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

Erik Brynjolfsson
MIT Sloan & NBER

Felix Eggers
University of Groningen

Avinash Gannamaneni
MIT Sloan

**ABSTRACT**

Digital technologies have transformed the nature of production and the types of goods and services consumed in modern economies. Yet our measurement framework for economic growth has not fundamentally changed since the 1930s. In principle, a better approach is now feasible. Specifically, changes in consumer surplus (compensating expenditure) are superior to changes in GDP as a measure of changes in consumer well-being, especially for digital goods. In practice, consumer surplus has been difficult to measure. We explore the potential for massively-scalable online Single Binary Discrete Choice experiments to measure changes in consumer surplus for digital goods. Through these experiments we seek to measure consumers' willingness to accept compensation for losing access to various digital goods and thereby estimate the changes in consumer surplus from these goods. Because very large numbers of Americans can now be reached electronically, changes in consumer surplus and other new measures of well-being derived from online choice experiments have the potential for providing cost-effective supplements to existing national income and product accounts. Our results indicate that digital goods have created enormous gains in well-being which are largely missed by conventional measure of GDP and productivity, and suggest that our approach can be scaled up to a broader set of goods and services. Two limitations of our methods are that they are much less precise than changes in GDP and they suffer from hypothetical bias. We estimate how much of an improvement in precision can be achieved with a larger sample size and various demographic controls and we document the direction and magnitude of bias present in our approach by conducting an incentive compatible study for Facebook. By periodically querying a large, representative sample of goods and services, including those which are not priced in existing markets, these methods could provide an estimate of annual changes in consumer well-being.

## Introduction

Gross Domestic Product (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. GDP is heralded as one of the greatest inventions of the 20th century by noted economists Paul Samuelson and William Nordhaus (Landefeld 2000). However many economists agree that GDP is a significantly flawed measure of well-being and attempts are being made to design alternative measures (Stiglitz et al. 2009).

Both economists and journalists routinely use GDP as if it is 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 derive links between productivity and improvement in living standards (OECD 2008). However many economists have long acknowledged that GDP is not a welfare measure. 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 economic progress.

GDP as a welfare measure is especially problematic in the emerging digital economy since most of the digital goods have nearly zero marginal cost and often a zero equilibrium price. Most of their contributions are therefore invisible in the GDP calculations (Brynjolfsson and Saunders 2009; Brynjolfsson and McAfee 2014). For instance, although information goods have become increasingly more ubiquitous and important in our daily lives, the share of the information sector as a fraction of the total GDP (~ 4%) has not changed in the last 30 years. Moreover, in many sectors (e.g. music, media, encyclopedias) people substitute paid (offline) goods with free online services (e.g., Spotify, YouTube, Wikipedia) so that the total revenue that shows up in GDP figures could fall even while consumers get access to better quality and more 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 on existing measures of GDP and productivity as a proxy for well-being.

The benefits of technological advance are distinct from the expenditures on goods and services. 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. Consumer surplus is defined as the difference between a consumer's maximum willingness to pay for a product and the actual price paid for it.

So far, the change in consumer surplus hasn't been widely used as a measure of well-being not because it is a poor measure of well-being, but because it is difficult to measure at scale. Estimating demand curves requires exogenous variations that shift the supply curve but not the demand curve and it is typically not practical to identify these variations for a large bundle of goods. However, with advancements in digital technologies it is now feasible to collect data about thousands of goods easily and quickly. These big data techniques have the potential to improve measurement of economic indicators. For example, Cavallo and Rigobon (2016) scrape the web to collect billions of prices for millions of

[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/).

[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).

products to construct prices indices and inflation measures for various countries as part of MIT's Billion Prices Project. Private companies like Amazon, Google and Facebook routinely conduct millions of online experiments to help understand consumer preferences and behavior. This scale of experimentation and inference would have been infeasible 20 years ago, but is now routine, at least in private industry.

In this research, we propose a way of measuring the changes in consumer surplus, first for the digital economy and then more broadly. We assess the feasibility of this approach with a small sample of goods. Through thousands of discrete choice experiments conducted online, we measure consumers' willingness to accept for losing access to various digital goods for a representative sample of the US population and estimate the demand curves for these goods in order to compute the changes in consumer surplus from a portion of the digital economy. We conclude that our approach is easily scalable and our end goal is to develop a system that tracks changes in consumer surplus of numerous goods and services in (near) real time via massive online choice experiments.

**Background**

One of the hypothesized explanations for productivity slowdown in US since the past decade is that existing economic indicators (including GDP) do not properly measure the advancements in information technologies. While average annual labor productivity growth was 2.8% per year over 1995-2004, it shrunk to 1.3% per year over 2005-2015 (Syverson 2016). 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 mismeasurement may be important in recent years, it was also likely important in the past, so its power to explain the slowdown is limited.

Our research does not aim to contribute directly towards this debate. Instead we focus on the more fundamental issue that GDP is not a direct measure of well-being in the first place. Thus, whether or not GDP mismeasurement has grown, is a distinct, albeit related, question from how well-being is changing. This has been an issue since GDP was invented and arguably an even bigger issue in the current digital era. 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 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 themselves contribute much to our understanding, if not to our GDP, and there have been a number of efforts to create a more comprehensive estimate of well-being.

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

[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)

In this paper, we are less ambitious and seek to stick more closely to a traditional microeconomic framework. In particular, we focus on the changes in consumer surplus generated by 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, e-commerce platforms but their revenues don't always reflect this increased use. 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 zero marginal cost and are hence priced much lower (often even zero price) 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 till now. Moreover, Waldfogel (2012) demonstrates that the quality of music has likely increased since 1999. Therefore, although the contribution of music industry to GDP statistics is shrinking consumers are better off than before; they are listening to more and better music.

In the case of classic physical goods such as cars, consumer surplus is more or less proportional to firm revenue (Figure 1). Keeping the supply curve fixed, as the size of the market increases as more consumers enter the market the demand curve shifts right. In this case, both consumer surplus and revenue increase proportionately. The increase quantity sold shows up in GDP statistics and hence both GDP and consumer welfare move in the same direction. Doubling the number of cars, apples or books sold is likely to roughly double revenues, GDP and consumer surplus. A similar logic applies for many services like haircuts, meals served or windows washed.

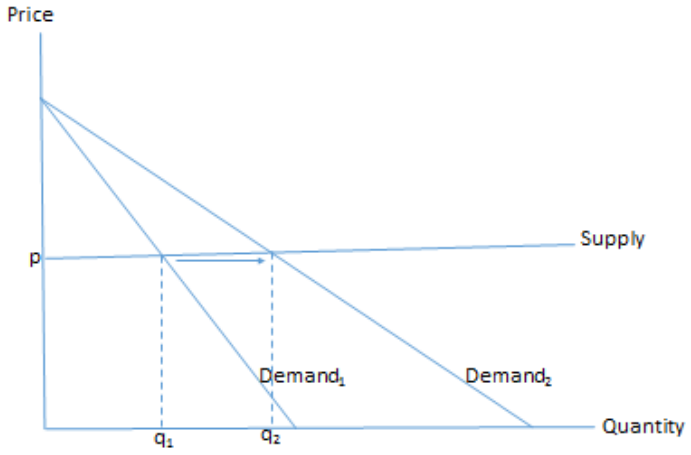



Figure 1: Consumer surplus and revenue for classic goods such as cars

A number of traditional goods and services are transitioning into digital goods and services. A good example of such a transition good is an encyclopedia. Previously browsing through a hard copy of an encyclopedia was often the best way to obtain information on a particular subject. Since 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 currently has over 60 times as many articles as Britannica and its accuracy has been found to be on a par with Britannica (Giles, 2005). Far more people use Wikipedia than ever used Britannica. 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. Similarly, many people now have maps, GPS and other services available for no extra cost once they are able to access the Internet on mobile device or home computers. For such transition goods, consumer surplus increases and revenue decreases as prices become zero (Figure 2). Hence GDP and consumer welfare move in opposite directions.

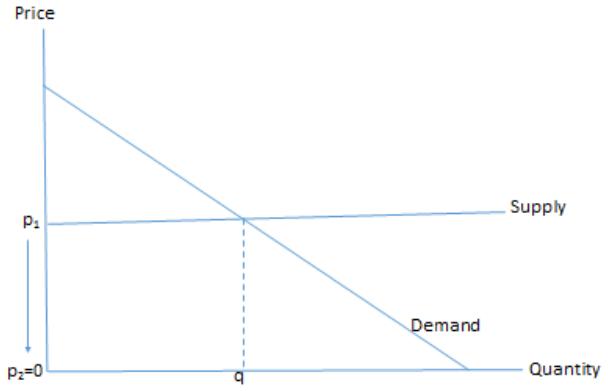



Figure 2: Consumer surplus and revenue for transition goods such as encyclopedias

A third category is the case of 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. In some cases, digital goods earn revenues from advertising and this can creates a significant contribution to GDP (Nakamura and Soloveichik 2015). However, changes in advertising revenues are generally not closely related to changes in consumer surplus (Spence and Owen 1977). Hence, as the demand for these free goods increases, consumer surplus will also increase but this change in well-being is not well-reflected in GDP (Figure 3). GDP may be completely unchanged due to this shift even though consumers are better off.

