Allocation of Company Research and Development Expenditures to Industries Using a Tobit Model

Christian Awuku-Budu[1] and Leo Sveikauskas[2]
November 18, 2015

Abstract

This paper uses Census microdata and a regression-based approach to assign multi-division firms' pre-2008 Research and Development (R&D) expenditures to more than one industry. Since multi-division firms conduct R&D in more than one industry, assigning R&D to corresponding industries provides a more accurate representation of where R&D actually takes place and provides a consistent time-series with the National Science Foundation R&D by line of business information. Firm R&D is allocated to industries on the basis of observed industry payroll, as befits the historic importance of payroll in Census assignments of firms to industry. The results demonstrate that the method of assigning R&D to industries on the basis of payroll works well in earlier years, but becomes less effective over time as firms outsource their manufacturing function.

Keywords: business R&D, industry classification, Tobit model, establishment payroll

---

[1] Christian.awuku-budu@bea.gov
[2] Leo. Sveikauskas@bls.gov. We thank Ian Mead, Cheryl Grim, and Ledia Guci for their useful comments and suggestions.

1. Introduction

Data on R&D expenditures in different industries have been a central input in the analysis of many economic phenomena. Information on R&D expenditures in various industries has been useful in understanding, among other areas, productivity growth (Griliches and Lichtenberg,1984; Scherer, 1982), the structure of the work force (Bartel and Sicherman, 1999; Allen, 2001), and U.S. performance in international trade (Sveikauskas, 1983). More recently, the Bureau of Economic Analysis began treating R&D expenditures as an investment, rather than as an intermediate input, in the National Income and Product Accounts (Smith and Holdren, 2013). Consequently, R&D expenditures help determine the official measure of output in each industry. Similarly, the Bureau of Labor Statistics now treats R&D stocks as an additional capital asset in determining measures of Multifactor Productivity growth in each industry.

Most of the data used to understand the impact of R&D or to measure R&D in official statistics come from the National Science Foundation (NSF) Surveys of Industrial Research and Development (SIRD), which the Census Bureau has administered for the NSF for many years. The Census Bureau determines industry R&D by assigning each firm to a particular industry, usually based on observed industry payroll, and then allocates the firm's entire R&D to the industry in which the firm is assigned[3]. Over recent years, the traditional methods widely used in the past would have assigned considerable proportions of the R&D found in multi-division firms to trade and services[4].

For large multi-division firms, assigning a firm's R&D expenditures to one industry assumes that these firms conduct little or no R&D outside their primary field. For example, if a large firm produces pharmaceuticals and medical devices, for the time period prior to 2008, the Census Bureau procedures would allocate the firm's entire R&D expenditures to the one industry with the greatest payroll. From 2008 forward, if the same firm reports one million dollars of R&D in pharmaceuticals and two million dollars in medical devices, Census procedures would assign the firm's entire R&D to medical devices. Since the 2008 Business Research and Development and Innovation Survey (BRDIS) shows that some firms conduct R&D in more than one industry, we argue that forcing a firm's entire R&D expenditures into a single industry distorts the industry distribution of R&D expenditures.

Because the BRDIS provides information on R&D spending by line of business, from 2008 forward, each firm's R&D expenditures can be readily assigned to different industries. However, there is no obvious way to allocate firm R&D expenditures to more than one industry prior to 2008 to generate a

---

[3] Firms are first assigned to the specific two-digit NAICS industry in which their payroll is the greatest. Firms are next assigned to the three-digit industry where payroll is the greatest among all the three-digit industries within the designated two-digit industry. Firms are similarly assigned to the four-digit industry with the greatest payroll within the designated three-digit industry. The process of successively assigning firms to more detailed industries is called hierarchical assignment. In general, a firm's entire R&D is assigned to the specific two, three, or four-digit industries determined through these procedures.

[4] Since 2004, the pitfalls involved in assigning firms to industries on the basis of payroll have become clearer (Foster and Grim, 2010; page 8). Therefore, from 2004 to 2007, the Census Bureau reassigned many firms to industries based on the main product rather than payroll. This resulted in a large decrease in R&D expenditures from services and trade, especially wholesale trade, management of companies and enterprises, and scientific R&D and related industries. Furthermore, in 2008 the Census Bureau began the Business Research and Development and Innovation Survey (BRDIS) which collects R&D expenditures on a line-of-business basis. The Census Bureau currently assigns all of a firm's R&D expenditures to the line of business in which the R&D expenditures are the greatest.

consistent time series of (private) R&D expenditures by industry.

This paper provides a method for assigning firm R&D expenditures to industries for years prior to 2008. The method allows for differences in R&D intensities across industries and utilizes all available information on each firm's total R&D expenditures and its distribution of payroll across industries. Data on domestic payroll from the Longitudinal Business Database (LBD) allow us to distribute R&D across industries in proportion to payroll, which is a generalization of the traditional Census methods of assigning firm R&D to a single industry according to payroll. Also, since payroll is available annually, we are able to estimate R&D intensity in each industry in every year between 1976 and 2007. One major advantage of the regression-based approach is that the estimates do not require ad hoc adjustments to account for a firm's switch from one industry to another. Instead, the gradual change in the distribution of payroll across industries automatically generates a corresponding natural distribution of R&D across industries.

One limitation of our approach is that observed firm payroll becomes a less effective method of allocating R&D to industries over time as payroll shifts from its manufacturing base to services and trade[5]. The BRDIS microdata show that many firms conduct most of the R&D in their primary field. This explains why previous and existing allocation approaches of industry R&D have been broadly accurate, even though they are ineffective in allocating the R&D which occurs in secondary industries. For instance, while many multi-division firms produce in many industries, far fewer firms actually conduct R&D in more than one line of business. Nevertheless, for the firms which actually do perform R&D in more than one line of business, assigning their R&D to different industries improves the industry distribution of R&D, provides a more consistent industry time series, and helps reduce the need for arbitrary adjustments. Our analysis also connects with important recent work on factory-less production which describes how firms outsource production domestically or abroad and therefore retain only a small proportion of payroll in their home base industry.

Section 2 presents evidence from the 2008 BRDIS. Section 3 outlines the proposed methodology. Section 4 applies the analytical framework outlined in Section 3 to create consistent measures of industry R&D expenditures from 1976 to 2008. Section 5 presents the empirical results, and explains how and why industry payroll has, over time, become a less effective method of assigning firm R&D to individual industries. Section 6 concludes.

2. Evidence of R&D in Secondary Industries from the 2008 BRDIS

The Business Research and Development and Innovation Survey (BRDIS) was introduced in 2008. Unlike the SIRD, the BRDIS collects information on total R&D expenditures and on R&D expenditures by detailed (6-digit NAICS) line of business. The additional data on the R&D which each firm conducts in

each line of business provide more information than ever before available on the exact areas in which each firm actually conducts R&D.

