

The Census Bureau Needs to Significantly Revise Reporting and Calculation of Its Online and Physical Retail Sales Figures and Commission an Independent Review

David S. Evans, Scott Murray, and Richard Schmalensee *

February 7, 2016

Abstract: This paper reports significant discrepancies in the Census's methods and reports of online and physical retail sales in the United States. We recommend material changes in how the Census Bureau reports e-commerce and physical sales reported to it. We also recommend that the Census commission an independent review of its methods and reporting.

* Evans is Founder, Market Platform Dynamics, Visiting Professor, University College London, and Lecturer, University of Chicago Law School. Murray is Managing Director of Data Analytics, Market Platform Dynamics. Schmalensee is Howard W. Johnson Professor of Management and Economics Emeritus at MIT and Dean Emeritus of the MIT Sloan School of Management. Email correspondence to david.evans@marketplatforms.com.

We released a [paper](#) on January 18, 2016 that reported that the Census Bureau significantly undercounts online retail sales in the US and presented estimates of the e-commerce share of retail under the assumption that the Census Bureau follows the methodology that it reported publicly and explained to us in a series of communications.¹ In fact, the Census Bureau does not follow the methodology that it reports publicly—or at least a reasonable interpretation of what it reports—and the Census tables of e-commerce at the three-digit level do not mean what a reasonable reader would expect them to mean. We believe the Census e-commerce figures are underestimates, as we said, but given what Census reports it is not possible to assess the magnitude of that underestimation or even verify its existence. We have therefore withdrawn the earlier paper and its estimates. This note explains what happened, identifies serious problems in how Census tabulates and reports the e-commerce data, and presents recommendations for reform.

I. Discovering and Resolving a Discrepancy in the Census Data

Our saga began in June of 2015. Evans and Schmalensee were working on the implications of the Internet, and in particular mobile, on the retail industry for a chapter in their forthcoming book *Matchmakers: The New Economics of Multi-Sided Platforms*. Like many economic researchers we were consumers of Census Bureau data. We weren't on a mission to find mistakes in the Census data; we just wanted data for our work. Scott Murray was working with us and developing tabulations based on the Census data on e-commerce for retail.

Murray alerted us to apparent discrepancies in the Census data. He was looking at the Census data by three-digit NAICS codes and found that they didn't make sense. Census reported

¹ David S. Evans, Scott Murray, and Richard Schmalensee, "Why Online Retail Sales Are Much Larger than US Census Data Report," January 28, 2016. Available at <http://ssrn.com/abstract=2716266> or <http://dx.doi.org/10.2139/ssrn.2716266>

that NAICS industry 452 (General Merchandise Stores), which would include Wal-Mart, only had \$88 million of e-commerce sales in 2013. That was obviously inconsistent with the fact that Wal-Mart had a multi-billion online retail division called walmart.com, as well as the fact that many other physical retailers also had substantial online sales.²

We then worked through November 2015 to get to the bottom of this discrepancy and to better understand what the e-commerce figures from Census, which we wanted to rely on because there wasn't any other reputable source, actually showed. For us, this was an unwanted digression that consumed a lot of time.

To understand what the Census data showed we poured through the documentation that they made available online. We also contacted the Census Bureau in June 2015 when we found the discrepancies. We were put in touch with the Chief, Retail Sales Branch, Economy-Wide Division. This individual was the person in charge of the e-commerce data and reports. He was extremely helpful and responsive in answering our emailed questions and helping us understand the Census methodology—or so we thought at the time.

Evans and Schmalensee, long-time consumers of Census data for their research, initially believed that a common confusion about the difference between establishments and firms was at the heart of the apparent discrepancy. We told Murray that we conjectured that physical retailers reported online sales for establishments and that the Census Bureau coded these establishments as part of non-store retailers. Murray did two things. He checked the Census documentation online, including the survey forms used to collect the data (see http://www.census.gov/retail/arts/get_forms.html), which indicated that the Census retail data were collected and reported at the firm level. He also e-mailed the Census official who both

² Wal-Mart Press Release, Feb 20, 2014. http://corporate.walmart.com/_news_/news-archive/investors/2014/02/20/walmart-reports-q4-underlying1-eps-of-160-fiscal-2014-underlying1-eps-of-511

confirmed that the data were collected at the firm level and explicitly ruled out the establishment explanation as the source of the undercount. During these email exchanges the Census official also confirmed that many large physical retailers do not report e-commerce sales and indicated that Census e-commerce sales were likely understated.

This left us in a quandary. We needed an estimate of the online share of retail sales for our book—both a current estimate and an estimate of growth over time. Discovery of a substantial error in the Census data was interesting, of course, but what we really needed for our research was a reliable estimate.