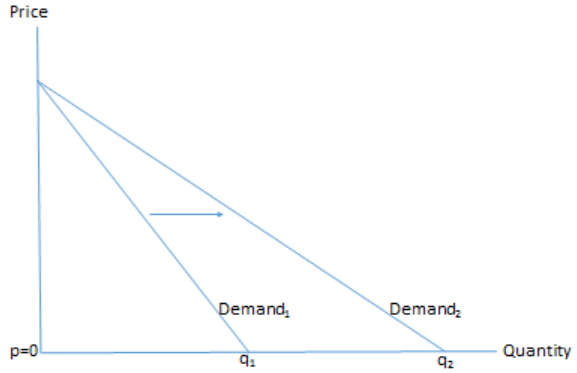



Figure 3: Consumer surplus and revenue for purely digital goods

Figures 2 and 3 suggest that changes in consumer surplus is 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 (robo advisors such as Wealthfront), customer service (AI powered services such as DigitalGenius) and law (AI powered bots such as DoNotPay).

While producer surplus cannot be inferred from consumer surplus, when it comes to technological advances, firms are able to appropriate only a small fraction of the social returns (Nordhaus 2005) so 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 accordingly. 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 if the producer surplus grows relative to the consumer surplus.

Recently there has been some 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 from 1999-2006. For 2015, this figure is estimated to be $55 billion (Syverson 2016). Although this 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 there is an opportunity cost associated with using free digital goods and this cost is equal to the wages lost due to not working. Therefore, the value of these digital goods is equal to these lost wages. Using this approach, Goolsbee and Klenow (2006) estimate the effect on consumer surplus for the median US resident to be $3000 for 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 between 2007 and 2011. If consumers value their time less than their wages, then the consumer surplus may be overestimated.

Nakamura and Soloveichik (2015) estimate the value of free media by computing the online advertising revenues generated by websites. Including free media increases real GDP growth by 0.019% according to their estimates. However, advertising revenues do not capture the entire value of the free digital goods. For e.g., 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. Moreover, advertising revenues need not be proportional to consumer surplus. Spence and Owen (1977) argue that advertisers pay for number of views (or clicks) 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 consumers might only be marginally interested in that program. However, for a niche program (less views), which 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 will use choice experiments to elicit consumers' valuation of goods. Specifically, we will ask consumers to make a choice between keeping a digital good or taking a monetary value when foregoing it. This approach experimentally varies the monetary values and therefore addresses the limitation that the actual market price of many digital goods is zero so that demand does not represent their 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; albeit at the expense of being "hypothetical", i.e., inconsequential (Carson and Groves 2007).

## Methodology

*Approaches to measure consumer value*

There are two general approaches to obtain input data to measure consumer value: 1) based on market data ("revealed preferences") and 2) based on choice experiments or survey techniques ("stated preferences").

Approaches based on revealed preferences relate variation in observed demand for a good to changes in its market prices in order to derive demand curves and prices elasticity (e.g., Cohen et al., 2016; Greenwood and Kopecky 2011). 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 observed market prices and differences in characteristics of the goods (Williams 2008). However, both of these approaches require variance in the observed market prices and are therefore not applicable to digital goods that are provided for free. Revealed preference approaches are therefore limited to goods for which a market price exists, e.g., (broadband) internet access fees (Greenstein and McDevitt 2011) or require a proxy for market price, e.g., time spent using the digital goods (Goolsbee and Klenow 2006; Brynjolfsson and Oh 2012).

Stated preference elicitation techniques provide more flexibility because they do not require a market price or real transactions to exist and can be applied to contingent scenarios (leading to contingent valuation studies). Questions of stated preference include asking consumers directly about their maximum willingness-to-pay (WTP) in monetary terms. This question reveals a ratio-scaled measure of a consumer's value of the good. However, this question has been shown to be less reliable and less valid, likely because consumers are not used to formulating own prices and do not have an incentive to reveal their true preferences (Miller et al. 2011; Carson and Groves 2007).

The introduction of non-hypothetical, incentive compatible variants to elicit WTP in form of auctions (e.g., Vickrey auctions, Vickrey, 1961) or lotteries (e.g., BDM, Becker, DeGroot, and Marschak 1964; Wertenbroch and Skiera 2002) mitigated some of these disadvantages, but at the expense of being complex, hard to understand, and by introducing (artificial) competitive pressure in auctions (Carson, Groves, and List 2014; Völckner 2006). These incentive compatible direct question formats are therefore applicable neither to digital goods, in which supply is not restricted, nor to large scale online choice experiments that consumers need to understand and answer quickly.

An alternative, indirect form of measuring stated preferences are discrete choice experiments (DCE) (Louviere, Hensher, and Swait 2000). DCEs ask consumers to choose between options and select the alternative that they value most. By experimental variation of the characteristics of the 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 synonym for choice-based conjoint experiments that typically involve about 8 to 12 sequential choice tasks that present multiple alternatives, e.g., three to five, and each alternative varies in 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 (e.g., valuation of preventing another Exxon Valdez type oil spill, 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. Apple-Samsung lawsuit).

*Proposed approach and implementation*

We propose to measure consumer value 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. We only ask one question per consumer and vary different 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 more reliably. Consumers only have to make a decision between two options instead of thinking about a monetary figure themselves. Moreover, we can compensate for the loss 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 massive data quantity to overcome some of the limitations of earlier research relying on smaller samples.

For the implementation, we use Google Consumer Surveys (GCS) as our primary platform. GCS allows us to run short one-question surveys cheaply 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). Survey creators pay per response, part of which goes to the publisher for hosting it. Users have to answer the survey in order to unlock premium content. 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).

*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 value 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 explicitly consider the influence of other attributes such as negative utility effects of advertising or limited privacy (as analyzed in IAB Europe; McKinsey analysis) as 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:

(1) $U(g^1, y - C^*) = U(g^0, y)$, or
(2) $U(g^1, y) = U(g^0, y + E^*)$,

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 $E^* > C^*$ (Hanemann 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 since 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 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 then define the status quo utility as $U(g^1) = 0$. Consequently, a consumer will (not) forego the good for amount E if $U(g^0, E)$ is positive (negative).

In order to estimate the equivalent measure $E^*$ we need estimates of how valuable consumers find using the good and how sensitive they react to changes 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 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 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) is:

$$(3)\ P(g^0, E) = exp(b_0 g^0 + b_1 E)/(1 + exp(b_0 g^0 + b_1 E))$$

or $1 - P(g^0, E)$, for keeping the service. The parameters can then 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.

*Criticism*

Such SBDC questions have several advantages compared to approaches that directly ask consumers about their WTP or WTA (i.e., C or E). SBDC valuations are based on consumer choices that are most similar to day-to-day (purchase or consumption) activities. They are natural manifestations of consumers' preferences and are easy to accomplish. DCEs have been shown to achieve good (external) predictive validity and produce valid estimates of WTP so that they should be favored over direct elicitations techniques (Carson and Groves 2007; Miller et al. 2011, Wlömert and Eggers 2016).

Moreover, SBDC questions are in line with economic theory and, based on the random utility model, can be used to estimate neoclassical Hicksian welfare measures (McFadden 1974, Carson and Czajkowski 2014). Moreover, a single "take-it-or-leave-it" referendum-like question has favorable incentive-compatibility properties, compared to multiple (e.g., "double-bounded"), sequential questions (Carson et al. 2003). Relatedly, single questions prevent the so-called starting point bias in follow-up questions, i.e., an anchoring effect in which subsequent prices are evaluated relatively to the first price the consumers was exposed to (Whitehead 2002).

However, the proposed SBDC questions, or rather contingent valuation questions in general, are not without criticism. Hausman (2012) identifies three major weaknesses of contingent valuation questions: 1) hypothetical bias; 2) differences between WTP and WTA; and 3) inconsistencies regarding scope and embedding (see a detailed rebuttal to his criticism from Haab et al. 2013). While the choice experiments we run in this paper differ somewhat from the contingent valuation approach that Hausman discusses, they have enough similarities that it is worth considering his critiques in some detail.