We therefore examine data from the 2008 BRDIS in considerable detail to assess the proportion of overall R&D that is conducted by multi-division firms which report R&D expenditures in more than one line of business. This assessment is not straightforward because the BRDIS data base includes categories such as 888 and 999 with R&D expenditures which cannot be assigned to a particular industry or are in balancing categories. R&D expenditures assigned to 888 or 999 represent only about 2 percent of all R&D expenditures. According to specialists at the Census, most of these 888 or 999 expenditures likely occur in the same industry in which the firm is classified.

As shown in Figure 1a, if we assign these expenditures to the firm's primary industry, 79 percent of all R&D expenditures occur in cases where the firm reports R&D and sales in only one line of business, about 5 percent in cases where the firm reports sales in more than one line of business but R&D in only a single line, and only 15 percent occurs in cases where firms report both R&D and sales in more than one line of business.

As Panel 1a of Table 1 shows, if the R&D expenditures in 888 and 999 categories are treated as occurring in separate industries, the corresponding percentages are 65, 5, and 29 percent, respectively. Therefore, somewhat more than 15 percent of research is carried out by firms conducting research in more than one line of business. Evidence that large multi-division firms are not as major a factor in R&D as might have been thought is useful information. Nevertheless, such firms still account for almost one of every six research dollars. This amount is well worth investigating.

Finally, the data reported in Panel 1b of Table 1 are important. They suggest that firms concentrated in a single industry are even more influential in world R&D than they are in U.S. R&D. Though these data cover only U.S. multinationals and foreign firms with operations in the U.S., these results suggest that R&D is heavily concentrated in the single home industry in production all over the world, not just in the U.S.

3. Analytical Framework

Let the ratio of research expenditures to payroll in industry i be a fixed amount which can be stated as $(RD/pay)_i$.[6] A similar expression describes the research to payroll ratio in industries j and k. Then the total research bill faced by firm 1 operating in industries i, j, and k can be expressed as:

$$RD_1 \equiv \left(\frac{RD}{pay}\right)_i pay_{1,i} + \left(\frac{RD}{pay}\right)_j pay_{1,j} + \left(\frac{RD}{pay}\right)_k pay_{1,k} \qquad (1)$$

where $pay_{1,i}$ is firm 1's payroll in industry i, and similar terms indicate firm 1's payroll in industries j and k. Equation (1) is an identity which holds true by definition. In practice equation (1) will not hold true exactly, so we also include an error term, u_1, for firm 1. Then:

---

[6] Throughout this analysis, we emphasize payroll because the Census Bureau has long assigned each firm's R&D to a specific industry on the basis of the payroll observed within each industry.

$$RD_1 = \left(\frac{RD}{pay}\right)_i pay_{1,i} + \left(\frac{RD}{pay}\right)_j pay_{1,j} + \left(\frac{RD}{pay}\right)_k pay_{1,k} + u_1 \qquad (1b)$$

An expression similar to equation (1b) also holds true for every firm, from firm 1 to firm n. Specifically:

$$RD_1 = \left(\frac{RD}{pay}\right)_i pay_{1,i} + \left(\frac{RD}{pay}\right)_j pay_{1,j} + \left(\frac{RD}{pay}\right)_k pay_{1,k} + u_1$$

$$RD_2 = \left(\frac{RD}{pay}\right)_i pay_{2,i} + \left(\frac{RD}{pay}\right)_j pay_{2,j} + \left(\frac{RD}{pay}\right)_k pay_{2,k} + u_2 \qquad (1c)$$

$$RD_n = \left(\frac{RD}{pay}\right)_i pay_{n,i} + \left(\frac{RD}{pay}\right)_j pay_{n,j} + \left(\frac{RD}{pay}\right)_k pay_{n,k} + u_n$$

In matrix notation, RD is a $n \times 1$ column vector indicating each firm's $R\&D$ expenditures, pay is a $n \times c$ matrix showing each of the n firms payroll in each of the c industries, and $RDtopay$ is a $c \times 1$ vector indicating the research to payroll ratio in each of the c industries. Then:

$$RD = pay \times RDtopay \qquad (2)$$

so that each firm's R&D is determined by the product of payroll in each industry multiplied by the unknown ratio of R&D to payroll in that industry, with all the product terms then summed across industries. The vector of estimates of RD to payroll in each industry, described in equation (2), is estimated by regression. Therefore:

$$RD = pay\left(\hat{\beta}\right) \qquad (2b)$$

where $\hat{\beta}$ is the regression estimate of $RDtopay$. As indicated in equation (1), the independent variables do not include a constant term.

The present discussion assumes that the data satisfy the appropriate conditions for estimation by linear regression. Several econometric considerations must be taken into account in the estimation of equation (2). These issues are discussed in greater detail in the next section. Since regression coefficients provide an unbiased estimate of the industry's R&D to payroll share, we can simply calculate the sum of observed total payroll in each industry, $pay_j = \sum_{i=1,n} pay_{i,j}$ and multiply the total payroll observed in each industry by the corresponding regression coefficient for that industry, $\hat{\beta}_j$, to determine the total R&D expenditures allocated to that industry. This procedure is used throughout this paper to develop estimates of the R&D expenditures occurring within each industry.[7]

Alternatively, if we are concerned with the industry distribution of R&D for a particular firm, we could assume that each firm is either a high research firm or a low research firm, and that this same characteristic is shared equally across industries. Specifically:

$$\widehat{RD} = pay\left(\hat{\beta}\right) \qquad (2c)$$

so that the adjustment factor for each firm is:

$$RD/\widehat{RD} \qquad (3)$$

To illustrate, if a firm has 25 percent greater research than implied by equation (2c), we assume the research to payroll ratio is uniformly 25 percent greater in every industry in which the firm operates.

---

[7] The same methods could be used to assign other items observed only at the firm level to industries and establishments. For example, information on advertising or on management expenditures is often collected only at the firm level, and similar techniques could be utilized to assign these important elements of intangible capital to industries.

Conversely, a firm with 90 percent of the research implied by equation (1) operates with 90 percent of the research to payroll ratio indicated for each industry. In this way, all research conducted by any firm is exhaustively and consistently assigned to a particular industry.[8] However, the present paper concentrates solely on industry measures of R&D and therefore does not address the distribution of R&D for a particular firm through equations (2c) and (3).

Econometric Issues

One important issue is heteroscedasticity, which is likely to occur in the error terms of equation (1), since the variance of R&D is expected to be much greater for firms which spend large amounts on R&D. All statistical tests are therefore based on robust standard errors.

A second point is that the data on R&D expenditures contain a large proportion of zeroes, especially, after the sample was expanded in 1992 to include a much greater number of firms, many of which are not likely to do much research. The frequency of zero observations suggests a linear probability model. Our estimates are based on the Tobit model, which is well suited for analysis of R&D, since it permits both zero observations and differences in the magnitude of the dependent variable.

Equation (2) provides no guidance as to what level of industry disaggregation is appropriate. The empirical analysis in Section 4, therefore considers results from several different levels of industrial disaggregation.

