The Importance of Reallocation in Cyclical Productivity and Returns to Scale: Evidence from Plant-level Data

Yoonsoo Lee*

Abstract

This paper provides new evidence that estimates obtained from aggregated data may not provide reliable estimates of average firm-level parameters, if the composition of producers with different levels of productivity changes over the business cycle. I examine plant-level data and show that, due to the countercyclical reallocation of output among heterogenous plants, productivity measures based on aggregate data do not reflect the actual path of productivity of a representative firm over the business cycle. Countercyclical reallocation may also bias aggregate estimates of returns to scale and help explain why decreasing returns to scale are found in industry-level data.

JEL Classification: D24, E32, O47
Keywords: Entry, Exit, Productivity, Returns to Scale

* Federal Reserve Bank of Cleveland, yoonsoo.lee@clev.frb.org. I am very grateful to Mark Bils for his valuable advice and encouragement. I thank Eric Bartelsman, Daniele Coen-Pirani, Yongsung Chang, Satyajit Chatterjee, Tim Dunne, John Fernald, Joao Gomes, Paul Gomme, Lance Lochner, Toshihiko Mukoyama, Peter Rupert, Mark Schweitzer, as well as seminar participants at the Rochester Wegamans Conference, the Federal Reserve Bank of Cleveland, the Federal Reserve System Committee Meeting in Macroeconomics, the Midwest Macro Meeting, Carnegie Mellon University, Concordia University, Seoul National University, the Wharton School, for useful comments. Randy Becker, Renee Fields, Shawn Klimek, and Christopher Kurz were especially generous with their time in helping me with the Census data. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Federal Reserve Bank of Cleveland or of the Board of Governors of the Federal Reserve System. The research in this paper was conducted while the author was a Special Sworn Status researcher of the U.S. Census Bureau at the Carnegie Mellon University Census Research Data Center and the Michigan Census Research Data Center. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed.

1 Introduction

Two important components of macroeconomic models are returns to scale and changes in productivity. To estimate a representative firm's production function or short-run productivity changes, researchers constructing business cycle models have used aggregate data. These estimates are based on the assumption of a representative firm, but a growing number of studies based on establishment-level data suggest a problem with this assumption. These studies have found large-scale, ongoing reallocations across individual producers with different productivity levels (for a survey, see Bartelsman and Doms, 2000).

What matters for macroeconomic models is the returns to scale across firms, appropriately aggregated. Studies based on aggregate data assume implicitly that the composition of producers remains the same over the business cycle, and the assumption has always limited the interpretation of the results. This limitation is well documented in earlier labor literature, where it has been shown that heterogeneous worker quality over the cycle creates composition effects in aggregate data of real wages: The entry and exit of low-wage workers creates a countercyclical composition bias in aggregate wages, which obscures the true cyclicality of real wages, as it is experienced by individual workers (Stockman, 1983; Bils, 1985; Solon et al., 1994; Chang, 2000).[1]

In this study, I show that aggregated data may not provide reliable estimates of average firm-level parameters if the composition of producers with different levels of productivity changes over the business cycle. I use detailed plant-level data on U.S. manufacturing to show that such composition changes do occur, that they are overall countercyclical, and that, as a result, estimates of productivity changes based on aggregate data may not reflect the true cyclicality of productivity that a representative firm experiences during the course of the business cycle. As predicted in

[1] In contrast to the labor market, in which changes in the composition of the workforce are mostly explained by changes that occur on the extensive margin, i.e., the entry and exit of workers, changes in the composition of producers are mostly explained by changes that occur on the intensive margin of production, i.e., reallocations between continuing plants. The effect of the entry and exit of plants is relatively small, because entering and exiting plants account for a small share of output in the industry.

theoretical models of the cleansing effect of recessions (Caballero and Hammour, 1994), I find that output shares are reallocated from less-productive to more-productive plants during recessions. Furthermore, plants entering and exiting during a recession are more productive than those entering and exiting during a boom. As less productive firms are driven out of or kept from entering the market during recessions, overall total factor productivity (TFP) may rise, reducing the procyclicality of aggregate TFP.

