^0) \cdot (q^1 - q^0)$, is the sum of the consumer surplus terms associated with the continuing goods; the next term, $p_0{}^1 q_0{}^1$, is the value of consumption of the new good in period 1, valued at the price for good 0 in period 1 (this is the usual price times quantity contribution term to the value of real consumption of the new good in period 1 which would be recorded as a contribution to period 1 GDP); and the last term, $-\frac{1}{2}(p_0{}^1 - p_0{}^{0*})q_0{}^1 = \frac{1}{2}(p_0{}^{0*} - p_0{}^1)q_0{}^1$ is the additional consumer surplus contribution of good 0 to overall welfare change, which would not be recorded as a contribution to GDP. Note that the first two terms are a measure of the welfare change we would get by just ignoring the new good in both periods. Thus the last two terms give the overall contribution to welfare change due to the introduction of the new good.

If we assume that the reservation price for the new good in period 0, $p_0{}^{0*}$, is equal to the observable price for the new good in period 1, $p_0{}^1$, then the last term in (12), the consumer surplus term for the new good, vanishes. However, it is likely that the reservation price for period 0, $p_0{}^{0*}$, is much higher than the corresponding actual price for good 0 in period 1, $p_0{}^1$.[15] Thus if we assume that $p_0{}^{0*} = p_0{}^1$ and evaluate (12), then the downward bias in the resulting Bennet measure of welfare change will be equal to a Harberger-type triangle, $-\frac{1}{2}(p_0{}^1 - p_0{}^{0*})(q_0{}^1 - 0) = \frac{1}{2}(p_0{}^{0*} - p_0{}^1)q_0{}^1$.

[15] Hausman (1996) argued that for cereals, the reservation price was about twice the price at the introduction of the new good, whereas Feenstra (1994) takes it to be infinity.

It is of interest to gauge the extent to which GDP growth is underestimated by not fully capturing the introduction of the new good. As comparisons may be made between periods far apart (e.g. before the introduction of the good and the most recent period), we will now be explicit about the point raised in footnote 13 of section 2; value change comparisons are difficult to interpret if the values are not expressed in comparable units. Hence, we recommend using real prices where, for example, the base period's prices are inflated by using the Consumer Price Index (CPI) to put them into comparable units with the current period's prices.[16]

Let γ denote one plus the rate of growth of the CPI between periods 0 and 1 (which may not be adjacent periods).[17] Then, adapting a result from Diewert (2005; 335), value change can be expressed as follows, where P and Q are price and quantity indexes, respectively, that satisfy $P \times Q = p^1 \cdot q^1 / p^0 \cdot q^0$.[18]

$$
\begin{aligned}
(13)\ p^1 \cdot q^1 - \gamma p^0 \cdot q^0 &= \gamma p^0 \cdot q^0 [p^1 \cdot q^1 / \gamma\, p^0 \cdot q^0 - 1] \\
&= \gamma p^0 \cdot q^0 [PQ/\gamma - 1] \\
&= \tfrac{1}{2}\, \gamma p^0 \cdot q^0 [\, 2\mathscr{P}Q - 2] \quad \text{where } \mathscr{P} \equiv P/\gamma,
\end{aligned}
$$

[16] Alternatively, we could deflate current prices to put them into the same units as the earlier period. Having units in a distant past period is, however, typically more difficult to interpret than using current period units. We recommend putting values into comparable units for both welfare and GDP growth adjustments, especially in high inflation environments or if periods are far apart in time. Similarly for spatial comparisons.

[17] We prefer to use the CPI rather than the GDP deflator for adjusting for general inflation, as the GDP deflator behaves perversely if import prices change. This is because the immediate effect of e.g. a fall in import prices is to increase the deflator; see Kohli (1982; 211). Also, Diewert (2002; 556, footnote 14) noted the following: "An example of this anomalous behavior of the GDP deflator just occurred in the advance release of gross domestic product for the third quarter of 2001 for the US national income and product accounts: the chain type price indexes for C, L, X and M decreased (at annual rates) over the previous quarter by 0.4%, 0.2%, 1.4% and 17.4% respectively but yet the overall GDP deflator increased by 2.1 %. Thus there was general deflation in all sectors of the economy but yet the overall GDP deflator increased. This is difficult to explain to the public!"

[18] That is, the formulae for the indexes P and Q are such that the product test from the axiomatic approach to index numbers is satisfied.

$$= \tfrac{1}{2}\, \gamma p^0 \cdot q^0 [(1+Q)(\mathscr{P} - 1) + (1+\mathscr{P})(Q-1)]$$

We can see that (13) can be decomposed into two components, a price change indicator, $\mathcal{I}_E$, and a quantity change indicator, $\mathcal{V}_E$:[19]

(14) $\mathcal{I}_E = \tfrac{1}{2}\, \gamma p^0 \cdot q^0 (1+Q)(\mathscr{P} - 1)$;

(15) $\mathcal{V}_E = \tfrac{1}{2}\, \gamma p^0 \cdot q^0 (1+\mathscr{P})(Q - 1)$

If P (in $\mathscr{P} \equiv P/\gamma$) and Q are replaced by superlative indexes,[20] such as the Fisher or Törnqvist, then the resulting indicators in (14) and (15) can also be called superlative. A corollary of Proposition 9 of Diewert (2005; 338) is that the Bennet indicator of quantity change approximates any superlative indicator to the second order at any point where the two quantity vectors are equal and where the two price vectors are equal.

The U.S. uses the superlative Fisher quantity index (the geometric mean of the Laspeyres and Paasche indexes given in section 2) for constructing real GDP, so we consider the following expression for the Fisher superlative quantity change indicator, $\mathcal{V}_E^F$:

(16) $\mathcal{V}_E^F \equiv \tfrac{1}{2}\, \gamma p^0 \cdot q^0 (1+\mathscr{P}^F)(Q^F - 1) \approx \tfrac{1}{2}(\gamma p^0 + p^1) \cdot (q^1 - q^0) = \mathcal{V}_B$,

---

[19] Diewert (2005; 333-337) derived these indicators in introducing the economic approach to indicators of price and quantity change, and called them "economic" indicators. Hence, the subscript "E" stands for "economic".
[20] See Diewert (1976) on superlative index numbers.

where $\mathscr{P}^F \equiv P^F/\gamma$, where $P^F \equiv [(p^1 \cdot q^0/p^0 \cdot q^0)(p^1 \cdot q^1/p^0 \cdot q^1)]^{1/2}$ is the Fisher price index, or GDP deflator in our context, γ is one plus the growth rate of the CPI between periods 0 and 1, and $Q^F \equiv [(p^0 \cdot q^1/p^0 \cdot q^0)(p^1 \cdot q^1/p^1 \cdot q^0)]^{1/2}$ is the Fisher quantity index, or real GDP growth in our context, and $\mathcal{V}_B$ is the Bennet quantity indicator where the price weights have been adjusted for general inflation.[21] Recall that the Bennet indicator of quantity change is the symmetric arithmetic average of first-order approximations to the Hicksian equivalent and compensating variations of equations (3) and (4). Alternatively, under the Diewert-Mizobuchi (2009) assumptions on the functional form for the consumer's cost function, the Bennet indicator is exactly equal to the arithmetic average of the equivalent and compensating variations. Hence, the Fisher superlative quantity change indicator, $\mathcal{V}_E{}^F$ in (16), can be interpreted as an approximation to a welfare change indicator, $\mathcal{V}_B$.

Re-arranging (16), we get an expression for an approximation to the Fisher quantity index:

$$(17)\ Q^F \approx [(\gamma p^0 + p^1) \cdot (q^1 - q^0)]/[\ \gamma p^0 \cdot q^0\ (1+\mathscr{P}^F)] +1$$

Note that the numerator is two times the Bennet variation, $\mathcal{V}_B$. Allowing for new goods, from (12) we have the following:

$$(18)\ 2\mathcal{V}_B = 2\gamma p^0 \cdot (q^1 - q^0) + (p^1 - \gamma p^0) \cdot (q^1 - q^0) + 2p_0{}^1 q_0{}^1 - (p_0{}^1 - \gamma p_0{}^{0*}) q_0{}^1$$

[21] If real GDP growth is not constructed using a superlative index such as the Fisher, but rather using e.g. a Laspeyres index as is standard in many countries, there will still be an approximation as in (16), but it may not be as accurate.

Then replace the numerator in (17) with (18). If Q^F omits the new good in period 0, then the (approximate) amount missing from Q^F is $(\gamma p_0{}^{0*} - p_0{}^1)q_0{}^1/[\gamma p^0 \cdot q^0 (1+\mathscr{P}^F)]$, which can simply be added to Q^F if $p_0{}^{0*}$ is known or can be estimated.[22]

Real GDP growth can then be adjusted, to a second-order approximation, for not fully capturing the introduction of a new good as follows:

(19) $\text{GDP-N} = Q^F + (\gamma p_0{}^{0*} - p_0{}^1)q_0{}^1/[\gamma p^0 \cdot q^0 (1+ \mathscr{P}^F))$

where GDP-N denotes GDP growth adjusted for the introduction new goods.

4. The Free Goods Problem

Consider a household whose preferences over N market goods and M goods that are available to the household with no visible charge can be represented by the utility function f(q, z) where $q \geq 0_N$ and $z \geq 0_M$ are vectors which represent the consumption of market goods and of free goods respectively. We assume that f(q, z) is defined over the nonnegative orthant in R^{N+M} and has the following properties: (i) continuity, (ii) quasiconcave in q and y and (iii) f(q, z) is increasing if all components of q increase and increasing if all components of z increase.

We define two cost or expenditure functions that are dual to f. The first cost function is the consumer's *regular cost function*, C(u, p, w), that is the solution to the following cost minimization problem which assumes (hypothetically) that the

[22] Note that this assumes that we are either able to adjust the GDP deflator, P^F, and the CPI, γ, for the price changes in new goods, or that such goods have negligible net impact on these inflation measures. This may depend on how these respective indexes have been constructed. See Diewert, Fox and Schreyer (2018) for expressions of biases in Laspeyres, Paasche, Törnqvist and Fisher indexes arising from not appropriately accounting for new and disappearing goods.

household faces positive prices for market and free goods so that $p >> 0_N$ and $w >> 0_M$ in (1):[23]

$$(20)\ C(u, p, w) \equiv \min_{q, z} \{p \cdot q + w \cdot z: f(q, z) \geq u, q \geq 0_N, z \geq 0_M\}.$$

We also define the household's *conditional cost function*, $c(u, p, z)$, which is the solution to the cost minimization problem defined by (21) below, where the household minimizes the cost of market goods needed to achieve utility level u, conditional on having the vector $z \geq 0_M$ of free goods at its disposal:

$$(21)\ c(u, p, z) \equiv \min_{q} \{p \cdot q: f(q, z) \geq u, q \geq 0_N\}.$$

It can be shown (using feasibility arguments) that $c(u, p, z)$ has the following properties where $u \in$ Range f, $p >> 0_N$, and $z \geq 0_M$: (i) for fixed u and z, $c(u, p, z)$ is nonnegative and linearly homogeneous, concave and nondecreasing in p and (ii) for fixed u and p, $c(u, p, z)$ is nonincreasing and convex in z. If in addition, $f(q, z)$ is linearly homogeneous in q and z (the homothetic preferences case), then $c(u, p, z)$ is linearly homogeneous in u, z for fixed p.

If the household faced positive prices $w >> 0_M$ for its "free" goods, then the regular cost function minimization problem defined by (20) could be decomposed into a two stage minimization problem using the conditional cost function c; i.e., we have, using definition (20):

$$\begin{aligned}(22)\ C(u, p, w) &\equiv \min_{q, z} \{p \cdot q + w \cdot z: f(q, z) \geq u; q \geq 0_N, z \geq 0_M\} \\ &= \min_{z} \{c(u, p, z) + w \cdot z: z \geq 0_M\}.\end{aligned}$$

[23] We assume u is in the range of $f(q, z)$.

Suppose $z^* \geq 0_M$ solves the cost minimization problem that is defined in the second line of (22) and suppose further that $c(u, p, z^*)$ is differentiable with respect to the components of z at $z = z^*$. Then the first order necessary conditions for z^* to solve the cost minimization problem imply that the following first order conditions hold:

$$(23)\ \nabla_z c(u, p, z^*) = -w\ .$$

With $z = z^*$, we can go to the cost minimization problem defined by (21) and find a q solution which we denote by q^*; i.e., q^* is a solution to:

$$(24)\ \min_q \{p \cdot q: f(q, z^*) \geq u, q \geq 0_N\}.$$

It can be seen that (q^*, z^*) is a solution to the regular cost minimization problem defined by (20) so that:

$$\begin{aligned}(25)\ C(u, p, w) &\equiv \min_{q, z} \{p \cdot q + w \cdot z: f(q, z) \geq u, q \geq 0_N, z \geq 0_M\} \\ &= p \cdot q^* + w \cdot z^*.\end{aligned}$$

Thus the imputed marginal valuation prices $w \equiv -\nabla_z c(u, p, z^*) \geq 0_M$ are appropriate prices to use when valuing the services of free goods in order to construct cost of living indexes or measures of money metric utility change.

Note that due to the fact that $c(u, p, z)$ is decreasing and convex in the components of z, the marginal price for an additional unit of z_m, $w_m(u, p, z) \equiv$

$-\partial c(u, p, z)/\partial z_m$, will be nonincreasing in z_m; i.e., it will usually decrease as we add extra units of z_m to the household's holdings of free goods.[24]

We define "global" willingness to pay measures for free goods using the conditional cost function. Consider a household that holds no free goods, has utility $u^* = f(q^*,0_M)$ where q^* is the observed market goods consumption vector and the household faces the vector of market goods prices p. We assume that the household minimizes the market cost of achieving its utility level so that $p \cdot q^* = c(u^*,p,0_M)$. Now suppose that the household acquires the vector of free goods $z^* > 0_M$. Since $c(u^*, p, z)$ is decreasing in z, the amount of income that the household would require to attain the same level of utility u^* is reduced to $c(u^*,p, z^*) < c(u^*,p,0_M)$. Thus in theory, the consumer should be willing to pay $c(u^*,p,0_M) - c(u^*,p, z^*)$ to acquire the bundle of free goods z^*. Thus define the "global" *willingness to pay function* for the acquisition of z^* as follows:

(26) $W_P(u^*,p, z^*) \equiv c(u^*,p,0_M) - c(u^*,p, z^*)$.

If the household holds the amount $z^{**} > 0_M$ of free goods, then we can develop an analogous willingness to accept measure as follows. Let q^{**} denote the household's observed market goods consumption vector and we again assume that the household faces the vector of market goods prices p. Let $u^{**} \equiv f(q^{**},z^{**})$. We assume that the household minimizes the market cost of achieving its utility level u^{**} so that $p \cdot q^{**} = c(u^{**}, p, z^{**})$. Now suppose that the household disposes of its vector of free goods z^{**}. The amount of income that the household would require

[24] If consumers can have the free good in unlimited amounts, then its price will be zero. However, even if the price is zero, if quality improves, the marginal willingness to pay for the improved quality will be positive, hence $w_m(u, p, z)$ will be greater than zero. We thank Marshall Reinsdorf for this point.

to attain the same level of utility u^{**} is increased to $c(u^{**},p,0_M) > c(u^{**},z^{**})$. Thus in theory, the consumer should be willing to sell its free goods for the amount $c(u^{**}, p, 0_M) - c(u^{**}, z^{**})$, i.e. the amount that they would accept for giving up the free goods. Thus define the *"global" willingness to accept function*, for the disposal of z^{**} as follows:

(27) $W_A(u^{**}, p, z^{**}) \equiv c(u^{**}, p, 0_M) - c(u^{**}, p, z^{**})$.

For welfare measurement purposes, it is useful to define *marginal* willingness to accept functions. Thus let e_m be a unit vector of dimension M with a 1 in component m and zeros elsewhere for $m = 1,...,M$. Assume that the household holds $z \geq 1_M$ units of the free goods, faces market prices p, has $q > 0_N$ units of market goods and $p \cdot q = c(u, p, z)$ where $u = f(q, z)$. Define the *m^{th} marginal willingness to accept function*, $W_m(u, p, z)$ as follows:

(28) $W_m(u, p, z) \equiv c(u, p, z-e_m) - c(u, p, z)$; $m = 1,...,M$.

Survey, experimental or indirect methods can be used in order to obtain approximate measures for these marginal willingness to accept functions. Let $W(u, p, z)$ denote the vector $[W_1(u, p, z),...,W_M(u, p, z)]$. It can be seen that $W(u, p, z)$ is a discrete approximation to the marginal valuation price vector $w \equiv -\nabla_z c(u, p, z)$ that was defined earlier by (23).[25]

Assuming that we have valuations for the free goods, we can extend the Bennet welfare change variation of (12) to include these goods. Following the set up for

[25] If $z_m = 0$, then we need to change the definition of $W_m(u, p, z) \equiv c(u, p, z-e_m) - c(u, p, z)$ to the corresponding marginal willingness to pay function, $W_m^*(u, p, z) \equiv c(u, p, z) - c(u, p, z+e_m)$.

regular goods in the previous section, let a new "free" good be indexed by the subscript 0 and let the N dimensional vectors of period t prices and quantities for the continuing goods be denoted by w^t and z^t for t = 0,1. The period 1 quantity of good 0 purchased during period 1 is also observed and is denoted by z_0^1. The period 0 reservation price for good 0 is not directly observed but we make an estimate for it, denoted as $w_0^{0*} > 0$. The period 0 quantity is observed and is equal to 0; i.e., $z_0^0 = 0$. Thus the price and quantity data (for the N+1 goods) for period 0 is represented by the 1+N dimensional vectors (w_0^{0*}, w^0) and $(0, z^0)$ and the price and quantity data for period 1 is represented by the 1+N dimensional vectors (w_0^1, w^1) and (z_0^1, z^1).

Then, in an extension of (12), welfare change including both new and free goods can be written as follows, where we again adjust period 0 prices by the one plus the growth rate of the CPI between periods 0 and 1, γ:

$$
\begin{aligned}
(29)\quad V_B = {} & \gamma p^0\cdot(q^1 - q^0) + \tfrac{1}{2}(p^1 - \gamma p^0)\cdot(q^1 - q^0) + p_0^1 q_0^1 - \tfrac{1}{2}(p_0^1 - \gamma p_0^{0*})q_0^1 \\
& + \gamma w^0\cdot(z^1 - z^0) + \tfrac{1}{2}(w^1 - \gamma w^0)\cdot(z^1 - z^0) + w_0^1 z_0^1 - \tfrac{1}{2}(w_0^1 - \gamma w_0^{0*})z_0^1,
\end{aligned}
$$

where the second line gives the contribution of the continuing and entering "free" goods.

If the concern is that real GDP omits the contribution from continuing free goods, then we can use the results of the previous section and re-write (19) to adjust real GDP growth, Q^F, as follows to reflect the welfare effects of free goods:[26]

[26] Welfare change in (29) should also be adjusted for general inflation, especially if inflation is high or if the periods being compared are far apart in time, and similarly for spatial comparisons.

(30) GDP-F = $Q^F + [2\gamma w^0\cdot(z^1 - z^0) + (w^1 - \gamma w^0)\cdot(z^1 - z^0) + 2w_0{}^1 z_0{}^1]/[\ \gamma p^0\cdot q^0\,(1+ \mathscr{P}^F))$,

where GDP-F denotes GDP growth adjusted for free goods.[27]

Including both regular and free new goods, we get the following expression for our adjusted real GDP growth:

$$
\begin{aligned}
\text{(31) GDP-B} = {} & Q^F + (\gamma p_0{}^{0*} - p_0{}^1)q_0{}^1/[\gamma p^0\cdot q^0\,(1+ \mathscr{P}^F)] \\
& + [2\gamma w^0\cdot(z^1 - z^0) + (w^1 - \gamma w^0)\cdot(z^1 - z^0) + 2\gamma w_0{}^1 z_0{}^1]\ /[\gamma p^0\cdot q^0\,(1+ \mathscr{P}^F)] \\
& + (\gamma w_0{}^{0*} - w_0{}^1)z_0{}^1/[\gamma p^0\cdot q^0\,(1+ \mathscr{P}^F)),
\end{aligned}
$$

where the first line of (31) is the adjustment arising from the entry of a new good, the second line is an additional contribution from accounting for continuing free goods, and the third line is the adjustment term arising from the entry of a new free good.[28] Thus GDP-B denotes GDP growth adjusted for new and free goods.[29] As GDP-B in (31) nests GDP-N from (19) and GDP-F from (3), we propose this as the generic term for these types of measures.

An alternative, simpler, approach to adjusting GDP for free goods is as follows. Using equation (25), we can define *total income* (T) as follows:

$$\text{(32) } T \equiv C(u, p, w) = p\cdot q^* + w\cdot z^*,$$

where (q^*,z^*) is a solution to the cost minimization problem with positive prices $w >> 0_M$ for the “free” goods. Hence, (32) gives the total income required so that a

---

[27] Note that this assumes that we are either able to adjust the GDP deflator, P^F, and the CPI, γ, for the price changes in continuing free goods, or that such goods have negligible net impact on these inflation measures.

[28] Obviously, (31) can easily be generalized to the case of multiple new regular and free goods.

[29] The “B” in GDP-B can be thought of as standing for the “benefits” arising from new and free goods, or “beyond”, as in the literature promoting broader measures of economic wellbeing “beyond GDP”.

certain level of utility can be attained through the consumption of market and free goods. Then $w \cdot z^*$ is the amount a consumer should be willing to pay to acquire the bundle of free goods z^*; see the willingness to pay function of equation (26). Alternatively, $w \cdot z^*$ is the amount of income needed to compensate for giving up the consumption of free goods, while maintaining the same level of utility; see the willingness to accept function of equation (27). Deflating the resulting nominal total income growth between periods 0 and 1, T^1/T^0, by the GDP deflator, P, gives real total income growth, GDP-B_T:

(33) $\text{GDP-B}_T \equiv (T^1/T^0)/P$

The GDP deflator will typically be the wrong deflator, as it does not take into account new (and disappearing) goods, which would usually mean that the deflator is too high.[30] The resulting quantity index then provides a lower bound estimate on the actual real growth rate.[31]

This total income approach has the advantage of not needing the period 0 reservation price for any new good, as the quantity consumed of the good in this period is 0 so that $T^0 = p^0 \cdot q^0$.

To summarize, GDP-B describes the extension of GDP to incorporate consumer benefits arising from digital goods, as measured through experiment evidence.

[30] This is because new goods typically fall in price after their introduction. Also, note that by using the GDP deflator here, there is an implicit assumption that an appropriate reservation price for the free good is the price observed in period 1 carried back to period 0. This is the carry-backward method discussed by Diewert, Fox and Schreyer (2018).

[31] Diewert, Fox and Schreyer (2019) have subsequently generalized this Total Income approach to consider non-free new goods. They show that, under some assumptions, the difference between GDP-B in (33) and standard GDP can be interpreted as the amount by which a maximum overlap quantity index (as typically calculated by national statistical offices) understates an approximate "true" Fisher index calculated using reservation prices for the new goods.

Our first method (equation 31) uses this experimental evidence on consumer valuations to derive an extension of GDP which is consistent with standard Hicksian concepts of welfare change. Our second ("total income") method (equation 33) extends GDP by including the extra income needed to achieve the same level of utility without the digital goods as with the digital goods.

Just as our approach makes it possible to calculate GDP-B in a way that accounts for new and free goods, it is straightforward to calculate an alternative measure of labor productivity by dividing GDP-B by hours worked. To distinguish it from conventionally-measured productivity, one can label this new metric *Productivity-B*.

5. Empirical Examples of GDP-B Applied to Free Digital Goods

In this section we apply our methodology to study the welfare gains generated by free digital goods. First, we consider the case of Facebook, using online choice experiments to elicit user valuations. Then we consider the valuation of a broader range of digital goods, using laboratory experiments in the Netherlands.

a) Valuing Facebook in the US

To estimate the consumer welfare created by Facebook, we conducted incentive compatible discrete choice experiments on a representative sample of the US internet population. Specifically, we set quotas for gender, age, and US regions to match US census data (File and Ryan 2014) and applied post-stratification for education and household income to obtain our sample. Because our focus is on Facebook users, we disqualified participants who did not use Facebook in the previous twelve months (but we can account for the overall number of Facebook users using secondary data).

In the experiment, each participant was asked to make a single discrete choice between two options: 1) keep access to Facebook or 2) give up Facebook for one month and get paid $E. We allocated participants randomly to one of twelve price points: E = (1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000). Before participants made the decision, we informed them that their decisions were consequential such that we would randomly pick one out of every 200 participants and fulfil that person's selection (Ding, Grewal and Liechty 2005; Ding 2007; Carson, Groves and List 2014). We also informed them about how we can monitor their Facebook online status remotely. In order to check if the selected participants gave up Facebook and qualified for the payment, we monitored their online status on Facebook for 30 days.[32]

We recruited respondents through an online professional panel provider, Research Now,[33] during the year 2016-17.[34] A total of 2885 participants completed the study including at least 200 participants per price point. We targeted consumers that were 18 years or older and lived in the US. We further asked consumers to select all online services they have used in the last twelve months from a list of 14 options, including a non-existent online service which we used as an attention check. We selected Facebook users for this study and disqualified users who selected the non-existent service. Participants were randomly allocated to one of the price points and we combine responses from all participants to estimate the demand curve.

[32] It is possible to remotely monitor when someone is last logged in on Facebook for any friend on Facebook.
[33] https://www.researchnow.com/
[34] These experiments are also reported in Brynjolfsson, Collis and Eggers (2019). In this paper, we combine the studies conducted in summer 2016 and summer 2017 to come up with estimates for the year 2016-17.

We fitted a binary logit model to the participants' decisions using the monetary values (in log scale) as predictors. Figure 1 shows the observed shares of participants willing to keep Facebook and the fitted line according to the logit model. According to the model, the median willingness-to-accept (WTA) price for giving up Facebook for one month is \$42.17 (bootstrapped 95% confidence interval = [\$32.53; 54.47]).[35]

Next, we provide an empirical illustration of the theoretical framework for free goods provided in Section 4. We consider the period from 2003 to 2017; Facebook was founded in 2003-04 and hence became a new free good that year. In our notation of the previous section, 2003 is then period 0 and 2017 is period 1. Assuming a simple linear relationship, the median WTA for Facebook in 2017 (\$42.17/month), translates to (w_0^1=) \$506.04/year ([390.36; 653.64]).[36] Note that this is price for giving up the 2017 version of Facebook, which includes all its attributes at the time, including the number of users, or size of the social network. We also need to determine the reservation price for Facebook in 2003 (w_0^{0*}); recall that the reservation price is the price which would induce a utility maximizing potential purchaser of a good to demand zero units of it. Here the good which is having its demand reduced to zero is the 2017 version of Facebook.

Following Hausman (1996), we could consider a reservation price of twice the median WTA (deflated to 2003 dollars); the reservation price for before the 2004 launch of Facebook is a then ($w_0^{0*} = 2w_0^1/\gamma \approx$) \$780. This is likely to be a very

[35] This "willingness to accept" price corresponds to the global willingness to accept function in equation (27) of Section 4. That is, it is the income needed in compensation for giving up the free good if the same utility level is to be maintained.

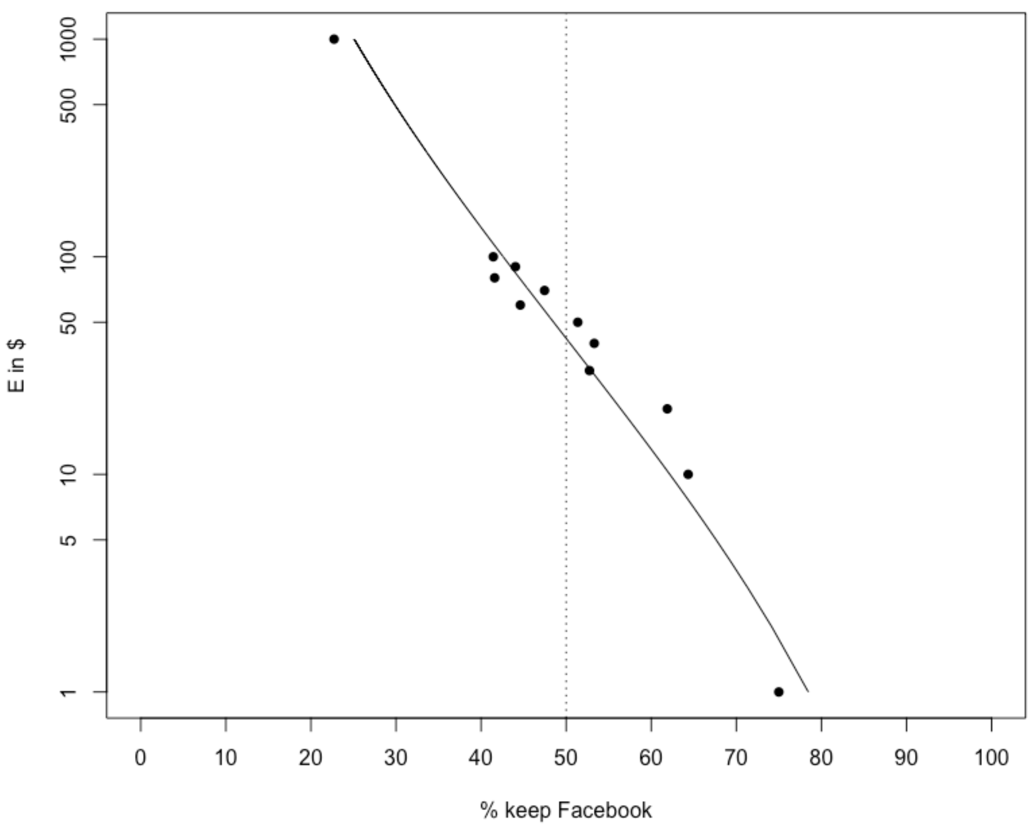
[36] Brynjolfsson et al. (2019), find that the relationship between valuation and time period is roughly log-linear and not linear, i.e. valuation for 1 year is a less than 12 times valuation for 1 month. Using hypothetical choice experiments, we find that it is closer to 10 times the valuation for 1 month. Here we assume a linear relationship for simplicity since it is not feasible to do a one-year incentive compatible study for Facebook.

conservative estimate. Note that the observed demand curve in Figure 1 reflects a much higher reservation price. In fact, there is a significant portion of the sample (>20%) which values Facebook at more than $1,000 per month. Apple-Cinnamon Cheerios, the product considered by Hausman, can be regarded as quite different to Facebook; it is a new variety of breakfast cereal with plenty of close substitutes, whereas Facebook can be characterized as a novel product.[37] In contrast to the low reservation price from applying Hausman's estimate, the approach of Feenstra (1994) uses a CES framework which requires that all reservation prices are infinity. This seems unreasonably high in our context.[38]

[37] Reinsdorf and Schreyer (2017, p. 5) note the following regarding the consequences for consumer price inflation of delaying the price measurement of such products: "…novel products may initially exhibit distinctive price change behaviour. The most common pattern is for prices of truly novel products to decline quickly at first, so the bias is upward."

[38] "Thus the CES methodology may overstate the benefits of increases in product availability." Diewert and Feenstra (2017, p.3).

Figure 1: WTA demand curve for Facebook



Hence, we focus on an alternative approach and estimate the intercept term in a linear regression of WTA on the corresponding share of users who keep Facebook, as plotted in Figure 1; this is the estimate of the monthly WTA that gives a share of zero. Our estimate is from a regression that omits the two extreme observations, for E = \$1 and E = \$1,000 (p-value = 0.0000, R^2=0.88).[39] At extreme values, even a small number of noisy responses will disproportionately affect the reservation

[39] We also estimated a regression using all observations. This resulted in a poorer fit (p-value = 0.0038, R^2=0.52) and a much higher estimate of the reservation price (\$8,126 in 2003\$). Using this higher estimate, we would find that the contribution to welfare change over the period 2003-17 is \$1,013 billion (in 2017\$) which translates to an average of \$72 billion per year. Per user, the welfare change over the period 2003-17 is \$5,018 which translates to \$358.48 on average per year.

price. Multiplying the estimate by twelve yields the 2017 annual reservation price and deflating, using the CPI, yields the reservation price in 2003 dollars. Using this approach, we estimate the reservation price (w_0^{0*}) to be $2,152 in 2003 dollars.

