



Using massive online choice experiments to measure changes in well-being

Erik Brynjolfsson^{a,b,1}, Avinash Collis^{a,1,2}, and Felix Eggers^{c,1}

^aSloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142; ^bNational Bureau of Economic Research, Cambridge, MA 02138; and ^cFaculty of Economics and Business, University of Groningen, 9747 AE Groningen, The Netherlands

Edited by Charles Bean, London School of Economics, London, and accepted by Editorial Board Member Paul R. Milgrom February 24, 2019 (received for review September 10, 2018)

Gross domestic product (GDP) and derived metrics such as productivity have been central to our understanding of economic progress and well-being. In principle, changes in consumer surplus provide a superior, and more direct, measure of changes in well-being, especially for digital goods. In practice, these alternatives have been difficult to quantify. We explore the potential of massive online choice experiments to measure consumer surplus. We illustrate this technique via several empirical examples which quantify the valuations of popular digital goods and categories. Our examples include incentive-compatible discrete-choice experiments where online and laboratory participants receive monetary compensation if and only if they forgo goods for predefined periods. For example, the median user needed a compensation of about \$48 to forgo Facebook for 1 mo. Our overall analyses reveal that digital goods have created large gains in well-being that are not reflected in 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 the existing national income and product accounts.

consumer surplus | digital goods | free goods | GDP | choice experiments

Digital technologies have transformed the types of goods and services consumed in modern economies. However, our national measurement framework for economic growth and well-being has not fundamentally changed since its invention in the 1930s. Gross domestic product (GDP) and derivative metrics like productivity (typically calculated as GDP/hours worked) dominate discussions of economic growth and performance. In principle, a more comprehensive approach is now feasible. By using massive online choice experiments we can estimate changes in consumer surplus, the primary component of economic welfare, and thereby supplement the traditional metrics based on GDP.

GDP measures the real 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. Some economists, policymakers, and journalists routinely use GDP as if it were a measure of well-being (1, 2). Nonetheless, while it is a good measure of production, GDP is a significantly flawed measure of well-being (2–4). [Indeed, Simon Kuznets, who was instrumental in developing our system of national accounts, said in 1934 that “the welfare of a nation can scarcely be inferred from [GDP].” In Brynjolfsson et al.^{*}, we more formally discuss the relationship of GDP to welfare in the context of choice experiments.] Attempts have been made to design alternative measures, typically focusing on measuring subjective well-being and life satisfaction. Despite progress, these measures remain very imprecise (5) and 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 (6).

The traditional national accounts are especially problematic as metrics of well-being when prices are zero and thus are absent or largely absent from GDP (7). This is increasingly the case for goods in the emerging digital economy because each user’s copy of a digital good, such as Wikipedia and most smartphone applications, has nearly zero marginal cost and often a zero market price. For instance, although information goods have unquestionably become increasingly ubiquitous and important in our daily lives, the official share of the information sector as a fraction of the total nominal GDP (~4 to 5%) was the same in 2016 as it was 35 y earlier. Moreover, in many sectors (e.g., music, media, and encyclopedias) people substitute zero-price online services (e.g., Spotify, YouTube, and Wikipedia) for goods with a positive price (e.g., CDs, DVDs, and *Encyclopedia Britannica*). 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 variety of digital goods (see *SI Appendix* for an exploration of when GDP and welfare are positively correlated and when they are uncorrelated or even negatively correlated).

To assess changes in living standards, and by extension the effects of policies that might affect living standards, it is necessary to properly measure the welfare gains from all goods. This includes goods without positive market prices, including many digital goods, public goods, and environmental goods. Because goods with zero price have zero contribution to GDP, the welfare

Significance

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. We propose a way of directly measuring consumer well-being using massive online choice experiments. We find that digital goods generate a large amount of consumer welfare 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.

Author contributions: E.B., A.C., and F.E. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. C.B. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹E.B., A.C., and F.E. contributed equally to this work.

²To whom correspondence should be addressed. Email: avinashg@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1815663116/-DCSupplemental.

Published online March 26, 2019.