In agreement with Baily et al. (2001)'s study of the cyclical behavior of labor productivity, I find the countercyclical effect of this reallocation (or the cleansing effect of recessions) is relatively weak for the manufacturing sector as a whole.[2] However, in certain industries at the lower level of data aggregation (e.g., durables), I find the effect is statistically significant. This finding is consistent with the view that reallocations are a within-industry phenomenon (Davis et al., 1996) and suggests that the cleansing effect of recessions may be more important within industries.

Having established the significant role that composition changes over the business cycle play in understanding the cyclical behavior of aggregate TFP, I further show that such changes may help explain why decreasing returns to scale have been found at the industry level in disaggregated data. Reallocations that have a countercyclical effect on productivity tend to bias estimates of returns to scale that are based on aggregate data for the following reason. As inputs are reallocated toward less-productive firms during booms, the marginal response of output to changes in input may appear lower in aggregate data than the marginal increase in the output of a typical firm, leading to smaller estimates of returns to scale in aggregate data. In most two- and four-digit SIC industries that are associated with significantly decreasing returns-to-scale estimates in the aggregate data, I find that as the plant-level data are aggregated to the industry level, estimates of returns to scale *decrease*.

[2] The lack of strong evidence of the cleansing effect of recessions has led researchers to focus on other aspects of recessions (e.g., sullying effect as in Barlevy (2002)).

This finding is at odds with previous research based on industry-level data, which documents higher returns-to-scale estimates at higher levels of data aggregation (e.g., Caballaero and Lyons, 1992; Basu and Fernald, 1997). In fact, my finding of a countercyclical effect of reallocations between plants contrasts sharply with other researchers' findings on the effects of between-industry reallocations, such as those of Basu and Fernald (1997, 2002) and Basu et al. (2004). Basu and Fernald (2002) claim that the reallocation between two-digit industries with different marginal products explains much of the cyclicality of aggregated productivity. However, the evidence from plant-level data clearly shows that, when corrected for aggregation effects running from the plants to the manufacturing industry, aggregated productivity appears more procyclical. Furthermore, the plant-level evidence in this paper points to some potential problems in studies based on industry-level data, namely that of the production function estimates, which are based on industry-level data.

Most of all, the previously mentioned studies ignore important reallocations that occur within a specific industry. While these studies typically assume that productivity change is exogenous at a lower level of data aggregation (two- or four-digit), I find that the effects of reallocations within industries on industry-level productivity measures are statistically significant and may have larger magnitude than in the manufacturing sector as a whole. While the effects of reallocations are statistically significant, such countercyclical effects within industries seem to be offset by the procyclical effects of between-industry reallocations (as in Basu and Fernald, 1997), and they become less significant at the higher level of data aggregation (e.g., manufacturing as a whole). As a result, this specific aggregation bias (i.e., that which is caused by reallocations across heterogeneous plants) can be bigger at the disaggregate level, causing industry-level estimates to diverge further from true firm-level parameters.

In fact, the plant-level estimates of returns to scale across two-digit industries suggest that the differences in true marginal products across industries, once corrected for composition bias, may not be as large as they appear in industry-level data. Compared to industry-level estimates, which vary substantially across industries, plant-level estimates are rather closer to constant returns to scale. While industry-level estimates show a wide variation of returns to scale across industries, a part of the variation in industry-level estimates reflects differences in the effect of the bias caused by within-industry reallocations, rather than the inter-industry variation in returns to scale of an average plant.

Differences in plant-level productivity must be interpreted with caution as well. Because plant-level prices are not observed, I use a measure that is based on real revenue, which reflects price or demand variation within an industry and may fail to distinguish such differences from productivity differences (e.g., Foster et al., 2006). Furthermore, it is not clear how much of the difference in TFP across plants is explained by differences in the quality of inputs. Although inputs are reallocated between plants with different (average) productivity levels, it does not necessarily imply that marginal products differ between plants.

Section 2 of this paper describes the data used in this study and the empirical evidence of composition changes over business cycles. Section 3 examines how changes in the composition of producers may affect returns-to-scale estimates for different levels of aggregation. Conclusions are presented in the last section.

