

Labor Productivity in Retail and Associated Trades as Measured by BLS: A Response to a Request for Advice from Committee on National Statistics (CNSTAT) of the National Academies of Sciences, Engineering and Medicine (NAS).

by

Roger R. Betancourt\*

Professor of Economics, Emeritus

University of Maryland, College Pk.

April 2020

- This response by the author is a reply to a request for advice from CNSTAT on a study they are currently undertaking. Since it forced me to think consistently about a wide range of measurement issues that I have encountered in my research over the last twenty-five years, I decided that it was worthwhile to make it widely available. There was a third section replying to a request for recommendations of experts who could or should participate in the study. I eliminated that section, which was included in the full response, to keep those recommendations private. I also made a few minor editorial changes.

## I. Key Issues involved.

Given the focus on labor productivity in the retail sector, i.e., output per unit of labor, the first issue that arises is how to measure output in the numerator and labor in the denominator both conceptually and empirically. Considering the numerator first. There is a useful introduction to the topic of whether to use the gross margin (Sales – Cost of Goods Sold) as measured by BEA or sales as measured by BLS in the literature (Triplett and Bosworth 2004, *Productivity in the US Services Sector*: Ch.8: pp. 238-242) as the output of retailing. The comment on the chapter by Ratchford (251-255) summarizes the state of the subject at that point in time. Succinctly put, **“Therefore, the practical question of output measurement for the official productivity measures is whether (deflated) [gross] margins or sales measures these services better.”** **What are these services? In subsequent joint work with other authors Ratchford (Kopalle et al. 2009, *Journal of Retailing*, 85 (1), p.56) lists five broad categories (...location, information, assortment, delivery and ambience.** (i.e., accessibility of location for acquiring the product; information about characteristics of the product and/or the store; assortment in terms of breadth and depth of products available; delivery in terms of assurance with respect to desired time and form; and ambience of the purchase and/or acquisition location.

1. Conceptually, the first key issue is whether deflated sales or deflated gross margins best measures these services. Under the assumptions typically made in the retail literature in economics and marketing, it has been shown that a store’s value added (gross margin) would be an exact measure of the store’s net (gross) service output [Ofer, G. *Review of Income and Wealth* 1973, p.366 (Betancourt 2004, *The Economics of Retailing and Distribution*: p.85; fn. 2)]. Under similar typical assumptions made in this retail literature, it has been shown that deflated sales is a measure of the store’s quantity of products sold (Smith and Hitchens, *Productivity in the Distributive Trades* 1985; Betancourt (cited above) 2004, pp.66-68). **Therefore, conceptually sales will never measure services correctly even under the best of all worlds where the assumptions are satisfied. At least in such an ideal world, the gross margin measures the services that constitute the economic activities of retailers.**
2. Empirically, the literature also provides guidance on where the main practical problems lie. BEA calculates the gross margin in real terms relying on the following assumption (Triplett and Bosworth (cited above) 2004, pp.238-239). The ratio of the nominal gross margin to nominal sales is fixed at the base year for thirty-four industry subgroups of the retail sector. The constant dollar value of the gross margin for any industry subgroup in any year equals this base year ratio times the constant dollar value of sales for the subgroup in that year. Hence, the aggregate constant value of the gross margin differs from the aggregate value of sales only due to variations in the constant dollar value of sales across subgroups with

different margins. Not surprisingly, since both BEA and BLS measure sales almost identically, average output growth per year is very similar whether one uses BLS deflated sales or BEA 'deflated' gross margins (e.g., compare rows 1 and 3 of table 8.2 in Triplett and Bosworth (cited above) 2004, p.237). **While BEA is on the right track conceptually, it punts empirically due to the use of an assumption for empirical implementation that makes the measure almost identical to the use of deflated sales.**