The estimated contribution to welfare due to Facebook in the U.S. over the period 2003-17 is $231 billion (in 2017$) which translates to $16 billion on average per year.[40] The per user welfare gain over the period 2003-17 is $1,143. Considering that this is a single new service, this estimate is substantial.[41] At the same time, given that the definition of users is that they access their Facebook account via any device at least once per month and the average user is Facebook for more than 40 minutes per day,[42] then this estimate does not seem excessive.

Next we turn to GDP-B growth to get an idea of the change that would result from extending the usual definition of GDP to include a free service such as Facebook. From the last line of equation (31) of Section 4, we have the following:

---

[40] Notes:
w_0^1 = $506.04 (95% C.I.: [390.36; 653.64])
γ = 1 + Growth rate of CPI = 1.3
Number of Facebook users in US in 2017 = 202 million
Data sources:
Chained CPI-All Urban Consumers, not seasonally adjusted, index for December 2003 to December 2017 is 1.2975, or 29.75%. https://www.bls.gov/cpi/data.htm
Internet users who access their Facebook account via any device at least once per month. https://www.statista.com/statistics/408971/number-of-us-facebook-users/

[41] Note that we are not accumulating benefits from the years in between 2003 and 2017. We are simply comparing the welfare change between two periods: 2003 when Facebook did not exist and 2017 when the 2017 version existed. The comparison between these two years, as opposed to any of the intervening years, is of interest as there was no close substitute to any subsequent version of Facebook in 2003. In the intervening years, if each version of Facebook, with increasing network size, is treated as a new good then we would need to also model the impact of the exiting versions of Facebook. We do not have the valuations required to do such a study.

[42] See https://www.emarketer.com/Chart/Average-Time-Spent-per-Day-with-Facebook-Instagram-Snapchat-by-US-Adult-Users-of-Each-Platform-2014-2019-minutes/211521

Adjustment to real GDP-B growth from accounting for Facebook over 2003-2017

$$= (\gamma w_0^{0*} - w_0^{1})z_0^{1}/[\gamma p^0 \cdot q^0 (1+ \mathcal{P}^F)]$$

$$= (\gamma w_0^{0*} - w_0^{1}) \text{ x No. of Facebook users in US in 2017 / } \gamma(\text{Nominal GDP in 2003})(1+ \mathcal{P}^F)$$

The GDP adjustment is a lower bound on the amount to add to GDP-B growth using this approach because we use official estimates of γ and P^F (which are unadjusted for the introduction of new goods) in the denominator. Normally, γ and P^F would be lower if we account for the fact that the price of the new goods typically fall following their introduction.[43]

From Table 1, for the reservation price of w_0^{0*} = \$2,152 in 2003, accounting for Facebook would increase real GDP-B growth by 1.54 percentage points from 2003 to 2017 (or, using the 95% CI estimates of w_0^1: [1.44, 1.62]). In other words, this amounts to an increase in real GDP-B growth of 0.11 percentage points on average per year over this period and an identical increase in Productivity-B. Real GDP grew by 28.82% and real GDP-B grew by 29.16% including the contribution from Facebook. Average real GDP growth over this period was 1.83% per year. Adding the contribution of Facebook means that GDP-B grew by 1.94% per year.[44] Considering that this is for just one product, including the benefits from Facebook results in a large impact on such an encompassing measure of economic activity as GDP-B and productivity-B.

Table 1: GDP-B Contributions, Facebook

	Total Income	Reservation Price

[43] See Diewert, Fox and Schreyer (2018) and Reinsdorf and Schreyer (2017).
[44] The corresponding growth estimate from using the reservation price estimated using all observations (\$8,126) is 2.20% per year on average.

Reservation Price w_0^{0*}, 2003$	—	$2,152
Percentage Points, 2003-2017	0.68	1.54
Per year	0.05	0.11
GDP-B Growth per year without Facebook (i.e. GDP growth)	1.83	1.83
GDP-B Growth per year with Facebook	1.87	1.94

Notes: w_0^1 = $506.04 (95% C.I.: [390.36; 653.64]), γ = 1 + Growth rate of CPI = 1.3, P^F = 1+ Growth rate of GDP Deflator[45] = 1.31, $\mathscr{P}^F = P^F/\gamma = 1.0078$, Number of Facebook users in US in 2017 = 202 million, Nominal GDP for 2003[46] = $11.5 trillion; The reservation price is 12 times the intercept from a linear regression of monthly WTA on the corresponding share of users who keep Facebook, dropping the observations for the two extreme observations, E=$1 and E=$1000 (p-value = 0.0000, R^2=0.88). "Per year" estimates are calculated using the arithmetic mean of the percentage point difference over the period. "Growth per year" estimates are calculated using geometric means.

Next we consider the total income approach of equation (33) in Section 4. We need the total nominal income (T) for both 2003 and 2017, which we calculate as follows:

T^0 = nominal GDP in 2003 + $w_0^{0*}z_0^0$ = $11.51 trillion + 0 ≈ $11.51 trillion

T^1 = nominal GDP in 2017 + $w_0^1 z_0^0$ = $19.39 trillion + $506.04 x No. of Facebook users in US in 2017 ≈ $19.49 trillion.

That is, total nominal income using GDP-B_T is higher by $102 billion in 2017 since the value of Facebook to consumers is taken into account. Recall, from Section 4, that this can be interpreted as the amount that consumers in aggregate would need in compensation in order to attain the same level of utility if access to Facebook foregone in 2017. This is for the 2017 version of Facebook, including

[45] GDP Implicit Price Deflator, annual, not seasonally adjusted, 2010=100: Growth for 2003 to 2017 = 112.05/85.69 = 1.31. https://fred.stlouisfed.org/series/USAGDPDEFAISMEI

[46] Gross Domestic Product, annual, not seasonally adjusted: https://fred.stlouisfed.org/series/GDPA. The beginning of year value for a 2004 product launch is the GDP of 2003.

all its characteristics, such as the size of the network. Hence, the result is independent of the changes in the characteristics of Facebook over the intervening years since its launch.

From equation (33), in our case $\text{GDP-B}_T = (T^1/T^0)/P^F = (19.49/11.51)/1.31 \approx 1.295$. Thus GDP-B grew by 29.50% between 2003 and 2017 using the total income approach, whereas conventionally-measured real GDP grew by 28.82%, giving a percentage point difference of 0.68 over the entire period, or 0.05 per year on average.[47]

Compared with conventionally-measured real GDP growth of 1.83%, our estimates of average GDP-B growth per year range from 1.87% for the total income approach to 1.91% for the approach using our estimate of the reservation price.

b) Valuing Free Digital Goods Using Participants in a Laboratory

We conducted similar incentive compatible discrete choice experiments in a university laboratory in the Netherlands in order to evaluate additional free digital services.[48] While the online status on Facebook can be monitored remotely to make sure that participants did not use this service, other digital goods do not offer this functionality so that we needed another approach to make the decisions consequential. For services that require a password-protected login, we informed the participants that, if selected, they will have to change the password to a computer-generated code that would be kept in a sealed envelope afterwards. If

[47] Recall that this is can be thought of as an underestimate of the additional growth from using GDP-B, as the deflator is not adjusted for the impact of new goods prices.
[48] These valuations are also reported in Brynjolfsson, Collis and Eggers (2019).

the seal was still intact and the password remained valid (not reset), we concluded that the participant in fact did not use this service. Additionally, we informed that we would check the usage statistics of the apps on the selected participants' devices. Therefore the laboratory setting was necessary in order to be able to contact participants in person after the study and make their decisions consequential.

We tested the valuation of the services Instagram, Snapchat, Skype, WhatsApp, digital Maps, Linkedin, Twitter as well as Facebook. We varied the monetary amount that we offered to participants to leave these services for one month within the range of €1 to €500. The respondents had to make decisions regarding each of these services, i.e., each respondent had to make eight decisions. One out of every fifty participants who completed the study got the chance to have their decision fulfilled. The specific service was determined randomly in this case.

The data collection took place at a large Dutch university in February and October 2017. Overall, 426 participants were available for the analysis, meaning that there were over 400 decisions for each digital service. The resulting estimated demand curves are given in Figure 2. The corresponding median WTA valuations and confidence intervals are given in Table 2.

We observe very high valuations for WhatsApp which all of the participants were using. No one was willing to give it up for €1, and the relative insensitivity of demand to price resulted in an estimated monthly median WTA of €535.73, far higher than for the other services. We interviewed participants after the study period to better understand these high valuations. They told us that WhatsApp had become a nearly indispensable focal platform for communicating with peers, co-workers and others in their community, leading to enormous disutility from lack of access.[49] Of course, the disutility for an individual would likely be much less if all members of the community could coordinate on switching to an alternative communications platform and the values should be interpreted accordingly. Such network effects are observed with many other goods as well, and do not mean that the valuations should be discounted but it may affect the value of other substitute goods.[50] Hence, the net contribution to welfare should account for changes in both the value on the focal good, and such substitutes.

In general, any good has a certain price/ valuation for every state of the world referred to as Arrow-Debreu state prices (for e.g. a bottle of water has a different valuation if you are thirsty versus not). In addition to network effects, digital goods can also have different valuations based on how long you have to give them up for and the availability of substitutes and complements. Specifying the state of the world in choice experiments lets us uncover the set of valuations for a single good across different states. For example, we could solicit valuations for giving up WhatsApp but letting them use substitutes or completely giving up all instant messaging services.

Facebook was used by almost all of the participants and had the next highest median WTA monthly valuation of around €100. The valuation for Facebook in this sample was thus significantly higher than that found for the US in the previous section ($42.17 ≈ €34.76). Maps (including Google, Bing, and Apple

[49] Some quotes from our interviews: 1. "Whatapp is the only communication tool I use to contact my friends here. Without it, I can do nothing." 2. "WhatsApp is crucial. I use the app every hour of the day to keep in touch with friends and family but also to discuss group projects or things about my work. I really need to keep access to this app. There is also not a very suitable alternative."

[50] The fact that most people now use telephones to communicate rather than telegrams does not mean that the price people are prepared to pay for calls should be discounted in any way. That said, the value is partly due to network effects and partly due to intrinsic differences between the two goods.

maps) were also highly valued, with WTA median values of almost €60 per month, followed by Instagram, Snapchat and LinkedIn.

For Skype and Twitter, we found very low median valuations of less than €1. Although 71% of the participants were using Skype, the majority were willing to give it up for one month for just €1, likely because other services offered very similar (video) calling possibilities and was not frequently used. Note that although Skype effectively provides access to a portion of the same network for 71% of sample, the valuation is massively different; €535.73 for WhatsApp and €0.18 for Skype. This suggests that it is not simply a valuation of the network that is being captured.

Twitter is only used by 33% of the sample which explains the low value for the median user, i.e., the utility maximizing strategy for those who do not use Twitter is, of course, to accept any money that was offered, and this encompasses the majority of users in our sample.

These WTA estimates are converted to annual figures by simply multiplying by twelve to get the annual estimates, as per the previous section, and these figures are then used to calculate annual GDP-B growth for the Netherlands. We use the total income method of equation (33), and hence avoid having to estimate a reservation price for each good. The results are reported in Table 3.[51] Since our sample for these laboratory experiments is not representative of the national population of Netherlands, we provide these figures solely to gauge the approximate magnitude of potential underestimation in welfare inferred from

[51] The welfare change estimates are available from the authors on request.

conventional GDP growth figures from not accounting for popular free digital services.

Figure 2: WTA demand curves for popular digital goods measured in a laboratory

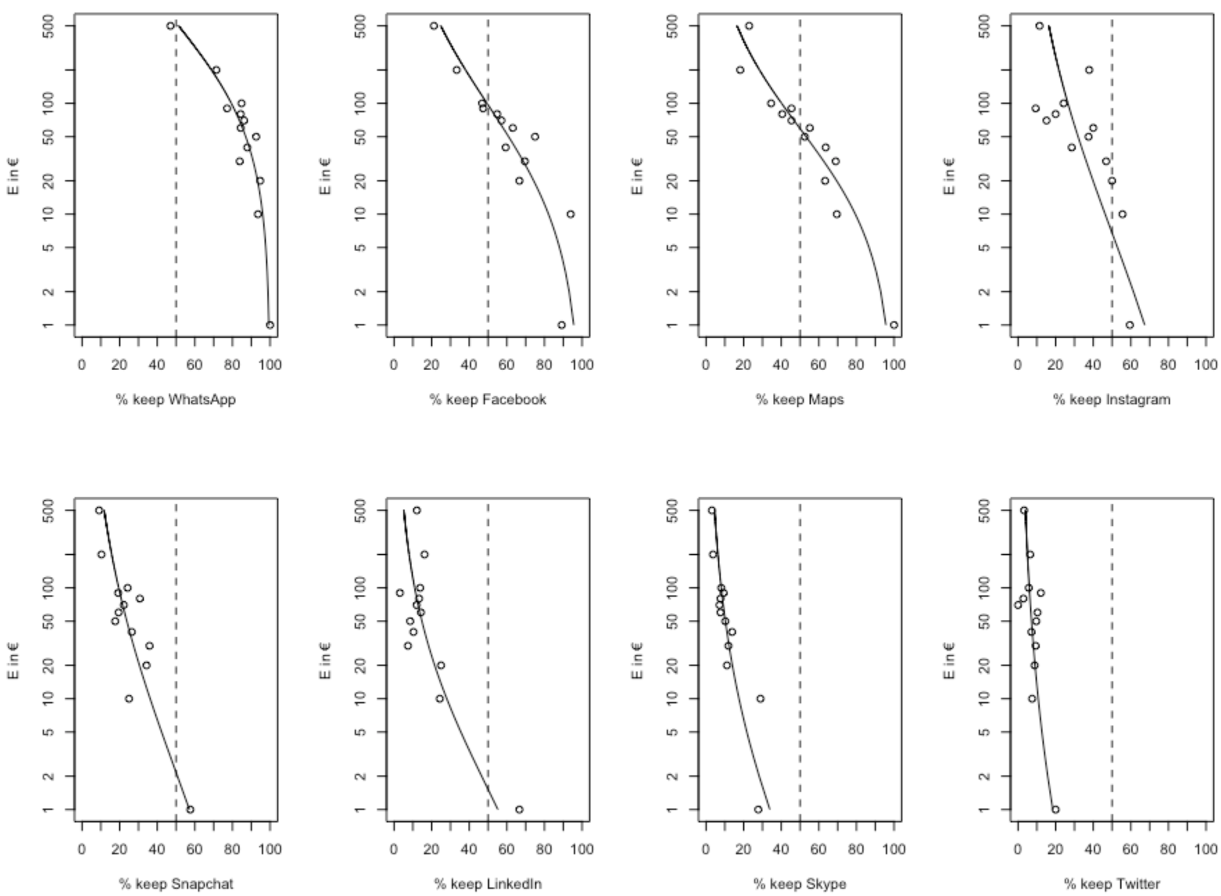


Table 2: Median Monthly WTA

Service	Launch Date	Median WTA	Lower CI	Upper CI
WhatsApp	January 2009	€535.73	€269.91	€1141.42
Facebook	February 2004	€96.80	€69.54	€136.68
Maps	February 2005	€59.16	€45.17	€78.31
Instagram	October 2010	€6.79	€2.53	€16.22
Snapchat	September 2011	€2.17	€0.41	€8.81
LinkedIn	May 2003	€1.52	€0.30	€5.84
Skype	August 2003	€0.18	€0.01	€2.58
Twitter	March 2006	€0.00	€0.00	€0.49

Table 3: Estimates of gross contributions of popular digital goods to real GDP-B growth in the Netherlands, percentage points, Total Income Method

Users Service	Average per year 10 million	Average per year 2 million
WhatsApp	4.10	0.82
Facebook	0.5	0.11
Maps	0.34	0.07
Instagram	0.07	0.01
Snapchat	0.02	0.00
LinkedIn	0.01	0.00
Skype	0.00	0.00
Twitter	0.00	0.00

Notes: Two alternative user populations are considered, 10 million and 2 million. The population in July 2017 was approximately 17 million, with around 2 million in the 15-24 age group (https://www.indexmundi.com/netherlands/demographics_profile.html), which is the age group of our laboratory sample. In January 2016, WhatsApp had 9.8 million (https://nltimes.nl/2016/01/25/dutch-people-leaving-twitter-en-masse-use-whatsapp-facebook). Quarterly data are used.[52] For products launched in the first half of the year, the period 0 values are taken to be those from quarter 4 of the preceding year. For products launched in the second half of the year, period 0 values are taken to be those of quarter 4 of the launch year. Per year estimates are calculated using arithmetic means of the percentage point difference in growth over the period that the respective goods were available.

From Table 3 we can see that WhatsApp, Facebook and digital maps contribute significantly towards GDP-B growth and hence conventional GDP estimates miss a great deal of value by not accounting for these goods. According to our estimates, if WhatsApp is used by only 2 million people in the Netherlands (the approximate population in the 15-24 years old age group in 2017 and the age group of our laboratory sample), its gross contribution to GDP growth over 2003 to 2017 would be 0.82 percentage points per year. This is large, especially when considering that (i) this is just one digital good, and (ii) that the actual using population of WhatsApp is likely to be much larger than 2 million. The actual Dutch number of users has been reported to be closer to 10 million, for both WhatsApp and Facebook.[53]

Hence, in Table 3 we report also report results for a user population of 10 million and find that, if accounted for, the annual average gross contribution of WhatsApp to GDP-B growth would have been a substantial 4.10 percentage points according to the total income method. It is important to note that if WhatsApp partially

---

[52] CPI: https://fred.stlouisfed.org/series/NLDCPIALLMINMEI;
Real GDP: https://fred.stlouisfed.org/series/CLVMNACNSAB1GQNL;
Nominal GDP: https://fred.stlouisfed.org/series/CPMNACNSAB1GQNL
The GDP Implicit Price Deflator is calculated as the ratio of the nominal GDP series divided by the real GDP series. This is because the official deflator series is annual (an average over the four quarters of each year), and we need to ensure that price times quantity equals value.

[53] According to an NL Times story on January 25 2016, “Whatsapp is the largest social network in the Netherlands with 9.8 million users. Facebook came in second place with 9.6 million....” https://nltimes.nl/2016/01/25/dutch-people-leaving-twitter-en-masse-use-whatsapp-facebook.
Given definitional uncertainty about what constitutes a “user”, and the potential for rapid change in user numbers, we consider potential bounds of 2 million to 10 million users out of a population of 17 million.

replaces conventional telephone calls and texting, then the traditional GDP captures the fall in disappearing value of these telephone services but misses the gains from WhatsApp. In contrast, the adjustment term to GDP-B growth due to WhatsApp could be very high because it captures these benefits from the introduction of WhatsApp relative to the counterfactual of lower valued telephone services.[54] This problem of GDP not reflecting benefits from free goods could become increasingly severe as more and more free digital goods are used as substitutes for traditional paid goods, such as Wikipedia replacing encyclopedias and various smartphone apps replacing a variety of traditional goods.

6. Applying GDP-B to adjusting for new features in smartphone cameras

Smartphone cameras are now the primary devices for taking photos. From the 1997 to 2017, the dominant photographic technology shifted from analog cameras to digital cameras to smartphone cameras. The total number of digital cameras shipped worldwide dropped from 121 million units in 2010 to 24 million units in 2016,[55] while worldwide smartphone sales increased from 297 million in 2010[56] to 1.5 billion in 2016.[57] Moreover, the marginal cost of taking a photo has fallen to approximately zero with smartphones, compared with up to 50 cents per photo for developing film in the analog era. Just between 2010 and 2017, the number of photos taken worldwide has increased from 350 billion to an estimated 2.5 trillion.[58] Furthermore, a photo taken on a smartphone today is typically superior

---

[54] In other words, in an alternative world without WhatsApp, the counterfactual GDP-B would drop by somewhat less than our estimate because users would probably have relatively higher valuations for telephone services.

[55] http://www.cipa.jp/stats/dc_e.html

[56] http://www.gartner.com/newsroom/id/1543014

[57] http://www.gartner.com/newsroom/id/3609817

[58] https://www.nytimes.com/2015/07/23/arts/international/photos-photos-everywhere.html

to a photo taken on an average camera twenty years ago, including its ability to be stored, shared or repurposed far more easily.

To illustrate the problem this change creates, we consider a simple case of two goods, each available in two periods: a digital camera and a feature phone[59] in period 0, and a smartphone with a digital camera in period 1.[60] Suppose that the value of the camera to the consumer is v_c, the value of the simple feature phone is v_f, and the value of the smartphone is v_c+v_f. Assume that a device fully depreciates in a time period, i.e., a consumer has to purchase new devices each period. Also assume that a consumer buys both the camera and the feature phone in period 0 and only the smartphone in period 1, and there are a total of x such consumers. Suppose that the price of the camera is p_c in period 0, the price of the feature phone is p_f in period 0, and the price of the smartphone is also p_f in period 1. Then we have the following consumer surplus measures, CS^0 and CS^1, for periods 0 and 1, respectively:

(34) $CS^0 = (v_c - p_c)x + (v_f - p_f)x \geq 0,$

(35) $CS^1 = (v_c+v_f - p_f)x \geq 0.$

Then the change in consumer surplus between periods 0 and 1 is $CS^1 - CS^0 = p_c x$. This is the cost saving of not buying the digital camera in period 1 because its functionality is now included in the smartphone. However, the contribution of these goods towards conventionally-measured GDP (i.e., the market price of final goods) is $(p_c + p_f)x$ in period 0 but only $p_f x$ in period 1. Hence the change in conventionally-measured GDP from period 0 to period 1 is $-p_c x$, which is exactly

---

[59] A feature phone is a phone defined as a phone with no camera for the purposes of this example.
[60] We thank Hal Varian for sharing his notes on GDP and welfare which contained a version of this example.

the opposite of the change in consumer surplus. Therefore, while conventionally-measured GDP goes down due to people not purchasing the digital camera, consumer surplus and GDP-B go up. The measured decrease in conventional GDP occurs because, even though it has the same market price (p_f) as the feature phone in this example, the smartphone is a higher quality product. That is, there is an implicit fall in price in shifting from the feature phone to the smartphone which is not being captured.

Hence, it is clear that GDP statistics should account for quality improvements in smartphones, including the introduction and improvements in smartphone cameras. While GDP-B does this, until January 2018, the BLS only incorporated quality adjustments for data plans offered by mobile network operators in the CPI.[61] Starting from January 2018, there is now quality adjustment of the CPI for telephone hardware, calculators and other consumer information items using hedonic modelling of the value of characteristics;[62] this is used by the Bureau of Economic Analysis (BEA) to deflate Personal Consumption Expenditures for telephone and facsimile equipment in constructing real GDP; see BEA (2014, Chapter 5, Table 5.A). Therefore, even though GDP statistics capture paid goods such as smartphones, they have failed for many years to completely capture quality adjustments in the US, and most countries still do not make any quality adjusts for smartphones; see e.g. Wells and Restieaux (2014, Table 1). Even when they do attempt to adjust for quality improvements, Groshen et al. (2017) state that hedonic techniques are not suitable for products such as smartphones when the set of relevant characteristics frequently change.[63] Note that quality

[61] https://www.bls.gov/cpi/factsheets/telephone-services.htm
[62] The methodology and characteristics used for the hedonic modelling are currently not published. https://www.bls.gov/cpi/factsheets/telephone-hardware.htm
[63] If we consider software features (including operating system and various apps) as part of the set of relevant characteristics for hedonic quality adjustments, then it is impossible to perform

improvements, such as the addition of a camera feature to a smartphone, can also be thought of as additions of new goods as described in our framework.

To demonstrate the importance of quality change as can be captured by GDP-B, we elicit the value generated of smartphone cameras for participants in a university laboratory in the Netherlands and compare that with the approximate price paid for them.

Specifically, we applied an incentive compatible BDM lottery (Becker, DeGroot, and Marschak 1964) in order to estimate the consumers' valuation of their smartphone camera. We asked participants to state the minimum amount of money they would request in order to give up their smartphone camera (both main camera and front camera) for one month. Participants were informed that this amount would serve as a bid in a lottery. If their minimum bid to forego their camera would be higher than a random price, drawn from a uniform distribution, they could keep access to their smartphone camera but would not receive any cash. If the random price exceeded their minimum requested amount, they would be paid the random price, provided that they would give up using the smartphone camera for one month. The utility-maximizing strategy of the participants in the BDM lottery is to provide a bid that matches their true valuation. Accordingly, we use the bids as measures of WTA to give up smartphone cameras.

In order to induce incentive compatibility and make the answers consequential, we provided further information that one out of fifty participants would be selected for the lottery and that if their bid was successful we would block their smartphone cameras with a special sealing tape (see Figure 3). The sealing tape

hedonic modelling because firms do A/B testing continuously and seek to improve these features as frequently as daily.

would break if the participants tried to peel it off so that it was not possible to re-apply it. We also signed the tape so that it was not possible to buy the same type of seal and re-apply a seal. If, after the one month period, the seal was still intact participants were rewarded with the money and the seal could be removed.

The study was conducted in the laboratory of a large Dutch university in November/December 2017 (to not cover the holiday season, respondents were allowed to postpone giving up their camera until January 2018). In total, 213 students participated.

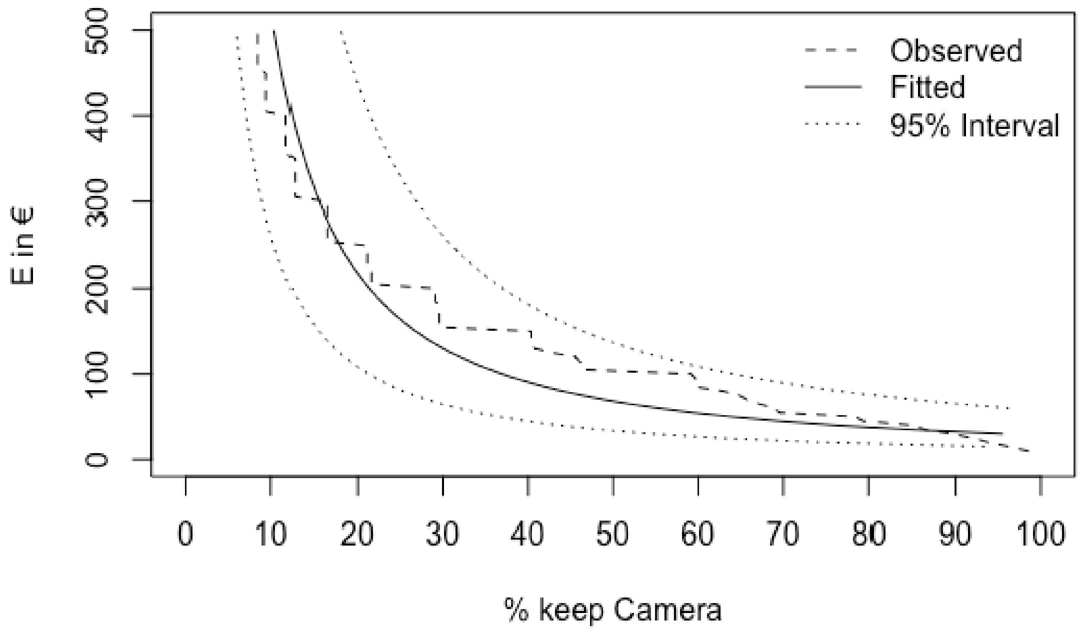
Figure 3: Sealed smartphone camera (intact and broken)

The sample was relatively balanced in terms of gender (54.5% were female) and represented the student population in terms of age (87.8% were between 18 and 24 years old). Participants reported that they use their smartphone cameras frequently and take, on average, 21.7 pictures (median = 10) and 2.3 videos (median = 1) per week. For 59% of the participants the smartphone camera is the only camera they possess. Only 16.4% own a separate point-and-shoot camera, and 18.8% a DSLR camera.

Directly eliciting monetary values in a survey leads to the observation of price thresholds, i.e., certain values that are stated more frequently. In our results, we observe that the bids 40, 50, 100, 150, 200 were each entered by more than 5% of the sample. The median bid that was given for the smartphone camera was €100. However, this median bid does not account for the price thresholds in the demand function. For example, the bids imply that 41% of the students would not give up their smartphone camera for €100, but 54% would at €100.01. To smooth the demand function, we therefore fitted a (multiplicative) function to the observed shares of students willing to accept the offer. This function explains 87.7% of the variation in demand and is depicted in Figure 4.

According to the fitted values, the median WTA for giving up the smartphone camera for one month is €68.13, albeit having a wide confidence interval (95%-CI = [€33.53; €136.78]). This implies a median annual WTA of over €800 for smartphone cameras, at least for the students in our sample, a value that is not captured in conventional GDP statistics.

Figure 4: Demand function for the smartphone camera



Analysts have estimated that it costs $20-$35 to manufacture the smartphone cameras present in current flagship models.[64] Similarly, a modular smartphone sold in the Netherlands can add front and back cameras for an additional charge to consumers of €70.[65] This study provides strong evidence that consumers obtain a significant amount of surplus from using their smartphone cameras and this surplus is an order of magnitude larger than what they actually pay.[66] Hence, there has been a large implicit price decline arising from quality change; the services received from the smartphone have increased due to quality change but

[64] E.g. http://www.techinsights.com/about-techinsights/overview/blog/cost-comparison-huawei-mate-10-iphone-8-samsung-galaxy-s8/, https://technology.ihs.com/595738/ihs-markit-teardown-reveals-what-higher-apple-iphone-8-plus-cost-actually-buys

[65] https://shop.fairphone.com/en/spare-parts (accessed January 2018)

[66] Of course, in a competitive market, most of the benefits from innovation go to consumers, not producers (Nordhaus, 2004)

this is not reflected in the measured price. Therefore, even for paid goods such as smartphones, it is crucial to adjust for quality improvements before estimating GDP statistics. This might not be an issue if consumers derived an equally large surplus from what they actually paid for while using digital or analog cameras previously. However it is hard to reconcile this hypothesis with advancements in smartphone cameras and the reduction in costs of taking photos.

We can use our total income approach for GDP-B in equation (33), which does not require calculation of a reservation price for the good in the period before it appears, to calculate an estimate of the contribution of accounting for value of the smartphone camera to consumers; we estimate an average contribution of 0.62 percentage points per year to GDP-B.[67]

7. Conclusion

This paper has developed a framework for measuring welfare change when there are new and free goods. This leads to a new measure, GDP-B, as well as its nested components, GDP-N and GDP-F, and corresponding productivity metrics.

[67] This is the arithmetic percentage point difference between the growth in GDP-B and official real GDP growth. It is calculated by assuming the following: (i) Smartphones with cameras appeared from July 2008, the date of the launch of the first iPhone in the Netherlands. Consistent with Table 3, period 0 is then taken to be Q4 of 2008. (ii) Based on EuroStat survey information on individuals who used a mobile or smartphone to access the internet (https://www.cbs.nl/en-gb/news/2018/05/the-netherlands-leads-europe-in-internet-access), the number of users of smartphones in 2017 was estimated to be 84% of the population of the Netherlands of age 15 and above (constituting 83.6% of the population). With a total population of 17 million this translates to approximately 12 million users in 2017. (iii) The annual median WTA is €817.56, and this is taken as the appropriate price for valuing the smartphone cameras; the purchase price of the camera component of the phone is assumed to be very small, so is treated as approximating zero for simplicity. With these assumptions, total income can be calculated for 2017 as nominal GDP plus the value of the smartphone cameras. The total income quantity index between the end of 2008 and 2017 can then be calculated by deflating by the official GDP deflator, and the difference with official real GDP calculated: 1.152-1.095=0.0563. That is, the difference with official real GDP is 5.63 percentage points over the nine years, or an arithmetic average of 0.63 percentage points per year.

These measures provide a means by which to understand the potential mismeasurement that arises from not fully accounting for goods which are new, free or both new and free. This is of increasing relevance in the modern digital economy given the frequent introduction of new goods and growing presence of free goods.

Appropriately, we drew on both old and new literatures to define a framework for measuring welfare change. We were able to use this framework to derive an explicit term that is the marginal value of a new good on welfare change. That is, we get a measure of the contribution to welfare of a new good, and hence the extent of welfare change mismeasurement if it is omitted from statistical agency collections that rely on conventional measures of GDP and productivity.

We also showed how to use GDP-B to derive an estimate of the addition to real GDP growth that would be required to account for the welfare gains from the introduction of a new good, without having to recalculate GDP numbers published by national statistical offices.