2 Composition Changes and the Cyclicality of Productivity

2.1 Measurement of Productivity and Data Description

The plant-level data used in this study are taken from the Longitudinal Research Database (LRD) maintained by the Center for Economic Studies at the U.S. Bureau of the Census. In this study, I use the Annual Survey of Manufactures (ASM) portion of the LRD for the years 1972 through 1997. Because the entire ASM comprises a representative sample of manufacturing plants (Davis et al., 1996), the survey allows me to assess the contribution of entering and exiting plants to the cyclical behavior of productivity, as well as the impact of output reallocation across plants.

Plant-level productivity is measured using a standard total factor productivity index similar to that used by Baily et al. (1992) and by Foster et al. (2001). The TFP index for plant j is computed as follows:

$$\ln tfp_{jt} = \ln Y_{jt} - \alpha_l \ln L_{jt} - \alpha_m \ln M_{jt} - \alpha_k \ln K_{jt}\,, \qquad (1)$$

where Y_{jt} is real gross output, L_{jt} is labor input, M_{jt} is real materials, and K_{jt} is real capital stock.[3] The input cost shares for four-digit industries are used as the measure of the corresponding factor elasticities.[4]

There are two problems in measuring cost shares in the ASM. First, the ASM only includes the wage and salary costs of labor. In calculating labor's share of total costs, I follow Bils and Chang (2000), magnifying each four-digit industry's wage and salary payments to reflect other labor payments, such as fringe payments and employer FICA payments.[5] Another problem is that capital costs are not available. Given that previous studies by Rotemberg and Woodford (1995) and Basu and Fernald (1997) find small profits in manufacturing, I ignore profit at the industry level so that total revenue is equal to total cost. Next, I calculate the share of costs for input *J* in the total revenue from the four-digit industry-level data, aggregated from the ASM panels. For these computations, I consider capital expenditure shares to be residuals.[6]

Real gross output is measured as the total value of sales, deflated by the four-digit industry-specific deflator. All output, materials, and investment deflators are from the NBER manufacturing productivity data set (Bartelsman and Gray, 1996).[7] Labor input is measured as

---

[3] Although it is possible to adjust output for the change in inventories, inventories for some plants (in particular, for small plants) are imputed (Baily et al., 2001). To avoid a possible measurement issue, I have chosen to use gross shipments as a simple measure.

[4] This procedure implicitly assumes that all plants in the industry operate with the same production technology, a common assumption in such studies.

[5] Bils and Chang (2000) use information from the National Income and Product Accounts to calculate the ratio of these other labor payments to wages and salaries at the two-digit industry level.

[6] An alternative way to measure capital's cost share is to use industry-level (two-digit) rental payments series. To avoid introducing measurement errors, I have chosen to measure capital's cost share as a residual. Basu et al. (2004) used the same strategy and found a result similar to direct attempts at measuring the share of capital expenditures.

[7] See Bartelsman and Doms (2000) for the drawback to using deflated production to measure productivity. Some caution is needed in interpreting the results. Ignoring any quality improvement in output that is not reflected in the

total hours for production and nonproduction workers. Because hours for nonproduction workers are not collected, I estimate the value for total hours by following the method in Baily et al. (1992), which is to multiply the total hours of production workers by the ratio of the total payroll for all workers to the payroll for production workers. Material input is measured as the cost of materials deflated by the four-digit industry materials deflator. Capital stocks for equipment and structures are constructed using the perpetual inventory method.[8]

2.2 Patterns of Entry and Exit over the Business Cycle

In this study, entering plants are either new plants, which appeared in the LRD for the first time, or plants that restarted production after a certain period of inactivity. Similarly, exiting plants include those that stopped producing during the following period and stayed inactive (i.e., they had zero employees or zero output) for a certain period of time, as well as those that permanently shut down.[9] As discussed in detail in Davis et al. (1996), samples in the ASM panels are rotated every five years. Only large “certainty” plants are continuously observed across different ASM panels; as a result, it is very difficult to measure entry and exit between the two years in which the panels are rotated. In order to avoid measurement errors caused by the panel rotations, entries and exits measured between two different ASM panels, namely for the years 1973–74, 1978–79, 1983–84, 1988–89, and 1993–94, are excluded from the results.