A hypothetical bias arises from contingent valuation questions and SBDC questions if consumer do not believe that they actually need to pay a price for any of the services in the near future. Hence, consumers might not consider the answer given to the stated preference question as consequential and have no incentive to reveal their true valuation so that a random response would be as a good as the true answer (Carson, Groves, and List 2014). According to the random utility model these answers increase the error variance so that preference estimates are less reliable. If these errors are unbiased around the true value, then they will tend to cancel out as sample size grows. However, apart from random error, it has been shown that the hypothetical bias often leads to a systematic bias, e.g., due to strategic incentives to understate or overstate the true WTP (Carson and Groves 2007). This is a more serious concern.

In order to quantify the magnitude of the hypothetical bias we test one digital good, i.e., Facebook, in which we provide consumers with real money if they stop using the service. In that way, choices are consequential and consumers have a clear incentive to provide their true valuation of the service (Carson, Groves, and List 2014). Similar applications of SBDC questions date back to a study by Bishop and Heberlein (1979) in which they measure the consumer surplus of goose hunting permits by providing hunters with actual money for returning their hunting permits. Similar approaches to mitigate the hypothetical bias include the incentive alignment procedure applicable to DCEs that has been shown to increase (external) validity substantially (Ding 2007; Ding, Grewal, Liechty 2005; Wlömert and Eggers 2016). As discussed in our results section, we confirm that a hypothetical bias exists and that this criticism is justified. However, in our study design, we measure WTA and not WTP. As a result, the hypothetical bias we observe leads to a relatively consistent and quantifiable understatement of value and therefore a conservative estimate of WTA.

While, theoretically, WTP and WTA should give (approximately) the same value, empirically they do not, which is recognized as being inconsistent with economic theory. Therefore, attempts have been made to extend (behavioral) theory in order to explain the disparities, e.g., with endowment effects, loss aversion, or uncertainty about the quality of the goods (Hanemann 1991; Plott and Zeiler 2005). Nevertheless, Hausman (2012) argues that the gap between WTP and WTA is "likely due to the reality that answers to contingent valuation surveys do not actually reflect stable or well-defined preferences but instead are opinions invented on the fly" (p. 47). This statement, however, is at odds with the fact (and Hausman's acknowledgment) that the differences between WTP and WTA are persistent and consistent, which rules out random effects of unstable preferences but rather results from systematic differences in these question formats. As we argue above, we find that WTA better represents consumer welfare since free digital goods require no upfront investment from consumers. Moreover, research from behavioral economics suggests that measuring WTP instead of WTA for free goods can lead to biased estimates since consumers may take the market price of zero as an informational "anchor" so that WTP estimates are also biased towards zero (Ariely, Loewenstein, and Prelec 2003).

The third major criticism, scope and embedding, refers to the proposition that consumers should be willing to pay more for a large effect than for a subset of that effect (or a good that is embedded in a larger package). Although these effects can be found empirically they are typically not considered large enough to be credible (Hausman 2012). Diamond and Hausman (1994) propose an adding-up test for the scope test. However, because digital goods can serve as substitutes or complements (e.g., social media can provide messaging functions or video can be used to listen to music) the adding-up test is not appropriate

in our context. Differences in scope and embedding therefore need to be critically addressed in terms of the substitutability or complementarity of the goods.

Much of the criticism refers to the nature of non-market goods that are typically subject of contingent valuation studies (e.g., clean air or clean water). With these goods, which may be used predominantly in a passive way, consumers have limited to no active experience or are rather unfamiliar with their true value (Carson 2012; Carson and Czajkowski 2014; Hausman 2012). Digital goods, on the other hand, are typically used on a day-to-day basis and required an active step in getting access to them (e.g., the deliberate choice to subscribe to Facebook) so that consumers should be familiar with them and are therefore better able to express and quantify their value. Morwitz et al. (2007) support this notion in their meta-analysis and empirical study. We therefore expect that our research context mitigates much of the above-mentioned criticism. However, arguably, it does not eliminate them completely. We will analyze the magnitude of the bias empirically.

**Results**

*Benchmark study: WTA Facebook*

Our first study focuses on consumer value of Facebook. The purposes of this study are to get an estimate of the magnitude of the hypothetical bias and to compare results from a representative sample of consumers invited by a professional panel provider with a GCS sample. In this study we apply the SBDC experiment by asking consumers if they would prefer to keep access to Facebook or go without access to Facebook for one month and get paid $E? We varied E between-subjects across twelve price points (E = 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000).

In order to measure the hypothetical bias we used a second group of consumers in which we asked the same question but with an additional incentive compatibility procedure. In order to make the question consequential for the consumer, we informed them that we will randomly pick one consumer out of every 200[5] respondents and fulfill his/her selection. Specifically, if they choose "Keep access to Facebook" nothing will change for them, however, they will also not receive the 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 about the procedure to monitor their Facebook online status and the requirement to provide their email address (see Appendix for the exact question wording). Thus, it is in their financial interest to accurately report their true value for Facebook by choosing the second option if and only if their value for Facebook is less than $E.

In the remainder of the questionnaire we collected demographic data and Facebook usage behavior.

*Sample*

We recruited consumers for this online study via Peanut Labs, a professional panel provider with 2.9 million active panelists and member of, among others, CASRO, ESOMAR, and MRA (Peanut Labs 2015). We targeted consumers that are 18 years or older and live in the US. Consumers who did not fulfill these criteria were screened out. 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. Consumers had to select Facebook in order to qualify for the survey; if they (also) selected the non-existent service they were disqualified. We controlled quotas for gender, age, and US regions to match US census data (File and Ryan 2014; we applied post-stratification for education and household income).

[5] 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.

Consumers who accessed the survey were randomly allocated to the incentive compatible (IC) or the non-incentive compatible (NIC) condition and, within each condition, to one of the tested price points. Allocation to the price points was also random, however we sampled the highest and lowest price points twice in order to obtain more reliable estimates for the endpoints of the demand function. Between June 23 and July 3, 2016 we received 2997 complete responses ($n_{NIC}$ = 1500, $n_{IC}$ = 1497). The average completion rate was 93.3%: 96.1% for the NIC survey and 90.4% for the IC condition. The lower completion rate for the IC group is typical for surveys with incentive alignment and is likely due to consumers who are not willing to participate in it. However, in this case the differences to the NIC group are minor and, since we actively controlled quotas, they are not likely to produce a non-response bias.

*Hypothetical bias*

Figure 4 shows the share of consumers who prefer to keep using Facebook at the different price points, detailed for both the IC and NIC condition. We focus on keeping the status quo option instead of being willing to accept the money in order to be consistent to demand curves, i.e., having a negative effect of price. For very low prices, i.e., a price of $1, the IC and NIC condition produce almost identical shares of around 75%. For higher prices the disparities increase leading to consistently higher shares in the IC condition.