As with other payroll-based allocation methods, one limitation of the regression method is that many firms have increasingly outsourced important activities to either domestic or foreign suppliers. Many such firms now have lower payroll in their basic industry. Due to such considerations, over time, the observed distribution of payroll becomes less effective in explaining R&D. Section 5 examines such issues.

Data Sources and Preparation

We link data on each firm's (nonfederal) R&D expenditures from the SIRD with that firm's corresponding establishments from the Census Longitudinal Business Database (LBD) for years 1976 to 2011[9]. For each of the firm's establishments, we gather information on domestic payroll and on the industry to which each establishment is assigned. Prior to linking the SIRD to LBD, we first impute missing nonfederal R&D expenditures in the SIRD and missing domestic payroll and industry codes in the LBD. For establishments' domestic payroll and industry codes, we use information from the 1997 Economic Censuses of Manufactures, Wholesale Trade, and Services; firm-level information on domestic payroll from the SIRD; and information from the adjacent years of the LBD. We impute missing

---

[8] In implementing such an approach, it is also possible to test whether firms primarily located in certain high technology sectors, such as computers or electronics, systematically adopt greater research intensity in their subsidiary industries.
[9] For more information about the LBD, see Jarmin and Miranda (2002).

nonfederal R&D expenditures in the SIRD by subtracting federal R&D expenditures from total R&D expenditures. The last step in data cleaning removes duplicate establishments from the sample.

After the SIRD and LBD data sets have been cleaned, we map firms in the SIRD to their establishments in the LBD using census unique firm identifiers in the LBD. We use a 5-year window in the mapping procedure. Foster and Grim (2010) and Awuku-Budu and Robbins (2014) similarly used a 5-year window to map firms in the SIRD and BRDIS to establishments in the LBD. Like these two previous studies, we match more than 99 percent of multi-unit firms, and more than 98 percent of single-unit firms, to their LBD establishments in every year of the SIRD.

In the next step, we assign NAICS industry codes to LBD establishments without a NAICS code; these represent less than 30 percent of LBD establishments[10]. NAICS codes are assigned as follows: In benchmark (Census) years, we match the establishment's Census permanent plant identifier (PPN) to the PPN in the Censuses of Manufactures, Wholesale Trade, or Services[11]. If there is a match, and the SIC code in the LBD matches the SIC code in the Censuses, we assign the corresponding NAICS code from the Censuses to the establishment in the LBD. For remaining establishments without NAICS codes, we use the official 1997 Census SIC-to-NAICS bridge table to assign NAICS codes in cases where the SIC code has a one-to-one match with the NAICS code. We use adjacent years LBD NAICS codes if the SIC codes match. In cases where no definitive assignment occurs, we use a simple probabilistic matching method to assign establishments to NAICS industries. Finally, we aggregate all establishment payroll, as determined through these procedures, into estimates of each firm's total payroll within each relevant industry.

4. Empirical Results from Regression Analysis

Once each firm's payroll in each industry has been established, we proceed with the empirical estimation based on equation (2). Table 2 shows results for a simple dichotomy between manufacturing and nonmanufacturing payroll. A separate cross-section regression is run for each year. Table 2 reports coefficients for both manufacturing and nonmanufacturing payroll and the associated standard errors. Data begin in 1976 because the LBD file, which makes it possible to determine each firm's payroll in each industry, begins in 1976.

The results in Table 2 are in many respects similar to the results obtained throughout Section 4. The coefficients typically show that payroll has a significant positive effect on R&D expenditures. The coefficients for manufacturing payroll are typically greater than those for nonmanufacturing payroll, and the impact of manufacturing payroll becomes relatively stronger over time. However, as discussed in detail below, the explanatory power of the regression declines substantially over time[12].

Figures 2a-2c show the distribution of R&D expenditures between manufacturing and

---

[10] We use an older version of the LBD. The LBD has been updated since our project ended. Teresa Fort, a CES researcher, has attached NAICS codes to all establishments in the LBD for the entire analysis period).
[11] PPN is the identifier prior to 2002. In 2002 and later, the primary Census identifier is the survey unit ID.
[12] The decline is not monotonic and there are some low R-squared values in earlier years (e.g., 1987).

nonmanufacturing industries using the regression results in Table 2 and compare these patterns with the corresponding official NSF series. Figure 2a compares R&D in manufacturing and total R&D as calculated through the regression method. Figure 2b reports R&D in manufacturing, as determined through the regression and NSF methods. Finally, Figure 2c shows the manufacturing share of R&D over time, again as implied by the regression and NSF methods. Our estimates consistently indicate that greater amounts of R&D than suggested by the official data occur outside manufacturing. As firms begin to shift payroll away from manufacturing, the regression method allows this shift to appear more quickly in the R&D data even before the trend away from manufacturing accelerates and entire firms begin to tip over to nonmanufacturing. After 2000, the regression method attributes less R&D to manufacturing, as the regressions find much less payroll in manufacturing.

The remaining portions of Section 4 report results from the regression method using several different levels of industry disaggregation. To conserve space and to present the essential results more clearly, in each case we present and compare only average results from the first four and last four years. Panel 3a of Table 3 reports results for the 1976-1979 and 2005-2008 periods for a specification which distinguishes between NAICS 31 (simple manufactures such as food and apparel), NAICS 32 (which includes wood, paper, petroleum and the research intensive chemicals industry), NAICS 33 (including metals, machinery, computer and electronics, and transportation equipment), and nonmanufacturing. Not surprisingly, payroll in NAICS 31 does not generate much R&D. Explanatory power continues to decrease over time.

Panel 3b of Table 3 reports similar results for NAICS 325 (chemicals), NAICS 334 (computer and electronics), and NAICS 336 (transportation equipment). In addition, other manufactures and nonmanufacturing are included. Payroll in chemicals and in computer/electronics has the greatest impact on observed R&D. Figures 3a and 3b compare the regression estimates with the official NSF data for chemicals and for computer/electronics.[13] Between 1999 and 2007 both the amounts and the trends are broadly similar, though the regression estimates appear to spike in 2001.

Panel 3c of Table 3 presents regression results for four-digit NAICS industry detail. The industries examined are NAICS 3254 (pharmaceuticals), NAICS 325 other (other chemicals), NAICS 3344 (semiconductors), NAICS 334 other (other computer and electronics products), NAICS 3364 (aerospace), and NAICS 336 other (other transportation equipment). In addition, the regression includes other manufactures and nonmanufacturing. Figures 3c and 3d compare R&D from the regression method with R&D from the NSF for 1999-2007 in two important 4-digit NAICS industries, pharmaceuticals and medical products and semiconductors and other electronic products. Here too the trends are broadly similar, and the regression method once

---

[13]We begin comparisons in 1999 because our firm data are on a NAICS basis, and 1999 is the first year in which the official NSF series are published on a NAICS basis.

again suggests a spike for pharmaceuticals in 2001. However, the regression method suggests much higher R&D in the computer/electronics area throughout the period. This result reflects the very high coefficients for computer/electronics payroll observed in recent years. Such patterns may in part have occurred because outsourcing reduced domestic payroll over recent years.