Table 1 summarizes the share of the total number of plants that entering and exiting plants

---

deflator may result in a downward bias in productivity growth. If new plants enter a market with new products having higher prices, and the number of new plants increases during a boom, the use of a single industry-level deflator may lead to overestimating the procyclicality of aggregate productivity.

[8] I follow Dunne et al. (1997) closely in constructing capital stock. For the benchmark, I use the book value of structure or equipment deflated by the two-digit industry capital deflator from the BEA (2-digit). I use the sum of structure and equipment as capital. The depreciation rate has been obtained from the BEA (2-digit).

[9] While I include temporary entry and exit to examine underlying changes in the aggregate data, those measures of entry and exit may be different from the measures of plant births (new startups) and deaths (permanent shutdowns) in the literature. The pattern reported in Table 1 did not change when I excluded temporary entry and exit and identified new startups and permanent shutdowns following Davis et al. (1996).

account for over the entire sample period, as well as the share they contribute to total employment and output, and their relative TFP indexes. ASM sampling weight is used in calculating statistics so that the sample is representative of all U.S. plants. Entering plants account for about 7 percent of plants in a given year, while 10 percent of plants in a given year stop producing during the following year. Entering and exiting plants tend to be smaller than continuing plants, as reflected in their generally smaller shares of employment and output (2–3 percent). The last column of Table 1 reports the relative TFP indexes for entering and exiting plants. These indexes consist of the weighted averages of TFPs for entering or exiting plants, divided by the weighted averages of TFPs for continuing plants in the same four-digit industry during the same year. I find that within the same four-digit industry, entrants are relatively more productive than continuing plants, while exiting plants are less productive than continuing plants.

The same statistics are separately reported for economic boom and recession periods in order to illustrate how the contributions of entering and exiting plants change over time. The second row of Table 1 summarizes the shares and relative TFP for entering and exiting plants for periods when the growth rate of real GDP exceeded 4 percent (i.e., a boom). The third row provides the same statistics for periods when the growth rate of real GDP fell below 1 percent (i.e., a recession). While the number of entering plants increases during a boom, the output share accounted for by entering plants does not increase to a significant degree.[10] This finding is partly explained by the relatively low productivity of entrants during a boom. Overall, plants that enter during a recession or in normal times are more productive than continuing plants. In contrast, plants that enter during a boom are less productive than continuing plants in the same industry. Although the magnitudes are relatively small, these differences in productivity over the business cycle are also found for exiting plants. Plants that exit during a recession are more productive than those that exit during a

---

[10] The exit rate is higher during booms because, in this study, the category of exiting plants includes those that stop production temporarily. Most of these plants enter during the boom part of the cycle, operate for a short period of time, and stop operating until they reenter the market. The employment losses due to exiting plants are higher during recessions.

boom.[11] This difference in the relative productivity of plants that move in and out of production suggests that aggregate productivity is subject to composition effects.

2.3 Decomposition of Aggregate Productivity Changes

Using plant-level data, I examine the extent to which such changes in the composition of producers or shifts in the share of outputs across plants affect the cyclical patterns of aggregate productivity. Following Baily et al. (2001), I have decomposed the time series changes in aggregate productivity into components that reflect a within-plant effect and other effects that reflect the reallocation of shares across plants including the effect of entry and exit:[12]

$$\Delta \ln TFP_t = \sum_{j \in Conti} \bar{s}_j \Delta \ln tfp_{jt} + \sum_{j \in Conti} \Delta s_{jt} (\overline{\ln tfp_j} - \overline{\ln TFP}) + \sum_{j \in Entry} s_{jt} (\ln tfp_{jt} - \overline{\ln TFP}) - \sum_{j \in Exit} s_{j,t-1} (\ln tfp_{j,t-1} - \overline{\ln TFP}), \qquad (2)$$