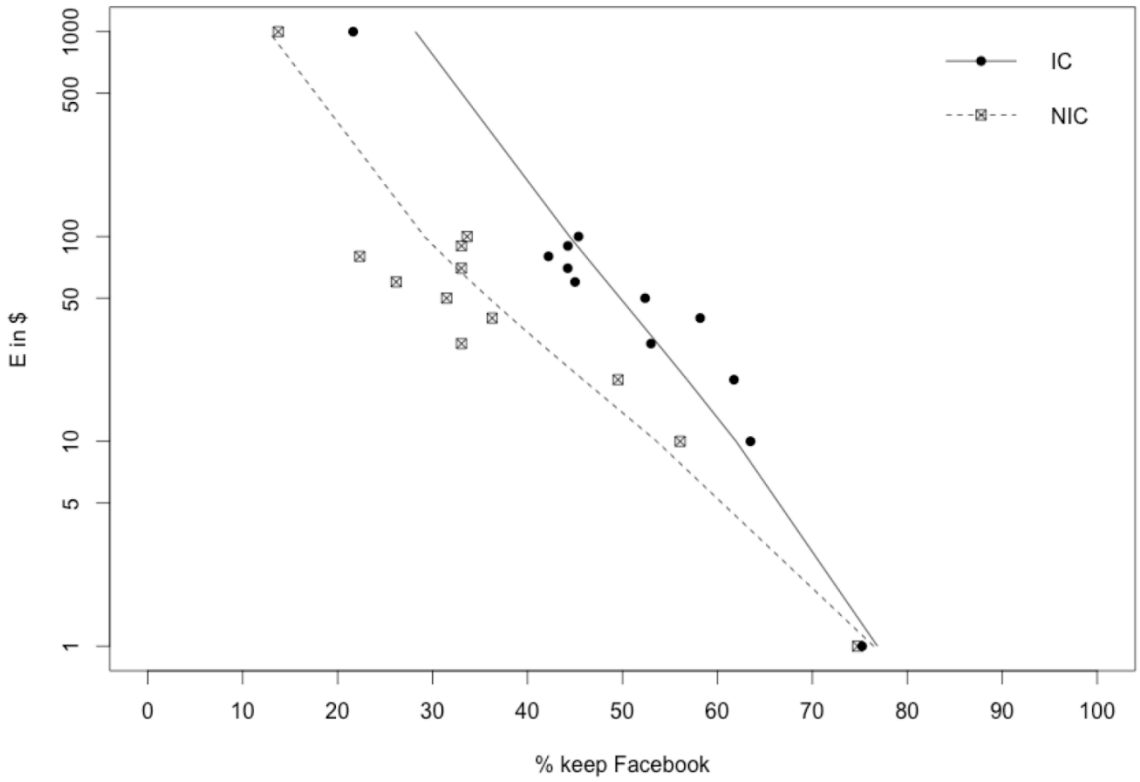



Figure 4: Assessment of hypothetical bias

In order to quantify the impact of the hypothetical bias we estimated a binary logit model that accounts for the magnitude of E (here, log(E) provided a better fit to the data), group membership (dummy variable for IC), and whether the IC group differs in sensitivity towards E. Table 1 shows the estimation results. The intercept represents the share of NIC consumers who prefer to keep Facebook at E = $1, i.e., exp(1.178)/(1 + exp(1.178) = 76.5%. The NIC group's utility decreases by 0.449 with every

one-unit increase in log(E), leading to a median WTANIC = exp(1.178/0.449) = $13.80 per month or $165.61 per year. The IC consumers do not differ in the intercept (Sig. = 0.905) but they react significantly less sensitive towards differences in E. A one-unit increase in log(E) results in a utility decrease of only -0.449 + 0.140 = -0.309. Consequently, IC consumers are willing to accept a substantially larger amount: WTAIC = $48.48 per month or $581.79 per year. Thus, the hypothetical WTA is understated in this research context and needs to be calibrated by a factor of 3.5. In their meta analyses of hypothetical biases, Murphy et al. (2005) and List and Gallet (2001) conclude that consumers typically overstate their hypothetical WTP by up to a factor of 3. An explanation for the difference to the meta analyses could be that consumers might have strategically understated their WTA in our NIC setting if they believed that the survey results could be used to determine a market price for the good (Carson and Groves 2007). However, the variance of the identified calibration factors in the meta analyses is large and studies that focus on WTA instead of WTP are scarce so that generalizations are limited.

| | **beta** | **Std. Error** | **z** | **Sig.** |
|---|---|---|---|---|
| **(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 |

*LL = -2105.16, df = 4, McFadden $R^2$ = 0.10, Hit rate = 0.66*

Table 1: Estimation results of binary logit model comparing IC and NIC group

*Selection bias*

Google Surveys might introduce a selection bias because selection criteria are less stringent compared to a professionally recruited sample based on representative quotas. To test this potential bias we compare the NIC group from the Peanut Labs sample to a Google Surveys 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 adjusting for the share of non-Facebook users. The SBDC experiment and its settings were identical in both samples.

Figure 5 compares the shares of users wanting to keep Facebook among both samples. The differences are minor and for the highest price point even non-existent (11.4% share in both samples). A binary logit model confirms that there are no significant differences between both samples, neither in terms of their intercept (Sig. = 0.991) nor sensitivity towards E (Sig. = 0.474).

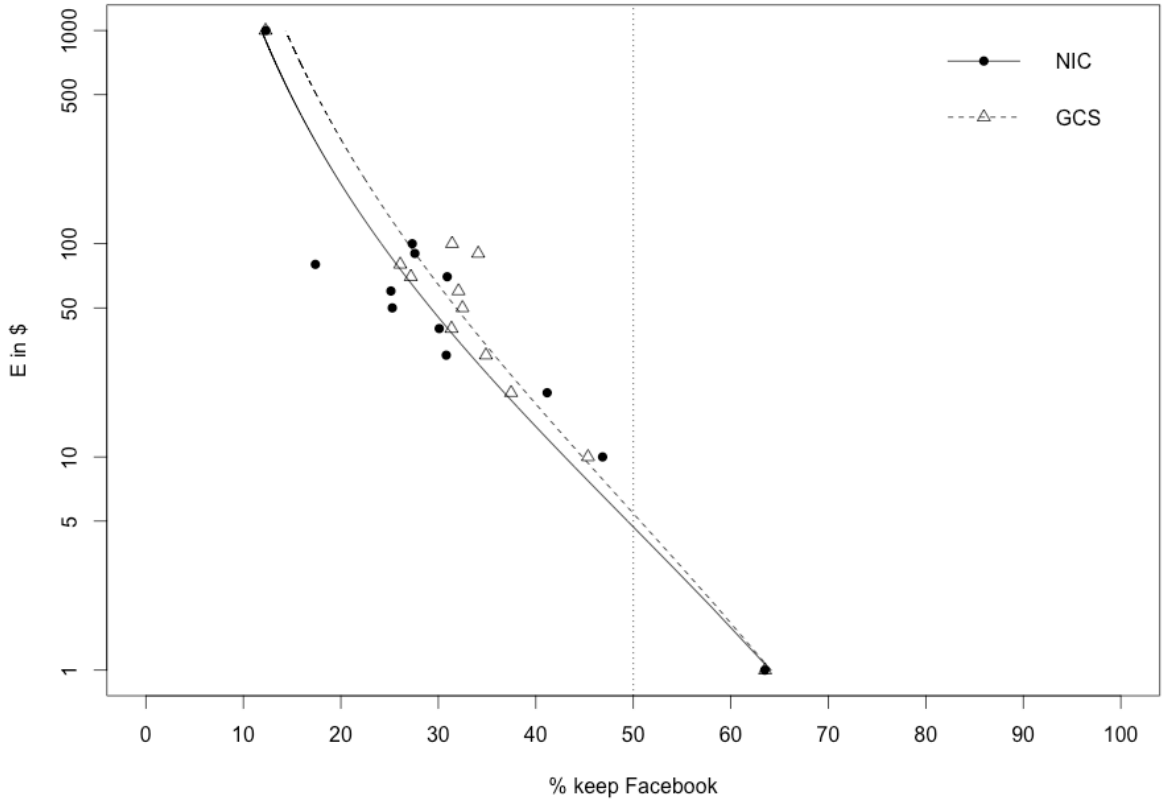



Figure 5: Assessment of selection bias

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

*LL = -2070.84, df = 4, McFadden R2 = 0.08, Hit rate = 0.71*

Table 2: Estimation results of binary logit model comparing NIC and GCS group

*Google Consumer Surveys*

We ran several SBDC surveys where we asked consumers if they would prefer to keep access to various digital goods or go without access to these goods for one year and get paid $E. We chose 10-15 price levels for each SBDC experiment and around 500 responses per price level. The available demographic variables (gender, age, income and urban density) were also added to an extended model to determine variation in valuations of different groups.

We identified the most widely used apps and websites on various devices and combined them into the following categories: Email, Search Engines, Maps, E-commerce, Video, Music, Social Media and Instant Messaging. Figure 6 plots the estimated WTA demand curves for these categories of digital goods. Table 3 provides the estimates of the extended logit model for these categories (including heterogeneity in valuation across different demographics) and Table 4 provides the median WTA values per year.

According to our median annual WTA estimates, Search Engines ($16,629) seems to be the most valued category of digital goods followed by Email ($6,896) and digital Maps ($2,790). One possible reason for this could be the lack of proper substitutes in the absence of search engines, email or digital maps. Since most consumers do not directly pay for these services, almost all of the WTA for maps contributes towards consumer surplus. What's more, for many people, these services are essential to their jobs. Next, Video streaming services (e.g. Youtube, Netflix) are most valued by consumers with a median WTA of $936 per year. Although, consumers do pay for some of these services, these amounts are of the order of $10-$20 per month (for those who pay) meaning the median consumer still enjoys an annual surplus of over $600. The remaining categories for which we estimated the median WTA are (in descending order) E-Commerce ($771), Social Media ($188), Instant Messaging ($145) and Music ($144).

All of these estimates are potentially biased downwards due to lack of incentive compatibility as shown previously. When taken together, the sum of these estimates suggests there is a significant amount of consumer surplus from digital goods.