Table 4 considers manufacturing as a single industry and includes other major sectors such as trade and services, as well as a residual all other industries category. Differences over the years are instructive. In pooled 1976-2008 data (Table 4a), manufacturing payroll (column 2) is clearly dominant, with a much higher estimated coefficient than either trade or services. Trade is important only in chemicals, where sales personnel are often central in pharmaceuticals. In contrast, by 2008 (Table 4b) manufacturing payroll is no longer the greatest coefficient and is in fact no longer significant (column 2). In contrast, trade and services are significant. These developments illustrate well how, for many U.S. firms, manufacturing payroll has become less important, and trade and services employment more influential, as the chain of production has become more fragmented. Finally, in both Table 4a and 4b observed payroll in the primary industry has a strong and significant influence on observed R&D and considerably increases observed R^2. Such patterns show how important large firms concentrated in one area are in the explanation of R&D.

Estimates of the regression coefficient, β, are usually, though not uniformly, positive in research intensive industries. $\hat{\beta}$ tends to be insignificant in low research industries. $\hat{\beta}$ also increases substantially over time, especially in research intensive sectors such as pharmaceuticals and computers and electronics. The most striking result is the decline in R^2 over time. In all the different cases considered, R^2 is substantially lower in 2005 to 2008 than it was from 1976 to 1979. In the earlier years, the distribution of payroll across industries explains firm R&D quite effectively; in recent years, information on payroll is much less effective in explaining firm R&D. As Table 4 shows, the substantial decrease in explanatory power which occurs over the period is associated with the decline in the importance of manufacturing payroll, and the increasing role of trade and services.

Changes in the Effectiveness of Payroll over Time

Our results show that the explained variance from equation (2) decreased substantially over time. In the late 1970s, observations on payroll in a few industries were in themselves sufficient to explain two-thirds or even three quarters of the variance in firm R&D spending. By the late 2000s, the same payroll variables explain only a third to 40 percent of observed R&D expenditures. The sharp decline in the effectiveness of R^2 raises important questions about the long-time Census procedure of classifying firms into industries based on observed payroll[14].

The decline in observed R^2 can be attributed to an increased prevalence of domestic or foreign outsourcing over time, whereby firms no longer manufacture their basic product but instead obtain their product from external producers. As firms increasingly rely on external supply chains, the observed

---

[14] This was a long-time procedure, but BRDIS did replace it in 2008.

industry distribution of firm payroll fragments, with a smaller proportion of payroll in the home industry and a greater proportion in various ancillary activities such as retail trade, wholesale trade, and management[15].

One method of investigating whether dispersion away from a firm's home industry has brought about a lower R^2 is to calculate, for each firm in every year, the Herfindahl Index of the distribution of payroll across industries. The central hypothesis is that if a firm has a lower Herfindahl value, indicating a broader dispersion of payroll across industries, this is associated with greater variance in the firm's error term u_i and therefore with a smaller contribution to R^2 in the cross-section. If a lower Herfindahl is consistently associated with lower R^2 in each cross-section, the cross-section relationship could then be used to determine how much the lower Herfindahl values which occur over time contribute to the observed decline in R^2.

Such empirical tests of whether firms with greater diversification of payroll contribute less to observed R^2 within each cross-section are an area of future work. If greater diversification successfully predicts lower explanatory power in each year's cross-section, such patterns may help us understand how the more diverse supply chains seen in more recent years have made payroll a less effective method of viewing each firm's activities, a matter of central concern to the Census[16].

5. Connections with Factory-Less Production

U.S. manufacturing employment declined substantially in recent years, especially since 2000. Economists typically attribute these employment losses to technological innovation and to the globalization of production. Factory-less production is probably only an extreme example of the more general long-term decline in U.S. manufacturing employment.

For example, Bayard, Byrne, and Smith (2014) provide a useful description of the prevalence of factory-less production. However, their work deals with large firms of the type included in the Standard and Poor's 500. It is therefore useful to supplement Bayard, Byrne, and Smith (2014) with a more representative sample of firms.

We have collected information on 503 firms in the 2008 BRDIS which perform R&D but report that their entire payroll occurs in establishments assigned to trade, with none in manufacturing, services, transportation, or other services. According to the payroll method, each of these firms should properly be assigned to trade. However, Census analysts (who determine the published data by assigning each firm to its main product) and the Census Industrial Register assign many of these firms to industries in manufacturing, especially in such high tech areas as computers, communications equipment, or pharmaceuticals. The Census methods also assign many of the firms in our 503 firm subsample to merchant wholesaling or to

---

[15] See Awuku-Budu and Carol Robbins (2014). They showed that, in recent years, firms' domestic payroll and R&D expenditures have shifted from establishments in manufacturing to establishments in services, trade, and management of companies and enterprises industries.
[16] This is no longer a central concern at least for BRDIS.

trade in computers, computer parts, semiconductors, and pharmaceuticals.

Data on these 503 firms provide useful information about what types of firms report their entire their entire payroll in trade. The data also include information on where different classification methods methods assign each such firms. These 503 firms represent a large variety of firms of different sizes, and are classified in several different industries. Such information therefore provides a different and complementary view of the prevalence of factory-less production from that given in Bayard, Byrne, and Smith (2014) or Bertrand and Fort (2013).

The firm ID is available for each of these 503 firms. By tracing these firms back in time, it is possible to understand how these firms became factory-less producers; whether they started out with domestic manufacturing establishments, which they later outsourced to foreign production, or whether they always purchased all their manufactured goods from abroad.

In addition, by preparing similar analyses for earlier years, it is possible to track the time dimension through which firms apparently engaged only in trade became important. Did firms concentrate in trade gradually from the 1980s, or did such patterns accelerate rapidly in the 2000s, when analysts first became greatly concerned about the classification of firms into trade? The addition of a time dimension should help us understand how and why factory-less production appeared. Furthermore, links to the corrections made by Census analysts and to the Census Business Register should help us understand what has occurred and how the Census has responded to these changes.

As in the discussion of the relationship between the Herfindahl index and r^2 in the previous subsection, the empirical work on these topics is a future venue for research.

6. Conclusions

Economists have been concerned about the validity of measures of the R&D expenditures occurring in each industry for many years. The main difficulty is that some multi-division firms are important R&D performers and the usual Census procedures assign all of such firm's expenditures to a single industry, which in some cases leads to industry misclassification. The 2008 BRDIS, which includes direct information on R&D by line of business, shows that firms conducting R&D in more than one line of business are not as prevalent as might have been anticipated. Nevertheless, the BRDIS does show that some firms do conduct R&D in more than one industry. By assigning research conducted in secondary industries to the appropriate industry, the industry distribution of R&D can be improved.

From 2008 forward, the amount of research actually conducted in each industry can easily be calculated from the BRDIS. However, no similar information is available for years prior to 2008, so no consistent time series is available for these improved industry designations.