where $\ln tfp_{jt}$ is the TFP index for plant j at time t, $\ln TFP_t$ is the aggregate TFP index at time t, s_{jt} is the share of output at plant j at time t, and a bar over a variable indicates the average of the variable over the base and end years ($t - 1$ and t). Because a sample of plants from the ASM is used, the share is further inflated by the ASM sampling weight. The first term in the equation reflects changes in productivity from continuing plants, holding output shares fixed (often interpreted as a "within" effect). The second term reflects changes in output shares from continuing plants for fixed levels of productivity (often interpreted as a "between" effect). The last two terms represent the contribution of entering and exiting plants, respectively. These two terms,

[11] The results were similar when I classified booms and recessions in a different way (i.e., boom (recession): years in which the growth rate of real GDP was greater (less) than the average growth rate of real GDP during the sample period (1972-1997).

[12] See Foster et al. (2001) for excellent reviews of previous studies using different decomposition methodologies and measurement issues. To focus on the effects of reallocation on the aggregate economy, the decomposition is done at the highest level of aggregation (e.g., the whole manufacturing sector). The decomposition thus includes reallocations between plants in different industries.

together constituting the net entry effect, along with the second term, the between effect, represent the effect of reallocations across plants on aggregate productivity changes.

In this decomposition, the change in shares in the second (between-plant) term is weighted by the deviation of plant-level productivity from the average of aggregate productivity, so that an increase in the share of output for a plant contributes positively only if the plant has higher productivity than the average aggregate productivity. In a similar manner, a new plant contributes positively to the aggregate change only if its productivity is above the average, while an exiting plant contributes positively only if its productivity is below the average.

The results of these decompositions are reported in Figure 1. As in Section 2.2, the results reported exclude the first ASM years in each panel in order to avoid measurement errors due to ASM sample rotations. The values for these missing years are interpolated in the figure. The decomposition components for total factor productivity reveal cyclical patterns similar to those found for labor productivity in Baily et al. (2001). The within-plant component shows clear procyclical behavior. It increased sharply during the booms of 1976 and 1983 and decreased markedly during the recessions of 1980 and 1991. Excluding the first ASM panel years, the correlation between the within-plant component and contemporaneous real GDP growth is 0.69. Whereas the within-plant term is very procyclical, the between-plant term moves in a countercyclical direction. The between-plant component increased during the recessions of 1975, 1982, and 1991 and decreased during the recovery years of 1976, 1983, and 1992. Although the contribution of plant entry and exit to the annual change in aggregate productivity growth was relatively small (with the exception of the 1990s), the net entry component also moved in a countercyclical direction. These countercyclical reallocation terms suggest that output shares shift from less-productive toward more-productive plants during recessions. As a result of these countercyclical tendencies, aggregate productivity may look less procyclical than true productivity at individual plants.

The magnitude of such countercyclical effects varies across industries from one level of data aggregation to another. Table 2 reports the results from ordinary least squares (OLS), when those reallocation terms are regressed on real GDP growth. For the manufacturing sector as a whole, the coefficients were negative but not statistically significant. Such a mild countercyclical reallocation effect is consistent with previous empirical studies, in which the effect overshadowed the cleansing view of recessions (see Barlevy, 2002). However, the results for disaggregated industries suggest that a cleansing effect can be statistically significant and potentially more important in certain industries. In durables for example, while within-plant TFP increases about 0.75 percent when real GDP grows 1 percent, reallocations (between effects and net-entry effects taken together) may decrease the response of aggregate TFP growth by as much as one-half percent. In six of the 20 two-digit SIC industries, I find that the countercyclical effects of these reallocations were statistically significant.

3 Composition Bias in Aggregate Estimates of Returns to Scale

The decomposition results suggest that countercyclical reallocations may cause a downward bias in the returns-to-scale estimates. Because the output shares of more-productive plants increase during recessions, the extent to which aggregate output decreases would be smaller than the decreases that would have been observed in a representative plant. Furthermore, because the shares of less-productive plants increase more during booms, the increase in output would look smaller in the aggregate data than the marginal increase in the output of a representative plant.