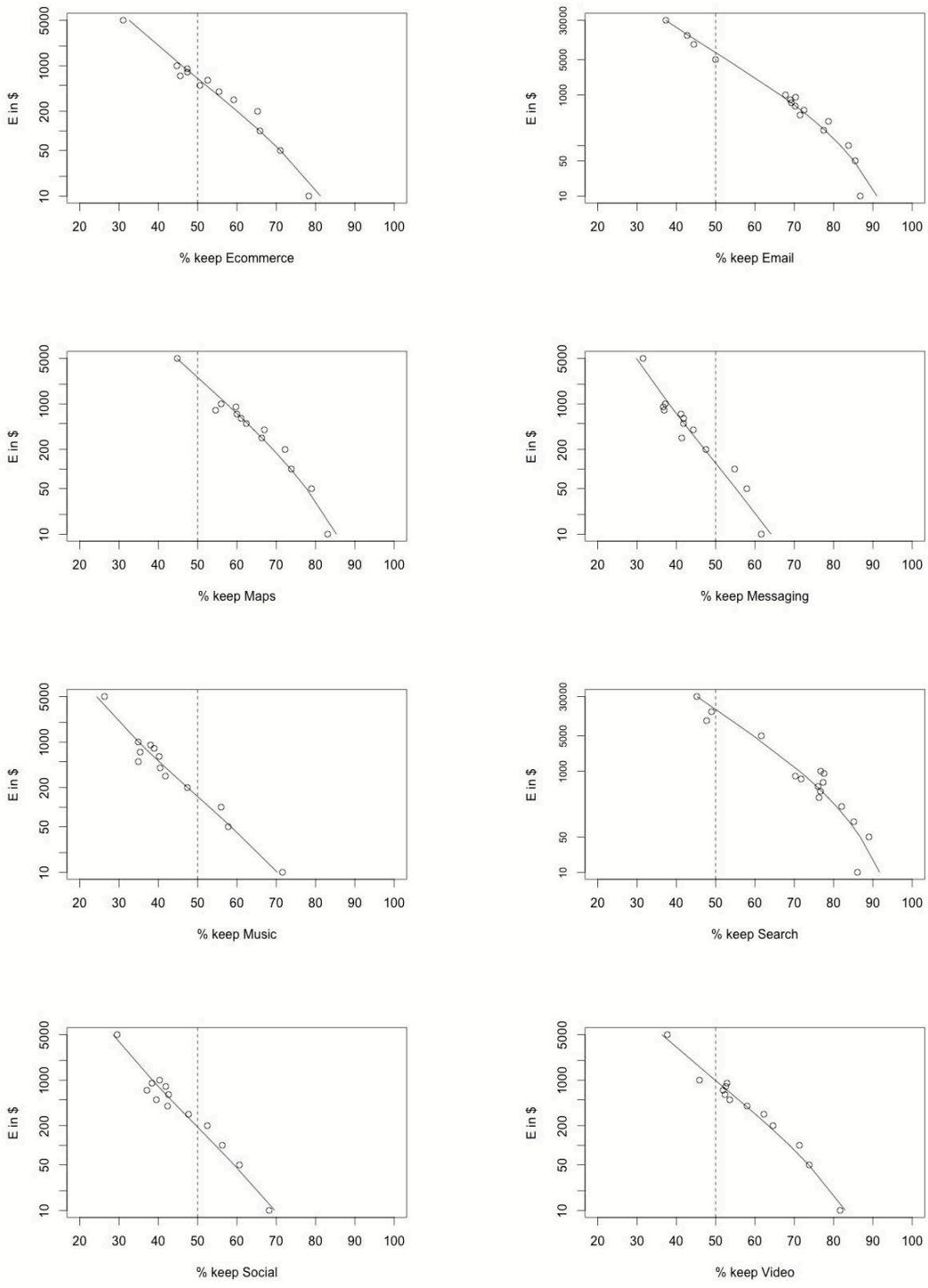



Figure 6: WTA demand curves for most widely used categories of digital goods

| | E-Commerce | Email | Maps | Messaging | Music | Search | Social | Video |
|---|---|---|---|---|---|---|---|---|
| **log(E)** | -0.352*** | -0.358*** | -0.321*** | -0.231*** | -0.324*** | -0.328*** | -0.285*** | -0.348*** |
| | (0.0190) | (0.0135) | (0.0198) | (0.0173) | (0.0181) | (0.0136) | (0.0178) | (0.0194) |
| **Male** | -0.110* | 0.0340 | -0.0565 | -0.165** | -0.0703 | -0.152** | -0.372*** | -0.0313 |
| | (0.0549) | (0.0545) | (0.0570) | (0.0532) | (0.0546) | (0.0557) | (0.0546) | (0.0564) |
| **Age 18-24 (Reference)** | | | | | | | | |
| **Age 25-34** | 0.154 | 0.221** | 0.272** | -0.00158 | -0.318*** | 0.169 | -0.132 | -0.0773 |
| | (0.0884) | (0.0842) | (0.0910) | (0.0862) | (0.0882) | (0.0883) | (0.0877) | (0.0944) |
| **Age 35-44** | 0.260** | 0.505*** | 0.415*** | 0.258** | -0.0122 | 0.263** | 0.172 | -0.192 |
| | (0.100) | (0.0933) | (0.103) | (0.0952) | (0.0975) | (0.0980) | (0.0982) | (0.102) |
| **Age 45-54** | 0.280** | 0.682*** | 0.368*** | 0.112 | -0.203* | 0.255* | 0.206* | -0.421*** |
| | (0.105) | (0.0996) | (0.106) | (0.102) | (0.101) | (0.102) | (0.103) | (0.106) |
| **Age 55-64** | 0.524*** | 0.947*** | 0.284** | 0.0397 | -0.403*** | 0.471*** | 0.00909 | -0.671*** |
| | (0.106) | (0.105) | (0.107) | (0.101) | (0.105) | (0.106) | (0.104) | (0.108) |
| **Age 65+** | 0.301** | 1.433*** | 0.119 | 0.121 | -0.499*** | 0.476*** | 0.0899 | -0.578*** |
| | (0.115) | (0.126) | (0.117) | (0.110) | (0.116) | (0.119) | (0.115) | (0.119) |
| **Age Unknown** | 0.0561 | 0.503*** | 0.0766 | -0.0426 | -0.160 | 0.132 | -0.249* | -0.427*** |
| | (0.0976) | (0.0934) | (0.104) | (0.0975) | (0.0990) | (0.0964) | (0.0980) | (0.101) |
| **Rural area (Reference)** | | | | | | | | |
| **Urban area** | -0.0154 | 0.216** | 0.167* | 0.0865 | 0.278*** | 0.175* | 0.103 | 0.154* |
| | (0.0775) | (0.0753) | (0.0795) | (0.0742) | (0.0768) | (0.0769) | (0.0766) | (0.0784) |
| **Suburban area** | -0.0240 | 0.103 | 0.164* | -0.0282 | 0.156* | 0.0791 | 0.0896 | 0.117 |
| | (0.0740) | (0.0715) | (0.0761) | (0.0712) | (0.0742) | (0.0735) | (0.0735) | (0.0752) |
| **Income 0-25k (Reference)** | | | | | | | | |
| **Income 25k-50k** | 0.0284 | 0.188* | 0.148 | 0.0521 | -0.00637 | 0.248** | 0.0623 | -0.121 |
| | (0.0943) | (0.0883) | (0.0960) | (0.0935) | (0.0936) | (0.0904) | (0.0958) | (0.0968) |
| **Income 50k-75k** | 0.00192 | 0.280** | 0.397*** | -0.0406 | 0.0244 | 0.409*** | -0.0124 | 0.0642 |
| | (0.101) | (0.0954) | (0.104) | (0.101) | (0.1000) | (0.0983) | (0.102) | (0.104) |
| **Income 75k-100k** | 0.220 | 0.258* | 0.135 | 0.0926 | -0.0488 | 0.432*** | -0.0297 | 0.0524 |
| | (0.131) | (0.127) | (0.136) | (0.128) | (0.132) | (0.128) | (0.133) | (0.134) |
| **Income 100k-150k** | 0.0730 | 0.104 | 0.592** | -0.128 | 0.0449 | 0.418* | -0.0128 | -0.140 |
| | (0.174) | (0.178) | (0.199) | (0.176) | (0.180) | (0.179) | (0.175) | (0.180) |
| **Income 150k+** | 0.523 | 0.120 | 1.201*** | -0.183 | -0.689* | 1.098** | 0.0355 | -0.255 |
| | (0.283) | (0.242) | (0.365) | (0.269) | (0.302) | (0.367) | (0.266) | (0.287) |
| **Income Unknown** | -0.0211 | 0.463 | 0.245 | 0.434 | 0.547 | 0.166 | -0.180 | -0.685* |
| | (0.249) | (0.254) | (0.275) | (0.278) | (0.279) | (0.268) | (0.287) | (0.277) |
| **Income Undisclosed** | 0.151 | -0.123 | -0.0201 | 0.0269 | -0.152 | 0.166 | -0.162 | -0.0139 |
| | (0.177) | (0.158) | (0.171) | (0.162) | (0.175) | (0.161) | (0.174) | (0.181) |
| **Constant** | 2.094*** | 2.273*** | 1.972*** | 1.087*** | 1.689*** | 2.631*** | 1.580*** | 2.665*** |
| | (0.164) | (0.142) | (0.170) | (0.154) | (0.158) | (0.149) | (0.160) | (0.172) |
| **Observations** | 6530 | 8079 | 6526 | 6600 | 6527 | 8081 | 6556 | 6572 |