This paper outlines an approach to assigning firm R&D to different industries for years prior to 2008. The method allows for industry differences in research intensity and utilizes all available information on payroll in each industry, which historically has been the main indicator used to assign firm R&D to industries. The empirical results are promising. The main difference is that R&D begins to

switch from manufacturing to services in earlier years, before services become sufficiently important that entire firms are reclassified.

References

Allen, Steven G. (2001), Technology and the Wage Structure, Journal of Labor Economics (April 2001), pp. 440-483.

Awuku-Budu, Christian and Robbins, Carol (2014), The Role of Industry Classification and Firm Structure in the Estimation of Research and Development Expenditures, Center for Economic Studies Working Paper 14-45, November 2014.

Bartel, Ann P. and Sicherman, Nachum (1999), Technological Change and Wages: An Interindustry Analysis, Journal of Political Economy (April 1999), pp. 285-325.

Bayard, Kimberly, Byrne, David, and Smith, Dominic (2014), The Scope of U.S. `Factoryless Manufacturing', presented at 2014 Research Data Center Annual Conference, U.S. Census Bureau, June 2014.

Bernard, Andrew B. and Fort, Teresa C. (2013) Factoryless Goods Producers in the U.S., Center for Economic Studies Working Paper 13-46, September 2013.

Cohen, Wesley M. and Klepper, Steven (1996), A Reprise of Size and R&D, Economic Journal (July 1996), pp. 925-951.

Crawford, Marissa J., Lee, Jennifer, Jankowski, John E., and Moris, Francisco A. (2014), Measuring R&D in the National Economic Accounting System, Survey of Current Business (November 2014), pp. 1-15.

Foster, Lucia and Grim, Cheryl (2010), Characteristics of the Top R&D Performing Firms in the U.S.: Evidence from the Survey of Industrial R&D, Center for Economic Studies Working Paper 10-33, September 2010.

Griliches, Zvi and Lichtenburg, Frank R. (1984), R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?', pp. 465-496 in Griliches, Zvi (Ed.), R&D, Patents, and Productivity, University of Chicago Press, Chicago, 1984.

Kelejian, Harry H. and Oates, Wallace E. (1974), Introduction to Econometrics, Harper & Row, New York, 1974.

Scherer, Frederic M. (1982), Inter-Industry Technology Flows and Productivity Growth, Review of Economics and Statistics (November 1982), pp. 627-634.

Smith, Shelly and Holdren, Alyssa E. (2013), Preview of the 2013 Comprehensive Revision of the National Income and Product Accounts, Survey of Current Business (March 2013), pp. 13-29.

Sveikauskas, Leo (1983), Science and Technology in United States Foreign Trade, Economic Journal (September 1983), pp. 542-554.

Sveikauskas, Leo (2007) R&D and Productivity Growth: A Review of the Literature, Bureau of Labor Statistics Working Paper 408, September 2007
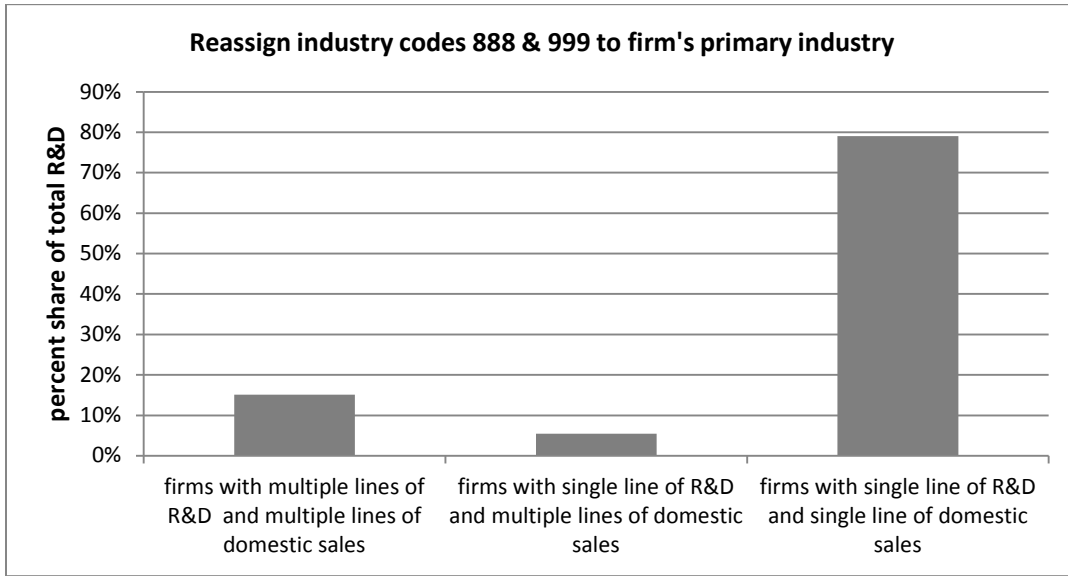
Figure 1a: Evidence from BRDIS

Reassign industry codes 888 & 999 to firm's primary industry

Table 1: Evidence of Firms with Multiple Lines of Business, R&D and Sales from BRDIS

Type of firm	Reassign Industry codes 888 & 999 to firm's primary industry	Keep industry codes 888 & 999 as industries outside of the firm's primary industry
Panel 1a (domestic R&D and sales)		
firms with multiple lines of R&D and multiple lines of domestic sales	15%	29%
firms with single line of R&D and multiple lines of domestic sales	5%	5%
firms with single line of R&D and single line of domestic sales	79%	65%
Panel 1b (worldwide R&D and sales)		
firms with multiple lines of R&D and multiple lines of domestic sales	13%	33%
firms with single line of R&D and multiple lines of domestic sales	4%	3%
firms with single line of R&D and single line of domestic sales	82%	63%