3. BLS started a new initiative (around 2000) focusing on the gross margin as the conceptual measure of output and developing a producer price index for the margin price measuring the difference between the retail price of a product and its acquisition cost, which could, in principle, allow for the construction of a deflated gross margin. There is no direct measurement of services within an establishment. The presumption is that holding constant the outlet in the PPI sample holds constant changes in these services. Early results comparing the margin price with the CPI index to use as a sales deflator for the same subcategories (food stores and auto dealers) led to the conclusion that "...the margin prices are quite different."(Triplett and Bosworth (cited above) 2004, p. 244). What has happened since? Additional subsectors of the retail trade were included in the PPI program. By 2004 there was near complete coverage and by 2006 even electronic retailing was included (e.g., Borgie 2014, *Beyond the Numbers: Prices and Spending* 3 (18), fn. 1) A 2019 listing of PPI's for retail trade industry subcategories and how the latter are defined is available at the six digit level in <https://www.bls.gov/ppi/ppiretailtrade.htm>. A 2019 listing of the ppi retail trade subindexes available at the three- digit level for net output of the retail trade is available in table 10 of [https://www.bls.gov/ppi/ppi\\_dr.htm](https://www.bls.gov/ppi/ppi_dr.htm) . It covers NAICS 441-448, 451- 452, 4532, and 454. These indexes aim to measure the margin prices of providing services to consumers by the retail sector with little transformation of the goods. In the U.S. data, it is stressed that PPI's track changes in margin prices on a percentage basis and, therefore, they are not suitable "...to compare differences in the absolute size of margins between store and non-store retailers." (Borgie 2014 (cited above), fn. 4). Ironically, the OECD-EU methodological guide on wholesale and retail trade states when defining uses for the margin prices "The margin price can be used to deflate the gross margin thereby providing a measure of real output for the distributive trade sector in the national income accounts." (Draper 2014, Ch. 5 in OECD-EU Methodological Guide for Developing Producer Price Indices for Services). **What is the reason the margin prices calculated by the BLS are implemented empirically in such a way that one of the main purposes stated in the OECD –EU methodological guide can't be attained? It defeats one of the main reasons for undertaking the effort!**
4. Taxes and transportation are excluded from the margin price by both the OECD-EU methodological guidelines (Draper (cited above) 2014) and the U.S. implementation mechanisms. Both are supposed to allow for standard discounts although not for clearance sales discounts. PPI's are supposed to represent prices from the perspective of the seller whereas CPI' are supposed to do so from the perspective of the purchaser. The first paragraph of the technical notes to the 2019 PPI detailed report cited above makes the point clearly. This implies that **any measure of labor productivity based on the ppi indexes calculated this way would then introduced a divergence between productivity and welfare**

since the same technical note indicates that prices calculated from both perspectives behave very differently. For, from the purchaser' perspective their welfare is affected by what happens to transportation and taxes. Both transportation and taxes would be included in the gross margin data as calculated by BEA but not as gathered by Census Bureau, which excludes sales taxes collected by firms for payment to local, state or federal authorities.