Based on what we understood the Census did given their published survey form, their published methodology, and a detailed e-mail exchange with the Census official responsible for their e-commerce estimates, we developed a methodology for producing a conservative estimate of the undercount, and applied it to estimate the correct online sales and share figures. We needed to cite something for our book. As a result we wrote a paper that documented the Census undercount and reported our methodology for developing better estimates.

We thought we were done, and our detour into Census data complete.

II. Census Bureau “Clarifies” How It “Really” Calculates Online and Physical Retail Sales

It was not to be. Around the time we posted our paper on SSRN we shared a copy with our colleague, Karen Webster, who used it for an [article](#) she published on the widely read PYMNTS.com website and in the PYMNTS daily newsletter. That led the Census Bureau to publish the following press release and to request Ms. Webster to issue a correction. The head of the Census Public Information Office wrote, “Please note that your article is misinforming the public about the accuracy of our online retail statistics. It's apparent that there is a lack of

understanding behind how we categorize E-Commerce Sales in our Retail Sales Data Sets. I would appreciate an update to your story with a link to our statement.” The statement read as follows.

“JAN. 21, 2016 — The U.S. Census Bureau's retail sales program provides statistics on all forms of retail sales, including e-commerce sales through the Monthly Retail Trade Survey and the Annual Retail Trade Survey. The Census Bureau categorizes e-commerce divisions of companies with physical storefronts as part of the electronic shopping and mail-order houses industry (NAICS 4541) as long as they do not primarily fulfill e-commerce orders from their stores. Companies provide separate information on their brick-and-mortar stores and their e-commerce divisions. For example, if an electronics store has online and brick and mortar sales, the online sales would be tabulated as part of the electronic shopping and mail-order houses industry and the brick and mortar sales would be tabulated in electronics and appliance stores (NAICS 443). This is similar to how companies would split reporting between two distinct brick-and-mortar divisions, such as a company that owns grocery stores and department stores. The Census Bureau *never discloses* information about specific companies. If companies do not provide data for a given month or year, the Census Bureau uses statistical methods to generate imputed sales and e-commerce sales values.” (Paragraph breaks eliminated.)

Disturbingly, the Census Bureau statement was flatly inconsistent with what the Census official had told us, in writing, inconsistent with what a reasonable researcher would interpret their published survey forms and methodology as saying, and inconsistent with what a reasonable reader would conclude from the Census reports. Nevertheless, we took the paper down from public distribution on SSRN and sent an email to the Census official for clarification. He said he wasn't allowed to respond to us and we had to go through the Public Information Office.

We did not believe having a technical discussion with the Census Public Information Office rather than the technical team was a productive avenue for resolving the discrepancy. Each of us—Evans, Murray and Schmalensee—then double-checked everything and went back to the email correspondence with the Census official We revised the paper to quote the e-mail

correspondence directly—including the e-mail confirming an undercount—and to provide more links to the Census public documentation. We reposted the paper on January 28, 2016

The Public Information Office sought to contact Murray and Webster after we posted the revised paper to explain why the statements about an error in the Census data were wrong. After an email exchange, Evans, Murray and Schmalensee agreed to have a conference call on these issues under the condition that the Census technical staff, not just the Public Information Office, participate in the call.

That call happened on February 5, 2016. That same day the Census Bureau sent us an annotated copy of the email exchanges from the Census official with “corrections”. The Census official was on the conference call, but he did not speak much, if at all, and didn’t answer questions posed directly to him.

If we accept what the Census Bureau claims on February 5, we reach the following conclusions.

- The e-commerce figures that are reported in the Census tables for physical retailers at the three-digit level are do not reflect meaningful estimates of anything.
- The Census obtains but chooses not to report separately the e-commerce sales actually made by firms that are mainly physical retailers.
- The Census chooses to include the e-commerce sales of physical retailers (e.g. Wal-Mart) in NAICS industry 454 (Non-Store Retailers), along with the sales of real non-store retailers (e.g. Amazon).

- The Census Bureau claims to include the online retail sales of manufacturers and other non-retailers with retail sales in the figures for total retail sales and in the e-commerce sales of industry 454.
- The Census Bureau does not deny the statements by the Census official that some large physical retailers do not report e-commerce sales and that the Census figures thus undercount e-commerce.

Given the discrepancies, however, between what the Census Bureau told us they did in the conference call, what the responsible Census official told us they did by e-mail, and what the Census public documentation says they did, we do not have a great deal of confidence that we really know what Census has done in the past to generate e-commerce figures or what they are doing now.