*$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$*

Table 3: Estimated logistic functions for most widely used categories of digital goods

| Category | WTA/ year | 95% CI lower | 95% CI upper | n |
|---|---|---|---|---|
| **All Search Engines** | $16,628.97 | $12,469.69 | $22,561.61 | 8081 |
| **All Email** | $6,895.80 | $5,602.04 | $8,508.91 | 8097 |
| **All Maps** | $2,789.88 | $2,008.42 | $3,906.18 | 6526 |
| **All Video** | $935.99 | $757.54 | $1,165.04 | 6572 |
| **All E-Commerce** | $770.91 | $621.50 | $973.99 | 6530 |
| **All Social Media** | $187.79 | $146.62 | $238.84 | 6556 |
| **All Messaging** | $144.75 | $107.12 | $193.47 | 6600 |
| **All Music** | $144.16 | $116.87 | $179.07 | 6527 |
| **Facebook** | $59.72 | $44.97 | $79.52 | 8029 |

Table 4: Median WTA Estimates

Next we conducted another study where we attempt to measure consumers' valuation associated with all of their internet at home. There have been recent attempts finding that increase in internet speeds in an area results in higher property values (Molnar et al. 2013, Ahlfeldt et al. 2016). Molnar et al. (2015), using hedonic techniques, find that the availability of fiber internet (high speed internet) is associated with a $5,437 increase in property value for the median house in US. Ahlfeldt et al. (2016) study the diffusion of broadband internet in England over the period 1995-2010. Using discontinuities in regions served by different broadband delivery points, they find that the average property price increased by about 2.8% when internet access changes from slow dial-up connections to first generation broadband (ADSL). Moreover, the gains were highest among richest people living in most densely-populated areas.

We conduct SBDC experiments involving identical houses with and without internet access (across all devices) and varying monthly house rents (house with internet access being more expensive to rent). The difference in rent for the house with no internet access can be thought of as a measure of WTA losing home internet access (since the status quo is to have internet). Figure 7 plots the WTA demand curve of home internet access. Table 5 provides the estimates of the extended logit model for home internet access.

Our SBDC experiments indicate a median WTA of $383 per month per household for home internet. Assuming internet access fee of $100 per month, the median household still enjoys a significant surplus of $283 per month. Table 5 also shows that younger people (up to age 35) report more surplus than older people. Moreover, people residing in urban and suburban areas report higher surplus than people living in rural areas.

The monthly WTA corresponds to $4,594 per year which is equivalent to about 8.8% of annual median household income. Assuming a simple linear adoption rate over a 20 year period, our estimates imply an annual growth in surplus of around 0.4% of annual median household income. The actual growth likely experienced significant discontinuities (bumps) when households first shifted from dial-up to broadband and later from broadband to fiber optic internet. For instance, the total difference between

going from no Internet at all to current generation broadband is about three times larger than the value for the transition from dial-up to ADSL broadband in England, as estimated by Ahlfeldt et al.

The estimated average annual consumer surplus for US is around $719 billion for home internet which corresponds to around 4.2% of US GDP. Consumer surplus is computed conservatively by integrating the area under the demand curve only between the lowest and highest price levels used in our choice experiments. As noted in Greenwood and Kopecky (2011) and Brynjolfsson and Oh (2015), the constant elasticity assumption for demand can imply infinite valuations as the demand curve approaches the y-axis so it is more realistic to truncate it. This will have little or no effect on changes in consumer surplus from year to year, which is likely to be of greater interest for most purposes.

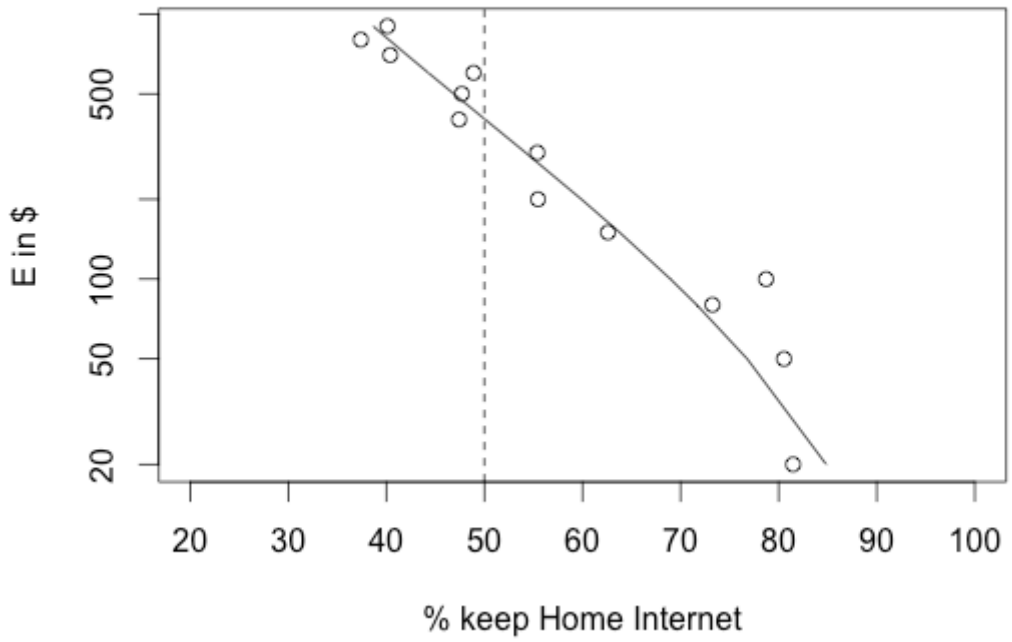



Figure 7: WTA demand curve for home internet

| | |
|---|---|
| **log(E)** | -0.346*** (0.0146) |
| **Male** | -0.386*** (0.0563) |
| **Age 18-24 (Reference)** | |
| **Age 25-34** | 0.00881 (0.0973) |
| **Age 35-44** | 0.454*** (0.102) |
| **Age 45-54** | 0.386*** (0.105) |
| **Age 55-64** | 0.581*** (0.107) |
| **Age 65+** | 0.663*** (0.120) |
| **Age Unknown** | 0.147 (0.104) |
| **Rural area (Reference)** | |
| **Urban area** | -0.108 (0.0789) |
| **Suburban area** | -0.119 (0.0755) |
| **Income 0-25k (Reference)** | |
| **Income 25k-50k** | 0.300** (0.105) |
| **Income 50k-75k** | 0.333** (0.111) |
| **Income 75k-100k** | 0.354* (0.141) |
| **Income 100k-150k** | -0.533* (0.221) |
| **Income 150k+** | -0.579 (0.357) |
| **Income Unknown** | -0.360 (0.292) |
| **Income Undisclosed** | -0.518* (0.211) |
| **Constant** | 0.248 (0.142) |
| **Observations** | 8029 |

*$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$*

Table 7: Estimated logistic function for home internet

*Facebook consumer surplus and growth calculation*

Based on the incentive compatible Facebook survey, we added usage and demographic variables to further understand differences in consumer value. The estimation results can be found in Table 8. As expected, the number of hours that consumers spend on Facebook per week (scale: 0 = "Less than 1 hour", 1 = "1 to 2 hours", 2 = "3 to 6 hours", 3 = "7 to 14 hours", 4 = "more than 14 hours") is a strong predictor for the value of Facebook. The same holds for the number of Facebook friends a consumer has (scale: 0 = "Less than 50", 1 = "51 to 100", 2 = "101 to 250", 3 = "251 to 500", 4 = "501 to 1000", 5 = "more than 1000"). These variables increase the overall probability to keep using Facebook in the SBDC question but do not moderate sensitivity towards E ($Sig._{Hours*log(E)} = 0.763$, $Sig._{Friends*log(E)} = 0.151$).