^{*}Brynjolfsson E, Collis A, Diewert WE, Eggers F, Fox KJ (2019) GDP-B: Accounting for the value of new and free goods in the digital economy. NBER Working Paper (National Bureau of Economic Research, Cambridge, MA).

gains from such goods are not properly captured in GDP statistics. Our approach uses massive online choice experiments to measure these welfare gains. In this paper, we focus on measuring welfare gains from digital goods in particular because of the rapid pace of innovation and adoption of these goods, which suggests that they may have a particular important effect on the changes in living standards. For instance, the average American spends 22.5 h per week online as of 2018 (8). Facebook, launched in 2004, had 2.27 billion active users worldwide as of September 2018 and the average user spent 50 min per day on Facebook and Instagram, up from zero in 2005 (9). WhatsApp, launched in 2009, had 1.5 billion active users worldwide as of January 2018 (10). These digital innovations either created completely new goods that did not exist before or replaced and significantly improved previously existing nondigital goods. For example, Google (11) and Wikipedia (12) have more quantity and better-quality results than libraries and physical encyclopedias. Therefore, the changes in welfare gains are likely to be larger for digital goods than for other goods which have not changed as radically. [Brynjolfsson and Oh (13), using an alternative approach to estimate consumer surplus based on time spent, find that the annual welfare gain in consumer surplus from free internet services was significantly higher than the welfare gain from television use.]

Consumer surplus is defined as the difference between the consumers' willingness to pay for a good and the amount that they actually pay. For instance, if a person were willing to pay up to \$100 for a pair of shoes but only had to pay \$70, then that person would gain \$30 of consumer surplus from that transaction. Economists consider the changes in consumer surplus as a measure of changes in consumer well-being (or welfare). Total well-being includes both consumer well-being and producer well-being. On average, producers are estimated to capture only 2.2% of the total welfare gains from innovation, with consumers capturing the remaining surplus (14). Thus, changes in consumer surplus are a good proxy for changes in overall well-being, especially when the ratio of consumer surplus to producer surplus is not changing rapidly. That said, for goods with market prices and appropriate quality adjustments of these prices over time, changes in real GDP can also be a good proxy for changes in well-being (7). However, GDP's usefulness as a proxy breaks down for goods which have a zero market price (15). [Some goods such as WhatsApp and Wikipedia do not have any advertising revenues, either. Other goods such as Google Search and Facebook have advertising revenues but the welfare gains from these goods need not be correlated with advertising revenues (16).]

Historically, changes in consumer surplus have not been widely used as a measure of economic progress. This reflects the fact that it has been difficult to measure consumer surplus at scale. Measuring consumer surplus typically requires estimating demand curves based on exogenous variations that shift the supply curve but not the demand curve, and it has not been practical to identify these variations using traditional market data for large sets of goods. However, with advances in digital technologies, it is now feasible to collect data about thousands of goods much more easily. In this research, we stick more closely to a traditional microeconomic framework than the subjective well-being research and propose a way of measuring changes in consumer surplus using experimental variation via online choice experiments. This approach is not only applicable to free goods and services in the digital economy but also more broadly to conventional goods (see *SI Appendix* for an example of measuring a nondigital good). Choice experiments provide more flexibility than market data because they do not require nonzero prices or market transactions to exist and are frequently applied in contingent valuation studies. These experiments allow us to estimate the demand curves for any good using data from thousands of consumers that are representative of the national

population. 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.

Methods and Results

We implement three distinct types of choice experiments, single-binary discrete-choice (SBDC) experiments (17), Becker-DeGroot-Marschak lotteries (BDM) (18), and best-worst scaling (BWS) (19), and find they have similar implications (see the end of this section and *SI Appendix* for details). For ease of exposition, we will focus on the SBDC approach, which involves consumers making a single choice among two options: whether to keep access to a certain good or to forego the good in return for receiving a specific amount of money. We only ask one question per consumer and vary the price points systematically across thousands of consumers for each experiment. We thereby obtain willingness to accept (WTA) valuations (i.e., the monetary compensation needed to compensate losing access to various goods).

We illustrate the method using Facebook to measure the consumer surplus with SBDC choice experiments. To avoid any bias that may affect consumer choices when the options are purely hypothetical choices, we applied the SBDC experiment in a nonhypothetical, incentive-compatible procedure to measure the consumer surplus. Incentive-compatible choice experiments make responses consequential. Specifically, it is in the best interest of respondents to reveal their true preferences. We asked consumers if they would prefer to (i) keep access to Facebook or (ii) give up Facebook for 1 mo in return for a payment of \$*E* (in *SI Appendix* we address sensitivity of the valuation depending on the time frame and show that our approach can detect consumers' sensitivity toward different time frames). We varied \$*E* systematically across several price points ranging between \$1 and \$1,000. To make the SBDC question consequential for the consumer, we informed them that we would randomly pick 1 out of every 200 respondents and fulfill that person's selection (i.e., they get the \$*E* cash at the end of the month after we verify that they have not been on Facebook for the month; see *SI Appendix* for more details on the experiment). We recruited a representative sample of US Facebook users from a professional market research firm in summer 2016 and again in 2017 to measure annual changes in consumer surplus obtained from Facebook.