5. Incidentally, the divergence between productivity and welfare affects the topic of productivity more generally in the case of retailing, but not necessarily in the case of wholesaling, due to retailing direct impact on the consumer and the cost-shifting possibilities it implies. For instance, some writers relying on increasingly available panel data have measured productivity in retail subsectors indirectly by comparing the productivity of entering establishments with that of exiting establishments and find, not surprisingly, that entering establishments have higher productivity than exiting establishments (e.g., Foster et al 2006 *Review of Economics and Statistics* 88 (4) ). While a lot of useful information arises from that approach, by definition it generates a divergence between productivity and welfare when applied to retailing for any productivity measure, including labor productivity. These studies often use deflated sales or real sales as a measure of output (e.g., Foster et al 2016 *Handbook on The Economics of Retailing and Distribution*, Ch.1) and discuss productivity in retailing without mentioning that by their measure increases in productivity could imply decreases in welfare. Using data from the contribution just cited (p.15 Table 1.1), namely the sum of establishments across their size categories in 1977 and 2007, one can calculate the number of establishments per person in both years. One obtains that in 1977 there were .00578 establishment per person in the U.S and in 2007 there were .00512 establishments per person in the US. This represents an 11% decrease in the number of establishments available to a typical consumer in the U.S. Meanwhile retail productivity exhibited substantial increases due to the growth of multi-establishment firms relative to single unit firms (e.g., Foster et al 2016 (cited above): Figure 1.4; p.26). The point of this example is to provide an illustration that productivity measures used by economists in retailing often can generate increases in productivity simultaneously with decreases in an important service that affects consumer welfare. This happens due to shifting the costs of generating a service output from the retailer to the retail customer, i.e., by forcing consumers to do the walking or the driving. **These studies never mention the potential divergence between their positive productivity results and their potential negative impacts on consumer welfare. Such a disclaimer needs to be included in these productivity studies. Why? Because it affects not only retailing but most service sectors where, in contrast to the retail sector, their often very substantial retailing activities are not separable from production or transformation ones from an accounting perspective.**
6. These issues also affect the BLS outreach to users, e.g., in their 'beyond the numbers' series. In an interesting recent article by J. Rudd on productivity in grocery stores the author uses as a measure of output - annual revenues deflated for inflation. Since this measure is similar to sales, it does not capture the services provided by retail establishments. **Yet, most of the technological advances discussed by the study are directly related to the production of**

services by grocery stores. <https://www.bls.gov/opub/btn/volume-8/checking-out-productivity-in-grocery-stores.htm>

7. This recent article is also of interest for what it implies above the proper measure to use in the denominator of a labor productivity measure. The retail sector has long had a tradition of employing part-time labor and productivity concepts incorporated this by using full time equivalent employees in the denominator. Recent technological advances in the employment of labor and its measurement by BLS have led to the use of hours work as the relevant measure in the denominator. This is a desirable feature. It manifests itself in the above article by their emphasis on the evolution of hours to measure the denominator in a productivity index. **Ironically, it makes great deal of sense if in the numerator or output index one was measuring primarily the production of these services in the retail sector through the gross margin not the output of products, which is what deflated sales or revenues captures in the numerator. Even the discussion of concentration in the article might be more relevant, if it related to the gross margins of food stores. For, profits of food stores represent a much larger share of the value of the gross margin than they are of revenues, since the latter also include profits of every participant along the supply chain, among other things.**
  
8. One issue arises with respect to a point of view both you and BLS have identified as important: namely, “ there is a concern that current measures don’t reflect the changing structure of the retail trade ...when auxiliary functions are carried out by a retail establishment , the related employment is separately counted by BLS to the appropriate industry.” The choice of what is the proper output measure goes a long way in addressing that concern. When the gross margin is the output measure, the answer is quite clear conceptually. Since it measures services actually provided by the retail sector, the warehousing done by the auxiliary establishments is part and parcel of the services provided by the retail sector, namely the storage necessary to produce assurance of product delivery at the desired time and in the desired form wanted by the customer. **Therefore, it is quite appropriate to attribute warehouse workers of auxiliary establishments of firms in the retail sector to the retail sector and not to a transportation and warehousing sector. Thus, what is inappropriate is the current procedure used by BLS! If you use deflated sales, or revenues, nothing makes sense so the answer to whatever BLS does with the warehouse workers is anything goes. In this case, you might as well save the waste of time and money of the satellite account.**
  
9. Finally, one issue you might want to consider is if it makes sense to do a satellite account in which it makes sense to calculate the output of these sectors (retail, wholesale, transportation and warehousing) you are concerned about in terms of the gross margin, i.e., as sales minus costs of goods sold. In retailing and wholesaling the gross margin captures the actual services provided by retailers and wholesalers and it is inappropriate conceptually to use sales or revenues of these goods as the relevant output in a productivity measure. **Similarly, in transportation and warehousing, the output of services provided relates directly to the margins generated by these activities. Indeed, in earlier writings, retailing,**

wholesaling, and transportation and warehousing are the distributive trades in a definitional sense (e.g., Yuskavage 2006).