We then introduced free goods into a standard microeconomic model of household cost minimization and re-worked our welfare change and real GDP growth adjustments terms to allow for there to be "free" goods (with an implicit or imputable price). Accounting for new and free goods in GDP gives us a new metric, GDP-B, which is a contribution to the literature on expanding GDP beyond the traditional definitions. Two empirical implementations of GDP-B are proposed. One requires (the estimation of) reservation prices, while the other, based on the concept of "total income" avoids this necessity. Hence, we have derived explicit adjustments for both welfare change and equivalent real GDP growth that account for new and free goods, both of which are new to the literature.

Following Brynjolfsson et al. (2019), we proposed a way of implementing these adjustments using incentive compatible discrete choice experiments. We quantify this adjustment for the case of an important example of a new and free good, Facebook, in the US using a representative sample of the US internet population. Under different assumptions, we provide two estimates for the impact of incorporating Facebook into GDP-B, ranging from 0.05 to 0.11 percentage points per year on average from 2004. What's more, since GDP is the numerator used to calculate both labor productivity and total factor productivity, both of these numbers would change by the same amount per year when accounting for new and free goods using GDP-B. These are significant changes, especially considering that Facebook is just one product, and a more comprehensive application of our approach would undoubtedly add to these estimates. Indeed, using laboratory experiments in the Netherlands, we find that the additions to GDP-B generated by many other digital goods is also quite large.

Finally, using another laboratory experiment for computing the welfare created by smartphone cameras, we also show how these methods can account for new features in smartphones and other products, thereby better capturing the value of rapid quality change and new features. To elicit the consumer valuations of quality attributes, the experimental approach proposed here is to block certain features of the goods (e.g. cameras in smartphones), or even take away the entire good, in exchange for monetary compensation. This is a practical alternative way to estimate the valuations of product characteristics for adjusting price indexes, as opposed to hedonic techniques, especially when the set of characteristics of goods changes rapidly.

The high valuations for WhatsApp and Facebook raise a host of interesting questions that can be explored in further. In future work, it would be insightful to delve deeper into these individual apps and study the sources of these valuations. In addition to product quality, network effects and focal point effects are also contributing factors towards these valuations. Furthermore, many of these digital goods are also associated with externalities and a parallel stream of research is needed to explore these issues in greater detail; for example, Allcott et al. (2019) explore the impact of Facebook on subjective well-being, news consumption and political polarization.

GDP-B and the related metrics proposed in this paper enable a more thorough exploration of the impacts of new and free goods on welfare, with significant potential policy implications. Not only can these metrics help us understand the much-documented and debated productivity growth slowdown experienced by industrialized countries since 2004, but they can also clarify which goods are contributing the most to economic growth and well-being as the economy evolves.

References

Ahmad, N. and P. Schreyer (2016), "Measuring GDP in a Digitalised Economy," OECD Statistics Working Papers, 2016/07, OECD Publishing, Paris.

Allcott, H., L. Braghieri, S. Eichmeyer and M. Gentzkow (2019). The Welfare Effects of Social Media (No. w25514). National Bureau of Economic Research.

Allen, R.G.D. (1949), "The Economic Theory of Index Numbers", *Economica* 16, 197–203.

BEA (2014), *National Income and Product Accounts*, Bureau of Economic Analysis, Washington, DC. https://www.bea.gov/national/pdf/allchapters.pdf

Becker, G. M., M. H. DeGroot and J. Marschak (1964), "Measuring utility by a single-response sequential method", *Systems Research and Behavioral Science*, *9*(3), 226-232.

Bennet, T.L. (1920), "The Theory of Measurement of Changes in Cost of Living", *Journal of the Royal Statistics Society* 83, 455-462.

Bishop, R. C., Boyle, K. J., Carson, R. T., Chapman, D., Hanemann, W. M., Kanninen, B., ... and Paterson, R. (2017), "Putting a value on injuries to natural assets: The BP oil spill", *Science*, *356*(6335), 253-254.

Brynjolfsson, E., A. Collis and F. Eggers (2019), "Using Massive Online Choice Experiments to Measure Changes in Well-being", *Proceedings of the National Academy of Sciences* 116(15), 7250-7255.

Brynjolfsson, E., Eggers, F., & Gannamaneni, A. (2018), "Measuring Welfare with Massive Online Choice Experiments: A Brief Introduction", *AEA Papers and Proceedings* 108, 473-76.

Brynjolfsson, E. and A. McAfee (2011), *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, Lexington, MA: Digital Frontier Press.

Brynjolfsson, E., and McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Brynjolfsson, E., and Oh, J.H. (2012), "The Attention Economy: Measuring the Value of Free Digital Services on the Internet," Thirty Third International Conference on Information Systems, Orlando.

Brynjolfsson, E., and Saunders, A. (2009), *Wired for innovation: how information technology is reshaping the economy*, MIT Press.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the Modern Productivity Paradox. *The Economics of Artificial Intelligence: An Agenda*, 23.

Byrne, D., J. Fernald and M. Reinsdorf (2016), "Does the United States Have a Productivity Slowdown or a Measurement Problem?" in J. Eberly and J. Stock (eds.), *Brookings Papers on Economic Activity: Spring 2016*, Washington, D.C.: Brookings Institute.

Carson, R. T., T. Groves and J.A. List (2014), "Consequentiality: A theoretical and experimental exploration of a single binary choice", *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 171-207.

Corrado, C., K.J. Fox, P. Goodridge, J. Haskel, C. Jona-Lasinio, D. Sichel and S. Westlake (2017), "Improving GDP: Demolishing, Repointing or Extending?", joint winning entry, Indigo Prize 2017, http://global-perspectives.org.uk/indigo-prize/indigo-prize-winners-2017/

Cowen, T. (2011), *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*, New York: Dutton.

Coyle, D. and B. Mitra-Kahn (2017), "Making the Future Count", joint winning entry, Indigo Prize 2017, http://global-perspectives.org.uk/indigo-prize/indigo-prize-winners-2017/

Diewert, W.E. (1974), "Applications of Duality Theory," in M.D. Intriligator and D.A. Kendrick (eds.), *Frontiers of Quantitative Economics*, Vol. II, 106–171. Amsterdam: North-Holland.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, Dan Usher (ed.), Chicago: University of Chicago Press.

Diewert, W.E. (1987), "Index Numbers", pp. 767-780 in J. Eatwell, M. Milgate and P. Newman, (eds.), The New Pa1grave: A Dictionary of Economics, London: The Macmillan Press.

Diewert, W.E. (1992), “Exact and Superlative Welfare Change Indicators”, *Economic Inquiry* 30(4), 565-582.

Diewert, W.E. (1998), “Index Number Issues in the Consumer Price Index”, *Journal of Economic Perspectives* 12:1, 47-58.

Diewert, W.E. (1992), “Harmonized Indexes of Consumer Prices: Their Conceptual Foundations”, *Swiss Journal of Economics and Statistics*138, 547-637.

Diewert, W.E. (2005), “Index Number Theory Using Differences Rather Than Ratios”, *American Journal of Economics and Sociology* 64:1, 311-360.

Diewert, W.E. (2009), “Cost of Living Indexes and Exact Index Numbers”, pp. 207-246 in *Quantifying Consumer Preferences*, edited by Daniel Slottje in the Contributions to Economic Analysis Series, United Kingdom: Emerald Group Publishing.

Diewert, W.E. and R. Feenstra (2017), “Estimating the Benefits and Costs of New and Disappearing Products”, Vancouver School of Economics Discussion Paper 17-10, University of British Columbia.

Diewert, W.E., K.J. Fox and P. Schreyer (2018), “The Digital Economy, New Products and Consumer Welfare”, ESCoE Discussion Paper 2018-16, Economic Statistics Center of Excellence (ESCoE), London, UK.

Diewert, W.E., K.J. Fox and P. Schreyer (2019), “Experimental Economics and the New Goods Problem”, forthcoming discussion paper, Vancouver School of Economics, University of British Columbia .

Diewert, W.E. and H. Mizobuchi (2009), “Exact and Superlative Price and Quantity Indicators”, *Macroeconomic Dynamics* 13: Supplement 2, 335-380.

Ding, M. (2007), “An incentive-aligned mechanism for conjoint analysis”, *Journal of Marketing Research*, *44*(2), 214-223.

Ding, M., Grewal, R., and Liechty, J. (2005), “Incentive-aligned conjoint analysis”, *Journal of marketing research*, *42*(1), 67-82.

Feldstein, M. (2017), “Understanding the Real Growth of GDP, Personal Income, and Productivity”, *Journal of Economic Perspectives* 31, 145-164.

Feenstra, R.C. (1994), “New Product Varieties and the Measurement of International Prices”, *American Economic Review* 84:1, 157-177.

File, T. and Ryan, C. (2014), “Computer and Internet Use in the United States: 2013”, U.S. Census Bureau. (Accessed at: http://www.census.gov/history/pdf/2013computeruse.pdf)

Goolsbee, A.D. and P.J. Klenow (2018), “Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce”, *American Economic Association Papers and Proceedings.*

Gordon, R. (2016), *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*, New Jersey: Princeton University Press.

Greenstein, S. and R.C. McDevitt (2011), “The broadband bonus: Estimating broadband Internet's economic value”, *Telecommunications Policy* 35(7), 617-632.

Groshen, E.L., B.C. Moyer, A.M. Aizcorbe, R. Bradley and D.M. Friedman (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.

Harberger, A.C. (1971), “Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay”, *The Journal of Economic Literature* 9, 785-797.

Hausman, J. (1981), “Exact Consumer Surplus and Deadweight Loss”, *American Economic Review* 71, 662-676.

Hausman, J.A. (1996), “Valuation of New Goods Under Perfect and Imperfect Competition” pp. 209-237 in T.F Bresnahan and R.J. Gordon (eds.), *The Economics of New Goods*, Chicago: University of Chicago Press.

Heys, R. (2018), “Challenges in Measuring the Modern Economy”, presentation at the ESCoE Conference, 16-17 May, Bank of England, London.

Hicks, J.R. (1940), “The Valuation of the Social Income”, *Economica* 7, 105–124.

Hicks, J.R. (1942), "Consumers' Surplus and Index Numbers", *Review of Economic Studies* 9, 126–137.

Hicks, J.R. (1946), *Value and Capital*, 2nd Ed., Oxford: Clarendon Press.

Hulten, C. and L. Nakamura (2017), "We See the Digital Revolution Everywhere But in GDP," presentation to the NBER/CRIW conference on "Measuring and Accounting for Innovation in the 21st Century," Washington D.C., March 10, 2017. http://conference.nber.org/confer/2017/CRIWs17/program.html (accessed March 10, 2017).

Jones, C.I, and P.J. Klenow (2016), "Beyond GDP? Welfare across Countries and Time", *American Economic Review* 106, 2426–2457.

Jorgenson, D. (2018), "Production and Welfare: Progress in Economic Measurement", *Journal of Economic Literature* 56, 867–919.

Kohli, U. (1982), "Relative Price Effects and the Demand for Imports", Canadian Journal of Economics 50, 137-150.

Kuznets, S. (1934). National income 1929–1932. A report to the US Senate. In 73rd Congress, 2nd Session. Washington, DC: US Government Printing Office

Mokyr, J., C. Vickers and N.L. Ziebarth (2015), "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" *Journal of Economic Perspectives* 29(3), 31–50.

Nakamura, L., J. Samuels and R. Soloveichik (2016), "Valuing 'Free' Media in GDP: An experimental approach," paper presented at the Society for Economic Measurement Conference, Thessaloniki, Greece, July 6-8.

Nordhaus, W. D. (2004). Schumpeterian profits in the American economy: Theory and measurement (No. w10433). National Bureau of Economic Research.

Reinsdorf, M. and P. Schreyer (2017), "Measuring Consumer Inflation in a Digital Economy", paper presented at the Economic Measurement Workshop 2017, 1 December, UNSW Sydney, Australia.

Samuelson, P.A. (1974), "Complementarity—An Essay on the 40th Anniversary of the Hicks–Allen Revolution in Demand Theory", *Journal of Economic Literature* 12, 1255–1289.

Sichel, D. (2016), "Two Books for the Price of One: Review Article of *The Rise and Fall of American Growth* by Robert J. Gordon", *International Productivity Monitor* 31, Fall, 57-62.

Syverson, C. (2017), "Challenges to Mismeasurement Explanations for the U.S. Productivity Slowdown", *Journal of Economic Perspectives* 31, 165-186.

Varian, H. (2016), "A microeconomist looks at productivity: A view from the valley", Presentation, Brookings. (Accessed at: https://www.brookings.edu/wpcontent/uploads/2016/08/varian.pdf)

Wells, J. and A. Restieaux (2014), "Review of Hedonic Quality Adjustment in UK Consumer Price Statistics and Internationally", UK Office for National Statistics.

Chapter 3 - Multi-Sided Platform Strategy, Taxation, and Regulation: A Quantitative Model and Application to Facebook

Abstract

Digital platforms, such as Facebook, Uber, and AirBnB, create value by connecting users, creators, and contractors of different types. Their supply and demand economies of scale make them natural monopolies, and have led to increasing calls for special regulations and taxes. We construct and illustrate an approach for modeling digital platforms. The model allows for heterogeneity in demand elasticity, disutility from advertising, and network effects across users of different types. We paramaterize our model using a survey of over 40,000 [Avi, new number here] US internet users on their demand for Facebook. We find Facebook has too low a level of advertising relative to their revenue maximizing strategy, confirming that they also value maintaining a large user base. We simulate [Seth confirm this] six proposed policies for government management of digital platforms, taking Facebook's optimal response into account. Taxes are mostly incident on Facebook profits, only slightly changing consumer surplus. More radical proposals, including 'data as labor' and nationalization, have the potential to raise consumer surplus by up to [Seth fix this] 42%. But a botched regulation that left the US with two smaller, non-competitive social media monopolies would decrease consumer surplus by [Seth fix this] 44%. [Something about how first best depends on shadow value and is odd because of ad disutility]

1 Introduction

Much of the value of many digital platform businesses comes from "network effects". A network effect is an externality that one participant in a market, digital platform, or similar system provides to others. But how exactly can one measure and exploit the value of network effects for any particular business or industry? In this paper we propose and implement a flexible strategy for the measurement and optimal harnessing of network effects. We then use the model to simulate the effect of several proposed or recently implemented digital platform regulations and taxes.

We make three main contributions. First, we provide a tractable framework for third-degree price discrimination on multi-sided platforms. This approach builds on traditional price discrimination models by taking into account network externalities. Second, we implement this model using data we collected on Facebook, introducing a novel methodology for the estimation of network effects. Using the calibrated model, we provide the first simulations of Facebook revenue and participation under counterfactual pricing policies. Finally, we use the model and data to estimate the social gains from proposals by politicians and academics to tax and regulate "Big Tech."

Our paper begins by introducing a model of platform participation that allows for several dimensions of heterogeneity. Users vary in their opportunity cost for using the platform, the value they get from other types of users using the platform, and the disutility they receive from advertising. It is a model of an n-sided platform in the sense that each individual or market segment can be thought of as a node of the network.[1] [Say something about how policies cascade and reach equilibria]

We show that the optimal pricing strategy for an n-sided platform entails decreasing fees or advertising for users who elastically demand the platform (the direct effect) and who create high amounts of network value for other profitable users who themselves demand the platform elastically (the network effect).

[Theoretical Results Here] ¿Perfect competition vs. perfect monopoly + the role of splitting ¿the role of price discrimination in the desirability of market power ¿potential for platform to NOT HAVE ENOUGH market power (will show this to be important) ¿person tax vs. income tax

After introducing and analyzing our model, we proceed to an empirical illustration. We collected information on US internet users' demand for Facebook across over 40,000 [Avi Fix] surveys conducted through Google Surveys. We categorize the surveyed into

[1] When conceived of this way, any platform, including a one-sided platform, can be thought of as an n-sided platform once we account for the heterogeneity in users within a side. For example, a telephone network, which is the classic example of a one-sided network, can be thought of as consisting of multiple sides that can be distinguished based on various characteristics including business vs. personal use, demographics, regional location, heterogeneity in activity (frequent users or not) and type of activity (always callers, callers and receivers, always receivers).

twelve demographic groups by their age and gender.

To collect information on demand for and network effects on the platform, we use an experimental choice approach in the spirit of (Brynjolfsson et al., 2019) and (Allcott et al., 2019). These papers measure the consumer surplus generated by digital goods by conducting discrete choice experiments where they offer consumers the choice to give up access to the good in exchange for monetary compensation. We build on these papers by asking about a new type of free good (the value of social connections) as well as by using information from the full distribution of responses to fit a demand curve (in our case logistic) rather than focusing on median and average responses. We adapt this approach to our case by giving consumers the choice to give up access to a subset of their network in exchange for monetary compensation. [Seth revise this paragraph]

Using this information about demand for Facebook, we estimate the parameters of a logistic demand curve for each demographic group, as well as the ten by ten matrix of their network externalities. We complement this with additional survey questions about friend frequency, the disutility of advertising, and publicly available data on Facebook's current advertising revenues by demographic group.

With this model of individual participation, we can calculate the effects of counterfactual pricing policies, government policies and demand shocks. We begin by simulating Facebook's revenue maximizing strategy. We find that Facebook could raise revenues by 2.65 [Seth Update] billion dollars a month (from a baseline of 1.57 billion) if it optimally price discriminated. It could raise revenues by only 2.08 [Seth Update] billion dollars a month if it increased monetization equally across users. This begs the question of why Facebook is 'leaving so much money on the table'. Implementing their revenue maximizing strategy entails squeezing value from their most inelastic users, reducing Facebook usage by 55.6% and lowering total consumer surplus by 82.8%. We infer that in addition to maximizing current revenues, Facebook values maintaining a large and happy user base. We impute the value Facebook places on maintaining a large user base as the one that justifies their current level of advertising as optimal. In subsequent simulations we take into account this non-monetary value when simulating Facebook's response to policy changes.

We then proceed to calculating the impact of changes in government policy on Facebook revenues, participation, and consumer welfare. We consider three taxation and redistributive policies. We show theoretically that a flatly applied tax on ad revenues would not change Facebook's optimal advertising level, so long as Facebook has no other considerations. However, if Facebook values a large user base, then a tax on advertising redirects it from raising high levels of advertising revenue to cultivating a large user base. A tax on the number of users has the opposite effect, leading Facebook to squeeze a smaller group of users with a higher level of fees. Quantitatively, we find that a 3% tax on advertising revenues would raise consumer surplus by 2% and that a ten dollar per month tax on users would lower consumer surplus by 3%. Another proposed policy for redistributing the wealth from Facebook is Weyl's "Data as Labor" framework, where internet users would be compensated for their 'labor' in viewing targeted advertisements (Posner and Weyl, 2018). We conceive of this policy as a rebate of Facebook's current advertising revenues to users. We find this policy would boost consumer welfare by 24%, about 58% of which is due to the direct transfer to current users with the remainder due to new users who join the platform, consuming more ads and providing more value to other users.

We also simulate three proposed regulatory interventions. The first is taking steps to enhance the competitiveness of the social media industry, by lowering barriers to entry and enforcing 'interoperability' (i.e. allowing users on a Facebook competitor to view posts by and communicate with users of Facebook and other Facebook competitors). We model this policy as creating perfect competition, and lowering the price of the platform to its marginal cost – i.e. forcing the elimination of advertising and other fees. A second policy we evaluate is the nationalization of Facebook for the purpose of maximizing its social welfare. If transfers can be frictionlessly distributed to Facebook users, the optimal policy is to implement an infinite subsidy for Facebook use – this is because transfers don't change social surplus, and every user of Facebook creates positive externalities for other users. In our simulation, we assume that a nationalized Facebook can create value for users as the inverse of how it creates disutility through monetization.[2] Finally we simulate the results of a 'botched' Facebook breakup which leaves America with two monopolies over half of the population each. We predict that perfect competition would raise consumer surplus by 9%, at the cost of eliminating all monetary profits, and that a social welfare maximizing Facebook would raise consumer surplus by 42%, at the cost of Facebook needing to go -255% into the red. Breaking Facebook into two non-competitive 'baby Facebooks' would be disastrous, lowering consumer surplus by 44%. It would also lower combined ad revenues by 93% as the baby Facebooks lowered advertising rates to retain even 82% of their original combined user base.

The paper concludes with a discussion of the strengths and weaknesses of this approach to modelling platform businesses and contemplates future work.

[2] In other words, if we estimate that a group experiences 20 cents in disutility from a dollar's worth of revenue in advertising, we assume that a nationalized Facebook can only increase the desirability of Facebook to a demographic by one dollar by spending five dollars.

2 Related Literature

A rich stream of theoretical literature studying network effects in the context of platform businesses has evolved over the past decade and a half. Following the seminal work of Parker and Van Alstyne (2005) and Rochet and Tirole (2003), platform researchers have extensively studied the impact of direct and indirect network effects on various strategic issues including pricing (Hagiu (2009)), launch (Evans and Schmalensee (2010)) and openness (Boudreau (2010)). The core insight of this research is that it can be optimal for a two-sided platform to subsidize one side and increase fees for the other side (Eisenmann et al. (2006)).

The above papers all focus on what are known as one or two-sided platforms. Examples of two-sided platforms are Uber (riders and drivers) and Ebay (sellers and buyers). In a two-sided platform, it can make sense to price discriminate based on side, because different types of users may provide different network externalities. For example, an additional Uber driver in a region provides a positive externality to riders (they will get a ride faster) but a negative externality to other drivers (they will have to wait longer in-between fares). However, a large literature suggests that even within a 'side' of a one or two-sided platform, users are heterogenous in the effect their actions have on the network. The empirical literature on network effects uses several techniques for their estimation, including studying exogenous shocks to the network (e.g. Tucker (2008)), using an instrumental variable approach (e.g. Aral and Nicolaides (2017)) and conducting field experiments (e.g. Aral and Walker (2012)).

There are several recent papers which model pricing in the presence of multi-dimensional network effects. For example, Bernstein and Winter (2012) determines a mechanism for optimally renting storefronts in a shopping mall where stores have heterogeneous externalities on other stores. Candogan et al. (2012) and Fainmesser and Galeotti (2015) consider monopolistic pricing of a divisible network good, where utility from the good is quadratic in the amount consumed and linear in the impact of neighbors' consumption. In (Candogan et al., 2012), the platform firm has perfect knowledge about all individuals' utility functions, but allows for individuals to vary in their utility from the platform good (although this utility must be quadratic). They show that the problem of determining profit maximizing prices is NP hard, but derive an algorithm guaranteeing 88% of the maximum. Fainmesser and Galeotti (2015) considers a similar model but assumes that all individuals have the same demand for the network good, while allowing for a random distribution of network connections. They find that allowing for the network to lower prices on 'influencers' must increase social welfare, but allowing firms to fully price discriminate might be harmful. The paper in this literature with a model most similar to ours is Weyl (2010). That paper, like ours, considers an indivisible platform good with network effects. It also, like this paper, allows for groups to vary in both their network effect on other groups and in their opportunity cost for using the platform. It finds that a wedge exists between the profit maximizing and social welfare maximizing pricing strategy.[3]

Our paper builds on these prior papers along several dimensions. First, our model features more realistic monetization, allowing for different types of users to face different levels of disutility from the firm increasing their level of advertising. This is in contrast to (Candogan et al., 2012) and (Fainmesser and Galeotti, 2015) which do not allow for such variation, and Weyl (2010) which features an unrealistic pricing scheme, where users are charged based on the level of participation of other users (i.e. an 'insulating tariff'). Weyl (2010) use of insulating tariffs in pricing forces users to immediately jump to a desired equilibrium in response to a price change, which prevents a dynamic analysis of a pricing change. Second, unlike (Candogan et al., 2012) and (Fainmesser and Galeotti, 2015) our model has a realistic amount of uncertainty within a side of a model, meaning that first degree price discrimination that drives consumer surplus to zero is impossible. [4] The most important contribution of our model is that it is the first one to allow for straightforward calibration. To the best of our knowledge, no previous paper has made quantitative model-based recommendations about multi-sided platform pricing, or quantitatively evaluated the welfare consequences of a platform regulation market structure change.

The illustration in our paper is of Facebook, a platform primarily monetized through advertising. Most platforms keep the quantity of ads ("ad load" to those in the industry) shown per user fixed while showing different ads to different users based on their characteristics and bid outcomes of ad auctions (e.g. Google (Hohnhold et al., 2015), Pandora (Huang et al., 2018a)). Platforms with a newsfeed, such as Facebook, WeChat and Linkedin, understand the trade-off between ad load and user engagement. Some of them show the same number of ads per person (see Huang et al. (2018b) for advertising on WeChat), while others fix the number of ads a user sees based on the expected revenue generated by the user in the long term (Yan et al. (2019) describe Linkedin's ad load strategy). While this optimization takes user engagement into account, network externalities generated by a user are not explicitly modeled and users generating different amounts of network externalities end up seeing the same number of ads.[5] In estimating structurally the impact of market structure on social welfare in

[3]The exact nature of this wedge – as a marginal, not an average distortion – was clarified in a published comment (Tan and Wright, 2018).

[4]The fact that platforms cannot fully first-degree price discriminate is testified to by papers which show that users benefit considerably on average from joining a platform. For example, Ceccagnoli et al. (2011) find that independent software publishers experience an increase in sales and a greater likelihood of issuing an IPO after joining a major platform ecosystem, and Brynjolfsson et al. (2019) find large consumer surplus from the use of digital platforms.

[5]Based on informal conversation with researchers who have worked with Facebook, our understanding

the presence of network effects, our paper is in the tradition of Rysman (2004). That paper has a model of an analog two-sided platform: the yellow pages. It uses instruments to find the spillover effects of additional advertisements on phone-book quality. Rysman that small *decreases* in competition might increase welfare, as there would be fewer better phone-books with more utilitous advertisements.

3 Analytic Model

The foundational element of a model of network effects is a stance on how agents connect to and gain welfare from the network. In our model, individuals with heterogeneous characteristics decide whether or not to participate in a network. Their desire to participate in the network is a function of their expectation of which other individuals will participate. For example, Jane Doe's desire to use Instagram is a function of which of her friends are also using Instagram. The key term in the model is the externality that users gain from others. Unlike other models of platforms, we allow for individuals of different characteristics to gain different amounts of value from the participation of others on the network. These market segments are the different sides of the platform.

We use the example of a social network, because our implementation section takes place in that setting. Therefore, in our baseline model, other incidental network characteristics mimic that of an internet social network. Once two users are using the network, there is no additional cost for them to form a connection. All connections where both users gain weakly positive value are immediately formed. We assume that the fee or subsidy faced by each participating network user is a binary function of their decision to participate on the network. This assumption is easy to modify for other contexts where fees are a function of the number or type of connections or interactions.[6]

The platform's monetization is also modeled. Users face disutility depending on how intensely they are monetized by the platform. This may correspond to the unpleasantness of advertisements or the disutility of knowing one's data will be harvested and resold. Alternatively, it may correspond to an explicit participation charge, such as WhatsApp's original $1 subscription cost.

This model is implementable in the sense that there is a clear strategy for measuring all the terms that appear in the model. It is scalable in the sense that these terms can be measured with as much precision and for as small a market segment as

is that in constructing its newsfeed, Facebook gives every potential entry a score, based on the amount of engagement the entry is expected to create in the user who sees the ad, the amount of revenue that might be generated (if it is an advertisement) and a penalty for being similar to a recently displayed entry.

[6] An example of an online platform with network effects and a non-binary fee structure would be an online auction house like eBay. eBay's main source of revenue is a progressive fee on the value of every transaction.

desired. As a first pass, a platform might distinguish between the network externalities and demand characteristics of broad user groups such as women and men. A more sophisticated platform with a larger research budget might estimate and incorporate into their optimization network externalities at the individual level. When calibrating the model, we make additional assumptions about the functional form of user demand for the platform.

3.1 Consumers

A consumer i chooses whether to participate in the platform ($\boldsymbol{P_i} = 1$) or not ($\boldsymbol{P_i} = 0$).[7] If the consumer i uses the platform ($\boldsymbol{P_i} = 1$)), they expect to receive

$$E[U_i(\boldsymbol{P_i} = 1)] = \mu_i(P_1, .., P_I, -\phi_i) \tag{1}$$

where P_j is the probability individual j participates on the platform. ϕ_i is the revenue the platform raises from individual i. A firm which monetizes using advertising might raise \$1 in revenue by displaying additional ads which create \$.20 in additional disutility (i.e. $\frac{\partial \mu_i}{\partial \phi_i} = .2$). Local telephone calls and pre-2016 WhatsApp monetized by charging a flat fee for participation (i.e. $\frac{\partial \mu_i}{\partial \phi_i}$ =\$1). [8] Note that users do not directly care about what other users are charged, but it is indirectly important to them insofar as it causes other users to participate on the network.

$\frac{\partial \mu_i}{\partial P_j}$ is the marginal utility of j being on the network to i (if i participates). For convenience, we will sometimes write the marginal value of a user j to a user i as

$$U_i(j) = \frac{\partial \mu_i}{\partial P_j} \tag{2}$$

and the marginal disutility of advertising as

$$a_i = \frac{\partial \mu_i}{\partial \phi_i} \tag{3}$$

In our theoretical analysis, our only assumption is that μ_i be continuously differentiable. In our calibration, we further assume that utility from the platform is linearly additive in the network effect from friends and disutility from ϕ. In other words, the

[7] Note that while demand functions are here defined at the individual level, as a practical matter firms may estimate them at the level of a demographic or social group. We consider an example with ten market segments in our calibration.

[8] In general, platforms monetize in many different ways. Some monetize by charging fees for transactions (Ebay, AirBnB, etc), some subsidize one side while charging others (Credit Cards), some by charging a flat fee for participation (Local telephone calls, pre-2106 WhatsApp), and some monetize by charging advertisers or selling advertisements (social networks). Our baseline model is best suited for evaluating the latter two approaches, but can be straightforwardly modified to handle other monetization methods.

parametric analysis assumes that $U_i(j)$ and a_i are constant.[9]

The value to a consumer of not using the platform, their 'opportunity cost', is an ex-ante unknown random variable.

$$U_i(\boldsymbol{P_i} = 0) = \epsilon_i \tag{4}$$

where ϵ_i are independent random variables (not necessarily symmetrical or mean 0). ϵ's distribution may vary by individual. This means that the probability of participating on a network, P, conditional on a given level of utility from the network good $U(\boldsymbol{P} = 1)$ is consumer specific.[10]

The distribution of ϵ_i determines how elastic i will be to changes in the platforms' attractiveness. Consider the case where ϵ_i is expected to be approximately equal to the utility of participation $U_i(\boldsymbol{P_i} = 1)$ – in other words, that it is likely that the user is 'on the fence' about using the platform. In this case, changes in ϕ_i or other consumers' participation will be highly likely to change i's participation.

Each consumer gets to see the resolution of their private outside option ϵ_i before participating, but not the resolution of anyone else's. Therefore, they base their decision to participate on the platform based on their beliefs in the likelihood of others participating. The *ex-post* consumer demand function is

$$\begin{cases} \boldsymbol{P_i} = 1 & \text{if } E[U_i(\boldsymbol{P_i} = 1)] > \epsilon_i \\ \boldsymbol{P_i} = 0 & \text{otherwise} \end{cases}$$

Note that P_i's are independent because ϵ_i's are independent.

We can write the *ex-ante* demand function (i.e. expected demand before ϵ_i is known) as:

$$P_i = Prob[E[(U_i(\boldsymbol{P_i} = 1)] > \epsilon_i] = \Omega_i(\mu_i) \tag{5}$$

[9]The assumption that the value of platform connections are linearly additive is not a harmless one, despite being made in all of the most similar papers extant ((Candogan et al., 2012), (Fainmesser and Galeotti, 2015), and (Weyl, 2010) all make this assumption). It means, for example, the additional value that Jane Doe gets from James Smith joining Instagram isn't a function of whether any third person is already on Instagram. This is a useful simplification in the context of social networks, but in the case of other networks it is likely unrealistic. Taking a food delivery platform as an example, it is likely the case that the 10th pizza delivery service joining the platform provides less platform value to the typical user than the 1st. A related simplification is the assumption that the value of a connection is only a function of the characteristics of the connected individuals. In general, the value of a connection to one individual may be a function of that individuals' connections to other individuals. We abstract from these possibilities in the calibration. The measurement of non-linearly additive network effects introduces large measurement challenges beyond the scope of this paper's illustration, but is something we plan to explore in future work.