Fig. 1 plots the estimated WTA demand curves, separated for 2016 and 2017. In 2016, the sample's median WTA was \$48.49 for giving up 1 mo of Facebook, and this valuation dropped to \$37.76 in 2017. We used bootstrapping to calculate 95% CIs for

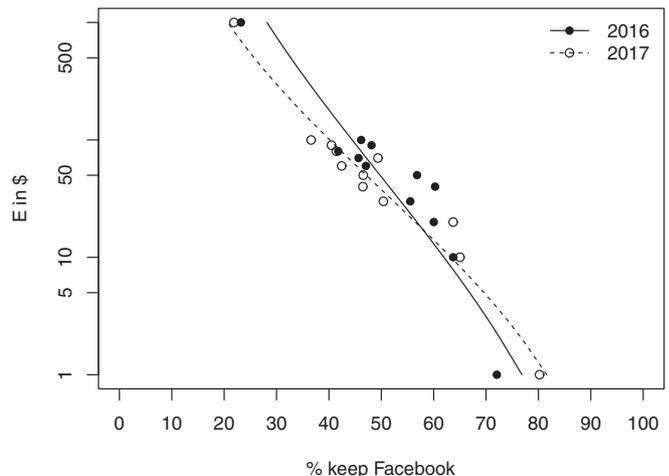


Fig. 1. WTA demand curves for Facebook in 2016 and 2017.

the median WTA values, that is, $CI_{2016} = [\$32.04, \$72.24]$, $CI_{2017} = [\$27.19, \$51.97]$. Although the median WTA values suggest a drop in value, the CIs are fairly broad so that we cannot establish, based on the data using this sample size from 2 y, whether the decrease in value follows a systematic trend (in *SI Appendix* we address the effect on precision by using larger sample sizes in the sensitivity analyses).

We added usage and demographic variables to further understand heterogeneity in consumer surplus. The usage of Facebook per week is a significant predictor for the value of Facebook, providing a reality check for our approach. Consumers value Facebook more if they spend more time on it or have more friends. Moreover, the more they post status updates or share pictures and videos and the more they like and comment and play games, the more they value Facebook. Consumers who reported using Instagram or YouTube value Facebook significantly less than those who do not use these services. Therefore, Instagram and YouTube can be considered to be substitutes for Facebook to an extent. In terms of sociodemographics, we find significant effects for gender and age of the respondent, as well as household income. Specifically, for any offer of payment of \$E, female respondents are more likely to keep Facebook than male users. The same holds for older consumers. The effects for household income are not monotonic. Compared with low-income households, households with an income between \$100,000 and \$150,000 perceive significantly less value in Facebook, while households with income above \$150,000 value Facebook more. Overall, these results indicate that Facebook provides substantial value to consumers. They would require a median compensation of \$40 to \$50 for leaving this service for a month.

We extend the experiment to additional popular digital goods in a laboratory setting in Europe. Although the sample consists of students and is not necessarily representative of the general population, we used a laboratory to have more control monitoring the usage of the goods to further explore the effects of incentive compatibility. We find that WhatsApp, Facebook, and digital maps on phones are highly valued by our subjects with median compensations for losing 1 mo of access of €536, €97, and €59, respectively. Other applications such as Instagram (€6.79), Snapchat (€2.17), and LinkedIn (€1.52) are valued an order of magnitude lower and Skype (€0.18) and Twitter (€0.00) have very low median valuations. (Average valuations or valuations for any given consumer will typically differ from median valuations.) In follow-up interviews, respondents reported that the strikingly high values for WhatsApp reflected its tight integration into their daily lives for coordination with family, friends, colleagues, schoolmates, and others and the high compensation needed for being digitally separated from this network.