## II. Essential Recent Research or Operational Changes You Should know About

I changed the order of your three broad questions for expositional convenience. An implicit but essential driver behind your study as well as BLS questions in requesting your study stems from the spread of the digital economy in general and in retailing in particular. This generates two recent research or operational issues you should know about and one that BLS seems to know about but that you might want to ignore. What is the impact of the digital economy on the proper conceptual definition of output of the retail sector? How is Census, BLS or BEA to measure e-commerce properly? What does it do to the NAICS 'similarity' principle that similar 'production' processes should be classified in the same sector. I address each of these questions below.

1. The answer to the first question is that the output of the retail sector remains conceptually the same. That is, it has to provide the same five broad services to their customers in a digital world as it did in the brick and mortar world. **What changes is how the provision of these services takes place. In turn, this implies that the feasible levels attainable due to technical or economic reasons change dramatically. Moreover, not in all dimensions is the digital technology capable of attaining higher levels. These results are available as Propositions 1 and 3 in a lead article published by the Journal of Interactive Marketing (Betancourt et al. February 2016).**
2. Currently, the U.S. Census Bureau has been the agency responsible for measuring e-commerce in the retail sector through ARTS as well as in services in general through the Annual Survey of Services. They have been publishing tables with figures since 1998. Both surveys have used exactly the same definition of e-commerce in their one and only question on the subject. Unfortunately, their definition leads to figures that overestimate e-commerce by close to 200% per year between 1998 and 2014 as a measure of output, defined as the gross margin, in the main retail sector affected by e-commerce, i.e., electronic shopping and mail order houses. Moreover, they overestimate e-commerce as a measure of online sales in this sector by between 122% and 187%, depending on whether you use beginning or end period weights. I published these results in **Economics Letters** (Betancourt November 2018). I sent a copy of the article to a mid-level professional at the census bureau and to the Deputy Director toward the end of 2018. I got no reply from either one. On September 23 2019 the Bureau issued a press release (**E-Stats 2017: Measuring the Electronic Economy**). It discontinued measuring e-commerce through the annual survey of services according to the old definition. Moreover, it would discontinue permanently all the tables previously published with e-commerce for the services sector. It introduced a new practical definition for this survey, better than the old one but not without

problems in my opinion, and it changed the name of the tables so they would no longer use the word e-commerce but instead use electronic revenues. In the case of the retail trade, however, it doubles down on their mistake by using the old definition exactly as before to gather data at higher frequencies, quarterly and monthly. BEA use their e-commerce data in their satellite accounts and BLS relies on their definition in their 'beyond the numbers series', e.g., Borgie (cited above) 2014 fn. 1. **One possible explanation of this behavior by Census is that the economists involved don't want to admit the serious mistake implied by their implementation of the definition of e-commerce in terms of revenues. Whatever the reason, it violates elementary ethical principles in public good provision for a statistical agency to persistently ignore their provision of mistaken information.**

3. Finally, in the case of printed and recorded products, the digital economy transforms physical products (books and journals or vinyl records, compact disks) previously distributed as goods into services through digital distribution. This transformation makes it impossible to separate production of the copies from their distribution. Once a copyright owner produces a master file, the digital process allows for production and distribution of copies to customers and/or distributors. Distributors can replicate the process, legally in exchange for a license fee, and produce and distribute copies to their customers. If a publisher undertakes the digital distribution, for example, the outcome appears in the information sector; if a non-store retailer undertakes the digital distribution, for example, the outcome appears in the retail sector. The production process, however, is the same in both cases. This result violates NAICS 'similarity' principle, i.e., classifying similar production processes in the same sector. **This issue is simply a consequence of how digital technology changes the distribution process by converting goods into services and there is nothing to do about it in a satellite account!**