3.1 Estimating Returns to Scale and the Effect of Composition Bias

In this section, I assess the size of the potential bias caused by composition changes. The key equation to estimate returns to scale follows Basu and Fernald (1997):

$$\begin{aligned} dy &= \gamma[c_L dl + (1 - c_L - c_M)dk + c_M dm] + dz \\ &= \gamma dx + dz, \end{aligned} \tag{3}$$

where *dy*, *dl*, *dk*, and *dm* are the growth rates of, respectively, output, labor, capital, and materials, and c_J is the share of costs for input *J* in the total cost. That is, the growth rate of output, *dy*, equals the returns to scale (γ) multiplied by the cost-share-weighted growth in inputs, *dx*, plus the productivity growth, dz. Cost minimization implies that returns to scale equal the ratio of average to marginal cost. Although inputs are plant-specific, I use industry-level input cost shares, averaged over the beginning and ending years of the period of change.

Many researchers, relying on a representative-firm framework, have used aggregated data and run regressions similar to Equation (3) to estimate returns to scale. This procedure implicitly assumes the existence of an aggregate production function, while dz now reflects aggregate productivity changes. As discussed in the previous section, aggregate productivity changes can be decomposed into components that reflect productivity changes within the plant and other components that reflect reallocations of shares across plants, including entry and exit:

$$\begin{aligned}
dy &= \gamma dx + dz \cong \gamma dx + d\ln TFP \\
&= \gamma dx + \sum_{j \in Conti} \bar{s}_j \Delta \ln tfp_{jt} + \sum_{j \in Conti} \Delta s_{jt} (\overline{\ln tfp_j} - \overline{\ln TFP}) \\
&\quad + \sum_{j \in Entry} s_{jt} (\ln tfp_{jt} - \overline{\ln TFP}) - \sum_{j \in Exit} s_{j,t-1} (\ln tfp_{j,t-1} - \overline{\ln TFP}) \qquad (4) \\
&= \gamma dx + \sum_{j \in Conti} \bar{s}_j \Delta \ln tfp_{jt} + \text{"}Reallocations\text{"} \\
&= \gamma dx + \sum_{j \in Conti} \bar{s}_j \Delta \ln tfp_{jt} + \delta dx + \varepsilon.
\end{aligned}$$

Given that reallocations are negatively correlated with aggregate input changes, a regression run on aggregated data may be subject to a bias caused by composition changes.[13]

In order to assess the size of the bias that may be present in studies using aggregated data, the "reallocations" term is regressed on aggregated input changes, dx, constructed from the LRD. For the manufacturing sector as a whole, the regression coefficient, δ is –0.043 (see Table 3). As expected, because of the relatively small contribution of plant entries and exits to aggregate productivity growth, the bias caused by net entry (Column 3) is much smaller than the bias caused by between-plant reallocations (Column 2).

Although the effects of composition bias on the returns-to-scale estimates for manufacturing

[13] In equation (4), $d \ln TFP$ is calculated assuming constant returns to scale. Therefore, the bias, δ should be interpreted with the understanding that the returns-to-scale coefficient, γ , is implicitly restricted to being equal to 1.

as a whole might not be large, the results for durable and nondurable manufacturing suggest that the effect of composition bias may be larger and possibly significant at more disaggregated levels of data.[14] Previous studies, such as Basu and Fernald (1997), which find decreasing returns to scale based on two-digit industry-level data, also suggest that composition bias may be more important at more disaggregated levels of data. Furthermore, Basu and Fernald find that reallocations between industries, in contrast with reallocations between plants, are procyclical in the sense that inputs are reallocated toward industries with higher returns to scale during a boom in the cycle. As a result, the estimate of returns to scale can be higher at the higher level of data aggregation.