[10]By adding a negative sign, this term can also be interpreted as the value or disutility of Facebook use in the absence of any friends or advertisements.

for more useful notation, define

$$\boldsymbol{U_i} \equiv E[U_i(\boldsymbol{P_i} = 1)] = \mu_i \tag{6}$$

The network is in equilibrium when individuals' decisions to participate are optimal responses to their beliefs about every other individuals' decision to participate. In our empirical illustration, we calculate the new equilibrium as a response to a shock through evaluating a series of 'cascades'. For example, if the firm were to raise ϕ_i we would first calculate the direct impact of only this change in price on user i. This is the first cascade. We would then calculate all individuals' decision to participate taking $i's$ new participation rate as given – the second cascade. Additional cascades estimate every groups' rate of participation, taking the previous cascades' rate of participation as an input. We calculate 1000 cascades in all of our simulations, but as a practical matter, the importance of cascades beyond the third or fourth is minimal.

For the symmetric network (i.e. where all individuals have the same ϵ distribution, A, and network externality), where utility is linearly additive in the network effects and disutility from advertising, the equilibrium is stable so long as

$$1 > \frac{\partial \Omega}{\partial \boldsymbol{U}} U(i)(I-1) \tag{7}$$

where $U(i)$ is the value from any consumer participating in the network to any other consumer, and I is the number of friends each user has. Intuitively, the network is unstable when users are very elastic and care a lot about the participation of others on the network. The derivation of this equation is in appendix B.

4 Optimal Platform Strategy

There are many questions you can ask about optimal platform strategy in this setting. Here we focus on the managerial implications for a revenue maximizing monopoly social network.

4.1 Monopoly Firm Price Setting

Consider a social network which can price discriminate among its users taking their demand functions (as well as the actions of competitors) as given. Platforms in this setting can price discriminate either by directly charging or subsidizing some users, or by giving some subset of users more or less advertisements.

Firms maximize expected total profits. After uncertainty is resolved, the firm's revenues are

$$\Phi = \sum_{i}^{I} \phi_i \boldsymbol{P_i} - F \tag{8}$$

Where ϕ_i is the revenue collected from or distributed to consumer i if they participate in the network. It is a choice variable from the perspective of the firm. $\boldsymbol{P_i}$ is a binary indicator of whether the consumer participates. F is the fixed cost of the platform firms' operation.[11]

P_i's are independent random variables, so firms maximize

$$E[\Phi] = \sum_{i}^{I} \phi_i P_i - F \tag{9}$$

where

$$P_i = E[\boldsymbol{P_i}] = \Omega_i(\boldsymbol{U_i}) = \Omega_i(\phi_1, \phi_2, ...) \tag{10}$$

the probability of a consumer participating P_i is an individual specific function of $\boldsymbol{U_i}$. Ω_i is the effective individual specific demand function. Ultimately the equilibrium level of participation is a function of preference parameters and the vector of ϕ's, and there are no variable costs, so the monopolist social media platforms' problem is to select the level of ϕ's that maximizes revenues.[12]

The firm seeks to maximize revenues

$$\max_{\phi_i} E[\Phi] = \sum_{i}^{I} \phi_i P_i \tag{11}$$

s.t.

$$P_i = \Omega_i(\boldsymbol{U_i}) \tag{12}$$

This yields the following first order condition

$$\frac{\partial \Phi}{\partial \phi_i} = P_i + \phi_i \frac{\partial P_i}{\partial \phi_i} + \sum_{j \neq i}^{J} \phi_j \frac{\partial P_j}{\partial \phi_i} \tag{13}$$

where

$$\frac{\partial P_i}{\partial \phi_i} = \frac{\partial \Omega_i}{\partial \boldsymbol{U_i}} \left(- \frac{\partial \mu_i}{\partial \phi_i} + \sum_{j}^{J} \left(\frac{\partial \mu_i}{\partial P_j} \frac{\partial P_j}{\partial \phi_i} \right) \right) \tag{14}$$

[11]We assume the platform faces no marginal costs, but adding a marginal cost does not change the qualitative results.

[12]Although hypothetically the function $a_i(phi_i)$ which relates user disutility from monetization to platform revenue might be thought of as being net of this fixed cost.

and,

$$\frac{\partial P_j}{\partial \phi_i} = \frac{\partial \Omega_j}{\partial \boldsymbol{U_j}} \left(\frac{\partial \mu_j}{\partial P_i} \frac{\partial P_i}{\partial \phi_i} + \sum_{k \neq i}^{K} \frac{\partial \mu_j}{\partial P_k} \frac{\partial P_k}{\partial \phi_i} \right) \tag{15}$$

This recursion is natural as P_i is a function of P_j, which is a function of P_i, etc. Equation (15) will converge to a finite value so long as each recursion of the network effect dampens out. This will occur so long as the equilibrium is stable. In our calibrated example evaluating only the first two recursions, or cascades, of this function tends to yield a decent approximation of the total change in revenues from a pricing change.

4.2 Profit Maximization Problem Simplified

Equation 13 gives conditions for the optimal schedule of fees (or other revenue raising monetization strategies) and subsidies for the general case. Even if not enough is known about the entire curve of functions to find the optimum, knowing the first derivative of the goal with respect to the choice parameters is useful. *An experimenting firm can simply use these equations to inch towards a local maximum via gradient decent.*

For simplicity in interpreting the first order condition, we retain only first term in brackets in 14 and 15. In other words, the following equations only take into account one cascade of network effects.[13] For clarity and parsimony, we also make the substitutions from equations 2 and 3

$$\frac{\partial P_i}{\partial \phi_i} = \frac{\partial \Omega_i}{\partial \boldsymbol{U_i}} \left(-a_i + \cancel{\sum_{j \neq i}^{J} \left(U_i(j) \frac{\partial P_j}{\partial \phi_i} \right)} \right) \tag{16}$$

and,

$$\frac{\partial P_j}{\partial \phi_i} = \frac{\partial \Omega_j}{\partial \boldsymbol{U_j}} \left(U_j(i) \frac{\partial P_i}{\partial \phi_i} + \cancel{\sum_{k \neq i}^{K} U_j(k) \frac{\partial P_k}{\partial \phi_i}} \right) \tag{17}$$

Then, substituting into 13, yields a new simplified first order condition

$$\frac{\partial \Phi}{\partial \phi_i} = \underbrace{P_i - \phi_i a_i \frac{\partial \Omega_i}{\partial \boldsymbol{U_i}}}_{\text{Direct Effect}} \underbrace{- a_i \frac{\partial \Omega_i}{\partial \boldsymbol{U_i}} \sum_{j \neq i}^{J} \phi_j \frac{\partial \Omega_j}{\partial \boldsymbol{U_j}} U_j(i)}_{\text{Network Effect}} \tag{18}$$

The simplified first order condition consists of two sets of terms. The first two terms report the direct effect of raising the amount of advertising on individual i by one dollar. This will raise revenue, based on that individual's current likelihood of participation, and lose revenue based on how elastic that individual's participation is.

[13] In the parametric section we will show that the first cascade of network effects is quantitatively much more important than subsequent cascades for a reasonable parameterization.

The two direct effect terms are what normal firms have to consider when pricing their products (note that when $U_j(i) = 0 \; \forall \; i, j$, i.e. when no network effects are present, 18 reduces to this pair of terms).

The last term in equation 18 is the network effect of an advertising increase. The increase in advertising makes i less likely to participate (in this approximation, by an amount $a_i \frac{\partial \Omega_i}{\partial U_i}$) which leads others to stop participating (by an amount $\frac{\partial \Omega_j}{\partial U_j} U_j(i)$). When these third parties stop participating, the platform loses on the current revenues that they were paying ϕ_j.

In other words, the fee or level of disutilitous advertising should be increased on user i if the increased revenue (P_i) is greater than the decreased revenue from the person directly impacted possibly dropping out (second term) plus the decreased revenue from all the charged person's friends potentially dropping out (third term).

This equation is a powerful tool for managers to think about monetizing their platform. While a similar equation might be able to be derived from the model in (Weyl, 2010) (i.e. that this result is latent in that model), a major contribution of this paper is a recasting of the firm's maximization problem in terms of elasticities of demand and other more interpretable terms.

This simplified first order condition can be made more precise by taking into account additional cascades of the network effect. In other words, because user i's fee increasing causes j to be less likely to participate, all those connected to j should be less likely to participate as well.

Unsurprisingly, the firms' profit maximizing strategy deviates from social welfare maximizing pricing. Appendix C reports the social welfare maximization problem. Intuitively, the wedge between the revenue and social welfare maximizing strategies arises from the fact that the platform only cares about monetization disutility and network spillovers that effect marginal users of the platform, wheras a social planner takes into account welfare changes for infra-marginal individuals who will use the platform in both scenarios.

5 Empirical Illustration – Facebook

The setting for our empirical illustration is Facebook. Facebook is an ad-supported social network. It was selected because it is used by a very large percentage of the US population, and previous research has demonstrated that many value it highly.

To illustrate how our method can be used by firms to price discriminate, we collected survey data to estimate our model. We conducted approximately 40,000 surveys on a representative sample of US internet population. Google Surveys provides information on a survey participants' gender and age group, so we distinguish market segments based on those characteristics. We divided Facebook users into ten market segments. These are a pair of genders and five age brackets. The market segments we consider are

- Gender: Male or Female
- Age: 25-34; 35-44; 45-54; 55-64; and 65+

we also intended to include individuals age 18-24 in our analysis, but we found it difficult to get a sufficient number of survey responses for this group. Individuals under the age 13 are not formally allowed to have Facebook accounts.

We asked the following sets of questions about individuals' demand for Facebook, combining responses within the ten market segments described. The full list of surveys conducted is documented in figure 1.

Question Text	Possible Responses	Number of Responses
How many friends do you have on *Facebook*?	**Survey 1:** 0-100; 100-200; 200-300; 300-400; 400-500; More than 500; I do not use Facebook **Survey 2:** 0-50; 50-100; 100-500; 500-700; 700-900; More than 900; I do not use Facebook	2509
Would you give up Facebook for *1 month* in exchange for $**[X]**? Choose Yes if you do not use Facebook.	Yes, I will give up Facebook; No, I would need more money	20050
"On Facebook, would you unfriend all your friends who are ***[demo_group]*** for 1 month in exchange for $**[X]**? Choose Yes if you do not use Facebook."	Yes, I will unfriend all these friends; No, I would need more money	13272
What *percentage* of your *friends on Facebook* are ***[demo_group]***?	0-10%; 10%-20%; 20%-40%; 40%-60%; 60%-80%; 80%-100%; I do not use Facebook	13037
What is the *maximum* amount of money (in US $) you would pay to personally *not see any advertisements on Facebook* for *1 month*? Select 0 if you do not use Facebook.	0; $1-$5; $5-$10; $10-$15; $15-$20; $20+	1001

Figure 1: List of surveys conducted

Figure 2 gives examples of how the surveys appeared to respondents. Respondents answered these surveys either as part of Google Rewards or to access premium content on websites.

5.1 Calibrating the Model for Facebook Market Segments

The very general utility function analyzed in section 4 is tractable enough to lead to some analytic results, some of which we have already elucidated. However, for the

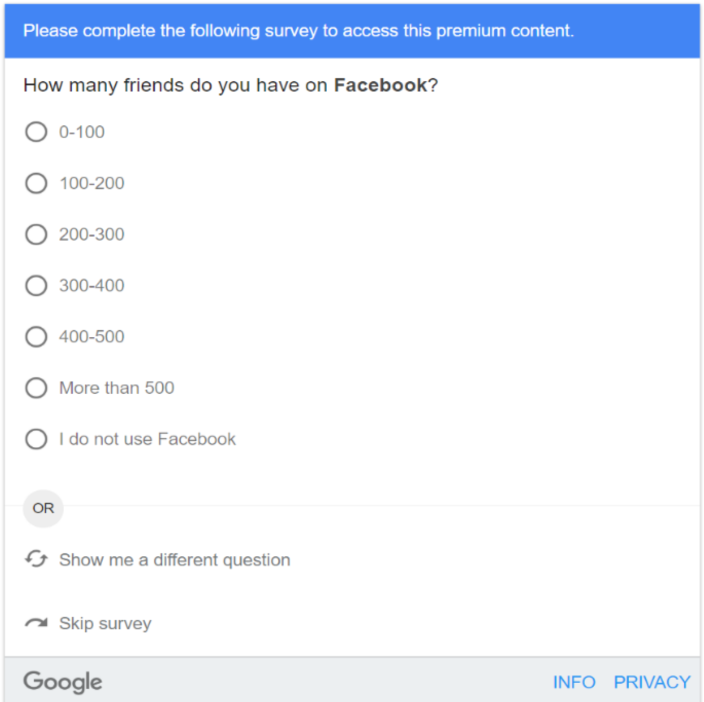


Figure 2: Google survey interface example. Note that each respondent only receives a single survey question, and that responses are limited to seven multiple choice options.

purposes of quantitative estimates, we need to select a more restrictive functional form for the utility function. We also need to make modifications to the model to account for the fact that we are estimating it over market segments, not individuals, and for the fact that not all individuals are friends.

We assume that the opportunity cost for using Facebook is distributed such that demand for Facebook Ω_i follows a logistic distribution. We estimate the parameters of Ω_i by running a logistic regression on responses to the question "Would you give up Facebook for 1 month in exchange for \$X? Choose Yes if you do not use Facebook.". Figures 9 through 18 report responses to these questions, and the logistic line of best fit. Table 1 reports the estimates underlying these curves. We convert from estimates of the CDF logistic equation to the PDF of the distribution of ϵ_i's using the equations

$$p(\epsilon_i) \sim \frac{e^{-\frac{\epsilon_i - \eta_i}{s_i}}}{s_i \left(1 + e^{-\frac{\epsilon_i - \eta_i}{s_i}}\right)^2} \tag{19}$$

where

$$s_i = (\text{Coef. on Cost}_i)^{-1} \tag{20}$$

and

$$\eta_i = (-\text{Intercept}_i)s_i \tag{21}$$

As an argument the function Ω_i takes the change in the value of Facebook due to an individual gaining or losing friends, or from experiencing a direct change in their advertising level ϕ_i. The parametric model of consumer utility we calibrate for each market segment i is linear in the number of friends of each type and in disutility from advertising, i.e.

$$\mu_i = \sum^{J} U_i(j) P_j z_i(j) D_j - a_i \phi_i \tag{22}$$

where $U_i(j)$ is the (linear) utility an individual i receives from having a friend in market segment j, P_j is the percentage of Americans in group j who use Facebook, $z_i(j)$ is the percentage of users of type j who i is friends with, D_j is the population of demographic group j, and a_i is the disutility caused by a level of advertising ϕ_i.

We estimate the parameters of 22 through a combination of survey questions, government sources and information publicly available through Facebook's ad API and quarterly reports.

D_j is taken from US Census reports for 2019. Our estimate of the current revenues that Facebook make from users by demographic begins by noting that Facebook raises \$11.62 dollars a month in revenue from US users through displaying them advertisements.[14] To calculate initial revenue per user $\overline{\phi_i}$ we take in data on the cost of advertising to users of different types from Facebook's advertisement API. After selecting which demographic group you would like to target, Facebook tells you how many impressions you are estimated to receive per dollar of spending. We take the inverse of this measure to be the relative value of a demographic to Facebook's ad revenue. By taking as given that the average value of a user per month is \$11.62, we can then calculate the revenue per user of a demographic using the following equations

$$\overline{\phi_i} = z\text{Relative Value}_i \tag{23}$$

and

$$11.62 = q\frac{\sum^{I} \text{Relative Value}_i \overline{P_i} D_i}{\sum^{I} \overline{P_i} D_i} \tag{24}$$

where q is a scaling term, $\overline{P_i}$ is the estimate of the initial participation rate on Facebook by the demographic group (taken as our estimate of $\Omega_i(\overline{\mu_i}) - \mu_i = 0$), and D_i is the total population of the group in the US.

[14] This is derived from Facebook's 2019 Q1 annual report, where they report \$ 34.86 in revenues per North American user per quarter.

To estimate the share of users by type that a user of type i is friends with, we combine the results of two sets of survey questions. First, we ask questions to solicit the total number of friends each demographic has on average. We then ask questions to solicit what percentage of their friends of of each demographic. We re-balance these responses to add to 100 percent (including a catchall category for individuals under age 25, who are not directly modeled). Figure 3 presents our estimate of the average number of friends by type for each demographic.

GENDER		Female	Female	Female	Female	Female	Male	Male	Male	Male	Male
	AGE	25-34	35-44	45-54	55-64	65+	25-34	35-44	45-54	55-64	65+
Female	25-34	71.0	47.2	43.2	24.0	21.5	64.2	34.9	18.3	19.7	15.3
Female	35-44	50.8	52.6	41.1	25.0	19.2	25.0	31.3	18.6	15.1	10.6
Female	45-54	42.3	38.7	34.0	29.1	14.6	21.7	23.6	25.4	23.7	10.7
Female	55-64	22.8	29.3	34.4	37.2	21.5	12.2	17.4	20.6	14.9	11.9
Female	65+	10.5	18.9	20.8	20.1	16.7	8.3	13.0	11.1	16.6	11.0
Male	25-34	46.1	41.0	12.6	25.9	23.8	51.6	31.0	31.0	21.4	14.7
Male	35-44	46.3	46.2	33.3	39.1	9.6	37.4	39.0	31.4	22.9	19.3
Male	45-54	20.7	32.0	30.3	20.9	16.5	22.1	27.4	22.1	20.6	13.4
Male	55-64	21.3	20.4	27.1	22.8	19.9	18.5	20.0	27.2	19.5	17.2
Male	65+	12.7	14.1	17.8	20.6	12.3	10.4	7.0	20.2	17.7	14.1

Figure 3: Average number of friends someone in Y-axis market segment has of the type in the X-axis market segment.

To estimate the value of friends by demographic group, we begin by asking users 'On Facebook, would you unfriend all your friends who are [gender] between ages [age bracket] for $X? Choose Yes if you do not use Facebook." We then rescale these responses by the estimated number of friends each ethnic group has, and our estimate of initial average welfare from Facebook (derived from our estimates of Ω_i) so that the sum of all friend network effects is equal to our estimate of the average initial utility per user from the platform. Finally, to estimate the disutility from advertising a_i we ask "What is the maximum amount of money (in US $) you would pay to personally not see any advertisements on Facebook for 1 month? Select 0 if you do not use Facebook." We divide this number by our estimates of initial revenues per user $\overline{\phi_i}$ to estimate a_i.

Figure 4 graphically represents Facebook usage and network externalities by market segment. The size of each node represents the relative current size of the Facebook user base by demographic. The thickness of the arrows corresponds to the relative value received by a Facebook user of the demographic the arrow is pointing towards from an additional Facebook user of the source demographic (i.e. the product of $z_i(j)$ and $U_i(j)$). As can be seen, there are more female users of Facebook overall and within each age group. The thickest lines in 4 flow from right to left. This is due in part due to older users having high valuations for connections to relatively non-abundant Facebook participants. The high value that older users glean from younger users is even more clear when restricting attention to the ten most valuable network effects, as figure 5 does. Appendix figures 19 through 22 restrict attention to the network effects experienced by and caused by other nodes of interest, displaying the rich heterogeneity of externalities on Facebook.

We calculate the impact of a change in advertising strategy, or some other change in Facebook's environment, over the course of multiple cascades. We denote the period when platform changes its advertising level as $t = 1$. The participation rate on the platform for a demographic group after cascade t is

$$P_{i,t} = \Omega_i(\sum^{J} U_i(j)z_i(j)D_j P_{j,t-1} - a_i\phi_{i,t}) \tag{25}$$

where $P_{i,0} = \overline{P_i}$, the initial rate of platform participation for the market segment.

We calculate the perceived welfare to a user of demographic i from the existence of Facebook after cascade t as

$$\int_0^{P_{i,t}} \left((\mu_i(\vec{P}_{j,t-1}, \phi_i) - e_i(\rho_i))\right) d\rho_i \tag{26}$$

where e_i is the inverse of Ω_i, giving the implied opportunity cost of Facebook use for every percentile of the population, i.e.

$$e_i = -s_i log(\frac{1 - p_i}{p_i}) + \mu_i \tag{27}$$

the total welfare to a demographic group from the existence of Facebook is the above amount times the number of users of that demographic group.

The revenue to Facebook from user participation of a given demographic after t cascades is

$$\Phi_{i,t} = \phi_{i,t} D_i P_{i,t} \tag{28}$$

we calculate 1000 cascades of the network effect but, as will be seen, most of the action occurs in the first few cascades.

6 Illustration Results

With the parameterized model in hand, we can now proceed to simulating counterfactual pricing strategies and potential government policies. We will begin by estimating Facebook's profit maximizing strategy. We then calculate the non-monetary value Facebook places on users which justifies their current monetization strategy as optimal – this is taken into account in the subsequent policy simulations. We then calculate, taking into account where appropriate Facebook's optimal response, the consumer welfare and Facebook revenue consequences of three tax and redistributive policies and three regulatory policies.

Figure 4: A graphical representation of Facebook usage and network externalities by market segment. The size of each node represents the relative current size of the Facebook user base by demographic. The thickness of the arrows corresponds to the relative value received by a Facebook user of the demographic the arrow is pointing towards from an additional Facebook user of the source demographic (i.e. $z_i(j)U_i(j)$)

.

Figure 5: A graphical representation of Facebook usage and network externalities by market segment. The size of each node represents the relative current size of the Facebook user base by demographic. The thickness of the arrows corresponds to the relative value received by a Facebook user of the demographic the arrow is pointing towards from an additional Facebook user of the source demographic (i.e. $z_i(j)U_i(j)$). Only the ten edges with the largest network externalities displayed.

	Cascade 1	Cascade 2	Cascade 3	Cascade 10	Cascade 20	Cascade 1000
Women 25-34	-0.94	-1.18	-1.25	-1.27	-1.27	-1.27
Women 35-44	-1.17	-1.47	-1.56	-1.59	-1.59	-1.59
Women 45-54	-0.50	-0.68	-0.73	-0.75	-0.75	-0.75
Women 55-64	-0.67	-1.03	-1.14	-1.18	-1.18	-1.18
Women 65+	-0.60	-1.17	-1.34	-1.41	-1.41	-1.41
Men 25-34	-0.29	-0.46	-0.51	-0.52	-0.52	-0.52
Men 35-44	-0.24	-0.37	-0.41	-0.43	-0.43	-0.43
Men 45-54	-0.21	-0.44	-0.51	-0.54	-0.54	-0.54
Men 55-64	-0.21	-0.47	-0.54	-0.57	-0.57	-0.57
Men 65+	-0.45	-0.87	-1.00	-1.05	-1.05	-1.05
Total CW Change	-5.28	-8.15	-8.99	-9.30	-9.31	-9.31
Profit Change	3.60	2.90	2.72	2.65	2.65	2.65

Figure 6: Changes in consumer surplus and Facebook profit after N cascades in billions of dollars per month, after Facebook implements its profit maximizing monetization strategy.

6.1 Facebook Profit Maximization

We begin by calculating Facebook's profit maximizing level of monetization. To calculate this, we iterate through guesses of different ϕ_i's for each demographic group until we identify a global maximum. We find that Facebook's profit maximizing strategy entails a large increase in the level of monetization. Therefore for this analysis we assume that the marginal disutility from increased monetization a_i is equal to 1 for each group.[15]

We find that Facebook's profit maximizing strategy entails increasing fees substantially across all groups. Both on a per-user level and in aggregate, the negative incidence of price increases falls mostly on women. This is because they demand Facebook more inelastically, and because they provide lower positive network externalities on average. They also currently provide less advertising revenue on average. These factors combine to make them relatively attractive targets for increased monetization. Figure 6 displays the change in Facebook ad revenues and consumer welfare after N cascades in billions of dollars per month.

Implementing this strategy would increase Facebook revenues by 2.65 billion dollars per month (from a baseline of 1.57 billion) at the cost of decreasing its user base by 55.6% and lowering consumer surplus by 82.8% (from a baseline of 11.2 billion). In other words, this strategy entails squeezing Facebook's most inelastic users for a much higher share of their surplus. If we restrict attention only to non-price discriminating strategies (i.e. requiring Facebook to raise ϕ_i for all groups proportionately), we find that the profit maximizing flat increase in ϕ is \$47.48 a month, which leads to an

[15] $a_i = 1$ is a logical upper bound, because Facebook could always simply charge a fee for use. In the policy simulations, which generally entail a reduction in advertising rates, we use our estimated a_i throughout.

increase in Facebook revenues of only 2.078 billion dollars a month, with even larger associated decreases in consumer surplus. This last result gives a sense of the importance of price discrimination to Facebook profit maximization, even when applied to potentially less important market segments such as age and gender.[16]

Why do our results imply that Facebook is leaving so much money on the table? There are two possible sets of answers. The first set of answers is that Facebook values having a large and happy user base. This could be because they value the data produced by a large user base (either for resale or for internal development), because they plan to monetize the user base further in the future (for example, keeping a marginal user on Facebook might increase the odds that they use Libra or some Oculus product in the future), or because it deters the entry of competitors. A second set of possibilities is that this is due to our model missing something important. For example, by taking into account only US users over age 25, we are potentially missing important network spillovers to and from other users of Facebook. If US users provide lots of value to users abroad, then it makes less sense to monetize them so intensely. Another possibility is that our surveys are soliciting a short-term demand elasticity for Facebook, whereas long-term demand for Facebook is more elastic.

6.2 The Impact of Tax and Redistribution Policies on Facebook Revenues and Social Welfare

We next simulate the consequences of three tax and redistribution policies. The two taxes we simulate are a tax on advertising revenues and a per-user tax. A tax on ad revenues has been proposed by leading economists such as Paul Romer (Romer, 2019). A three percent tax on sales of ads by large online platforms has recently been passed by France, but has not yet been implemented (CNBC, 2019). Grauwe (2017) proposes a 10 dollar per user tax.[17] A more radical proposal is the “Data as Labor” framework proposed in Weyl (2010). In this framework, perhaps through a collective bargaining process, users would be compensated for their ‘labor’ in providing data and viewing advertisements. We operationalize this last policy as Facebook maintaining its current level of advertising, but rebating to each demographic group the full revenue it collects from displaying them ads.

Before we proceed to simulations, our model has a novel theoretical point to make about the incidence of taxes on digital platforms. So long as a tax is flatly applied to all platform sources of revenue and utility, it will not distort the platform’s optimal vector of ϕ_i’s. To see this, consider the maximization problem 9. A tax that is equally

[16] Likely, even more gains from price discrimination could be achieved by price discriminating over more important attributes for demand like occupation or income.

[17] Professor Grauwe proposes this as an annual levy, but we consider a per-month tax.

	Initial	$10 Per User Tax	3% Ad Revenue Tax	"Data as Labor"
Net Ad Revenue (millions)	$ 1,565.7	-58%	-30%	-100%
Consumer Suplus (millions)	$ 11,225.5	-3%	2%	24%
Number of Users (millions)	134.741	-2%	1%	13%
Addl Tax Revenue (millions)	$ -	$ 1,326.1	$ 32.7	$ -

Figure 7: Estimated Facebook revenue, usage and consumer surplus changes as a result of three different redistributive reforms: a three percent tax on ad revenues, a ten dollar per a month per user tax, and a rebate of ad revenue to users in the spirit of Posner and Weyl's 'Data as Labor' proposal. The first two policies assume Facebook responds optimally to the tax change. The final proposal assumes advertising levels for all demographics are kept fixed at their current level.

applied to all ϕ_i revenues would add a multiplier term to this equation. It would not change the firm's first order maximization equation. In other words, under the assumptions of our model, including the assumption that platform's only source of revenues are advertisements and the platform faces no marginal cost, a flat tax on advertising revenues is fully incident on the platform's profits.[18] We have found however, that Facebook derives non-monetary utility from maintaining a large user base. Therefore a tax on ad-revenues will cause it to shift between it's two tasks from making revenues from selling advertising to increasing utility by cultivating a large user base. On the other hand, a tax on the amount of users will lead the platform to adopt a strategy that tries to squeeze a smaller share of users for more of their surplus.

Figure 8 summarizes the results of these three simulation experiments. As our theory suggested, a per-user tax slightly decreases the number of users and consumer surplus, while raising a large amount of revenues. On the other hand, a 3 percent tax slightly boosts consumer surplus and participation rates. However, it does not raise much revenue, and it has a disproportionate negative effect on Facebook net revenues, because Facebook reduces it's level of advertising in response. The "Data as Labor" policy has the most positive implications. Advertising, which is productive in the sense that it raises more revenue than the direct disutility it causes, is used to fuel a transfer to users. This directly makes users better off, and has a knock on effect of attracting additional users to the Facebook platform, who themselves provide positive spillovers to inframarginal users. About 58% of the welfare increase is due to the direct transfer to current users with the remainder due to new users who join the platform, consuming more ads and providing more value to other users.

[18] In the case of an ad tax that only applied to certain jurisdictions or demographic groups, there would be an incentive for the firm to increase monetization of users who provide value to the taxed group.

6.3 The Impact of Regulations on Facebook Revenues and Social Welfare

The final set of policies we simulate are regulatory. Many proposals have been made for regulation of online platforms and social media, some of which (especially those regarding "Fake News" and political manipulation) are beyond the scope of this current study.

Here we consider a set of three potential reforms, two assumed to be implemented perfectly, and one a 'worst case scenario' for a botched reform. The two positive reforms are a move to increase the competitiveness of social media and nationalization by a benevolent social planner.

In principle, it is not obvious whether decreasing the market power of a digital platform is a good or bad thing for social welfare. On the positive side, completely eliminating market power would force platforms to 'price' at their marginal cost – here assumed to be zero. It might also have positive political implications.

Increasing competition might be bad for a few reasons. First, and most theoretically interesting, a monopolist can cross-subsidize different sides of a market in a way that a competitive firm cannot. In the same way that a government might subsidize an infant industry for the good of the total economy in the long-run, a monopolist platform is a sort of 'stationary bandit' who has an interest in taking into account at least some network effects. This incentive differs from the social planners' interest in that the monopolist only cares about the network effect on marginal platform users (rather than on all platform users). Another reason market power might be good in this setting in particular is that it might prevent 'production' through advertising. Because advertisements raise more revenue than the disutility they directly cause, the social welfare optimum may include a positive, rather than zero, amount of advertising. Of course, this argument is null if advertising revenues can be rebated (as we assumed in the 'Data is Labor' case above), but one can imagine several frictions that might cause this.

Perhaps the most important reason increasing platform competition is not an obvious win is that it has the potential to destroy network effects by splitting the market. If multi-homing is costly and network effects do not spillover across platforms, then increasing the number of platforms may decrease the positive network effects that are the main draw and purpose of digital platforms. To resolve this last concern, a recent study of anti-trust and regulation in the context of digital platforms, (Scott Morton et al., 2019), has called for mandated 'interoperability' alongside other policy changes that would lower barriers to entry. Interoperability would require Facebook to share posts and other communiques with competitor social networks, who would then be allowed to display them on their platforms. We consider 'perfect competition' as en-

	Initial	Perfect Competition	Facebook Breakup	Welfare Maximization
Net Ad Revenue (millions)	$ 1,565.7	-100%	-93%	-255%
Consumer Suplus (millions)	$ 11,225.5	9%	-44%	42%
Number of Users (millions)	134.741	5%	-18%	20%
Change in Non-Rev Value (millions)	$ -	-	-$8,155	$ 5,019

Figure 8: Estimated Facebook revenue, usage and consumer surplus changes as a result of three different regulatory reforms: mandating interoperability and lowering barriers to entry to create perfect competition, a botched Facebook breakup that creates two uncompetitive monopolies for half of the US, and nationalization by a benevolent social planner.

tailing this interoperability, and model it as the elimination of all advertisements on Facebook.