Nevertheless, the following is a summary of what we were told by the Census Bureau in the conference call and through the “corrections” of the e-mail exchange with the Census official.

According to the Census Bureau on February 5, firms that are engaged to a significant extent in distinct activities, such as e-commerce and physical retail or manufacturing, are in principle asked to file separate reports for the sets of establishments or divisions primarily engaged in each activity. Thus, for instance, a large manufacturer with significant e-commerce sales would in principle report those sales separately from other revenues, and those sales would appear in Census reports under NAICS industry 454, Non-Store Retailers. Similarly, the total revenue of a large general merchandiser with physical stores and an e-commerce division would in principle be divided between NAICS industry 452, General Merchandise Stores, and industry

454, with the latter having all the revenues of the e-commerce division and the former all the revenues of the physical stores.

Census nonetheless reports e-commerce sales for General Merchandise Stores and for all the other three-digit physical retail industries (NAICS industries 441-453) except for Gasoline Stations (NAICS industry 447). Census claimed on February 5 that e-commerce sales are allocated to those industries when a firm or group of establishments cannot easily be divided on the basis of organizational structure or establishment locations. An auto dealer, for instance, might deliver cars that have been ordered online under terms negotiated online. Revenue from the sales of those cars is e-commerce revenue, but the dealer may fill out only a single report listing total revenue and e-commerce revenue. In this case, the e-commerce revenue would apparently show up in NAICS industry 441, Motor Vehicle and Parts Dealers, rather than industry 454.

Unlike other e-mail exchanges, the following email from the Chief, Retail Sales Branch, Economy-Wide Division was not corrected by the Census Bureau orally or in writing on February 5:

“The root issue here is getting all companies not just in retail but in all sectors of NAICS be it wholesale, manufacturing or retail to accurately report E-commerce. It is not a reflection on the data collection instrument it is just a willingness to report or whether their accounting systems can breakout E-commerce sales for our purposes. For those non-store retailers specializing in E-commerce it is relatively easy but for large companies it is often times difficult to get them to report, and when they do, they do not necessarily report in the manner that we at Census and researchers would prefer.”

Based on the Census statements of February 5, Census figures for e-commerce revenues by three-digit NAICS industries have little meaning. For instance, even though Morningstar confirms that Wal-Mart is classified as in NAICS industry 452 (General Merchandise Stores) at

the firm level, as are many other large retailers, and WalMart's 10K statements report billions of dollars of online sales, Census reports only \$88 million in online sales for industry 452 in 2013. Presumably, its online sales, if reported to Census, are reflected in the total for Non-Store Retailers, as are the online sales reported to Census of non-retailers like Apple or Nike.

Assuming that the Census Bureau does what it said it does in the February 5 call and email "corrections" our attempts to estimate the Census undercount were based on a wrong premise and are therefore wrong. That doesn't mean there isn't an undercount, but it does mean that we don't have any ability, as researchers outside the Census Bureau, to verify the existence of that undercount or to provide estimates of its magnitude.

III. Recommendations for Reform for the Census Bureau

We conclude by offering several recommendations to the Census Bureau:

- First, the Census Bureau should commission an independent audit of the methodology used for calculating e-commerce figures for retail trade and probably for other areas as well. Ideally, this audit should be conducted by economists or statisticians with experience in collecting and analyzing commercial data, with participation by businesspeople actually involved in e-commerce.
- Second, the Census Bureau should report e-commerce sales for entities that are primarily physical retailers separately from entities that are primarily online retailers. It should revise the historical data on e-commerce sales by three-digit

NAICS code since a reasonable interpretation of those data is inconsistent with what Census actually collected and reports.

- Third, going forward, given the move of historically pure online retailers into physical retail, the Census should also report brick and mortar sales by primarily online retailers.
- Fourth, the Census Bureau should report e-commerce retail sales by non-retail firms separately. The movement of manufacturers into direct retail as a result of the possibility of making online sales is an important economic phenomenon that should be measured.
- Fifth, the Census Bureau should publish detailed and transparent methodologies for e-commerce and provide e-commerce tables with meaningful estimates. It is clear from our experience and the need for the Census to correct itself so extensively that it currently does not do this, at least with respect to retail trade.

We have long viewed Census Bureau as an exceptionally professional organization, but we must admit that our confidence in it has been shaken by this experience. The problem isn't that the Census Bureau made a mistake; that happens all the time in quantitative research. A more serious problem is that the Census Bureau's e-commerce statistics are much less useful than they could be given the data they collect. But the most serious problem is that it was so hard for even careful researchers to learn how Census actually collects and compiles its data and thus to discover whether or not the published data are accurate.