Since the effect of between-industry reallocations may offset the effect of the composition bias caused by within-industry reallocations across plants, it is more relevant to examine the effect of composition changes within a narrow industry (e.g., four-digit).[15] In the next section, I discuss the effects of reallocations on the estimates of returns to scale at the four-digit industry level. However, in consideration of space limitations, I report the estimates at the two-digit level. In comparing the plant-level estimates to the two-digit industry-level estimates, the reader should keep in mind that two-digit industry-level estimates will reflect a bias caused by between-industry reallocations (i.e., reallocations across four-digit industries with different *returns to scale*) in addition to the bias caused by reallocations, on which this study is focused (i.e., reallocations between plants with different *productivity*).

[14] I found a significant reallocation effect in durables when the reallocation term is regressed on dx, constructed from the NBER manufacturing database. Although I report results based on the aggregated LRD to keep the data consistent, dx based on aggregated LRD may understate aggregate input changes, due to sample attrition (Appendix of Davis et al., 1996).

[15] Such differences in the effects of reallocations (i.e., between-plant vs. between-industry) explain why the countercyclical effect of between-plant reallocations is statistically significant at the disaggregate industry level but not in manufacturing as a whole. If resources are reallocated toward plants in industries with higher productivity levels, such procyclical effects of between-industry reallocations will offset the countercyclical effect of between-plant reallocations within industries.

3.2 Estimates of Returns to Scale: Industry- vs. Plant-level Data

One straightforward method of avoiding composition bias is to measure returns to scale at the plant level, giving fixed weights to the exact same plants over time. In this section, I estimate the baseline model in Equation (3), using an ordinary least squares (OLS) regression at two different levels of aggregation, i.e., the plant level and the two-digit SIC industry level.[16]

A potential problem of plant-level analysis is the attenuation bias caused by measurement errors. Previous studies suggest that plant-level returns to scale might be understated by measurement errors present in plant-level hours or capital stocks.[17] Because the specification requires measuring changes in inputs and outputs, first-differencing variables may magnify the attenuation bias, leading to a much smaller returns-to-scale estimate. As a standard response to errors in variables, I introduce an instrument: the cost-share-weighted growth in inputs, dx, measured over $t + 1$ and $t - 2$. Given that a firm's input decisions are highly correlated, plant-level input changes between t and $t - 1$ and those between $t + 1$ and $t - 2$ should be highly correlated as well. If measurement errors are not serially correlated, an IV estimation using the instrument will yield consistent estimates of returns to scale. Although IV estimation may help reduce the attenuation bias caused by measurement errors, it does not take into account the endogeneity of inputs.

Column (1) of Table 4 presents the plant-level OLS results for a pooled sample, which includes plants that have operated for two consecutive years in which they produced nonzero

[16] As pointed out by a number of researchers since Marschak and Andrews (1944), production-function estimates obtained by the OLS method are subject to a simultaneity bias generated by the relationship between productivity and input demands. However, because this study focuses on the effects of composition bias and its primary concern is the difference between estimates at different levels of aggregation, OLS estimation serves the purpose. Assuming that the specification is correct for both the plant- and industry-level regressions, a direct comparison of plant- and industry-level estimates allows an assessment of the size of aggregation bias in estimates obtained from industry-level data.

[17] See Westbrook and Tybout (1993) for evidence of measurement errors in capital. In the appendix, I present evidence of the attenuation bias caused by measurement errors, following the method of Griliches and Hausman (1986) and Goolsbee (2000). In Table A6, returns-to-scale estimates rise (even within the same set of plants), as changes in inputs and outputs are measured over a longer period of time.

output. All plant-level regression results are obtained from weighted regressions using the ASM sampling weight, so that the sample is representative of U.S. manufacturing as a whole. The IV estimates appear in bold if the Hausman specification test rejects the null hypothesis, i.e., consistency of the OLS at the 5 percent level of significance.