Gender and region of the consumer do not have an significant effect on the overall valuation of Facebook. There is a significant effect of age such that, conditional on using Facebook, older consumers are more likely to keep Facebook and, accordingly, value it more. A possible explanation is that younger consumers are more likely to switch to substitutes (e.g., Instagram, Snapchat), should Facebook not be available. Household income appears to have a negative effect. However, the negative trend stops for households with more than $150,000 income per year and for consumers who preferred not to disclose their income.

We can use the number of hours of Facebook usage as a proxy to sketch consumer valuation over time. Average Facebook usage has increased from 31 minutes per day in 2012 to over 50 minutes per day in 2016 (see Figure 8). Calculating the change in WTA due to increased intensity of usage (ceteris paribus) suggests that consumers' valuation of Facebook doubled from a median of $24.80 per month per user in 2012 ($297 per year) to $48.34 per month in 2016 ($580 per year, see Table 9). Integrating the probability function within the interval of analyzed price levels and multiplying with the number of active users in the US over time suggests that the national increase in consumer surplus from Facebook was $70 billion from 2015 to 2016.

| | beta | Std. Error | z | Sig. |
|---|---|---|---|---|
| **(Intercept)** | 0.402 | 0.272 | 1.475 | 0.140 |
| **log(*E*)** | -0.342 | 0.032 | -10.764 | <0.001 |
| **Facebook hours usage per week** | 0.306 | 0.049 | 6.246 | <0.001 |
| **Facebook number of friends** | 0.152 | 0.044 | 3.423 | 0.001 |
| **Gender (male)** | -0.050 | 0.118 | -0.420 | 0.675 |
| **Age 18-24 (reference level)** | (0.000) | | | |
| **Age 25-34** | 0.011 | 0.206 | 0.052 | 0.959 |
| **Age 35-44** | 0.223 | 0.204 | 1.089 | 0.276 |
| **Age 45-54** | 0.438 | 0.202 | 2.164 | 0.030 |
| **Age 55-64** | 0.485 | 0.206 | 2.353 | 0.019 |
| **Age 65+** | 0.972 | 0.225 | 4.323 | <0.001 |
| **US region Northeast (reference)** | (0.000) | | | |
| **US region Midwest** | -0.222 | 0.178 | -1.248 | 0.212 |
| **US region South** | -0.296 | 0.160 | -1.847 | 0.065 |
| **US region West** | -0.171 | 0.177 | -0.968 | 0.333 |
| **Income less than 25K (reference)** | (0.000) | | | |
| **Income 25K to 50K** | -0.063 | 0.192 | -0.330 | 0.741 |
| **Income 50K to 100K** | -0.260 | 0.176 | -1.474 | 0.140 |
| **Income 100K to 150K** | -0.545 | 0.209 | -2.609 | 0.009 |
| **Income 150K or more** | 0.313 | 0.214 | 1.461 | 0.144 |
| **Income "prefer not to answer"** | 0.704 | 0.354 | 1.988 | 0.047 |

LL = -1057.6, df = 18, McFadden $R^2$ = 0.11, Hit rate = 0.66

Table 8: Facebook value diagnostic

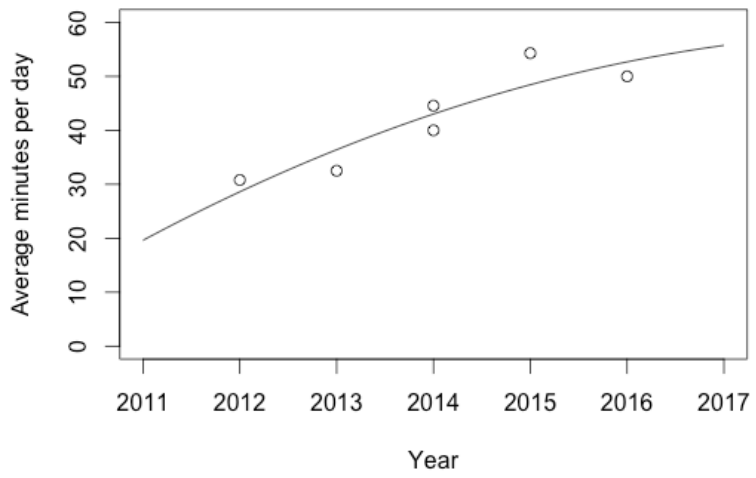



Figure 8: Facebook hours of usage over time

| Year | Number of active US users (in millions) | Median WTA per user per year | Annual Change in Consumer Surplus (in millions) |
|---|---|---|---|
| 2012 | 183 | $297.58 | |
| 2013 | 195 | $369.80 | $76,699 |
| 2014 | 202 | $444.61 | $57,057 |
| 2015 | 210 | $517.17 | $57,895 |
| 2016 | 222 | $580.14 | $69,706 |

Table 9: Development of consumer surplus for Facebook

*Sensitivity analysis*

*Effect of random answers*

Error variance in the consumer's decisions has a negative effect on the precision, i.e., scale of the estimates S (Hauser, Eggers, Selove 2016). Specifically, the scale S is inversely proportional to the error variance. The scale S cannot be separately identified from the estimates, such that:

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

Lower scaled estimates (more error), i.e., estimates with lower magnitude, cause the logit function to become linear. Higher scaled estimates (less error) lead to a stepwise function that allows to predict decisions and identify the the median WTA more precisely (see Figure 9).

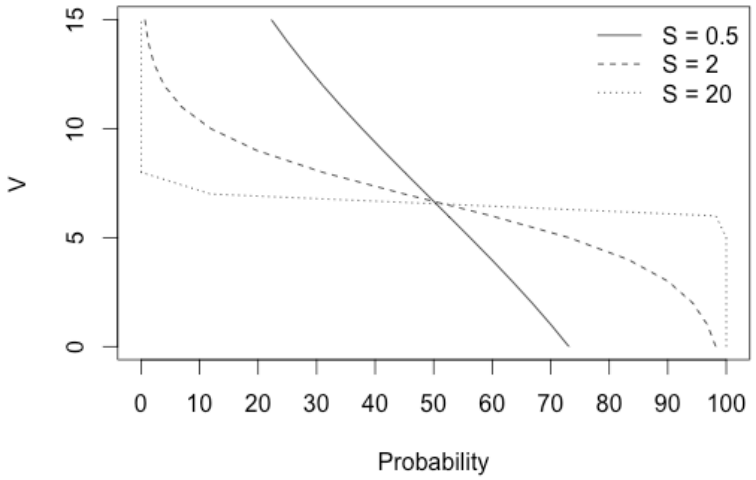



Figure 9: Effect of scale of the estimates on logit function

The effect is demonstrated in Table 10. The table shows the result of a modified bootstrapping procedure in which subsamples were drawn from the IC Facebook sample. In each subsample we replaced R random original responses with the same amount of random answers and estimated the logit model. This process was repeated 1000 times in order to be able to observe the variance of the results. The results show that more random answers decrease 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 mean WTA as well as the standard error of the estimates remain largely unaffected. The same results are obtained for the NIC group.

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

Table 10: Effect of random answers

To get an estimate about the magnitude of error in the overall sample, we compared how many choices can be recovered correctly based on the estimated probabilities and a cutoff value of 0.5. Overall this measure results in the internal hit rate of correct predictions. For the IC group the hit rate is 62.1%, i.e., 37.8% of the choices cannot be predicted correctly based on the amount of E alone. Interestingly, the hit rate for the NIC group is higher, 69.9%. This suggests that consumers in the IC group faced a decision that was more difficult to make, likely because their choices were consequential.

*Effect of sample size*

The reliability of the WTA estimates 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. Table 11 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. We therefore show how the 95% confidence interval develops when increasing the sample size. There is uncertainty in the measure even with a sample size of 1500. Simulating additional responses shows that a sample of 6000 consumers would reduce the range of the confidence interval from more than $40 to $20.

| Sample size | Mean intercept | Mean beta log (E) | **Std. error Intercept** | **Std. error beta log(E)** | mean 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 11: Effect of sample size

**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[6] and Adobe's Digital Price Index[7] 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 due to technological advancements. We explore a potential way of measuring changes in consumer surplus through SBDC experiments. Our method is highly scalable and relatively inexpensive. Therefore, it can be run at very frequent, regular intervals to keep track of changes in consumer surplus. As argued previously, this measure can be an important complementary indicator of consumer well-being for the digital economy.