One component of many plans to increase platform competition includes mandatory 'breakups'. For example, an essay by a leading presidential candidate calls for, among other things, Facebook to be split from Instagram and Whatsapp (Warren, 2019). To the extent that these are separate platforms that do not allow for network effects across them, such a breakup is sensible. But one can imagine a botched breakup of Facebook that both destroyed network effects and failed to increase competition (e.g. by dictating that users must use only one of the two platforms). We model such a 'worst case scenario' Facebook breakup as the creation of two Facebook monopolies each serving half of the US population.[19]

The final scenario we simulate is one in which a benevolent social planner takes over Facebook, and runs it to maximize social welfare. Such a planner would internalize all network effects. Here, we also model the planner as taking into account the platforms' desire to have a large user base (as this might represent the future value of data collection). This simulation entails the platform running 'negative advertisements' (i.e. expending money to boost the welfare of users on the platform). [20]

Results from these three simulations are summarized in figure 8. We find that perfect competition would raise consumer surplus by 9%, at the cost of eliminating all monetary profits. Taking only Facebook's monetary revenues into account, perfect competition actually lowers social surplus, because the reduction in ad revenues is larger than the reduction in consumer welfare. However, if a large user base is still assumed to create social surplus at the same rate as for Facebook today, the policy creates a clear social welfare improvement. A social welfare maximizing Facebook would raise consumer surplus by 42%, at the cost of Facebook needing to go -255%

[19]Such a scenario is not that far-fetched. The breakup of 'Ma' Bell Telephone led to the creation of several regional monopolies and one 'long-distance' monopoly.

[20]We assume that these negative advertisements symetrically create utility at the rate a_i. If we assume that $a_i \geq 1$, then the social welfare optimum comes at a transfer of negative infinity.

into the red. However, breaking Facebook into two non-competitive ‘baby Facebooks’ would be disastrous, lowering consumer surplus by 44%. It would also lower combined ad revenues by 93% as the baby Facebooks lowered advertising rates to retain even 82% of their original combined user base.

7 Conclusion and Managerial Implications

Building on Rochet and Tirole (2003), Parker and Van Alstyne (2005) and Weyl (2010) we construct and illustrate an approach for firms to incorporate network effects in their monetization strategies. The specific example we emphasize is a firm which can discriminate in its advertising to profit maximize. Taking the first order condition for profit maximization with respect to the advertising schedule yields a recursive equation that can be evaluated to the desired decree of precision. The managerial insight is that platform owners should increase advertising on market segments which inelastically demand the platform (the direct effect), don’t have much disutility from advertisements, and don’t create much network value for others. Platforms should decrease advertisements on those who elastically demand the platform and create high amounts of network value for other profitable users who demand Facebook elastically (the first cascade of the network effect).

We use this model to estimate, in the case of Facebook, the revenue and welfare consequences of different pricing strategies, taxes, regulations and market structures. As far as we know this is the first paper to produce such predictions. Hopefully these findings will be useful in guiding policy makers, and will serve as one approach among many for projecting the impact of policy changes.

That being said, our approach is not without weaknesses. One important issue is trickiness in soliciting the necessary data to estimate the model. Consumers may not fully understand or reliably answer questions about their valuations for different friend groups. Poor memory may also be an obstacle. There may also be important differences between short and long-term elasticities of demand. Similarly, if individuals have very high variance or skewness in their platform valuations, network effects, or number of friends, the average of these values within a group may be a poor summary statistic – especially if these measures are correlated within a side of the market/demographic group. Relatedly, in our parameterization we currently assume that the value from friends is linearly additive and that the disutility from advertising revenues is linear. Both are clearly simplifications. However, with a larger budget, incentive compatible experiments, smaller market segments or within-platform proprietary data, each of these concerns could be addressed, and the nature of utility functions measured more precisely. Another limitation of the current approach is that advertisers are treated as price setters, rather than as a side of the market. A more complete model would treat advertisers as a heterogenous mix of agents as well.

Finally, our model conceives of consumers as atomistic price takers. This ignores the possibility that highly valuable users with market power might bargain with the platform or that users might unionize to demand a better equilibrium. However, the implications of such a scenario could be estimated in an extension of the model. In any case an intriguing area for future investigation is to actually conduct experiments on platforms to see how well real world phenomena match our predictions.

References

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The welfare effects of social media," Technical Report, National Bureau of Economic Research 2019.

Aral, Sinan and Christos Nicolaides, "Exercise contagion in a global social network," *Nature communications*, 2017, *8*, 14753.

_ **and Dylan Walker**, "Identifying influential and susceptible members of social networks," *Science*, 2012, *337* (6092), 337–341.

Bernstein, Shai and Eyal Winter, "Contracting with heterogeneous externalities," *American Economic Journal: Microeconomics*, 2012, *4* (2), 50–76.

Boudreau, Kevin, "Open platform strategies and innovation: Granting access vs. devolving control," *Management science*, 2010, *56* (10), 1849–1872.

Brynjolfsson, E, A Collis, and F Eggers, "Using massive online choice experiments to measure changes in well-being.," *Proceedings of the National Academy of Sciences of the United States of America*, 2019.

Candogan, Ozan, Kostas Bimpikis, and Asuman Ozdaglar, "Optimal pricing in networks with externalities," *Operations Research*, 2012, *60* (4), 883–905.

Ceccagnoli, Marco, Chris Forman, Peng Huang, and DJ Wu, "Co-creation of value in a platform ecosystem: The case of enterprise software," *MIS Quarterly, Forthcoming*, 2011.

CNBC, "France targets Google, Amazon and Facebook with 3% digital tax," Mar 2019.

Eisenmann, Thomas, Geoffrey Parker, and Marshall W Van Alstyne, "Strategies for two-sided markets," *Harvard business review*, 2006, *84* (10), 92.

Evans, David S and Richard Schmalensee, "Failure to launch: Critical mass in platform businesses," *Review of Network Economics*, 2010, *9* (4).

Fainmesser, Itay P and Andrea Galeotti, "Pricing network effects," *The Review of Economic Studies*, 2015, *83* (1), 165–198.

Grauwe, Paul De, "Why Facebook Should Be Taxed And How To Do It – Paul De Grauwe," Oct 2017.

Hagiu, Andrei, "Two-sided platforms: Product variety and pricing structures," *Journal of Economics & Management Strategy*, 2009, *18* (4), 1011–1043.

Hohnhold, Henning, Deirdre O'Brien, and Diane Tang, "Focusing on the Long-term: It's Good for Users and Business," in "Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining" ACM 2015, pp. 1849–1858.

Huang, Jason, David Reiley, and Nick Riabov, "Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio," *Available at SSRN 3166676*, 2018.

Huang, Ms Shan, Sinan Aral, Jeffrey Yu Hu, and Erik Brynjolfsson, "Social Advertising Effectiveness Across Products: A Large-Scale Field Experiment," 2018.

Jackson, Matthew O, *Social and economic networks*, Princeton university press, 2010.

Morton, Fiona Scott, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A Douglas Melamed, and Jamie Morgenstern, "Committee for the Study of Digital Platforms: Market Structure and Antitrust Subcommittee Report," 2019.

Parker, Geoffrey G and Marshall W Van Alstyne, "Two-sided network effects: A theory of information product design," *Management science*, 2005, *51* (10), 1494–1504.

Posner, Eric A and E Glen Weyl, *Radical markets: Uprooting capitalism and democracy for a just society*, Princeton University Press, 2018.

Rochet, Jean-Charles and Jean Tirole, "Platform competition in two-sided markets," *Journal of the european economic association*, 2003, *1* (4), 990–1029.

Romer, Paul, "A Tax that Could Fix Big Tech," *The New York Times*, May 2019.

Rysman, Marc, "Competition between networks: A study of the market for yellow pages," *The Review of Economic Studies*, 2004, *71* (2), 483–512.

Tan, Hongru and Julian Wright, "A Price Theory of Multi-Sided Platforms: Comment," *American Economic Review*, 2018, *108* (9), 2758–60.

Tucker, Catherine, "Identifying formal and informal influence in technology adoption with network externalities," *Management Science*, 2008, *54* (12), 2024–2038.

Warren, Elizabeth, "Here's how we can break up Big Tech," Mar 2019.

Weyl, E Glen, "A price theory of multi-sided platforms," *American Economic Review*, 2010, *100* (4), 1642–72.

Yan, Jinyun, Birjodh Tiwana, Souvik Ghosh, Haishan Liu, and Shaunak Chatterjee, "Measuring Long-term Impact of Ads on LinkedIn Feed," *arXiv preprint arXiv:1902.03098*, 2019.

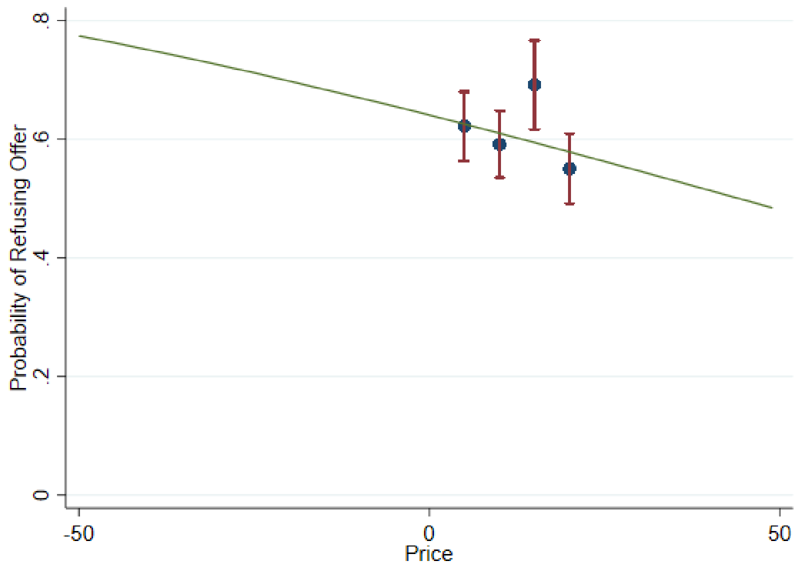


Figure 9: Underlying data and estimate of the the demand curve (Ω_i) for women age 25-34. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for \$X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

A Additional Tables and Figures

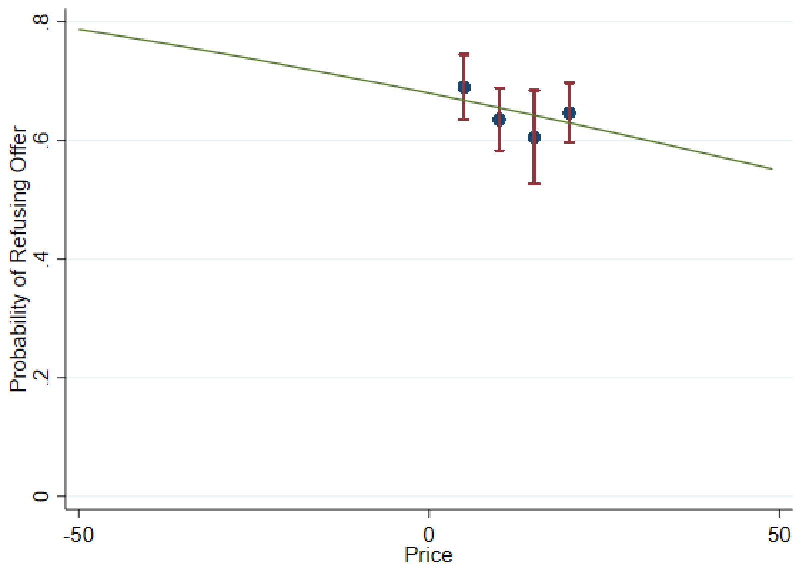


Figure 10: Underlying data and estimate of the the demand curve (Ω_i) for women age 35-44. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

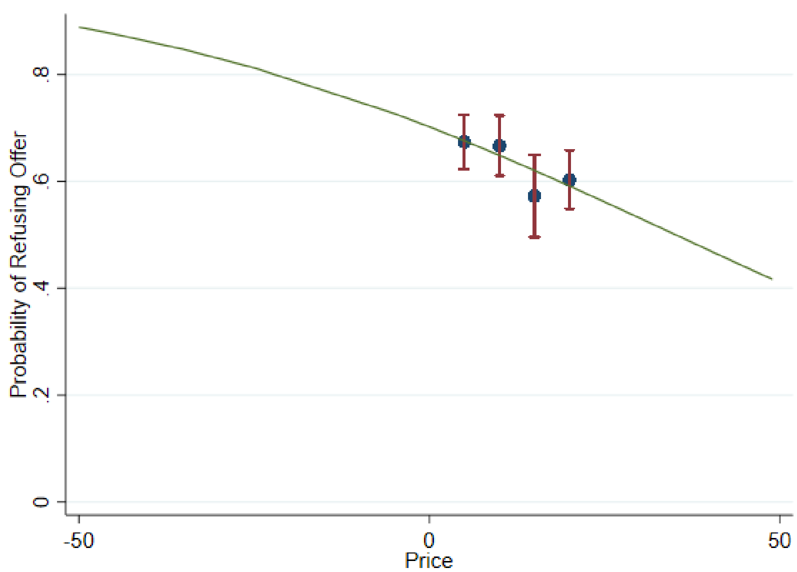


Figure 11: Underlying data and estimate of the the demand curve (Ω_i) for women age 45-54. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

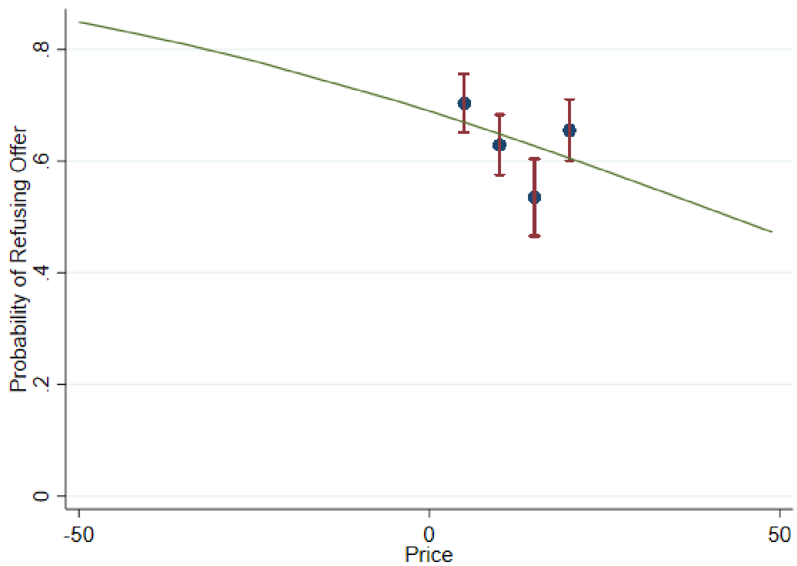


Figure 12: Underlying data and estimate of the the demand curve (Ω_i) for women age 55-64. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

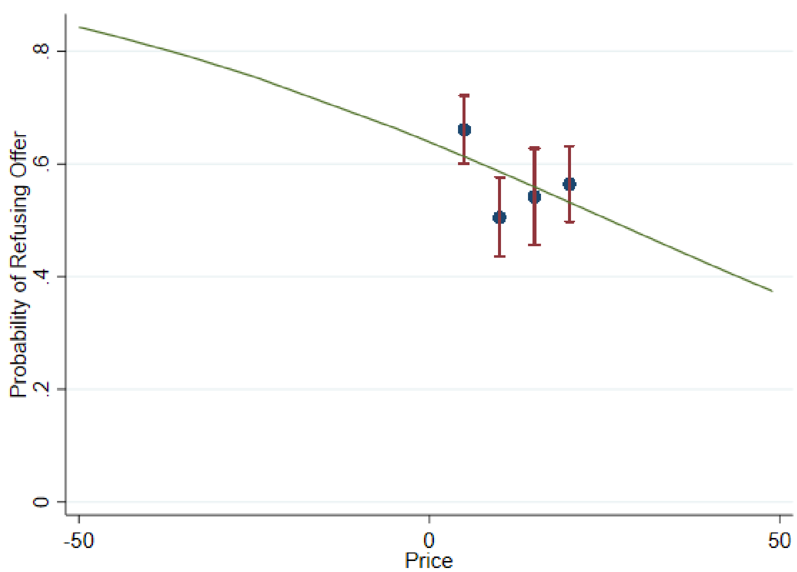


Figure 13: Underlying data and estimate of the the demand curve (Ω_i) for women age 65 or older. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

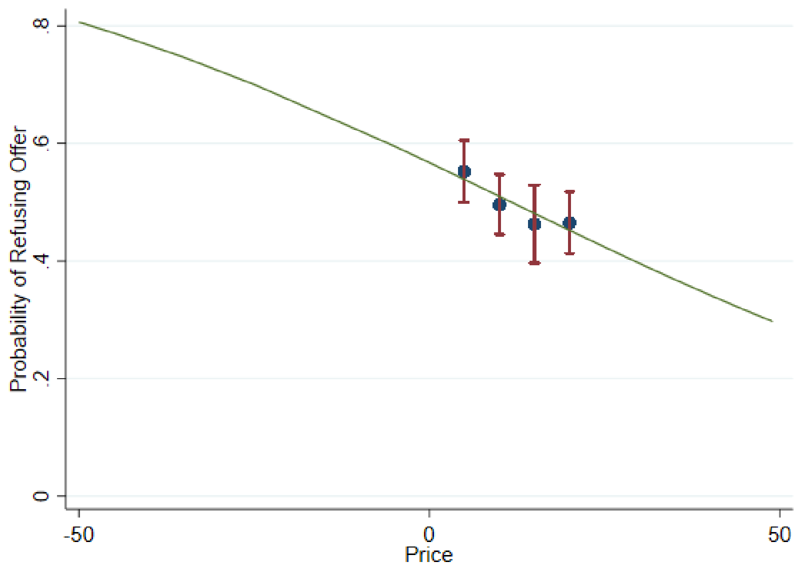


Figure 14: Underlying data and estimate of the the demand curve (Ω_i) for men age 25-34. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

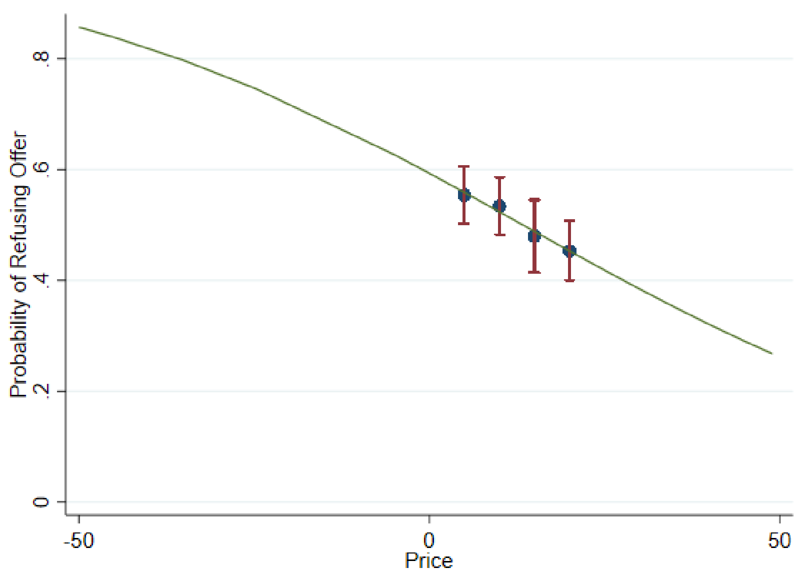


Figure 15: Underlying data and estimate of the the demand curve (Ω_i) for men age 35-44. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for $X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

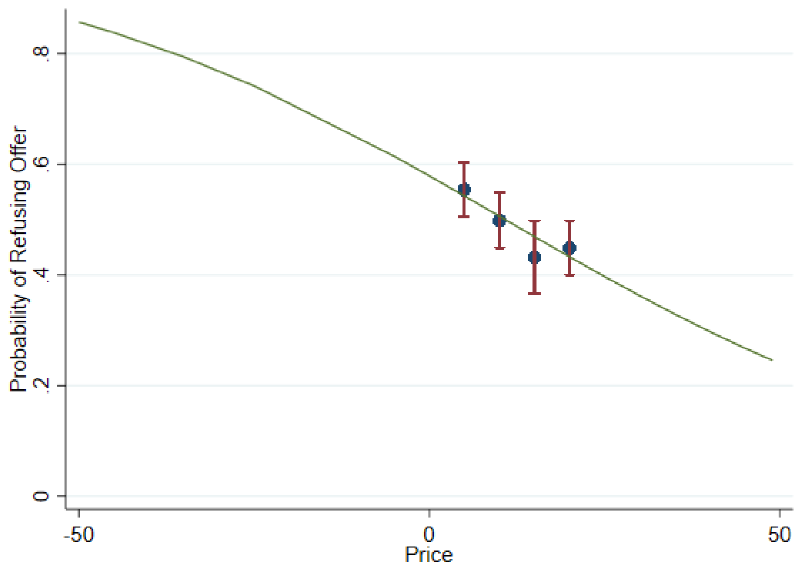


Figure 16: Underlying data and estimate of the the demand curve (Ω_i) for men age 45-54. The points are the mean response to the question “Would you give up Facebook for 1 month in exchange for \$X? Choose Yes if you do not use Facebook.” for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

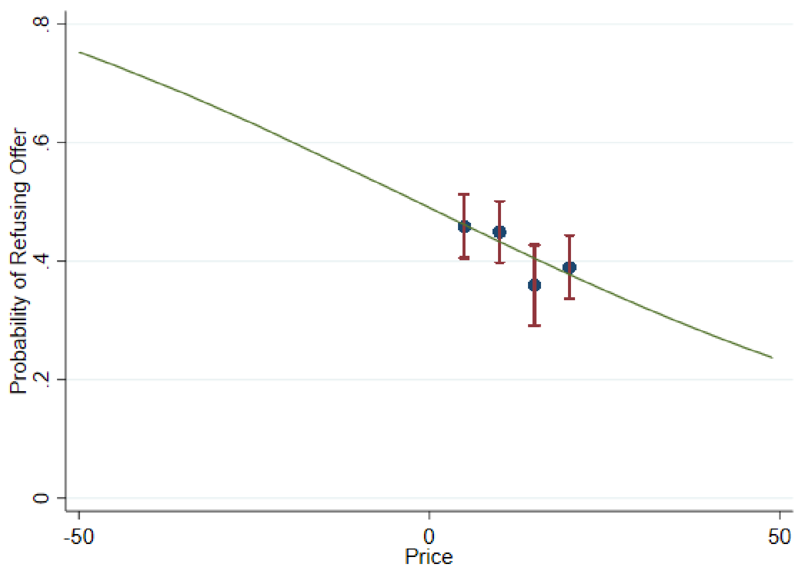


Figure 17: Underlying data and estimate of the the demand curve (Ω_i) for men age 55-64. The points are the mean response to the question “Would you give up Facebook for 1 month in exchange for \$X? Choose Yes if you do not use Facebook.” for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

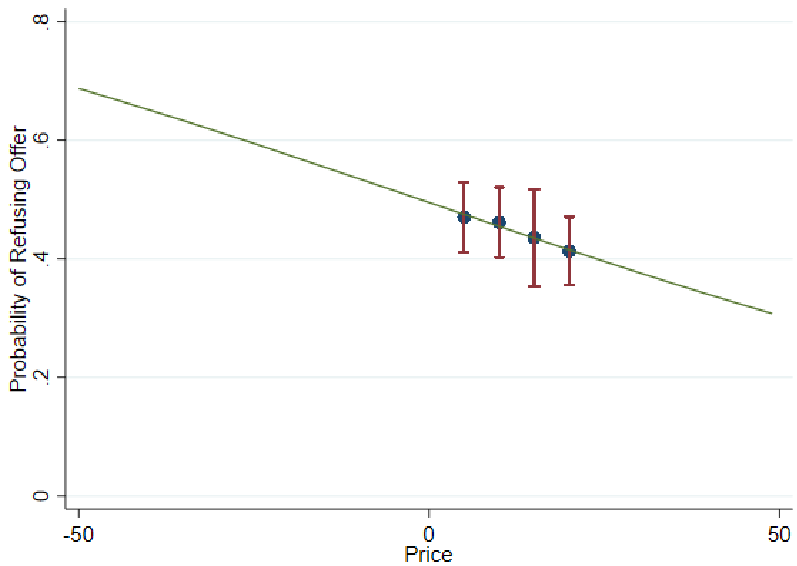


Figure 18: Underlying data and estimate of the the demand curve (Ω_i) for men age 65 or older. The points are the mean response to the question "Would you give up Facebook for 1 month in exchange for \$X? Choose Yes if you do not use Facebook." for individuals of the group. Confidence intervals are based on binomial statistics. The curve, in green, is the logistic line of best fit.

Figure 19: A graphical representation of Facebook usage and network effects. Relative value of Females 65+ to users of different demographics displayed.

Figure 20: A graphical representation of Facebook usage and network effects. Relative value of other users to Females 65+ displayed.

Figure 21: A graphical representation of Facebook usage and network effects. Relative value of Males age 25-34 to users of different demographics displayed.

Figure 22: A graphical representation of Facebook usage and network effects. Relative value of other users to Males age 25-34 displayed.

Intercept	Coefficient on Cost	Demo Group
.578603	.0130975	Female 25-34
.753299	.0111212	Female 35-44
.856898	.0243794	Female 45-54
.798354	.0185573	Female 55-64
.570240	.0221589	Female 65+
.270967	.0231238	Male 25-34
.378010	.0282876	Male 35-44
.319277	.0294697	Male 45-54
-.03944	.0230791	Male 55-64
-.01939	.0161522	Male 65+

Table 1: Coefficient estimates from a logit regression of willingness to stop using Facebook on cost of Facebook proposed (equal to negative of the Price offered to stop using Facebook).

B Network Stability

B.1 Stability of Equilibria

An important first question is whether the network just described is stable. We define a network as stable at equilibrium $\vec{P}$ if the derivative of a connected individual's best response function with response to these probabilities is less than 1.[21] This is a version of a 'trembling hand' perfect equilibrium, meaning that the equilibrium is robust to small fluctuations in each individual's likelihood of participation.

For a symmetric network (i.e. every individuals' demand function Ω is identical), assuming that utility is linearly additive in the network effects and disutility from advertisements, the probability of participation for any individual is

$$P = \Omega\Big(\sum^{I} U(i)P - a\phi\Big) \tag{29}$$

where $U(i)$ is the value of any connection.[22] Then the best response function is

$$\frac{\partial\Omega}{\partial P} = \frac{\partial\Omega}{\partial \boldsymbol{U}}\Big(\sum^{I} U(i) - a\phi\Big) \tag{30}$$

And so a network equilibrium is stable so long as

$$1 > \frac{\partial\Omega}{\partial \boldsymbol{U}} U(i)(I-1) \tag{31}$$

In other words, a network equilibrium is stable so long as the average user doesn't have too many connections, is too elastic in their individual participation, or gains too much value from every additional connection. If the inequality is violated, small deviations from an equilibrium are liable to send participation to a boundary condition of 100% participation or zero participation.

B.2 Stability of Equilibrium to Demand Shock

Relatedly, we can also consider the resilience of a network equilibrium to a shock in preferences.

Theorem 1. *Consider a symmetric network where Ω is continuously differentiable and utility is linearly additive in network effects and the disutility from advertisement. Then for any stable equilibrium (as defined above) $\frac{P_i}{\phi_j}$ and $\frac{P_i}{\phi_j}$ are finite*

[21] This concept of equilibrium stability borrows from Jackson (2010) section (9.7.2). In that model, only some individuals are connected in the network, but in our model all are connected. In that model p corresponds to the percentage of neighbors who participate, but in our model it corresponds to the likelihood of anyone who participates.

[22] the value of a 'connection to oneself' is assumed to be 0

Proof. Rewriting equation 15 with the assumption all nodes are identical, before i gets hit with a fee, yields:

$$\frac{\partial P_j}{\partial \phi_i} = \frac{\partial \Omega}{\partial \boldsymbol{U}}\left(U(i)\frac{\partial P_i}{\partial \phi_i} + \sum_{k \neq i}^{K} U(i)\frac{\partial P_k}{\partial \phi_i}\right) \tag{32}$$

Substituting in 14 and summing yields

$$\frac{\partial P_j}{\partial \phi_i} = \frac{\partial \Omega}{\partial \boldsymbol{U}}\left((I-2)\frac{\partial P_j}{\partial \phi_i}U(i) + \frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)\Big(U(i)(I-1)\frac{\partial P_j}{\partial \phi_i} - \frac{\partial A}{\partial \phi_i}\Big)\right) \tag{33}$$

Solving for $\frac{\partial P_j}{\partial \phi_i}$ yields

$$\frac{\partial P_j}{\partial \phi_i} = \frac{-\big(\frac{\partial \Omega}{\partial \boldsymbol{U}}\big)^2 U(i)\frac{\partial A}{\partial \phi_i}}{1-\Big(\frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)(I-2) + \big(\frac{\partial \Omega}{\partial \boldsymbol{U}}\big)^2\big(U(i)\big)^2(I-1)\Big)} \tag{34}$$

The network will not unravel due to a welfare change so long as 34 is not infinite. This is equivalent to showing that the denominator is not equal to zero (as all other terms are finite).

However, the denominator never takes the value 0 when the network stability criteria is satisfied. Rearranging terms, the denominator can be written as

$$1 - \frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)\Big((I-2) + \big(\frac{\partial \Omega}{\partial \boldsymbol{U}}\big)\big(U(i)\big)(I-1)\Big) \tag{35}$$

From the assumption that the network is stable, we have

$$1 > \frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)(I-1) \tag{36}$$

This implies

$$I - 1 > (I-2) + \frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)(I-1) \tag{37}$$

and applying B again implies

$$1 > \frac{\partial \Omega}{\partial \boldsymbol{U}}U(i)\Big((I-2) + \big(\frac{\partial \Omega}{\partial \boldsymbol{U}}\big)\big(U(i)\big)(I-1)\Big) \tag{38}$$

And if $\frac{\partial P_j}{\partial \phi_i}$ is finite, clearly so to is $\frac{\partial P_i}{\partial \phi_i}$. So long as the network is stable in the normal sense, it is stable to welfare shocks.

□

Lemma 2. *In a symmetric network,* $\frac{\partial P_i}{\partial \phi_j} = 0$ *if* $\big(\frac{\partial \Omega}{\partial \boldsymbol{U}}\big)^2\frac{\partial A}{\partial T}U(j) = 0$

Proof. Directly from (34) □

C Social Welfare Maximization

The increase in welfare due to the platform's existence for a given individual i is

$$W_i = P_i E[\mu_i(\vec{P}, \phi_i) - \epsilon_i | U_i > \epsilon_i] \tag{39}$$

in other words, welfare from the platform is the odds an individual participates on the platform, multiplied by their expected surplus from platform use. This expected surplus is equal to the value of platform use less opportunity cost.

Evaluating this equation yields

$$W_i = P_i \int_{-\infty}^{U_i} \frac{\mu_i(\vec{P}, \phi_i) - \epsilon_i}{Prob(U_i > \epsilon_i)} f(\epsilon_i) d\epsilon_i \tag{40}$$

where $f(\epsilon_i)$ is the pdf of ϵ_i. There is an upper bound on the integral, because an individual only participates – and pays the opportunity cost – if the value of participation exceeds the opportunity cost. Now, μ_i is a constant with reference to the integral, so this reduces to

$$W_i = P_i \frac{\mu_i(\vec{P}, \phi_i) F(\epsilon_i)}{Prob(U_i > \epsilon_i)} \Big|_{-\infty}^{U_i} - P_i \int_{-\infty}^{U_i} \frac{\epsilon_i}{Prob(U_i > \epsilon_i)} f(\epsilon_i) d\epsilon_i \tag{41}$$

where $F(\epsilon_i)$ is the CDF of ϵ_i. Now, $Prob(U_i > \epsilon_i) = F(U_i) = P_i$ so

$$W_i = P_i \mu_i(\vec{P}, \phi_i) - \int_{-\infty}^{U_i} \epsilon_i f(\epsilon_i) d\epsilon_i \tag{42}$$

Social welfare maximization needs to take into account both consumer surplus and platform surplus. Using the same equation for platform profits as used above, this means social welfare maximizaiton entails

$$\max_{\phi_i} \sum_{i}^{I} [P_i\big(\phi_i + \mu_i(\vec{P}, \phi_i)\big) - Q_i(\epsilon_i)\Big|_{-\infty}^{U_i}] - F \tag{43}$$

s.t.

$$P_i = \Omega_i(\boldsymbol{U_i}) \tag{44}$$

Where $Q_i(\epsilon_i)$ stands for the indefinite expectation integral $\int \epsilon_i f(\epsilon_i) d\epsilon_i$.