We also ran larger-scale choice experiments on a representative sample of the US internet population using Google Surveys,

which are well-suited to implement our one-question SBDC experiments. Although these experiments are not incentive-compatible, we are able to access a much larger population at a fraction of the cost and get a more precise surplus estimate. Moreover, by focusing on relative changes between 2016 and 2017 rather than the absolute magnitude of the valuations, any hypothetical bias from lack of incentive compatibility is likely to be mitigated as long as the bias in one year is similar to the bias in the following year. Thus, a change in measured valuations should reflect a real difference. We identified the most widely used applications 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 also in summer 2016 and 2017. In these studies, we asked consumers to consider giving up access to these categories for 1 y. The counterfactual that consumers are provided in these studies is to lose access to all options within a category (e.g., all search engines or all social media) for a year. We applied 6 to 15 price levels for each product category and gathered ~500 responses for each price level for each year ($n = 64,940$). According to the median estimates for 2017 (Table 1), 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 is that for many people these services are essential to their jobs, making them reluctant to give up these goods even in exchange for high monetary values. What's more, we asked for the value of the entire category such that there are no effective between-category substitutes. Because most consumers do not directly pay for these services, almost all of the WTA for these goods contributes toward consumer surplus.

Video streaming services (e.g., YouTube and 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 to \$20 per month, or \$120 to \$240 per year. Our measure suggests that the surplus the median consumers receive from these goods is a 5 to 10 multiple of what they actually pay. Recall that the payment is visible in national accounts, but not the consumer surplus. 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; see *SI Appendix* for CIs and additional WTA percentiles).

As a benchmark to these hypothetical SBDC experiments, we conducted additional choice experiments based on the BWS approach (19). BWS asks consumers to repeatedly select the best and worst options from experimentally varied sets of alternatives. Consumers are required to make a trade-off when deciding which goods they perceive as most and least valuable. This may

Table 1. Median WTA estimates for most popular digital goods categories

| Category | WTA per year 2016, \$ | WTA per year 2017, \$ | 95% CI WTA per year 2016, \$ | | 95% CI WTA per year 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 |

media (22)], negative effects on the users' own mental or physical health, which users do not fully appreciate (23), or varying degrees of lock-in. Of course, the same can be true for any goods purchased in markets, as reflected in GDP and related statistics. An advantage of the choice-experiment approach is that it is possible to make externalities or other aspects of user choices salient, vary time periods, and observe the resulting effect on the valuations. Thus, we can better understand the distinctions between private and social valuations or short- and long-term valuations. For example, we find that the median valuation for giving up Facebook for 2 wk is 2.7 times the median valuation for giving it up for 1 wk, and the median valuation for giving up Facebook for 1 mo is 4.5 times the median valuation for giving it up for 1 wk, suggesting a nonlinear relationship between valuations and time periods. This positive nonlinear relationship is confirmed when keeping the cash amount fixed and varying only the time period (see *SI Appendix* for more details on the effect of time periods on valuations).

Our method is highly scalable and relatively inexpensive. Market research products such as Google Surveys let us reach representative samples of internet populations, and a single response to an SBDC experiment costs as little as 10 cents. Therefore, we can run these SBDC experiments at frequent, regular intervals to track changes in consumer surplus for many (digital) goods and categories. This measure provides valuations at a more detailed level than aggregate measures of subjective well-being and can be an important complementary indicator of consumer well-being for the digital economy. Moreover, we can also use the same approach to estimate the welfare gains for physical goods (e.g., in *SI Appendix* we give the example of breakfast cereal) and other nonmarket goods (such as environmental goods and public goods provided by the government). Because GDP is a measure of production and not welfare, this can help address an important and long-standing gap in our understanding of the economy.

Hypothetical choice experiments, as shown above using Google Surveys, can be easily and cheaply conducted online, but the stated preferences might suffer from hypothetical bias. However, the differences between hypothetical and incentive-compatible approaches are much less severe when analyzing annual changes in valuations, rather than absolute levels. GDP growth is often considered to be more relevant to policymakers than absolute levels. Similarly, changes in consumer surplus valuations across time are more relevant than absolute valuations. To address concerns of bias associated with answering hypothetical questions, we also conducted incentive-compatible choice experiments for Facebook using a representative sample of the internet population and other popular digital goods in a laboratory setting. Incentive-compatible choice experiments are

harder to conduct online but they provide accurate estimates of revealed preferences.