Apart from the limitations associated with choice experiments which were mentioned previously, a major limitation of our study is the lack of precision in our estimates. While the BEA is able to measure GDP very precisely (e.g. US GDP was reported as $16,514,593,000 on the first day of 2016), we are only able to provide a relatively coarse estimate of changes in consumer surplus. That said, our approach is at least attempting to directly measure a concept that is not correctly measured by other official data. In short, we believe it is better to be approximately correct than precisely wrong. Another limitation of our study is that it is biased towards people using the internet. 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[8]. Accordingly, our results must be interpreted as relevant to this audience, but not necessarily others.

We have several next steps lined up in this research project. First, to mitigate the concerns associated with hypothetical bias involved in choice experiments, we plan to run several more incentive compatible choice experiments for other digital goods in the laboratory and in the field. Documenting the hypothetical bias for a variety of goods will help us in identifying the correction factor that needs to be applied to our consumer surplus estimates. Second, apart from SBDC experiments, we are also exploring several other types of choice experiments including best-worst scaling (BWS) and pairwise comparisons of goods to determine the best approach for measuring WTA estimates of digital goods. Finally, we are planning to expand our study to several countries worldwide to compare the well-being of consumers across these countries.

**(Note to reader: Results of several more incentive compatible experiments and studies comparing discrete choice experiments with pairwise comparisons/ conjoint analyses will be available by 17 Feb 2017 when finished papers are due.)**

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

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

[8] http://www.pewresearch.org/fact-tank/2015/07/28/15-of-americans-dont-use-the-internet-who-are-they/

**References**


Aeppel, T. (2015). Silicon Valley Doesn't Believe U.S. Productivity Is Down. The Wall Street Journal. (Accessed at: http://www.wsj.com/articles/silicon-valley-doesnt-believe-u-s-productivity-is-down-1437100700)

Ahlfeldt, G., P. Koutroumpis, T. Valletti, (2016) Speed 2.0: Valuing universal access to digital highways, *Journal of the European Economic Association*.

Ariely, Dan; Loewenstein, George and Prelec, Drazen. " 'Coherent Arbitrariness': Sta- ble Demand Curves without Stable Preferences." *Quarterly Journal of Economics*, February 2003, 118 (1), pp. 73-105.

Balistreri, E., McClelland, G., Poe, G., & Schulze, W. (2001). Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values.*Environmental and Resource Economics*, *18*(3), 275-292.

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. and Heberlein, T.A., 1979. Measuring values of extramarket goods: Are indirect measures biased?. *American journal of agricultural economics*, *61*(5), pp.926-930.

Brynjolfsson, E. and Oh, J., 2012. The attention economy: measuring the value of free digital services on the Internet.

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

Brynjolfsson, E., & Saunders, A. (2009). What the GDP Gets Wrong.(Why Managers Should Care). *MIT Sloan Management Review*, *51*(1), 95.

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

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., Groves, T., 2007. Incentive and informational properties of preference questions, *Environmental and Resource Economics*, 37(1), pp.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), pp.171-207.

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), pp.257-286.

Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *The Journal of Economic Perspectives*, *30*(2), 151-178.

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), pp.45-64.

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

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

File, T. and Ryan, C. 2014. Computer and Internet Use in the United States: 2013, U.S. CENSUS BUREAU, http://www.census.gov/history/pdf/2013computeruse.pdf

Furman, J., & Orszag, P. (2015). A Firm-Level Perspective on the Role of Rents in the Rise in Inequality.

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. and Klenow, P.J., 2006. *Valuing consumer products by the time spent using them: An application to the Internet* (No. w11995). National Bureau of Economic Research.

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

Greenstein, S., & McDevitt, R. C. (2009). The broadband bonus: Accounting for broadband Internet's impact on US GDP (No. w14758). National Bureau of Economic Research. Chicago

Greenwood, J. and Kopecky, K.A. 2011. "Measuring the Welfare Gain from Personal Computers: A Macroeconomic Approach" Federal Reserve Bank of Atlanta Working Paper No. 2011-05. Available at SSRN: http://ssrn.com/abstract=1800266

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), pp.635-647.

Hatzius, J. (2015). Productivity paradox 2.0. Goldman Sachs Global Macro Research, 39, 6-7. (Accessed at: 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, Cambridge, MA: MIT Sloan School of Management

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

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), pp.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), pp.229-254.

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

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), pp.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)

Molnar, G., Savage, S., & Sicker, D. (2015). Reevaluating the Broadband Bonus: Evidence from Neighborhood Access to Fiber and United States Housing Prices. *Available at SSRN 2628396*.

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

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, http://news.peanutlabs.com/rs/peanutlabs/images/PL-PanelBook_2015.pdf

Pepitone, J. 2012. Encyclopedia Britannica to stop printing books. CNN. Available at: http://money.cnn.com/2012/03/13/technology/encyclopedia-britannica-books/ [Accessed March 21, 2016].

Plott, C.R. and Zeiler, K., 2005. The Willingness to Pay–Willingness to Accept Gap, the. *The American Economic Review*, *95*(3), pp.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), pp.29-31.

Spence, M., B. Owen. 1977. Television programming, monopolistic competition, and welfare. *Q. J. Econ.* 103–126.

Stephens-Davidowitz, S., H. Varian. 2015. A Hands-on Guide to Google Data. Available at: http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf [Accessed March 21, 2016].

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2010). 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.

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

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

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), pp.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.

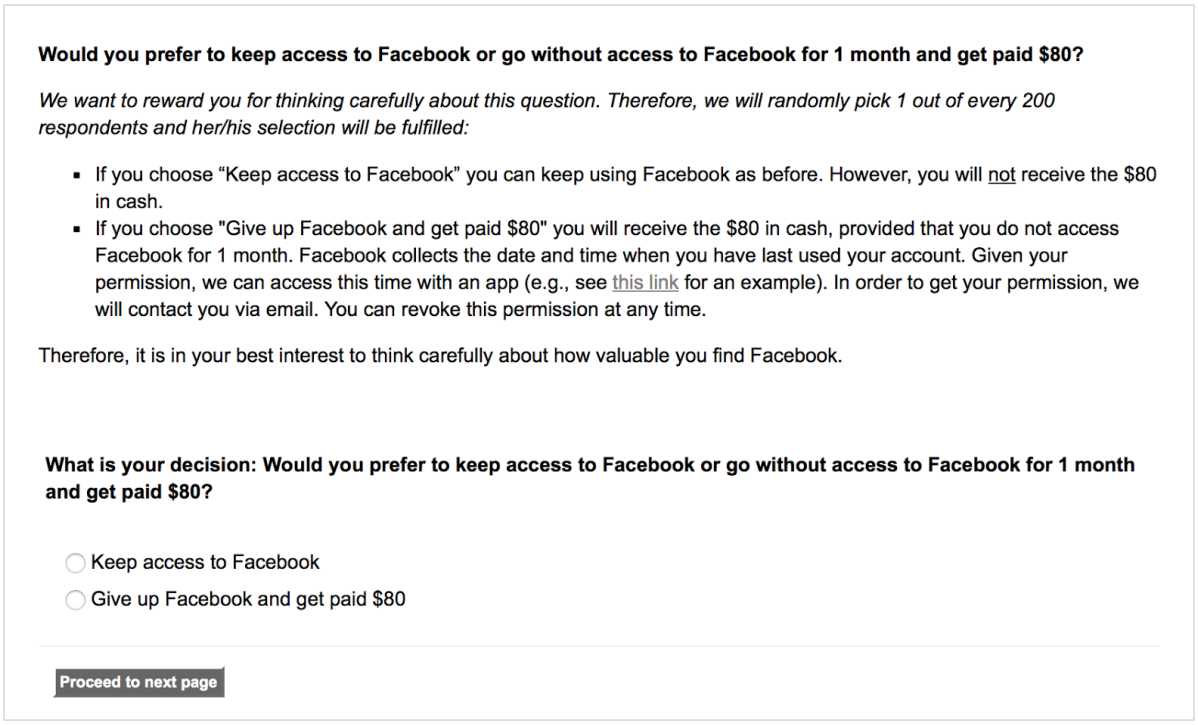
Whitehead, J.C., 2002. Incentive incompatibility and starting-point bias in iterative valuation questions. *Land Economics*, *78*(2), pp.285-297.

Williams, B. 2008. " A Hedonic Model for Internet Access Service in the Consumer Price Index." Monthly Labor Review, July, pp.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), pp.195-210.

## Appendix

*Example of IC Questionnaire WTA Facebook*



**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 10: Screenshot incentive compatible SBDC question (for E = $80)

[Follow-up if second option was chosen]
Since you have chosen "Give up Facebook and get paid $E" in the previous question, please enter your email address in the following field. We need to know your email address in order to contact you if you are the randomly chosen respondent. We will not contact you otherwise. We will keep this information private and will not share it with others.

Email: _____