| Table 2: Regression Coefficients | | | | | |
| | Manufacturing | | Non-manufacturing | | |
Year	Coefficient	Std. Error	Coefficient	Std. Error	R-squared
1976	0.118***	(0.01)	0.0612***	(0.01)	0.70
1977	0.0903***	(0.03)	0.0648***	(0.02)	0.61
1978	0.0517**	(0.02)	0.0738***	(0.02)	0.47
1979	0.127***	(0.02)	0.0370**	(0.02)	0.64
1980	0.141***	(0.02)	0.0673***	(0.01)	0.67
1981	0.126***	(0.02)	0.0651***	(0.01)	0.65
1982	0.0503**	(0.02)	0.0938***	(0.02)	0.46
1983	0.0244*	(0.01)	0.250**	(0.10)	0.42
1984	0.0869***	(0.03)	0.0937	(0.06)	0.31
1985	0.139***	(0.03)	0.273**	(0.11)	0.68
1986	0.231***	(0.02)	0.0024	(0.00)	0.60
1987	0.109	(0.07)	0.0029	(0.00)	0.27
1988	0.807***	(0.11)	0.320*	(0.17)	0.78
1989	0.229***	(0.02)	0.168**	(0.07)	0.71
1990	0.233***	(0.02)	0.154**	(0.06)	0.69
1991	0.253***	(0.03)	0.161***	(0.06)	0.71
1992	0.290***	(0.03)	0.0819**	(0.04)	0.47
1993	0.268***	(0.04)	0.154***	(0.05)	0.63
1994	0.318***	(0.03)	0.0613**	(0.03)	0.66
1995	0.351***	(0.04)	0.0626**	(0.03)	0.68
1996	0.376***	(0.04)	0.0777***	(0.03)	0.66
1997	0.360***	(0.04)	0.0725***	(0.02)	0.62
1998	0.329***	(0.04)	0.0835***	(0.02)	0.55
1999	0.336***	(0.06)	0.0555***	(0.02)	0.44
2000	0.370***	(0.06)	0.0542**	(0.02)	0.39
2001	0.401***	(0.07)	0.000791	(0.00)	0.32
2002	0.284***	(0.07)	0.100***	(0.03)	0.37
2003	0.314***	(0.07)	0.102***	(0.04)	0.36
2004	0.297***	(0.06)	0.116***	(0.04)	0.39
2005	0.371***	(0.06)	0.0633***	(0.02)	0.36
2006	0.407***	(0.06)	0.0599***	(0.02)	0.38
2007	0.453***	(0.06)	0.0718***	(0.02)	0.40
2008	0.273***	(0.08)	0.0610***	(0.02)	0.20
All years	0.242***	(0.02)	0.0125*	(0.01)	0.28

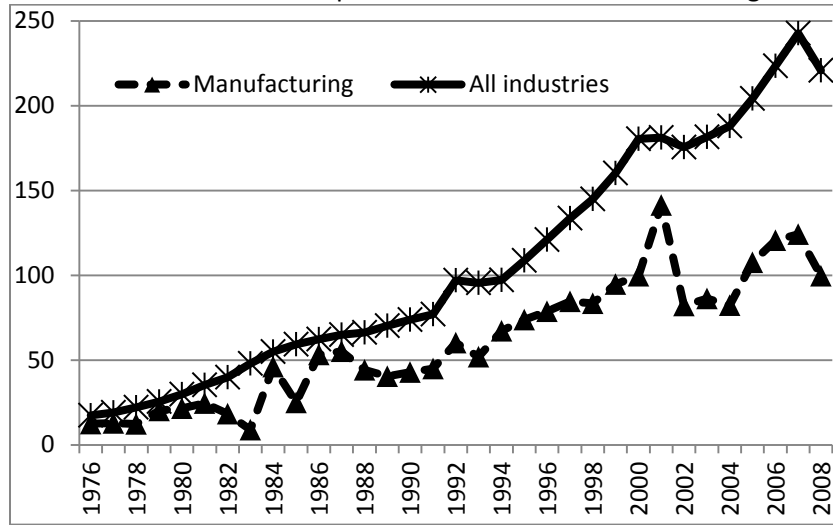Figure 2a: Distribution of Nonfederal R&D Expenditures between Manufacturing and Nonmanufacturing

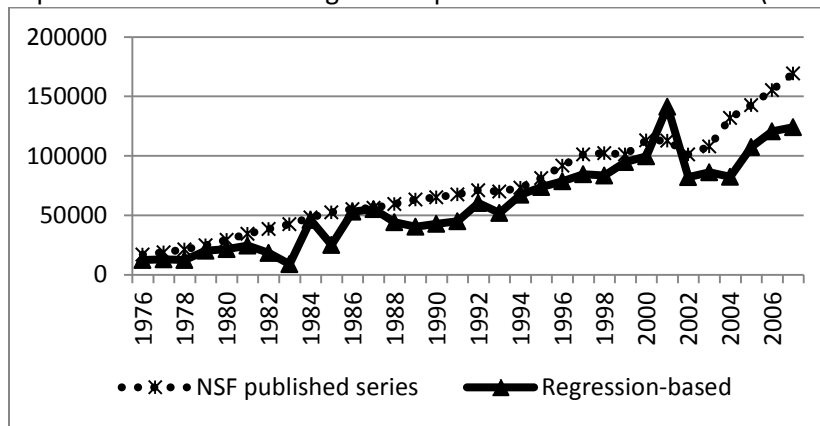Figure 2b: Comparison of Manufacturing R&D Expenditures with NSF Series (levels), 1999-2007

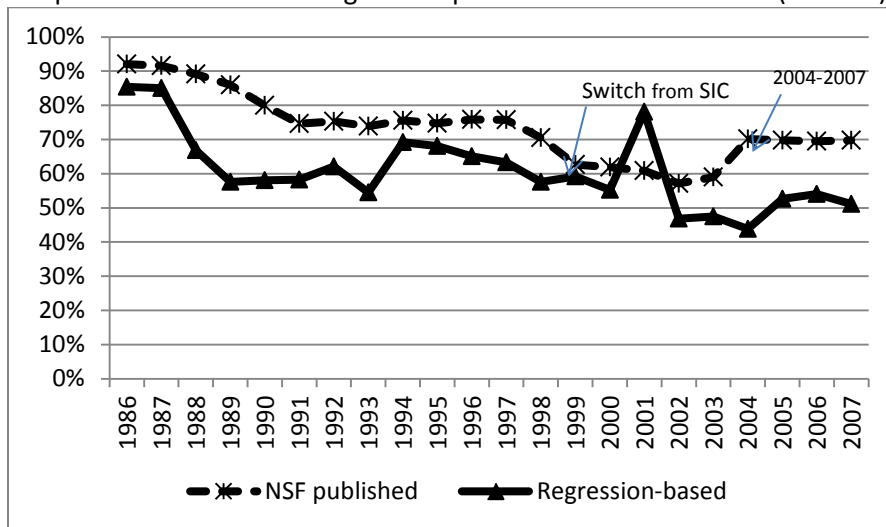Figure 2c: Comparison of Manufacturing R&D Expenditures with NSF Series (Percent), 1999-2007

Panel		1976-1979	2005-2008
	Table 3 : Regression Coefficients for Manufacturing		
	Panel 3a: 2-Digit NAICS Industries		
A	NAICS 31 simple manufactures such as food and apparel	0.007	-0.011
B	NAICS 32 which includes wood, paper, petroleum and the research intensive chemicals industry	.149**	.476***
C	NAICS 33 including metals, machinery, computer and electronics, and transportation equipment	.105***	.384***
D	Non-manufacturing	.058***	.064***
E	R-squared	0.63	0.35
	Panel 3b: 3-Digit NAICS Industries		
F	Chemical products	.253**	1.100***
G	Computer and electronic products	.308**	.693***
H	Transportation equipment	.149***	.316**
I	All other manufactures	.032**	0.059
J	Non-manufacturing	.039**	.060***
K	R-squared	0.75	0.41
	Panel 3c: 4-Digit NAICS Industries		
M	Pharmaceutical and medicine products	.440***	1.339***
N	Other chemical products	.242**	.516***
O	Semiconductor & other electronic products	0.329	1.773***
P	All other computer & electronic products	.338**	.442*
Q	Aerospace product & parts	.144**	.254**
R	All other transportation equipment	.149***	.487**
S	All other manufactures	.029**	.080*
T	Non-manufacturing	.039***	.056***
U	R-squared	0.76	0.47

Note: *** indicates significantly different from zero at the 99 percent level. ** is significant at the 95 percent level, and * indicates significant at the 90 percent level. The number of *s reported is the average number of *s for that variable over the four-year time period.