The first term is the utility from participation to users of the platform μ_i and to the firm ϕ_i. These are both multiplied by the odds of participation. The next term is the expected opportunity cost to an individual from participating.

Chapter 4: Effects of restricting social media usage

Abstract

Recent research has shown that social media services create large consumer surplus. Despite their positive impact on economic welfare, concerns are raised about the negative association between social media usage and performance or well-being. However, causal empirical evidence is still scarce. To address this research gap, we conduct a randomized controlled trial among students in which we track participants' digital activities over the course of three quarters of an academic year. In the experiment, we randomly allocate half of the sample to a treatment condition in which social media usage is restricted to a maximum of 10 minutes per day. We find that participants in the treatment group substitute social media for instant messaging and do not decrease their total time spent on digital devices. Contrary to findings from previous correlational studies, we do not find any impact of social media usage on well-being and academic success. Our results also suggest that antitrust authorities should consider instant messaging and social media services as direct competitors before approving acquisitions.

INTRODUCTION

Social media increasingly plays an important role in our daily lives. Ever since the launch of major modern social media platforms such as Facebook, users have adopted them at an explosive pace and adoption continues to increase to this day. Over 2.7 billion users worldwide are projected to use social media services in 2019[1]. This corresponds to over a third of the global population and 72% of internet users. This figure is expected to grow at around 4-5% every year for the next few years. The average adult spends over 45 minutes every day on social media platforms[2].

Given this rapid adoption and usage of social media platforms, it is essential to study the impact of social media on the well-being of users. Brynjolfsson, Collis and Eggers (2019) find that digital technologies, including social media, generate a large amount of consumer surplus. More specifically, they conduct incentive compatible choice experiments to measure the consumer surplus generated by Facebook and find that the median US Facebook user obtains around $48/month of value from using Facebook in 2017 as measured from their willingness to accept to give up access to Facebook for a month. They also conduct a similar experiment with students at a large European university and find that the median student in their sample obtains €97/month of value from using Facebook.

While Facebook and other social media services seem to generate a large amount of consumer surplus and contribute towards the economic well-being of their users, questions are raised about the negative externalities generated by social media. There is an active debate in media and academic research about the impact of social media on subjective well-being (including happiness and life satisfaction) and productivity. Current empirical results

[1] Source: eMarketer (https://www.emarketer.com/Article/eMarketer-Updates-Worldwide-Social-Network-User-Figures/1016178, accessed on May 6, 2019)
[2] Source: Nielsen (https://www.nielsen.com/us/en/insights/news/2018/time-flies-us-adults-now-spend-nearly-half-a-day-interacting-with-media.html, accessed on May 6, 2019)

are ambiguous. Across different studies, correlational evidence points towards a positive, neutral (null results) and negative relationship between social media use and well-being (see Haidt (2019) for a comprehensive literature review of social media use and mental health). However, most of this evidence suffers from issues related to reverse causality (Cheng, Burke and Davis (2019)) and inaccurate measures of self-reported social media use (Orben, Dienlin and Przybylski (2019)). Rigorous causal evidence on long term impacts of social media use on well-being is lacking.

Concerns are also raised in the field of Education policy on the impact of screen time (including social media use) on academic performance of students. Critics contend that social media use on smartphones distracts students from focusing in classes and affects their grades. Motivated by these concerns, the French education ministry banned smartphones in schools from first through ninth grades[3]. The American Academy of Pediatrics also recommends parents to limit the time spent by children and adolescents on social media so that they have enough time left to study[4]. However, a rigorous analysis of the data used in previous correlational studies that were used as evidence to support these policies suggests that the effects of social media use and screen time on adolescent well-being are too small to warrant policy changes (Orben and Przybylski (2019)).

Given these widespread concerns and conflicting correlational evidence on the impact of social media on well-being, it is necessary to obtain causal evidence in a timely manner before policies are implemented hastily. We seek to fill this research gap by conducting a first of its kind randomized controlled trial to measure the causal long term impact of social media use on academic performance and well-being.

[3] Source: The New York Times (https://www.nytimes.com/2018/09/20/world/europe/france-smartphones-schools.html, accessed on May 11, 2019)

[4] Source: American Academy of Pediatrics (https://www.aap.org/en-us/about-the-aap/aap-press-room/Pages/American-Academy-of-Pediatrics-Announces-New-Recommendations-for-Childrens-Media-Use.aspx, accessed on May 11, 2019)

We recruit students at a large European university to be part of our study over the course of three academic terms (quarters). The subjects install a software on their personal computers and mobile devices. This software tracks all of the digital activities of the subjects during the entire duration of the study period. The first term serves as the baseline period. In the second term, subjects are randomized into treatment and control groups and the treatment group has social media use (Facebook, Instagram and Snapchat) restricted to a maximum of 10 minutes per day across all devices. We then measure the post-treatment effects in the third term.

We observe the entire space of digital activities performed by our subjects that covers online and also offline activities on their devices, including activities related to learning (such as writing in Microsoft Word or reading a PDF). Our social media use metrics are computed based on the actual time spent on social media and are not based on self-reported metrics of time spent, which is predominantly used in the existing literature. In addition to the digital activities, we obtain objective metrics of performance (grades) in addition to subjective well-being scores solicited through surveys.

Contrary to results from previous studies using observational data, we do not find evidence that social media causes a positive or negative impact on well-being (including life satisfaction and mental health). Moreover, we also do not find any evidence that social media usage impacts academic success. However, we find significant substitution effects. Specifically, we see that participants in the treatment group substituted their use of social media services for instant messaging apps (e.g. WhatsApp). In total, these participants do not spend less time on their digital devices (computers and mobile phones) as those in the control group.

Our paper makes three main contributions. First, we test the popular media narrative portraying social media as the villain responsible for negatively affecting well-being of society. We do not find any evidence supporting this hypothesis. Second, educators and parents are increasingly concerned about the impact of digital distractions on academic performance and are restricting the online activities of students (for example through parental control software or by taking away their devices). While previous evidence seems to suggest that device usage in class might negatively affect academic performance, our results show that restricting social media usage from the lives of students (inside and outside class) might not have the intended effect. Finally, our paper is the first to provide evidence of substitutability between social media and instant messaging apps. This has major implications for antitrust authorities analyzing the market power of major social media platforms such as Facebook which owns Instagram (another social media service) and WhatsApp (instant messaging service).

The paper proceeds as follows. In the next section, we provide a brief review of existing literature on the impact of social media use on well-being and academic performance. In the following section, we describe the design of our experiment and data collected over the course of the study. We then show the main results and conclude with a discussion of the limitations of this study and directions for future research.

RELATED LITERATURE

The impact of the internet in general, and social media in particular, on well-being has attracted the attention of a number of researchers in the fields of psychology, epidemiology and human-computer interaction (HCI) over the past decade. Almost all of this literature uses self-reported metrics of technology use and provides cross-sectional correlational evidence. Kraut and Burke (2015) provide a review of this literature and express skepticism regarding cross-sectional and survey-based studies due to the presence of several confounding factors.

Moreover, correlational studies might suffer from an abundance of researcher degrees of freedom and the file drawer problem such that only significant results are published, inevitably leading to the implication that social media usage either has a positive or negative effect. However, a null result or insignificant findings regarding social media usage might be a plausible outcome.

Orben and Przybylski (2019) rigorously analyze popular large scale social datasets (n=350k) used in previous correlational studies studying the impact of technology use on well-being by conducting a specification curve analysis of the data. This analysis involves running all possible analytical models using various combinations of the covariates. Instead of selective reporting, results from all of these analyses are reported. They find a small negative association between digital technology use and adolescent well-being. However this effect is economically insignificant explaining at most 0.4% of the variation in well-being. For comparison, the authors show that seemingly neutral activities such as eating potatoes have the same negative association with well-being as technology use. Given these concerns with correlational analyses involving cross sectional data, Kraut and Burke (2015) call for experimental evidence paired with tracking data to provide reliable evidence on the relationship between internet use and well-being.

The subset of literature focusing on the association between social media use and well-being has found a wide range of effects (negative, mixed, positive and null). Using a longitudinal survey, Shakya and Christakis (2017) found a negative association between Facebook use and well-being. In contrast, Burke, Marlow and Lento (2010) find a positive association between directed communication on Facebook and social well-being due to subjects reporting improved feelings of social bonding and reduced loneliness. Similarly, Hobbs et al. (2016) match Facebook profiles with public health records and find that being more socially integrated online (by accepting more Facebook friends) is associated with reduced risk of mortality. Burke and Kraut (2016) find that targeted messages from strong ties is associated with positive improvements in well-being while viewing messages from friends broadcasted to all of their friends and receiving one-click feedbacks were not associated with any improvement in well-being. Cheng, Burke and Davis (2019) combine a survey of Facebook users with their Facebook activities and find that subjects reporting problematic use of Facebook were also going through a major life event such as a breakup. This shows that confounding variables could be a major concern in previous studies associating social media use and well-being.

Orben, Dienlin and Przybylski (2019) use a large scale longitudinal dataset and conduct a specification curve analysis to rigorously analyze the relationship between adolescent social media use and well-being. Most of the analyses report tiny, trivial and insignificant results. Moreover, they provide evidence for reverse causality showing that social media use predicts well-being in the future and vice versa.

Another major concern related to existing studies is the use of self-reported usage data. Survey respondents are typically asked to report the average time they spend on the internet, social media and digital devices. Several papers show that self-reported measures of technology use (including social media usage) are poorly correlated with actual usage and contain systematic patterns of misreporting (Junco 2013, Scharkow 2016, Ellis et al. 2019, Ellis 2019).

Given this inconclusive evidence and lack of objective technology use data in existing literature, it is essential to obtain reliable causal evidence in a timely manner to inform policy makers. We aim to resolve this gap by obtaining evidence through a randomized controlled trial and using objective technology use metrics tracked by a software installed on the digital devices of our experimental subjects. In terms of outcome variables, we track measures of subjective well-being (life satisfaction and mental health) and performance (grades and number of credit points) over the duration of three quarters of an academic year (8 months) with the actual treatment lasting 2.5 months. To the best of our knowledge, this study is the first that tracks all of these components of well-being and over a long period of time.

There is a small stream of literature using experiments to study the relationship between social media or computer usage and well-being or performance. Verduyn et al. (2015) conduct a lab experiment where subjects are primed to passively use Facebook for 10 minutes and find that passive use is associated with a decline in subjective well-being. However, it is not straightforward if results from a 10 minute treatment can be generalized to long term effects.

Marotta and Acquisti (2018) conduct an experiment with workers recruited from Amazon mechanical turk and offer productivity enhancing tools to subjects. One of the treatment groups has popular social media sites blocked during work hours. They find that workers in this group completed more tasks and increased their earnings. Carter, Greenberg and Walker (2017) conduct a randomized controlled trial in a US university where classes in the treatment group prohibited the use of computers in the class. They find that average exam scores were higher in the treatment group compared to the control group classes where students were allowed to use their computers. Using causal inference methods on observational data, Belo, Ferreira and Telang (2014) and Beland and Murphy (2016) study the impact of broadband access and banning mobile phones in schools respectively on academic performance and also find evidence suggesting that digital distractions during class reduce academic performance. Taken together, evidence seems to suggest that digital device use in class or at work is harmful for student or worker performance. However, the overall causal impact of social media usage in life (inside and outside class or at work) on performance and well-being still remains an open question. Our study complements this research by analyzing the overall long term impact of social media on well-being and performance as the subjects in our treatment group has restricted use of social media throughout their day for a long period of time.

The closest paper to our research is the experiment conducted by Allcott et al. (2019). They conducted a randomized controlled trial of Facebook users where subjects in the treatment group had to deactivate their Facebook account for 1 month. They find that this treatment reduced total online activity including other social media and this reduction persists after the end of the experiment. However, they use self-reported metrics of usage of online activities which are weakly correlated to objective usage metrics according to previous research. They measure 11 different metrics of subjective well-being and find that deactivating Facebook led to increase in subjective well-being for 4 out of the 11 metrics. Overall, the magnitude of the effects are small and it is not clear if these effects would have persisted for a treatment of longer duration. For a longer treatment duration, subjects could learn to live in a world without Facebook by discovering alternative substitutes providing similar use cases and their subjective well-being scores could go back to pre-experiment levels.

EXPERIMENT

Procedure

We recruited students in the faculty of economics and business of a large European university to take part in an academic study. We used a flyer to invite students in lectures and from the pool of participants of the behavioral research lab of the university. The flyer informed students about the subject of the study, the required activities, the reward, and about measures to protect the participants' privacy. Specifically, we let the students know that the study required to install a software on their computers and mobile devices that keeps track of their digital activities and that allows them to analyze how much time they are spending on various categories of activities. We also stated that the study tracks their academic performance and well-being. Moreover, we informed the students that, in order to qualify for the reward, they need to keep the software running during the time of the study and to take part in four online surveys; one at the beginning of the study and one after each quarter. In addition to getting the software for free, we offered students €20 and a one out of 100 chance to win €1,000 if they take part until the end of the study.

The sign up link forwarded interested students to a registration form that provided a more detailed privacy statement, informed consent, and asked students for their student email address, basic information about their studies (program, year), and the number and type of computers and mobile devices. The registered students were then invited to the study according to the experimental design detailed below.

Experimental Design

The recruitment of students took place in the first quarter of the academic year 2018/19. We scheduled the experiment to run for the remaining three academic quarters. We will refer to these three terms as block 1, 2, and 3 of the study (which are quarter 2, 3, and 4 of the academic year). Each block consists of seven weeks of teaching and two examination weeks. The specific timeline was:

- Block 1: from mid-November to end of January, with holidays from December 24 to January 3.
- Block 2: from February to mid-April.
- Block 3: from mid-April to end of June, with holidays from April 19 to 24.

We used the first block to establish a baseline of the students' digital activities. In block 2, we randomly assigned participants to one of two conditions: a control group without specific instructions and a treatment group that received an incentive to use social media as little as possible. Specifically, we instructed them to use Facebook, Instagram and Snapchat (the most popular social media services according to block 1) for a maximum of 10 minutes per day. We did not block these services completely because not having access to social media at all might have a negative effect on students[5], e.g., if they use it to exchange important information about their studies. The 10 minute limit enables students to still access relevant information while not allowing them to waste a longer period of time. The software would inform students in the treatment group when they reached the limit and automatically block Facebook, Instagram and Snapchat afterwards. Students could disable this feature if they needed to use these services for longer. We informed students that we gave away another €1,000 among all students who achieved to stay under the 10 minute limit throughout block 2. Block 3 served to assess post-treatment effects.

We invited students to four surveys in total. We have sent the first survey in the first week of block 1. This survey asked students to give informed consent and, after referring them to the privacy statement, their agreement to use their academic grades for the purpose of the study. Moreover, we asked them about basic demographic information, their study program, and additional work activities next to their studies. Moreover, we provided measures of subjective well-being (see specific measures below). Upon completion of this first survey, we gave students the installation and registration instructions for the tracking software and asked them to keep this software running henceforth on all their computers and mobile devices[6]. Surveys 2, 3, and 4 followed after each block and repeated the subjective well-being measures in order to track students' well-being over time. We gave students a one week deadline to fill out each survey.

---

[5] This is consistent with the Goldilocks hypothesis according to which moderate digital use may be advantageous compared to no use or overuse (Przybylski and Weinstein 2017).

[6] While the software was supported by Windows, OS X, and Android devices, it was not compatible with iOS devices (iPhone or iPad). In order to make sure that students with iOS devices complied to the 10 minute limit in the treatment condition, we informed them that we will ask them at a random time to hand in a screenshot of the Screen Time feature of iOS that reports similar information.

Survey Measures

As measures of subjective well-being we use the satisfaction with life scale (SWLS; Diener et al., 1985) that consists of five items (In most ways my life is close to my ideal; The conditions of my life are excellent; I am satisfied with my life; So far I have gotten the important things I want in life; If I could live my life over, I would change almost nothing)[7]. These items are measured on a 7-point scale (1 “strongly disagree” to 7 “strongly agree”). SWLS is the most widely used scale to measure subjective well-being and is also used in previous studies studying social media use and well-being (e.g. Kross et al. 2013, Verduyn et al. 2015).

For measuring mental well-being, we adopted the shortened Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Tennant et al., 2007; Stewart-Brown et al. 2009) with seven items (I’ve been feeling optimistic about the future, I’ve been feeling optimistic about the future, I’ve been feeling useful, I’ve been feeling relaxed, I’ve been dealing with problems well, I’ve been thinking clearly, I’ve been feeling close to other people, I’ve been able to make up my own mind about things). These items are assessed on a 5-point scale ranging from “None of the time” to “All of the time”. The SWEMWBS is a popular scale to measure mental well-being and is used in previous studies studying technology use and mental well-being (Przybylski and Weinstein 2017).

Overview of Data Sources

[7] In addition to SWLS, we also collected direct measures of happiness and life satisfaction through standard questions widely used in previous literature. Besides numerous other studies, the happiness question is used in the World values survey (Inglehart et al. 2014) and the life satisfaction question is used by Gallup (Kahneman and Deaton 2010) to calculate its well-being index. These questions are highly correlated with objective measures of well-being such as brain activity, emotional expressions and suicide rates as well as decision utility (Perez-Truglia 2019). We obtain qualitatively similar results using these happiness and life satisfaction scores as we found using SWLS.

Overall, our study makes use of three data sources: digital activities tracked by the software, self-reported measures via surveys, and academic grades from the educational administration. Table 1 shows an overview of these data types.

The software tracks users' activities on each device in 5-minute intervals and records how many seconds a user has actively used a specific program, app, or website in this interval, ranging from 1 to 300 seconds. Specifically, it records the user id, the name of the activity, the system name (Windows, Mac OS, or Android), and a timestamp. Since we are specifically interested in social media activities of the three most used social network services Facebook, Instagram and Snapchat, we used lookup tables to classify activities accordingly. For example, Facebook usage could appear in the activities as "facebook.com", "fb.com", "messenger.com", "Facebook for Android", "Facebook for Windows", etc. We gathered this list of activities in block 1 and used each of these activities to count toward the 10 minute limit for the treatment group in block 2.

The European university at which this study took place uses a grading system that ranges from 0 to 10. Any grade below 6 represents a fail. A grade of 7 is most common and often referred to as "standard", a 6 as "below standard", and an 8 or higher as "above standard". The grade for a lecture typically consists of a combined grade of the final exam and assignments that have to be completed during the course.

=== TABLE 1 ===

SAMPLE

A total of 191 respondents completed the first survey. As is typical for longitudinal studies, some students dropped out over time such that 157 students completed survey 2, 144 survey

3, and 121 the final survey. The survey participation corresponds to the number of participants who reported digital activities using the software (see Table 2).

The following results will be based on the sample that recorded activities for at least 30 days in block 1 and 2 and completed surveys 1, 2, and 3. We will analyze the post-treatment data from block 3 and survey 4 separately. From the 134 students who recorded activities in block 1 and 2, we were able to match 122 from all data sources, i.e., twelve students did not answer (one of) the surveys or did not follow courses in at least one of the blocks.

Despite the dropouts, most importantly, there are no significant differences between the treatment and the control group, in terms of gender ($p = 0.471$), age ($p = 0.961$), mobile device operating system ($p = 1.000$), number of years studying at the university ($p = 0.541$) or whether students are working next to their studies in block 1 ($p = 0.974$) or block 2 ($p = 0.594$) (see Table 3 for details). There are also no significant differences between those who started the study and those who dropped out in terms of gender ($p = 0.701$), age ($p = 0.113$), mobile operation system ($p = 0.975$), of work status in block 1 ($p = 0.109$) or block 2 ($p = 0.169$). However, there is a significant difference between these samples regarding the study year ($p = 0.027$) such that those who dropped out are more likely to be Bachelor degree students than Master's students. One potential explanation is that Bachelor degree students are more likely to quit their studies and not have any courses or grades registered ($p = 0.041$).

=== TABLE 2 ===

=== TABLE 3 ===

RESULTS

Digital Activities

Social media usage. On average, students tracked 223.7 minutes of digital activities per day across the entire study (SD = 115.1 minutes). Students who use an Android mobile device recorded significantly more activities (265.6 minutes; $p < 0.001$), compared to students with an iOS device (182.5 minutes) as iOS was not supported by the software. While our activity estimates are more accurate for Android users we expect the treatment condition to be equally effective for both of these segments because we informed participants to also inspect their iOS tracked activities (see above).

Figure 1 shows the total number of minutes tracked by day, averaged for students with Android mobile devices in the treatment (black dots) and control groups (white dots). We report the activities for users with Android devices because the tracking is more accurate as it captures activities on desktop/laptop computers and their mobile devices (figures corresponding to the overall sample are in the Appendix, Figure A-1). The solid vertical lines separate blocks 1, 2, and 3 and the dashed vertical lines indicate the start of the examination period. Overall, digital activities remain on a high level each day but are reduced during the winter holiday season and during the examination periods.

As a manipulation check, the bottom part of Figure 1 shows activities for social networking (Facebook, Instagram, and Snapchat combined) for the Android sample. The mean daily usage in minutes is 21.1 minutes (27.9 minutes) for users in the treatment (control) group in block 1, which is not significantly different ($p = 0.310$). The incentive to reduce social media activities was effective as students in the treatment condition significantly reduced their social media usage in block 2 compared to the control group ($p = 0.009$). The horizontal line represents the 10-minute limit imposed on the treatment group. The average usage per day is close to the limit in the treatment group with 8.1 minutes. Within the control group, the average daily usage of 24.2 minutes in block 2 is on the same level as in the first block (p = 0.245).

=== FIGURE 1 ===

Remarkably, although students in the treatment group significantly reduced their social media activities, their overall digital activities overall are not affected but, in fact, exceed those of the control group in block 2 (p = 0.026). This result indicates that students substituted or even overcompensated their social media usage with other activities.

Substitution. Figures 2.1 and 2.2 show the time series of activities of users with an Android device for the most used categories of services (we exclude the categories of general utilities, which holds mostly operating system activities, and uncategorized services for which there are no significant differences between the groups; activities for all users are in the Appendix, Figures A-2.1 and A-2.2). We find significant substitution for social networking with instant messaging (p = 0.008). Accordingly, more students in the treatment condition used instant messaging in block 2 when social media was restricted compared to block 1 and to the control group (difference-in-differences). The activities increased from an average daily use of 25.1 minutes in block 1 to 28.8 minutes in block 2 in the treatment group, while the usage decreased from 21.8 minutes to 15.2 minutes in the control group. Most activities (92.9%) in this category are related to WhatsApp.

We also find a significant increase in usage of music for Android users (p = 0.027) in the treatment group in block 2. However, average daily activities in this category are rather low (below five minutes) and the difference is mostly driven by two outliers who listen to music for more than 30 minutes each day on average. Other activities show plausible patterns, e.g., the reference and learning category (activities include the university intranet, PDF reader, Wikipedia, Mendeley, Google scholar, EBSCO, etc.) shows peaks before the exam period. However, these and other activities are not affected by reduced social media usage (an overview of significance tests comparing treatment and control groups in block 2 vs. block 1 is given in the Appendix, Table A-1).

=== FIGURE 2.1 ===

=== FIGURE 2.2 ===

Subjective Well-being

For the subjective well-being measures, the SWLS and SWEMWBS scores are calculated as the sum of their items (with SWEMWBS being transformed according to a defined conversion table). The students score averages (SD) in SWLS of 25.0 (5.5), 25.0, (5.4), 25.1 (5.3) in the three surveys at the beginning of the study and after block 1 and 2. That means they are between an “average” and “high” score of satisfaction. Treatment and control group are not significantly different at the beginning of block 1 ($p = 0.182$; survey 1), at the end of block 1 ($p = 0.212$; survey 2), or, most importantly, at the end of block 2 after the exposure to the treatment ($p = 0.167$; survey 3). The same implications hold for the SWEMWBS scores that shows average scores of 22.8, 22.5, and 22.4 in the three surveys (see Table A-2 in the Appendix for details). Figure 3 plots the differences between survey 3 and survey 2 (before and after the social media restriction) in terms of SWLS and SWEMBS. The distributions are centered on zero, illustrating the non-significant difference between the treatment and control group.

=== FIGURE 3 ===

Table 4 shows correlations with the subjective well-being measures and the digital activities in block 1 (i.e., activities that are not affected by the treatment condition) for users with an Android device. We detail the correlations of the subjective well-being measures at the beginning of the first block (in survey 1) and at the end of the block (in survey 2) to study potential reverse causality, e.g., increased well-being leads so more/less social media activities or vice versa.

Satisfaction with life and mental well-being are positively correlated and are also significant predictors over time, i.e., subjective well-being in survey 1 is positively correlated with the same measure in survey 2. Regarding digital activities, we see, on average, negative correlations between subjective well-being and all digital activities, albeit not being significant. Similarly, we do not find significant correlations with social media use. (We will address causality in the section below.) To address potential non-linear effects we also report correlations with categories of social media usage. Specifically, we used dummy variables relating to low usage with an average of less than 2 minutes per day (36.1% of users), medium usage of 2 to 20 minutes (39.5%), and high usage of 20 minutes or more (24.4%). A non-linear relationship is likely as low usage generally shows the most negative subjective well-being scores, while medium usage and not high usage scores the highest well-being. However, we cannot rule out reverse causality regarding these findings as the only significant relationships are between satisfaction with life measured in survey 1 and the social media activities measured after the survey has taken place.

Activities related to communication, i.e., instant messaging and email, show significant positive correlations in the second surveys (for instant messaging regarding mental well-being and for email in terms of satisfaction with life). A consistent significant negative correlation can be observed for activities in the video category and satisfaction with life (both surveys) and mental well-being (survey 1). This suggests that less satisfied students and those with lower mental well-being at the beginning of the block increasingly watch videos in the subsequent block.

=== TABLE 4 ===

Academic Grades

The students participating in our study scored an average (SD) of 7.105 (1.078) in block 1 and 7.122 (0.954) in block 2. Most grades can be classified as "standard". The average (SD) sum of credit points per block is 13.571 (5.754) in block 1 and 12.353 (5.118) in block 2. Differences between the treatment and control group are not significant for grades ($p = 0.113$) but for the number of credit points such that the treatment group attempted to score significantly more credits ($p = 0.035$). This is visualized in Figure 4 as the difference in grades and credits in block 1 and 2. Note that the number of ECTS represents the courses that the student *attempted* to pass but they are also stored if the student failed the exam. A comparison of the number of successfully passed courses shows no significant differences between the groups ($p = 0.383$).

=== FIGURE 4 ===

Table 5 shows the correlations of academic performance in block 1 with the subjective well-being measures and digital activities. Accordingly, grades in block 1 are positively related to grades in block 2. The grades are not significantly correlated with the number of credit points, possibly due to a trade-off regarding a good grade and completing more courses. Credits are also not positively related over time, which is reasonable as more credits in one term means that the students have to obtain fewer credits in subsequent terms. The average grade is positively and significantly correlated with satisfaction with life measures. This holds for SWLS measures at the beginning[8] and end of the block.

Regarding correlations with digital activities we can observe significant positive effects on the average grade for writing and presentation activities, which are required to complete assignments (that are part of the grade for the majority of courses). We also see a positive effect of instant messaging on the grade, however, a negative effect on the number of credit points. Social media usage is not significantly correlated with the academic performance in block 1. Only when categorizing students based on their social media usage we see significant effects such that medium usage is negatively correlated with the number of credits.

To what extent these findings can be interpreted as causal evidence will be addressed in the following section by analyzing the complete randomized control trial across the two blocks using difference-in-differences analyses with regression models.

=== TABLE 5 ===

Regressions

Table 6 shows the results of a regression of the average grade and number of credit points. We use data from the two teaching blocks with "block 2" being a dummy variable indicating the block in which the treatment took place. Similarly, the "treatment group" refers to a dummy variable that identifies students that were exposed to the treatment. We use gender, age, number of years at the university, and whether the student is working next to the studies (dummy variable) as control variables.

[8] We have also obtained grades for the first quarter (i.e., block 0). These grades are also significantly correlated with SWLS measures in Survey 1 such that a reasonable explanation would be that academic grades positively affected subjective well-being measured in the subsequent survey.

There is no evidence that the treatment group achieves higher grades in block 2 (Treatment group * Block2 interaction), in which social media usage was restricted, than the control group ($p = 0.239$). This also holds for other subsets of the sample such as students with an Android device. Overall, the amount of variance explained in the grades remains very low with 3.4%. Only the number of years that the student has spent at the university is a significant positive predictor, i.e., Master's students achieve higher grades than Bachelor's.

In terms of number of credit points, i.e., number of courses attended, we do see a significant effect of restricting social media. While students overall attended courses with fewer credits in block 2 ($p = 0.004$) this is not the case for students in the treatment group such that they targeted significantly more credits ($p = 0.023$). With this dependent variable, the number of years at the university has a significant negative effect ($p = 0.002$) and, overall, 11.7% of variance in the number of credit points can be explained. The subset of Android users replicates these results, albeit generally with lower levels of significance due to the smaller sample size.

However, as noted above, the number of credit points does not necessarily show that students successfully passed more courses as also failed courses are included. Using the number of courses passed as the dependent variable shows that the students in the treatment group in fact do not differ from the control group (a Poisson model replicates these results). Thus, it appears that the treatment group attempted to pass more courses or courses with more credits compared to the control group but did not necessarily succeed.

Regarding subjective well-being measures SWLS and SWEMWBS, we do not see any significant effects due to using Facebook, Instagram, and Snapchat less (see Appendix Table A-3).

=== TABLE 6 ===

These null results of the effects of restricting social media usage on academic performance and subjective well-being raises the question of whether our study is underpowered to detect economically significant effects. The maximum difference in life satisfaction (on the SWLS scale) that our sample cannot detect is 3 on a scale of 5-35 (average life satisfaction score in our sample is 25 with a standard deviation of 5). A score of 25-29 is considered as a "high score", therefore even if the treatment group's life satisfaction score increased by 3 it is not sufficient to change the classification of the score from "high" to "very high". The maximum difference in average grade that our sample cannot detect is 0.7 on a scale of 1-10 (average grade in our sample is about 7 with a standard deviation of 1). Given that students receive grades which are whole numbers, this threshold is still below the value that would increase the treatment group's average grade by a full point.

To further address potential concerns about statistical power we applied a hierarchical Bayes ANOVA (BANOVA) model that includes between and within subject effects and accommodates unobserved heterogeneity by including a normal distribution of the parameters across individuals (Wedel and Dong 2019). All models converge and generally replicate the results above (details are available from the authors upon request).

Post-treatment Effects

The formal analysis of the post-treatment effects is based on a sample of 106 students who provided activity data throughout all three blocks and in all four surveys. While the treatment condition significantly reduced their social media usage in block 2 compared to the control group, this effect was not permanent. After we suspended the limit in block 3 the social media activities of users in the treatment group increased again, showing no significant differences to the control group any longer ($p = 0.668$). We further do not see any significant differences between the treatment and control group in block 3 in terms of grades (p = 0.152), number of credit points (p = 0.923), satisfaction with life (p = 0.499), or mental well-being (p = 0.966), i.e., there is no lagged effect of reduced social media usage.

DISCUSSION AND CONCLUSION

In this paper, we analyzed the effects of restricting social media usage. We did not find significant causal effects of social media usage on academic performance, other than students attempting to pass more courses or courses with more credits. However, we found robust evidence of substitution effects that can potentially explain the null finding. Specifically, we showed that social media and instant messaging apps can be substitutes. The European Commission approved Facebook's approval of WhatsApp in 2014 based on Facebook's claim that it operates in a different market and does not compete directly with WhatsApp[9]. Our results indicate that they are in fact direct competitors. After acquiring WhatsApp, Facebook started automatically matching its users' profiles with their WhatsApp accounts[10] and in the near future plans to integrate WhatsApp, Instagram and Facebook user accounts[11]. Antitrust authorities should consider the market power of this combined entity if the world's biggest social media platforms are integrated with the world's biggest instant messaging platform.

While we found null results estimating the causal impact of social media usage on well-being and academic performance, and not all null results matter, we believe that null results are

[9] Source: European Commission (http://europa.eu/rapid/press-release_IP-14-1088_en.htm, accessed on June 1, 2019)

[10] The European Commission fined Facebook €110 million in 2017 for this practice because Facebook had provided misleading information about the feasibility of automatically matching profiles during its acquisition of WhatsApp. However while announcing this fine, the Commission still maintained its belief that Facebook and WhatsApp do not directly compete with each other (http://europa.eu/rapid/press-release_IP-17-1369_en.htm, accessed on June 1, 2019).