A major limitation of our study remains the relative lack of precision in our estimates. Compared with GDP, we are only able to provide a relatively coarse estimate of changes in consumer surplus given our sample size. Although the median WTA is robust to random noise in the data (see *SI Appendix* for sensitivity analyses of the effect of random noise and sample size on media WTA), the overall demand schedule, including very high or low values, is not: A small fraction of consumers with 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 population or requires targeting a sample of users of these goods (in *SI Appendix* we also explore other key percentiles such as the 25th and 75th percentiles). Future work should use more massive sample sizes to narrow the CI of the WTA estimates. Moreover, 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 toward people using the internet. The choice experiments are only accessible online, and therefore people not using the internet at all [around 11% of the US population (24)] are excluded.

Despite their limitations, the choice experiments we conduct are 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.

Materials and Methods

See *SI Appendix* for a detailed description of all materials and methods used within this study as well as additional discussion on why GDP and welfare need not be correlated and summaries of previous research on measuring welfare gains from digital goods. Our studies were approved by the Institutional Review Boards (IRB) of the Massachusetts Institute of Technology and the University of Groningen. For the short hypothetical SBDC experiments run on Google Surveys, informed consent was not required as the IRB determined that these studies were exempt. For all of the remaining studies, including the incentive-compatible SBDC experiments, informed consent was obtained on the first page of the study.

ACKNOWLEDGMENTS. We thank Susanto Basu, Carol Corrado, Erwin Diewert, Kevin Fox, Jana Gallus, Robert Hall, John Hauser, Leonard Nakamura, Hal Varian, and participants of the Conference on Research on Income and Wealth (2016) and the American Economic Association annual meeting (2018) for helpful comments. We thank the Massachusetts Institute of Technology Initiative on the Digital Economy, via a grant from the Markle Foundation, for generous funding.

1. Federal Reserve Bank of Boston (2003) How do we measure standard of living? Available at <https://www.bostonfed.org/-/media/Documents/ledger/ledger2003/measure.pdf>. Accessed January 19, 2019.
2. Feldstein M (2017) Underestimating the real growth of GDP, personal income, and productivity. *J Econ Perspect* 31:145–164.
3. Stiglitz J, Sen A, Fitoussi JP (2009) Report by the commission on the measurement of economic performance and social progress. Available at <https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report>. Accessed January 19, 2019.
4. Bean C (2016) Independent review of UK economic statistics (Cabinet Office, London). Available at <https://www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-final-report>. Accessed January 19, 2019.
5. Krueger AB, Stone AA (2014) Psychology and economics. Progress in measuring subjective well-being. *Science* 346:42–43.
6. den Haan W, Ellison M, Ilzetzki E, McMahon M (2017) Happiness and well-being as objectives of macroeconomic policy: Views of economists. Available at <https://voxeu.org/article/views-happiness-and-wellbeing-objectives-macroeconomic-policy>.
7. Dynan K, Sheiner L (2018) GDP as a measure of economic well-being. Available at <https://www.brookings.edu/research/gdp-as-a-measure-of-economic-well-being/>. Accessed January 19, 2019.
8. Center for the Digital Future at USC Annenberg (2019) The 2018 digital future report. Available at <https://www.digitalcenter.org/wp-content/uploads/2018/12/2018-Digital-Future-Report.pdf>. Accessed January 8, 2019.
9. Facebook (2019) Facebook company information. Available at <https://newsroom.fb.com/company-info/>. Accessed January 8, 2019.
10. Constine J (2018) WhatsApp hits 1.5 billion monthly users. \$19B? Not so bad. Available at <https://techcrunch.com/2018/01/31/whatsapp-hits-1-5-billion-monthly-users-19b-not-so-bad/>. Accessed January 8, 2019.
11. Chen Y, Jeon GY, Kim Y-M (2014) A day without a search engine: An experimental study of online and offline searches. *Exp Econ* 17:512–536.
12. Giles J (2005) Internet encyclopaedias go head to head. *Nature* 438:900–901.
13. Brynjolfsson E, 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*. Available at <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1045&context=icis2012>.
14. Nordhaus WD (2004) Schumpeterian profits in the American economy: Theory and measurement. NBER Working Paper No. 10433 (National Bureau of Economic Research, Cambridge, MA).
15. Groshen EL, Moyer BC, Aizcorbe AM, Bradley R, Friedman DM (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. *J Econ Perspect* 31:187–210.