Figure3a:3-Digit NAICS, Comparison of Chemical Product R&D Expenditures with NSF published Series, 1999-2007
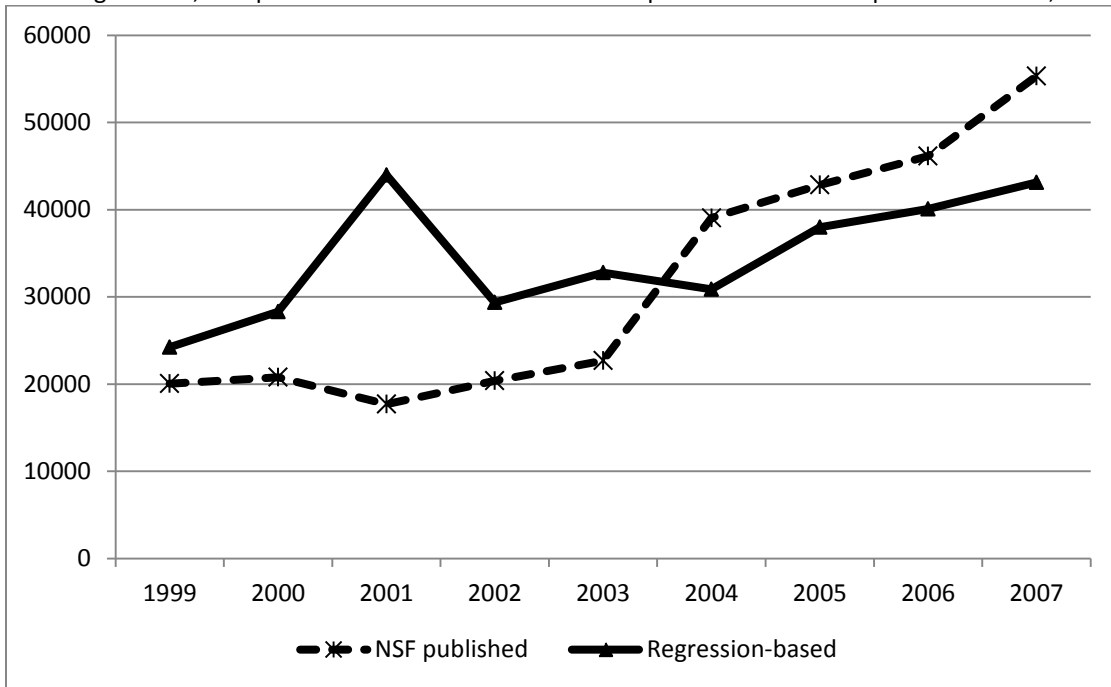

Figure3b:3-Digit NAICS, Comparison of Computer and electronic Products R&D Expenditures with NSF published Series, 1999-2007
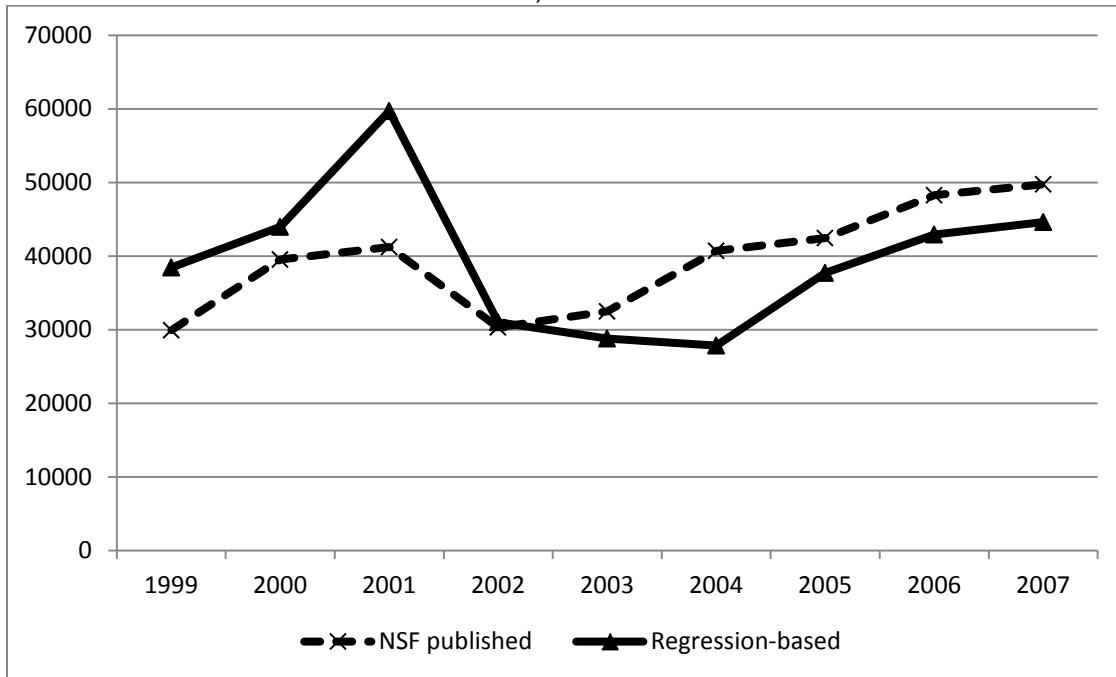
Figure 3c: 4-Digit NAICS, Comparison of Pharmaceutical and Medicine Products Manufacturing R&D with NSF Published Series, 1999-2007
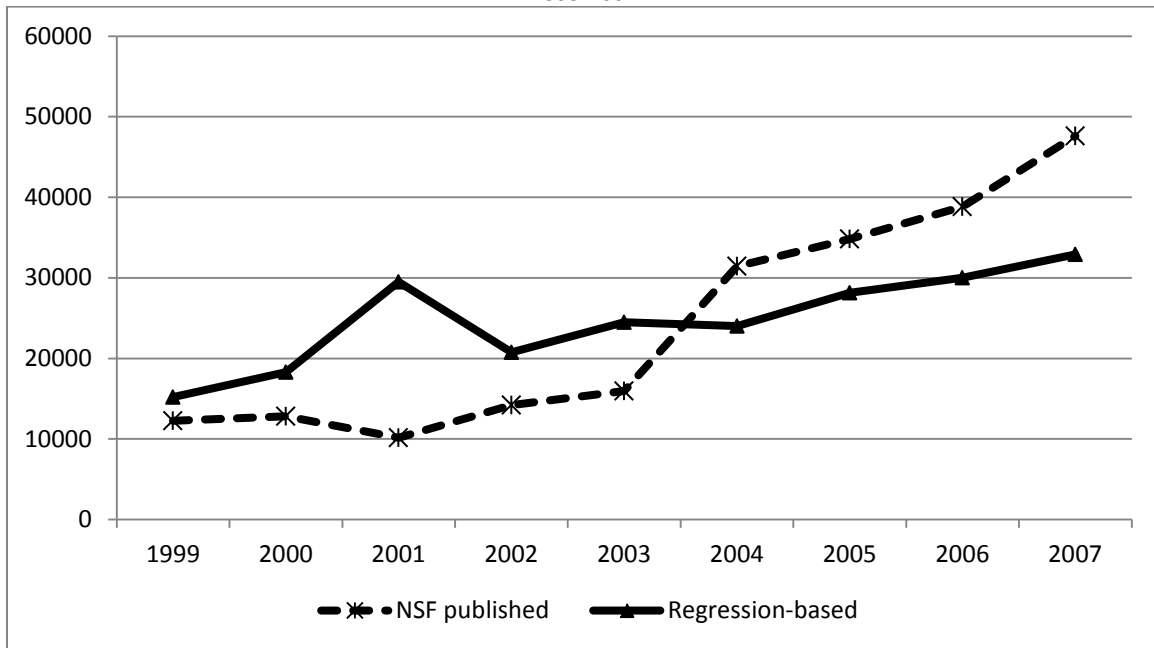