[11] Source: The New York Times (https://www.nytimes.com/2019/01/25/technology/facebook-instagram-whatsapp-messenger.html, accessed on June 1, 2019)

interesting and important in this context. The media has hyped correlational studies showing a negative association between social media usage and well-being and it is important to balance this narrative through causal evidence.

Moreover, it is interesting to notice that while social media generates large amount of consumer surplus (Brynjolfsson, Collis and Eggers 2019), it doesn't seem to affect the subjective well-being of users. Future research can explore this wedge between consumer surplus and subjective well-being and see whether they are correlated for some products and uncorrelated for others. Future research should also explore the addictiveness of social media in more detail. Our findings in block 3 show that the students in the treatment condition go back to their old habits and do not adopt a lower social media usage that they experienced in block 2. On the other hand, showing students how much time they are spending on social networks via the software seems to have an overall negative trend on its usage (comparing usage in block 1 and block 3). Curing social media addiction (if it is indeed addiction) might therefore be a longer process.

A limitation of our study is the lack of a larger sample size to detect smaller effects. While these small effects might not be economically significant, more research is needed using massive samples. Moreover, due to our student sample implications for the general population are limited. It could be that students use social media mostly for communication purposes and therefore show significant substitution effects with instant messaging. We might see different effects for users who visit social media for entertainment, e.g., watching videos. However, it is challenging to recruit a large number of subjects from a respresentative sample for a long term study. Direct collaborations with social media platforms or internet service providers (which control internet traffic) could be a way of obtaining data from larger samples. Moreover, while we only study the impact of social media on students and academic performance, future research can look at workplace settings and study the impact of social media and its substitutes on worker productivity and well-being. We believe that rigorous causal evidence through randomized controlled trials and objectively measured time spent is the way forward in addressing questions regarding the impact of technology on well-being. The widespread adoption of most major technologies in the past such as radio, television, video games and computers was followed with unfounded fears about their impact on well-being. This story repeats again with social media. We find that social media usage does not cause lower well-being or poor academic performance. Rather, we demonstrate that students find other means of social networking using instant messaging when exogenously restricting their social media usage. To conclude: You can take social networking away from the students, but you cannot take students away from their social network.

REFERENCES

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2019). The Welfare Effects of Social Media (No. w25514). National Bureau of Economic Research.

Beland, L. P., & Murphy, R. (2016). Ill communication: technology, distraction & student performance. *Labour Economics*, 41, 61-76.

Belo, R., Ferreira, P., & Telang, R. (2013). Broadband in school: Impact on student performance. *Management Science*, 60(2), 265-282.

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

Burke, M., & Kraut, R. E. (2016). The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of computer-mediated communication*, 21(4), 265-281.

Burke, M., Marlow, C., & Lento, T. (2010, April). Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1909-1912). ACM.

Carter, S. P., Greenberg, K., & Walker, M. S. (2017). The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review*, 56, 118-132.

Cheng, J., Burke, M., & Davis, E. G. (2019, April). Understanding Perceptions of Problematic Facebook Use: When People Experience Negative Life Impact and a Lack of Control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 199). ACM.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49, 71-75.

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*.

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior?. *International Journal of Human-Computer Studies*.

Haidt, J. (2019). Social Media Use and Mental Health: A Review. Available at: https://docs.google.com/document/d/1w-HOfseF2wF9YIpXwUUtP65-olnkPyWcgF5BiAtBEy0/edit (accessed on May 11, 2019).

Hobbs, W. R., Burke, M., Christakis, N. A., & Fowler, J. H. (2016). Online social integration is associated with reduced mortality risk. *Proceedings of the National Academy of Sciences*, 113(46), 12980-12984.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: All Rounds - Country-Pooled Datafile 1981-2014. Madrid: JD Systems Institute.

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, 29(3), 626-631.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the national academy of sciences*, 107(38), 16489-16493.

Kraut, R., & Burke, M. (2015). Internet use and psychological well-being: Effects of activity and audience. *Communications of the ACM*, 58(12), 94-100.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., ... & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PloS one*, 8(8), e69841.

Marotta, V., & Acquisti, A. (2018). Interrupting Interruptions: A Digital Experiment on Social Media and Performance. Available at SSRN 3283951.

Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences*, 116(21), 10226-10228.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, 3(2), 173.

Perez-Truglia, R. (2019). The effects of income transparency on well-being: Evidence from a natural experiment (No. w25622). National Bureau of Economic Research.

Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the Goldilocks Hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. *Psychological Science*, 28(2), 204-215.

Scharkow, M. (2016). The accuracy of self-reported Internet use—A validation study using client log data. *Communication Methods and Measures*, 10(1), 13-27.

Shakya, H. B., & Christakis, N. A. (2017). Association of Facebook use with compromised well-being: A longitudinal study. *American journal of epidemiology*, 185(3), 203-211.

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a Rasch analysis using data from the Scottish health education population survey. *Health and quality of life outcomes*, 7(1), 15.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of life Outcomes*, 5(1), 63.

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., ... & Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*, 144(2), 480.

Wedel, M., Dong, C. (2019). BANOVA: Bayesian Analysis of Experiments in Consumer Psychology. *Journal of Consumer Psychology*, 1– 21.

TABLES

Table 1: Overview of data sources

Type	**Measure**	**Source**	**Data collection**
Digital activities	Usage in number of seconds	Tracked by software	On each participant’s device throughout the entire study
Subjective well-being	Rating scales	Self-reported in surveys	At the beginning of the study and after each teaching block
Academic grades	From 0 to 10, with <6 = failed, 6 = below standard, 7 = standard, >8 above standard;	Educational administration	Once at the end of the academic year

Table 2: Number of participants

Part	Number
Completed survey 1	191
Used software in block 1 (calibration)	149
Completed survey 2	157
Used software in block 2 (treatment)	134
Completed survey 3	144
Used software in block 3 (post-treatment)	125
Completed survey 4	121
Took courses in block 1, 2, and 3	158

Table 3: Descriptive statistics of the sample

	Treatment	Control
Number of students	60	62
Gender: Female (vs. male)	0.467	0.548
Age (SD)	22.1 (3.3)	22.1 (3.1)
Mobile device operating system: Android (vs. iOS)	0.500	0.500
Studying in first to third year (vs. more than three years)	0.667	0.629
Working next to studying (block 1)	0.400	0.419
Working next to studying (block 2)	0.500	0.435

Table 4: Correlations of subjective well-being measures and digital activities in block 1 (Android users)

	SWLS Survey 1 (start of block 1)	SWLS Survey 2 (end of block 1)	SWEMWBS Survey 1 (start of block 1)	SWEMWBS Survey 2 (end of block 1)
SWLS Survey 2 (end of block 1)	**0.80**			
SWEMWBS Survey 1 (beginning of block 1)	**0.76**	**0.63**		
SWEMWBS Survey 2 (end of block 1)	**0.63**	**0.72**	**0.77**	
All digital activities	-0.06	-0.11	-0.15	-0.05
Social media usage	-0.03	-0.02	-0.13	-0.02
Social media usage low	**-0.31**	-0.23	-0.08	-0.13
Social media usage medium	**0.26**	0.19	0.14	0.15
Social media usage high	-0.02	-0.01	-0.08	-0.05
General Reference & Learning	-0.13	-0.14	-0.20	-0.06
Instant Message	0.18	0.24	0.14	**0.27**
Browsers	-0.09	-0.13	-0.01	-0.11
Video	**-0.28**	**-0.27**	**-0.28**	-0.21
Writing	0.24	**0.27**	0.10	0.09
Search	-0.19	-0.19	-0.22	-0.10
Email	0.21	**0.28**	0.11	0.10
General News & Opinion	0.19	0.04	0.14	0.13
Games	0.06	-0.17	0.04	-0.04
Presentation	0.14	0.10	-0.01	0.03
General Shopping	-0.05	0.02	-0.12	0.02
Music	0.12	0.10	0.12	0.24

(Correlations in bold font are significant on a 5% level)

Table 5: Correlations of academic performance with measures of subjective well-being and digital activities in block 1 (Android users)

	Average grade in block 1	Sum of credit points in block 1
Sum of credit points in block 1	-0.16	
Average grade in block 2	**0.46**	-0.19
Sum of credit points in block 2	0.06	0.01
SWLS Survey 1 (beginning of block 1)	**0.27**	-0.10
SWLS Survey 2 (end of block 1)	**0.31**	0.02
SWEMWBS Survey 1 (beginning of block 1)	0.16	-0.07
SWEMWBS Survey 2 (end of block 1)	0.22	0.01
All digital activities	0.16	-0.16
Social media usage	0.02	-0.03
Social media usage low	-0.08	0.21
Social media usage medium	0.06	**-0.26**
Social media usage high	0.00	0.09
General Reference & Learning	0.13	-0.21
Instant Message	**0.28**	**-0.28**
Browsers	-0.03	-0.18
Video	-0.04	0.15
Writing	**0.25**	0.05
Search	-0.13	-0.21
Email	0.12	-0.21
General News & Opinion	-0.07	-0.10
Games	-0.06	-0.08
Presentation	**0.28**	0.17
General Shopping	-0.02	-0.01
Music	-0.08	-0.13

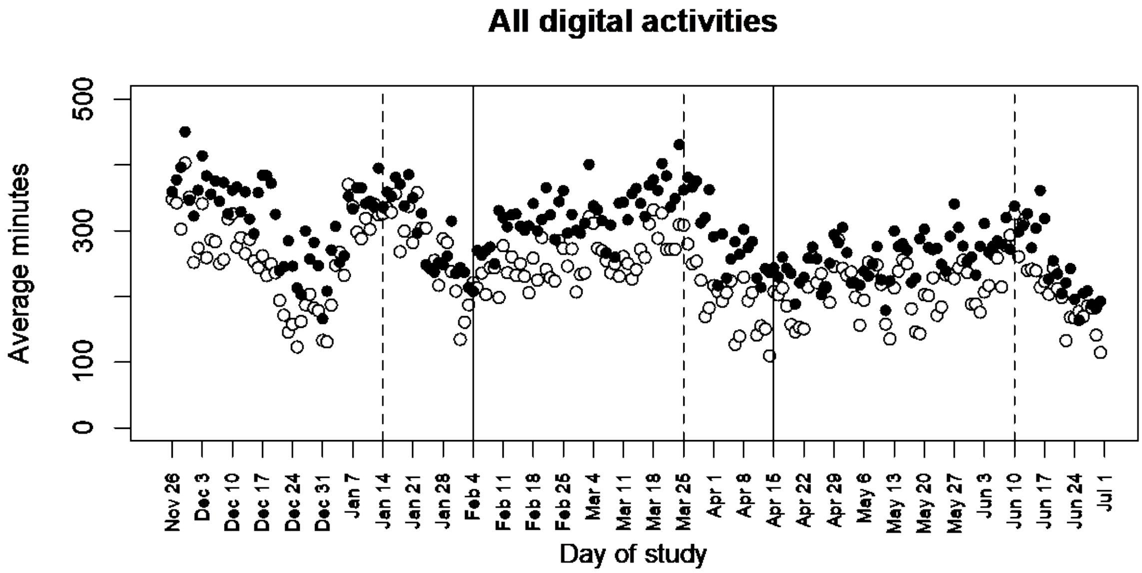
(Correlations in bold font are significant on a 5% level)

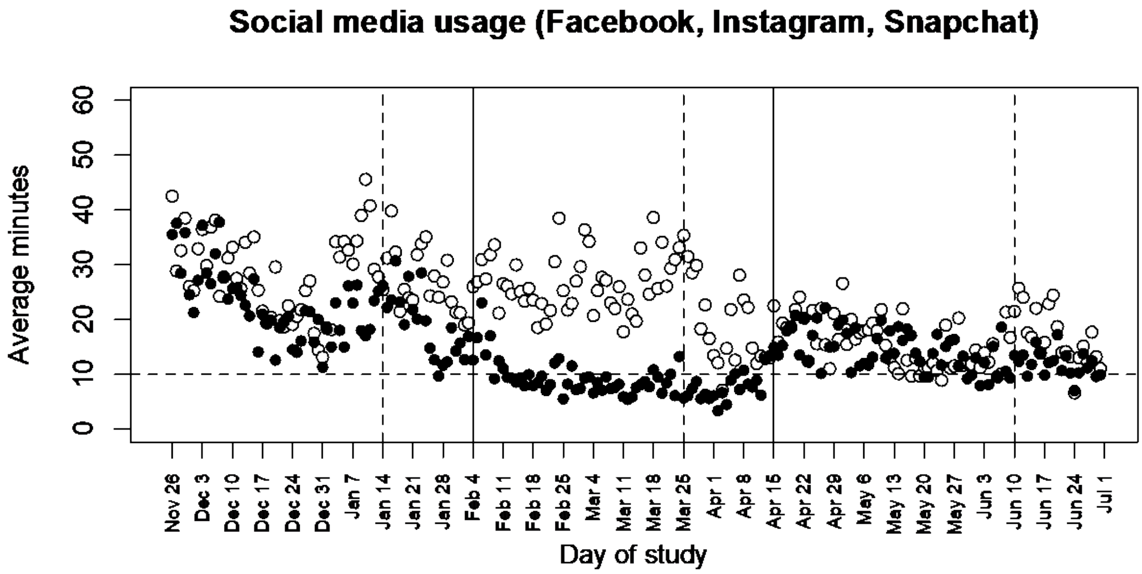
Table 6: Regression of academic performance

	Regression of average grade				Regression of number of credit points				Regression of number of courses passed			
	All users		Android users		All users		Android users		All users		Android users	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
(Intercept)	**6.950**	**<0.001**	**6.578**	**<0.001**	**19.406**	**<0.001**	**13.393**	**0.009**	**3.904**	**<0.001**	**2.616**	**0.007**
Treatment group	0.237	0.205	0.536	0.067	-1.740	0.071	-1.159	0.388	-0.064	0.721	0.124	0.625
Block 2	0.173	0.350	0.321	0.272	**-2.750**	**0.004**	**-2.931**	**0.031**	-0.347	0.051	-0.345	0.178
(Treatment group*Block2)	-0.310	0.239	-0.592	0.149	**3.090**	**0.023**	3.431	0.070	0.220	0.382	0.211	0.554
Gender (female)	0.142	0.281	0.137	0.507	0.706	0.300	0.388	0.684	0.201	0.113	0.090	0.618
Age in years	-0.016	0.534	0.002	0.969	-0.130	0.311	0.164	0.483	**-0.055**	**0.022**	0.003	0.949
Years at the university	**0.114**	**0.028**	0.083	0.294	**-0.854**	**0.002**	**-1.232**	**0.001**	**-0.109**	**0.028**	-0.131	0.060
Working next to studies	-0.043	0.750	-0.202	0.358	-0.009	0.990	1.439	0.156	-0.082	0.523	0.134	0.484
R-squared	0.034		0.054		0.117		0.159		0.113		0.083	

FIGURES

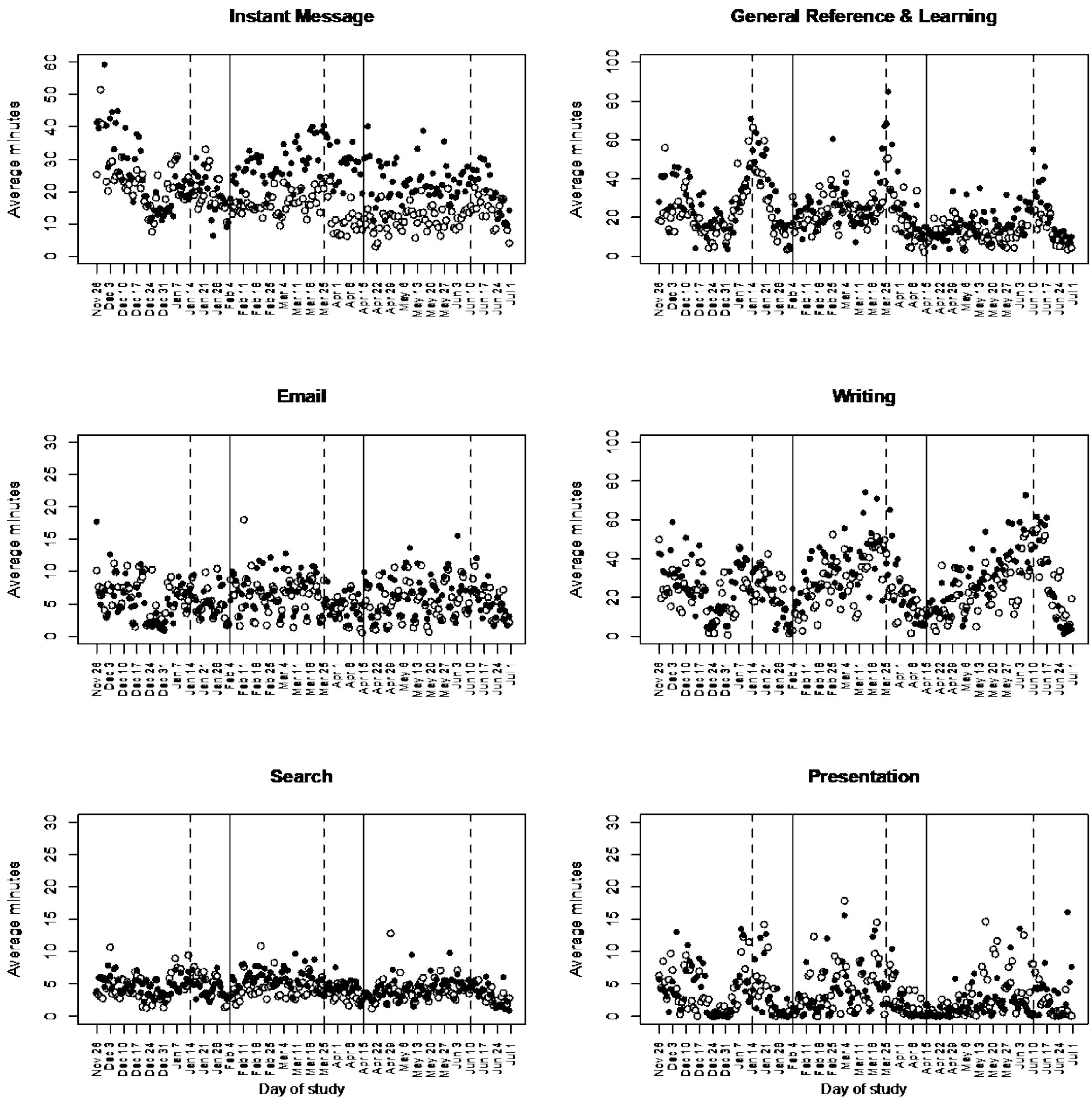
Figure 1: All digital activities and social media usage over time (users with Android devices)





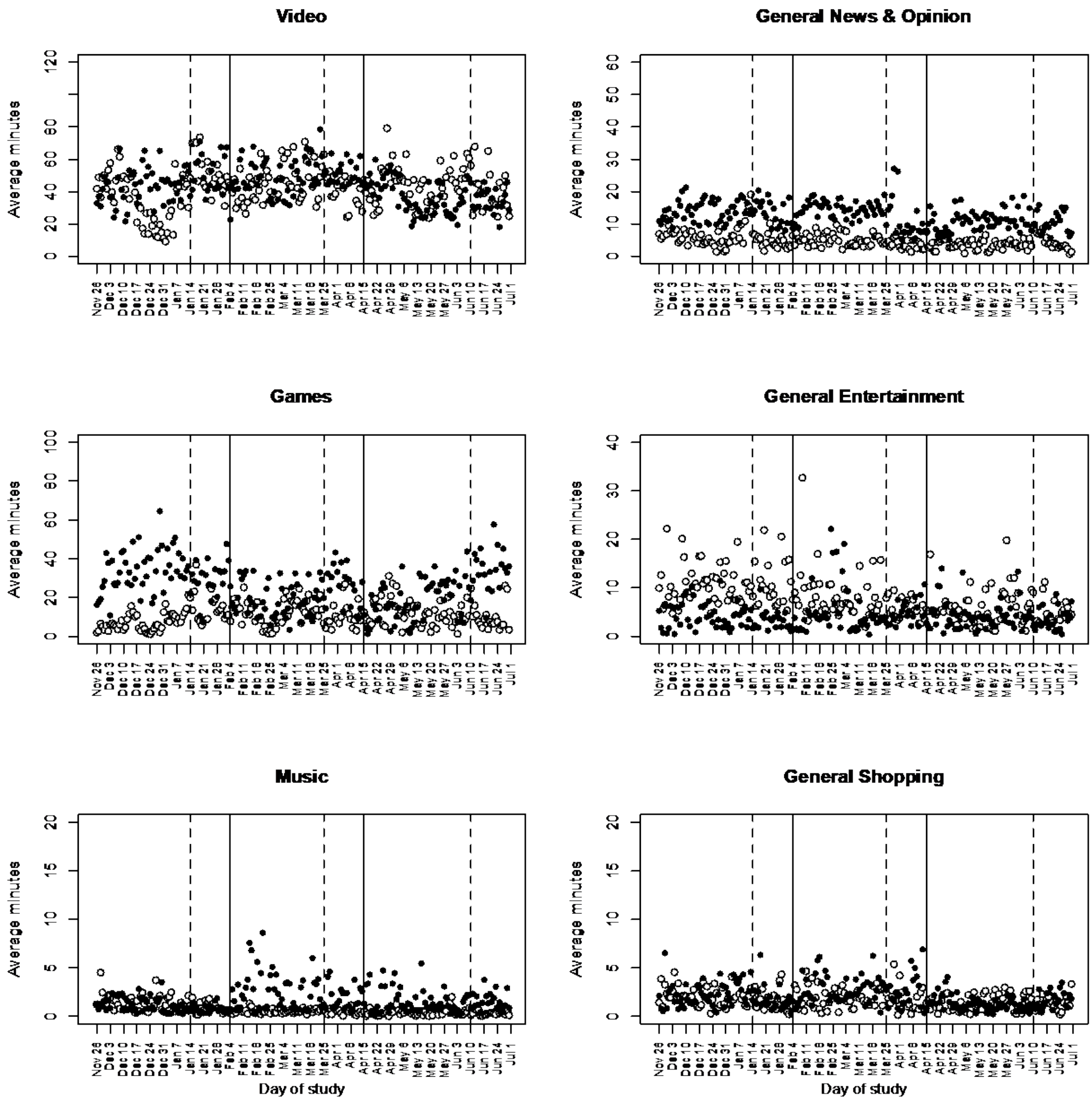
black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Figure 2.1: Tracked digital activities over time (users with Android devices)



black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Figure 2.2: Tracked digital activities over time (users with Android devices)



black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Figure 3: Differences in subjective well-being measures

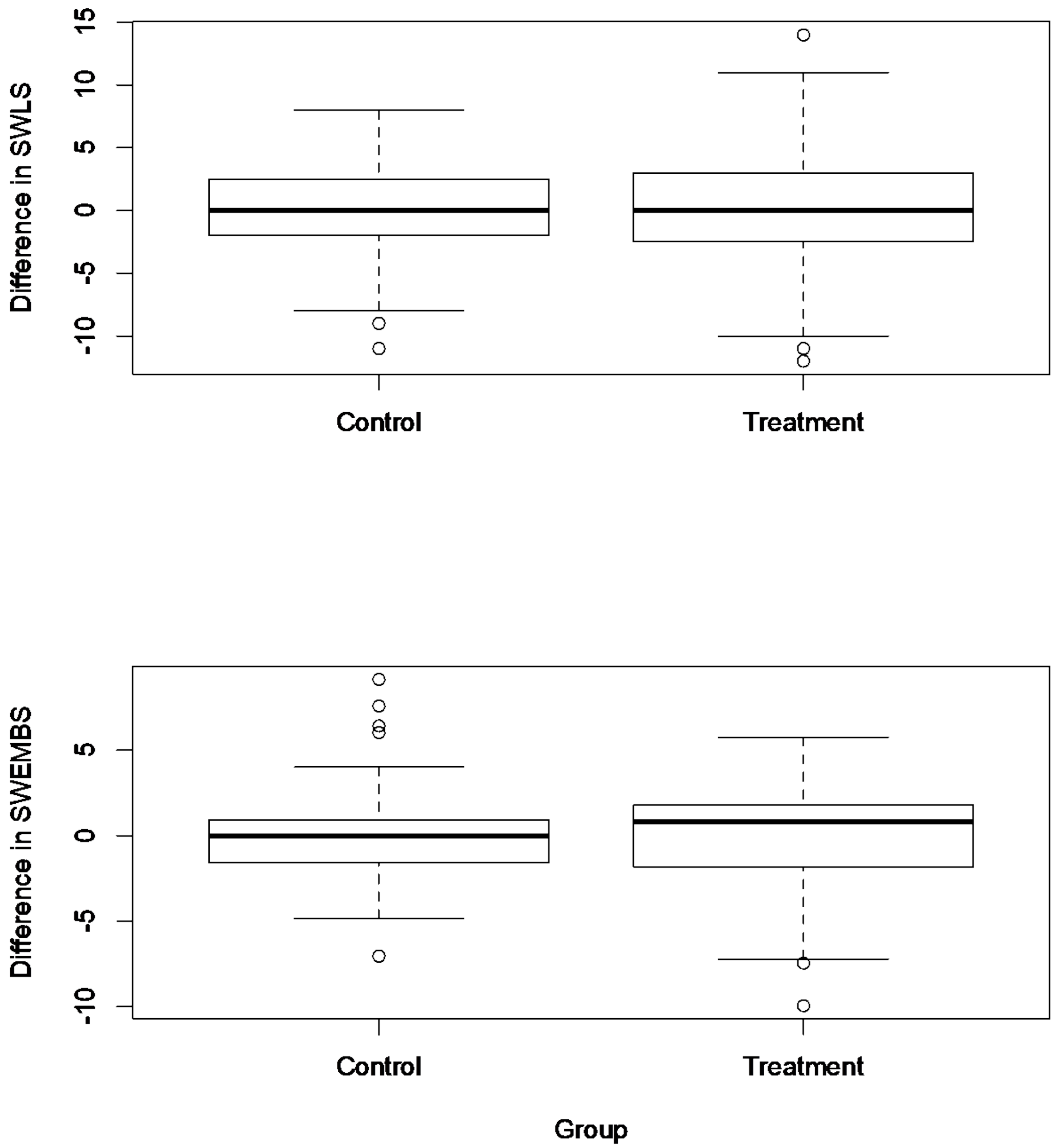
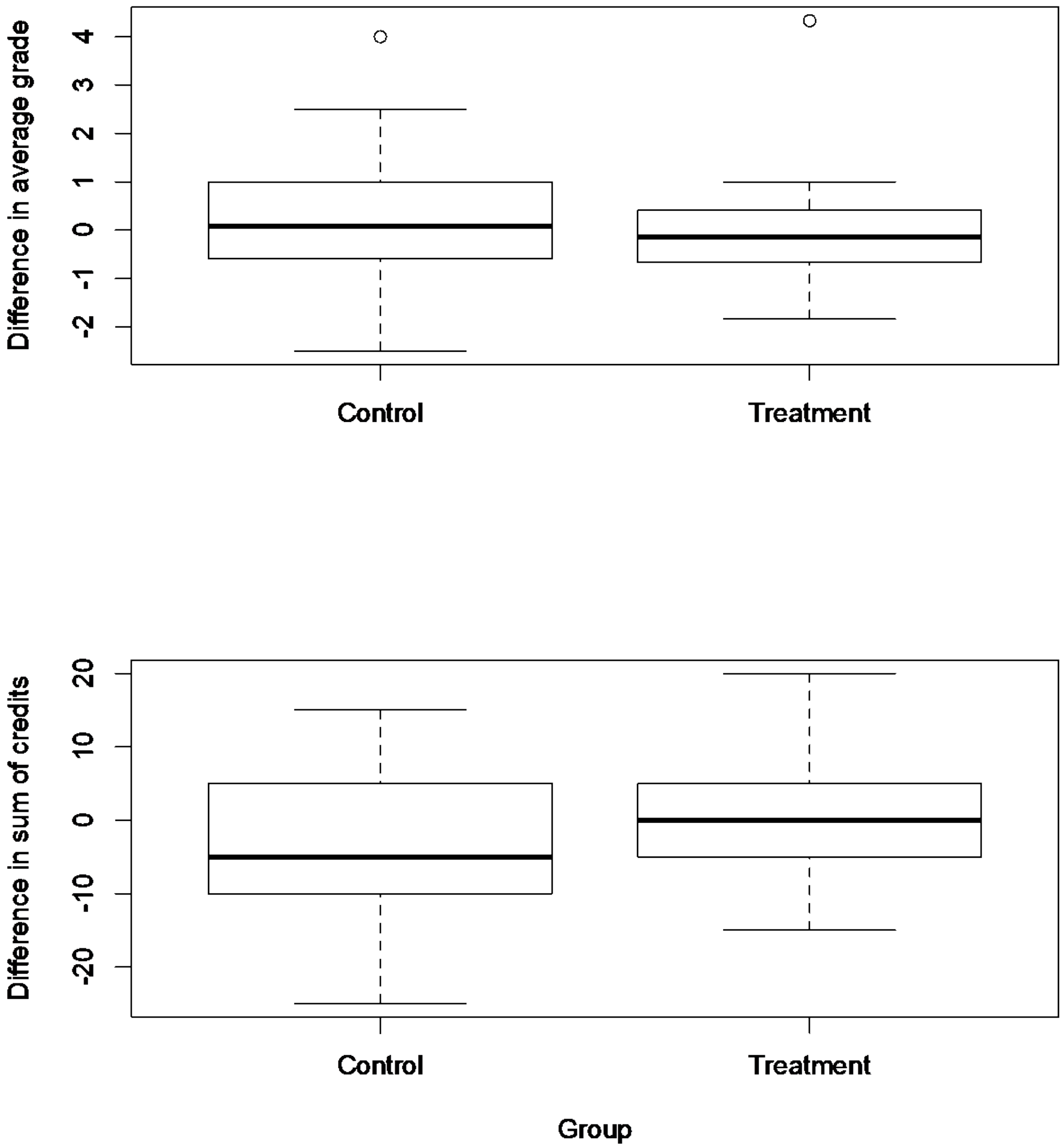
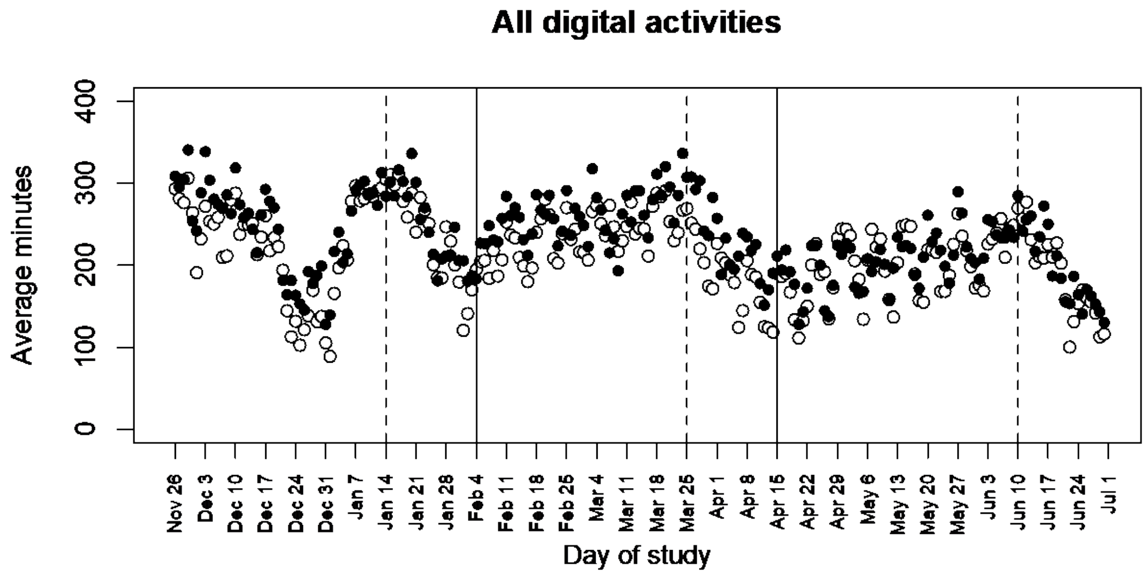