Figure 3d:.4-Digit NAICS, Comparison of Semiconductor and Other electronic Product Manufacturing R&D with NSF Published Series, 1999-2007
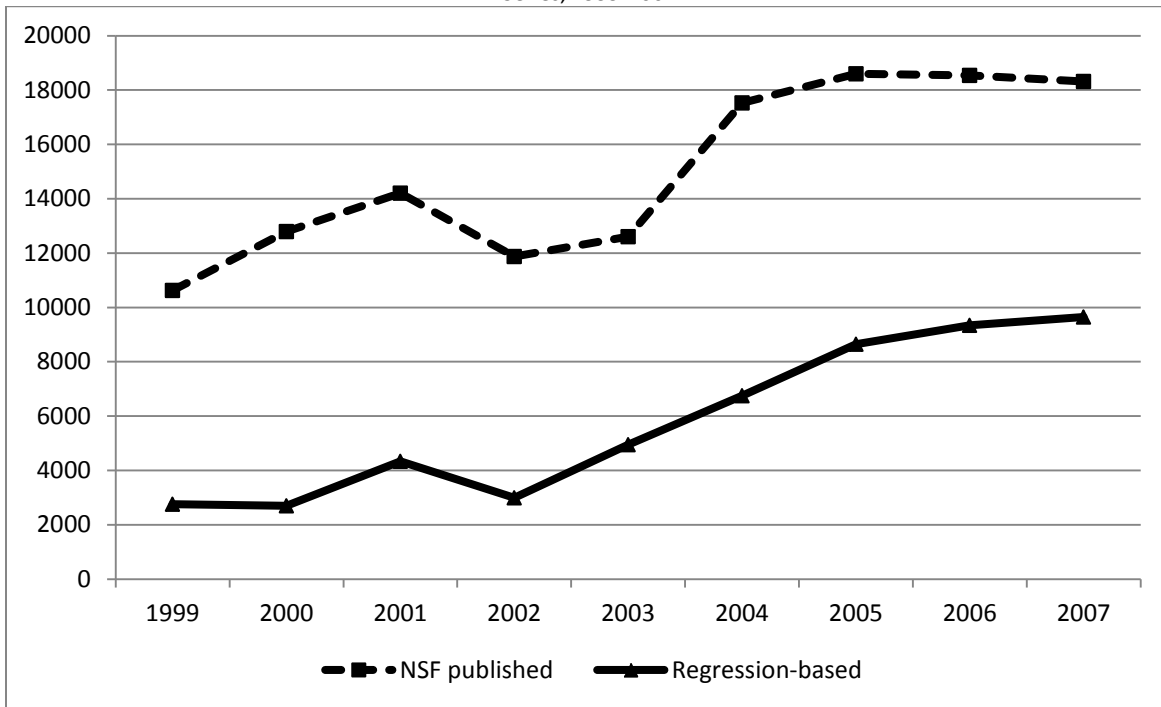
Table 4a: Regression of Nonfederal R & D Expenditures on Payroll by Primary and Secondary Industry Groups, 1976-2008						
	All firms	All firms exclude primary industry	Primary industry is computer and electronic products	Primary industry is chemical manufacturing	All manufacturing firms	All manufacturing firms excluding payroll in primary industry
Variables	(1)	(2)	(3)	(4)	(5)	(6)
payroll in manufacturing	0.366***	0.589***	0.0302	0.373***	0.235***	0.588***
	(0.048)	(0.067)	(0.055)	(0.142)	(0.059)	(0.069)
payroll in all other industries	0.0693	0.0725	0.111	-0.679***	0.0731	0.0764
	(0.055)	(0.060)	(0.487)	(0.118)	(0.058)	(0.064)
payroll in services	0.0928**	0.111**	0.300***	0.696***	0.0964**	0.112**
	(0.039)	(0.045)	(0.046)	(0.087)	(0.041)	(0.047)
payroll in trade	0.0136	0.015	0.00874	0.863***	0.0131	0.0149
	(0.009)	(0.010)	(0.006)	(0.115)	(0.008)	(0.010)
payroll in primary industry	0.145***				0.218***	
	(0.016)				(0.027)	
Year dummies	yes	yes	yes	yes	yes	yes
Number of observations	161300	161300	23400	14000	121300	121300
R-squared	0.323	0.237	0.391	0.702	0.389	0.267

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4b: Regression of Nonfederal R & D Expenditures on Payroll by Primary and Secondary Industry Groups, 2008						
	All firms	All firms exclude primary industry	Primary industry is computer and electronic products	Primary industry is chemical manufacturing	All manufacturing firms	All manufacturing firms excluding payroll in primary industry
Variables	(1)	(2)	(3)	(4)	(5)	(6)
payroll in manufacturing	0.214	0.216	0.128	-0.432	-0.242	-0.145
	(0.155)	(0.162)	(0.141)	(0.326)	(0.183)	(0.230)
payroll in all other industries	-0.106	-0.0933	-0.946	-2.435*	-0.372	-0.315
	(0.121)	(0.111)	(0.985)	(1.358)	(0.249)	(0.223)
payroll in services	0.0932*	0.116**	0.0344	1.206***	0.186	0.284*
	(0.052)	(0.058)	(0.152)	(0.394)	(0.123)	(0.148)
payroll in trade	0.480*	0.506*	1.382***	0.683	0.578***	0.758***
	(0.255)	(0.269)	(0.419)	(0.432)	(0.190)	(0.234)
payroll in primary industry	0.0893*				0.497***	
	(0.052)				(0.090)	
Number of observations	8800	8800	1100	900	6400	6400
R-squared	0.273	0.236	0.596	0.681	0.517	0.402

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.