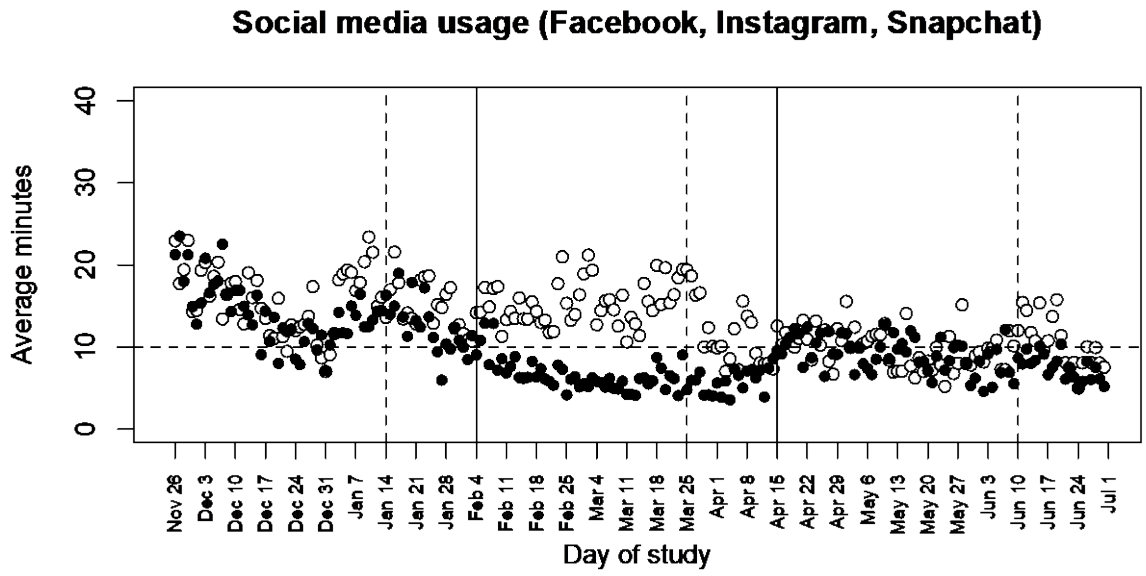
Figure 4: Differences in academic performance



APPENDIX

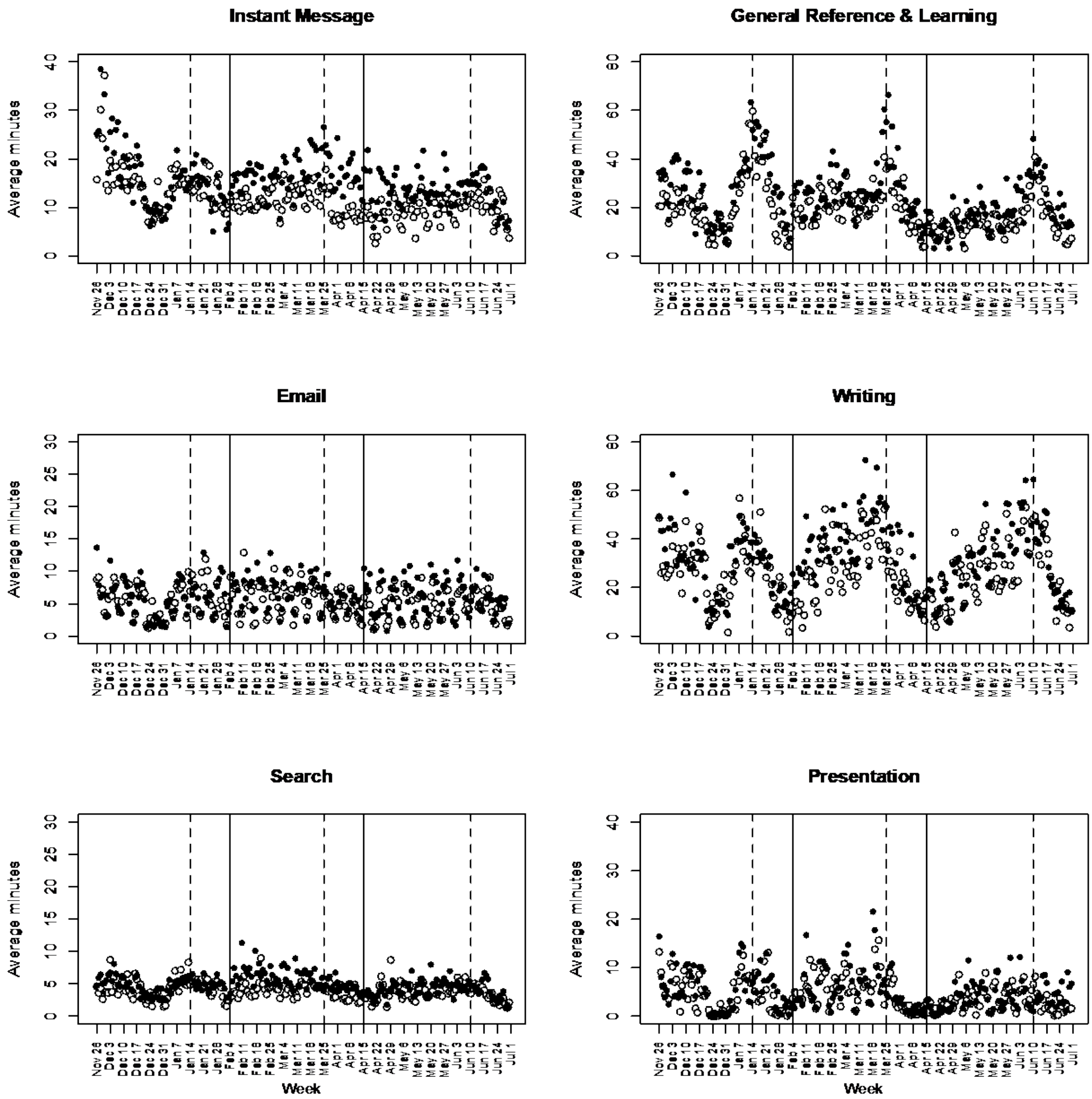
Figure A-1: All digital activities and social media usage over time (all users)





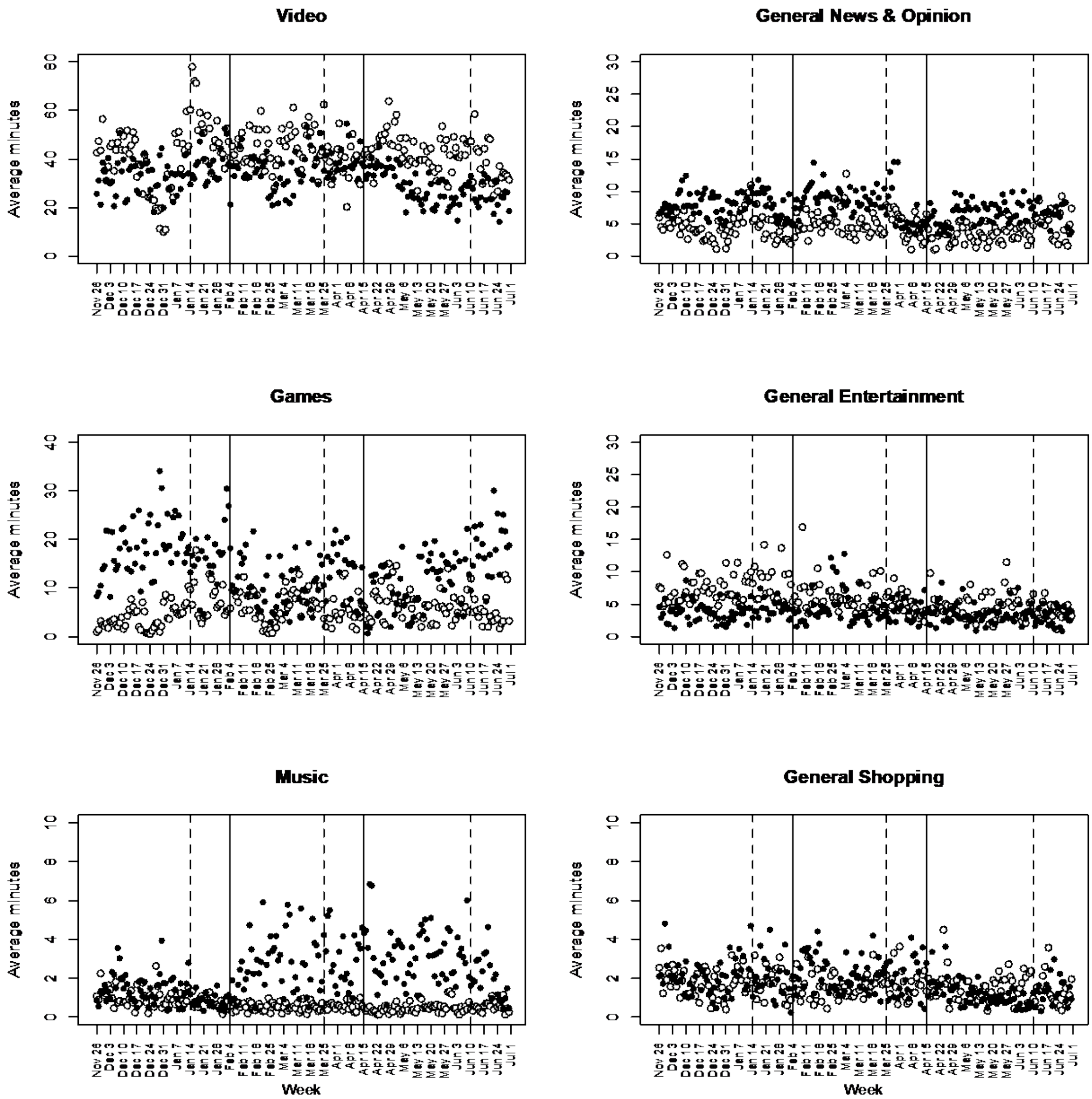
black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Figure A-2.1: Tracked digital activities over time (all users)



black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Figure A-2.2: Tracked digital activities over time (all users)



black = treatment, white = control group
solid vertical lines: start of a new teaching block
dashed vertical lines: start of the exam period

Table A-1: Significance of differences between treatment and control group in block 2 vs. block 1 (difference-in-differences)

Category	p-value all users	p-value Android users
Social media	**<0.001**	**0.013**
General Reference & Learning	0.811	0.491
Instant Message	0.123	**0.008**
Browsers	0.628	0.257
Video	0.513	0.715
Writing	0.532	0.304
Search	0.745	0.494
Email	0.352	0.104
News & Opinion	0.864	0.543
General Entertainment	0.679	0.506
Games	0.411	0.662
Presentation	0.857	0.803
General Shopping	0.833	0.611
Music	0.099	**0.027**

All variables measured on a log scale.

Table A-2: Summary statistics for well-being measures

Measure	Min	Mean	Max
SWLS survey 1	9.0	25.0	35.0
SWLS survey 2	11.0	25.0	34.0
SWLS survey 3	9.0	25.1	35.0
SWEMWBS survey 1	15.3	22.8	30.7
SWEMWBS survey 2	14.1	22.5	30.7
SWEMWBS survey 3	12.4	22.4	35.0

Table A-3: Regression of satisfaction with life and mental well-being measures

	Satisfaction with life (SWLS)				Mental well-being (SWEMWBS)			
	All users		Android		All users		Android	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
(Intercept)	**29.028**	**<0.001**	**36.967**	**<0.001**	**24.409**	**<0.001**	**30.497**	**<0.001**
Treatment group	-1.205	0.221	-1.350	0.380	-0.233	0.676	-0.063	0.942
Block 2	0.207	0.831	-0.552	0.720	0.025	0.963	-0.342	0.691
(Treatment group*Block2)	-0.253	0.855	1.285	0.552	-0.126	0.872	0.505	0.676
Gender (female)	-1.021	0.143	-0.818	0.455	**-0.991**	**0.013**	-0.768	0.210
Age in years	-0.161	0.221	-0.494	0.068	-0.057	0.446	**-0.338**	**0.026**
Years at the university	0.166	0.542	0.107	0.798	-0.139	0.370	-0.089	0.705
Working next to studies	0.280	0.692	-0.036	0.975	0.724	0.072	1.020	0.117
R-squared	0.032		0.051		0.055		0.142	

Chapter 5 - How Should We Measure the Digital Economy?

Suppose we make you an offer. Would you give up access to Google search for one month if we paid you $10? How about $100? $1,000? What do we need to pay you to lose access to Wikipedia? Your answer can help us understand the value of the digital economy.

We've seen an explosion of digital goods and services over the past 40 years: Not just Google and Wikipedia, but social networks, online courses, maps, messaging, videoconferencing, music, and all the other apps on your smartphone. Americans spent an average of 6.3 hours a day on digital media in 2018, a large and growing share of our waking lives.

Are we getting any value from these goods? They largely go uncounted in official measures of economic activity such as GDP and productivity (which is simply GDP per hour worked). In fact, while see more photos, listen to more and better music, and have myriad other benefits that we couldn't imagine 40 years ago, if you only had access to the GDP numbers you'd think that the digital revolution never happened. The contribution of the Information sector as a fraction of total GDP has barely changed since the 1980s, hovering between 4-5% during most of those years and reaching a high of only 5.5% at the end of 2018. To paraphrase Robert Solow, we see the digital age everywhere except in the GDP statistics.

The reason that the value of digital goods is largely missing from GDP is that the measure is based on what people *pay* for goods and services. With few exceptions, if something has a price of zero, then it has zero weight in GDP. But, of course, even free goods can create value. In fact, most of us get *more* value from zero-price goods like Wikipedia and digital maps than we did from their more expensive paper versions.

If we want to understand how the internet is contributing to our economy, we need better ways to measure free goods and services. With our current measurement tools, the benefits of digitization are being dramatically underestimated, and as a result policymakers are likely to make mistakes when they decide how to invest in everything from infrastructure and R&D to education and cyberdefense. When it comes to regulating technology firms, competition authorities and antitrust regulators might decide on the wrong course of action if, when weighing the effects of regulation, they look only at prices and not benefits. In short, the way we measure the economy matters. And while GDP has always undercounted the benefits of new innovations, that's no reason not to improve upon it. Our ability to manage the growing digital economy depends on doing so.

That's why we developed a set of new techniques to estimate the contribution of digital goods to the economy. Our research with Felix Eggers confirms that the economic benefits of Internet search, online encyclopedias, social networks, digital maps, and other internet and mobile services are enormous. Facebook alone created over $225 billion of value for consumers since 2004, according to our estimates. To account for that value, we propose that governments start measuring how much people *benefit* from goods and services – not just how much they pay for them.

What GDP doesn't measure

Traditionally, economists, policymakers, and journalists look at changes in GDP over time and as a proxy for how the economy is doing. It's a relatively precise number that comes out every quarter and that says how much the economy is growing or shrinking. However, GDP measures the monetary value of all final goods *produced* in the economy. It's a measure of how much we paid for things. It does not measure how much we *benefit*. In fact, well-being might go down when GDP goes up or vice-versa.

The good news is that economics does provides a way, at least in theory, to measure consumer well-being. That measure is called consumer surplus. The concept of consumer surplus represents the difference between the maximum a consumer would be willing to pay for something and what they actually have to pay for it. If you would have paid up to $100 for a shirt but only have to pay $40, then you have gained $60 of consumer surplus. Changes in consumer surplus are considered as a measure of changes in consumer well-being.

To understand why GDP can be a misleading proxy for consumer well-being consider the example of Encyclopedia Britannica and Wikipedia. Britannica used to cost several thousand dollars, meaning its customers considered it to be worth at least that amount. Wikipedia is free and has a far greater quantity of articles at comparable quality than Britannica ever did. In fact, Britannica went out of business in 2012 as consumers abandoned it. Measured by spending, the encyclopedia industry is shrinking. Measured by benefits, consumers have never been better off: they get a tremendous amount of consumer surplus from Wikipedia. In our research, we find that the median U.S. consumer values Wikipedia at around $150 per year but they pay $0. This one good translates roughly to around $45 billion of consumer surplus in US that doesn't show up in GDP.

Historically, assessing consumer surplus has been tricky, which is one reason it hasn't been used much for measuring the economy. Consumer surplus is not normally directly observed. In contrast, GDP depends on what we actually pay for goods and services so it can be observed at the cash register and shows up on companies' revenue statements.

Fortunately, just as the digital revolution created some tough measurement challenges, it also provides some wonderful new measurement tools. We have been able to use digital survey techniques to run massive online choice experiments on hundreds of thousands of consumers about their preferences. Using these new tools, we get estimates of the consumer surplus for a variety of goods, including free ones that are missing from the economic statistics. In our research, we show how this method can directly measure consumer surplus in a scalable way by asking consumers to make choices. In some cases, we have them chose between various goods (e.g. Would you rather lose access to Wikipedia or Facebook for the next month?). In other experiments they choose between keeping access to a good or giving it up in exchange for monetary compensation (E.g. Would you give up Wikipedia for a month for $10?). To make sure that consumers reveal their true preferences, we can enforce at least some of their choices and give them the money that we offered if and only if they forgo the good we are asking about.

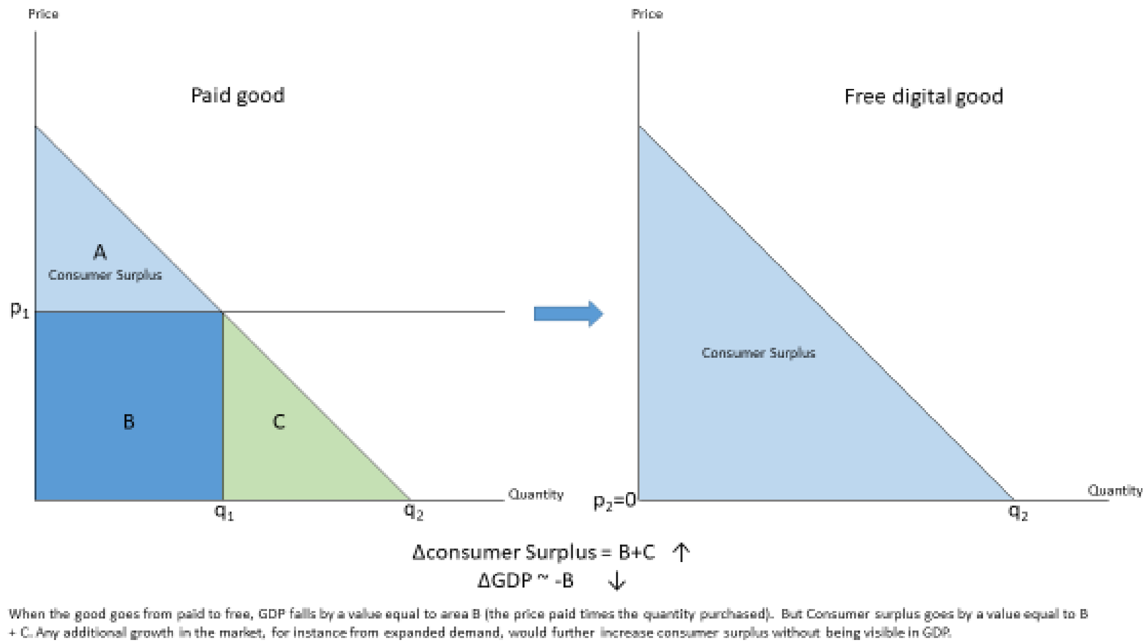


Figure 1: Changes in GDP vs. Consumer Surplus for goods transitioning from paid to free

How much do we value digital goods?

Here's an example of how this works. To measure the consumer surplus generated by Facebook, we recruited a representative sample of US-based Facebook users and offered them varying amounts of money to give Facebook up for one month. To make sure that they responded truthfully, some of these respondents were randomly selected for these payments and asked to give Facebook up for the month. Because we temporarily added them as a Facebook friend — with their permission, of course -- we could verify that they indeed didn't log in for that month and gave them the cash that we offered.

We found that the median Facebook user in the US would need compensation of $48 to give up the service for one month. 20% would give it up for as little as $1, but a significant chunk of users (20%) refused to give it up for less than $1,000. In total, we estimate that consumers derived $16 billion of value per year from Facebook since its inception in 2004 up to 2017.

We conducted a similar study in Europe in a university laboratory and found that the median user there needed a compensation of €97 to give Facebook up for one month. We also found that users who have more friends value Facebook more, reflecting the fact that network effects are a key factor contributing towards this high consumer valuation. People who also use Instagram and YouTube value Facebook less, implying that they might be substitutes to Facebook. In terms of demographics, we found that on average, women value Facebook more than men. We also found that older people value it more than younger people. This may reflect the fact that older people often lack good substitutes for Facebook while younger people are more likely to migrate to alternative social media platforms (e.g. Snapchat, Instagram) if they give up Facebook.

One might think that the value generated by Facebook is already accounted for in GDP via its advertising revenues. However, our estimates indicate that Facebook generates over $500 of consumer surplus per year for the average user in the US and Europe. In contrast, average revenue per user is only around $140 per year in US and $44 per year in Europe. Even for one of the most skilled advertising platforms, advertising revenues are only a fraction of the total consumer surplus generated by it. More fundamentally, research has shown that advertising revenues and consumer surplus need not be correlated with each other – people can get a lot of value from content that doesn't generate much advertising, and vice-versa. So it would be a mistake to use advertising revenues as a substitute for actually measuring consumer surplus.

We conducted more studies to measure the consumer surplus generated by most popular categories of digital goods in the US (Figure 2) and some popular digital goods in a controlled setting in a university laboratory in the Netherlands. We asked our respondents for the amount of money they would need to be compensated with to give up a single good or an entire category for one month or one year. For goods that weren't free, this monetary compensation was beyond the money they'd save by not purchasing, so they are an estimate for the consumer surplus generated by these goods. In laboratory studies, we gave respondents a chance to get real cash after we verified that they actually gave up the good.

Overall, our results indicate that digital goods have created a tremendous amount of economic well-being as indicated by our measures of consumer surplus. Search engines are the most valued category of goods in the US with a very high valuation of over $1400 per month for the median user, followed by Email and Maps. These are categories which do not have comparable offline substitutes. For many people, they are virtually essential for work or everyday life. In general, we found that it is harder for users to give up an entire category of goods than giving up a single good and switch to a substitutes and that's reflected in higher valuations for whole categories than for the sum of individual applications. For example, search engines are the first stop online for work or personal browsing before we navigate to an address on the web. Video streaming and E-commerce platforms are also highly valued by consumers. Social media, music streaming and instant messaging are not as highly valued as the other categories. Users pay to access some of these services. For example, users pay $10-$20 per month or $120-$240 per year for video streaming services (e.g. Netflix, Hulu, HBO etc.). That said, the consumer surplus generated from video streaming services is still 5-10 times what users pay to access them. (These numbers come from surveys described below. When respondents had to actually give up the goods in our incentive-compatible experiments the amounts were typically even higher.)

In our experiment in Europe, we found that WhatsApp had five times the valuation of Facebook. We interviewed our subjects to understand the reasons for this high valuation and found that WhatsApp has become a practically indispensable communication tool used to stay in touch with friends, family, and co-workers. It is the focal choice for coordinating activities within groups, setting up appointments and staying in the loop regarding meetups and events. In contrast, WhatsApp has a relatively low valuation in the US because most US users use other instant messaging tools.

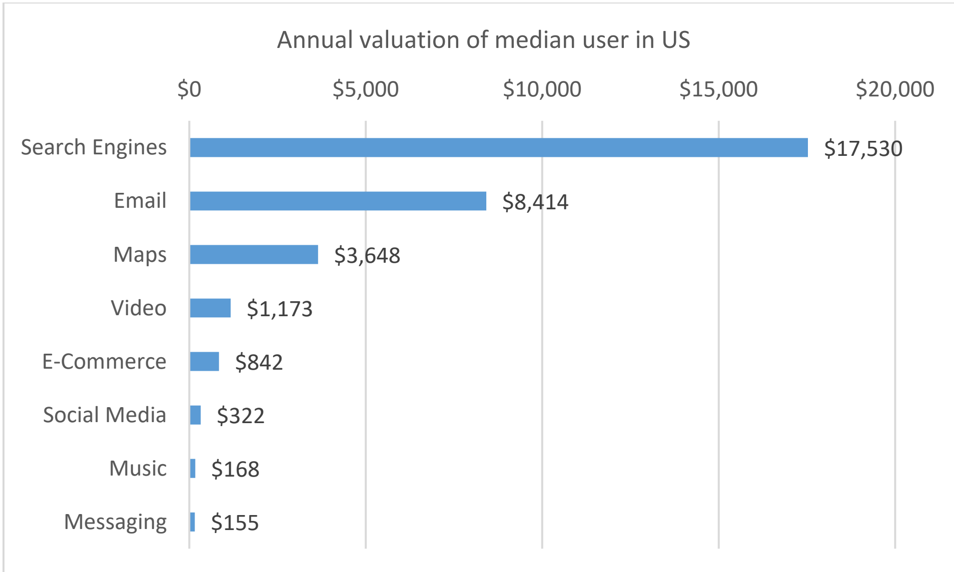


Figure 2: Valuations of popular categories of digital goods

Implications for measuring the economy

Working with Erwin Diewert, Felix Eggers, and Kevin Fox, we used these insights to develop a new metric for measuring the benefits associated with the digital economy. We call this new metric GDP-B because it builds upon GDP to account for the *benefits* (not costs) of new and free goods. Our GDP-B metric supplements and extends the traditional GDP framework, using it as a base to which we add the contributions to well-being from new and free goods. Our choice experiment approach estimates these contributions in two ways. First, we use large scale hypothetical surveys where we simply ask respondents how much they'd need to be paid to give up a given good for a given period of time. Second, we validate those survey results by running smaller scale incentive-compatible studies with real monetary incentives – like the ones we've described above. By combining these two approaches, our method offers a relatively inexpensive way for policymakers, managers, and economists to measure the well-being of consumers rather than just the production side of the economy.

To put the economic contributions of digital goods in perspective, we find that including the benefits of Facebook would have added about 0.11 percentage points to GDP-B growth per year on average in the US from 2004, when Facebook was launched, through 2017. During this period, the traditional measure of GDP grew by 1.83% per year. Although GDP and GDP-B are not directly comparable, our measure suggests that undercounting even just one good – Facebook – means that GDP substantially underestimated growth in consumer well-being over that timeframe. Of course, there were many other digital good introduced during this period and in ongoing work, we are doing a fuller accounting that includes each of them as well.

While it is tempting to conclude from our work that the slowdown in productivity over the past decade and a half might disappear if we properly account for the benefits of the digital revolution, our work is not sufficient to support that claim. While the unmeasured benefit of free goods are important today, it's also true that earlier waves of free and nearly free goods like antibiotics, radio, and television clearly delivered substantial uncounted consumer surplus. In fact, Nobel-prize winning economist William Nordhaus estimated that firms were able to capture only 2.2% of the total surplus generated from technological innovations during the 20th century while the remaining 97.8% of the surplus went to consumers. Were these innovations less important than the current wave? It's hard to say since no one did studies like ours back then. Undercounting the benefits of technology isn't a new problem, but that doesn't make solving it less urgent.

Our findings have two important limitations. First, our GDP-B estimates are still far from comprehensive and are not as precise as the GDP measures. We will need to include far more goods and do more online choice experiments for each good to get a fuller and more precise measure how the economy is generating benefits from new and free goods.

Second, like traditional GDP, our measures do not capture some of the potential negative externalities associated with online platforms. For example, Hunt Alcott Luca Braghieri, Sarah Eichmeyer, and Matt Gentzkow have explored the potential for Facebook to lead to addictive behavior and there is widespread debate about the impact of internet use and smartphones on happiness and mental health. Others have argued that some digital goods are damaging to social cohesion or political discourse. For now, our GDP-B metric treats people as rational decision makers and only captures the private benefits associated with goods, not the social costs and benefits. While traditional GDP is subject to the same criticisms, we are working on addressing both these limitations, as are others.

Furthermore, other researchers have developed useful methods to quantify aspects of subjective well-being including happiness and life satisfaction (See Sidebar). However, a survey of leading macroeconomists suggests that these subjective well-being metrics are not yet as precise, comparable or reliable as we would like.

On a spectrum ranging from current macroeconomic indicators such as GDP and productivity to well-being indicators such as happiness, our GDP-B metric lies in somewhere in the middle. GDP is relatively precise but doesn't capture much of what we'd ideally measure; happiness assessments have the opposite problem. GDP-B strikes a balance between those extremes. In our view, it is important for policy makers to have an understanding of this entire spectrum of measures and focus on the relevant metrics for any particular policy objective.

Although we believe GDP-B can be relevant for a large number of policy considerations, it's particularly important for those affecting the digital economy. More and more critical consumer services are free or nearly free, and it's essential that we understand their impact on our lives. Questions of how to regulate tech, how much to subsidize digital infrastructure, and even what sort of new digital products and services entrepreneurs ought to create depend on understanding how much we all benefit from the digital economy.

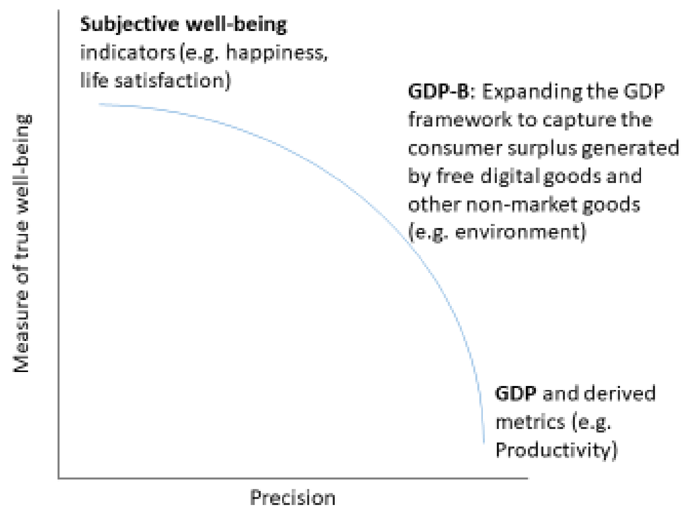


Figure 3: Spectrum of well-being measures

Our approach to measuring consumer surplus can be scaled up to estimate the contributions or not only thousands of digital goods, but also conventional goods from breakfast cereal to jet travel. More ambitiously, we may be able to get better estimates of the benefits associated with other non-market goods such as the environment and public goods like healthcare and infrastructure. Ultimately, as governments, managers and researchers in a variety of countries around the world adopt this approach, we will get meaningful estimates of how both digital and non-digital goods contribute to our well-being, and with better measurement, comes better management.

Appendix:

Short history of GDP

Gross domestic product (GDP), first developed in the 1930s, is rightly heralded as one of the greatest inventions of the 20th century. It measures the monetary value of all final goods produced in the economy. Although it is today widely used as a metric of well-being, the leader of the team that created it Simon Kuznets warned that "the welfare of a nation can scarcely be inferred" from GDP and that was not its purpose when it was created. Among its weaknesses is that fails to capture negative externalities associated with growth such as pollution or congestion. Moreover, non-market activities such as household production (when people do unpaid tasks for themselves at home) are not included in GDP. Since its conception in the 1930s, GDP has been updated, revised and extended a number of times. For instance, better measures of computer prices and software investments were introduced in 1999. Meanwhile, "satellite" accounts have been introduced to track particular aspects of economic activity such as household production and R&D. We see GDP-B as a possible path toward providing a more complete dashboard of economic indicators.

Alternative measures of well-being

Since GDP is a flawed measure of well-being, several attempts have been made to design alternative measures of well-being. For example, the United Nations developed Human Development Index (HDI) which is composed of life expectancy, education and income per capita. Chad Jones and Pete Klenow have put forward a summary measure of well-being consisting of consumption, leisure, mortality and inequality. Several countries including UK and Bhutan quantify subjective well-being by surveying citizens on questions related to happiness and life satisfaction. The OECD Better Life Index is an interactive tool that allows users to compare countries across 11 dimensions of well-being including environment, health and life satisfaction. Michael Porter, Scott Stern have developed the social progress index comprising of 54 indicators measuring the extent to which countries provide for the needs of their citizens (basic human needs and needs related to well-being and opportunity).