In order to examine the effects of aggregation, the same equation, (3), is estimated using the industry-level data, created by aggregating all plants in the industry for a given year. The ASM sampling weight is used to make the aggregated data mimic the data used in aggregate studies representing the entire industry. Because the industry-level estimation is less likely to be subject to measurement errors, the OLS estimates are used as the industry-level estimates.[18]

Before I discuss the results at the two-digit industry level, I want to mention the relationship between the composition bias and the difference between industry- and plant-level estimates. Overall, I find that industry-level estimates are larger than plant-level estimates in most two-digit industries in the durables sector. This result may appear inconsistent with the prediction of the productivity decomposition in the durables sector as a whole. However, as noted in previous studies, estimates of returns to scale at the higher level of aggregation may rise as resources are reallocated toward plants in industries with higher returns to scale during booms. Since reallocations within a two-digit industry include reallocations across four-digit industries with different returns to scale (as in Basu and Fernald, 1997), such reallocations across industries may dwarf the countercyclical effect of reallocations across plants with different productivity levels.[19]

In order to exclude the potentially countervailing effects of between-industry reallocations and to focus on the effect of reallocations between plants with different productivity levels, I need to focus on reallocations within very narrowly defined industry (i.e., at the four-digit level).

---

[18] As it turns out, the Hausman test suggests that the OLS estimates are not statistically different from the industry-level estimates obtained from IV estimations.

[19] Although Basu and Fernald's studies focus on reallocations at the two-digit industries, Wilson's (2000) study, which finds that estimates of returns to scale are higher at the two-digit than at the four-digit level, suggests that reallocations across four-digit industries within a two-digit industry are also procyclical and potentially important in explaining the rising returns–to-scale estimates.

Overall I find the direction of the composition bias is consistent with the difference between industry-level estimates and plant-level estimates at the four-digit industry level.[20] The correlation between the composition bias (δ) and the difference in the estimates (i.e., plant-level estimates – industry-level estimates) was negative and statistically significant (–0.18). This finding suggests that the composition bias explains the difference between plant-level and industry-level estimates.

Compared to the industry-level estimates, the plant-level estimates are smaller in industries with industry-level estimates larger than 1 and larger in industries with industry-level estimates smaller than 1. In Table 4, the two-digit industry-level estimates (Column 3) show wide variation, ranging from 0.413 for tobacco (SIC 21) to 1.621 for electrical machinery (SIC 36), whereas the plant-level estimates are rather closer to constant returns to scale.

The bias implicit in aggregated data might help resolve the puzzling finding of the decreasing returns to scale in previous studies that used industry-level data. For example, the statistically significant, decreasing returns-to-scale estimates found in the petroleum (SIC 29) and leather (SIC 31) industries suggest the existence of relatively large positive profits, which seems to contradict empirical evidence of a low profit level, as found in previous studies (Rotemberg and Woodford, 1995; Basu and Fernald, 1997). However, this does not necessarily imply that a typical plant in these industries has decreasing returns to scale, making positive pure profits. Even if the production of an average plant in these industries exhibits constant returns to scale, aggregation may create a bias in the aggregate estimates and lead to a different implication than would the true returns to scale of an average plant. This finding suggests that differences in industry-level estimates of returns to scale across industries may reflect differences in the size of the bias caused

[20] In 102 four-digit industries, I found decreasing returns to scale using industry-level data. However, in 66 of these, plant-level estimates suggested constant returns to scale. In 27 of these 66 industries, this difference can be explained by the composition bias (δ), which was negative and statistically significant. On the other hand, I found a significantly "positive" composition bias in some four-digit industries with industry-level estimates of increasing returns to scale. In 39 of 46 four-digit industries with industry-level estimates significantly greater than 1, plant-level estimates suggested constant returns to scale. In 11 of these 39 industries, the difference could be explained by the composition bias, which was positive and statistically significant.

by within-industry reallocations, rather than the between-industry differences in the returns to scale of an average plant.

Implication and Caveat

In order to address the question of what the estimates of the average firm-level returns to scale are, Table 5 presents returns-to-scale estimates in the manufacturing sector when various instruments are used to deal with measurement error. While I find slightly decreasing returns to scale for continuing plants, the returns-to-scale estimate is not very different from 1.[21]

However, appropriate caution should be used when the parameters in this paper are used for the calibration of a macroeconomic model. First, the production function at the plant level may be different from the aggregate production function at the industry level. While the plant-level is likely to be the better level for measuring the production technology of a representative agent, parameters estimated from plant-level data may fail to capture important macroeconomic mechanisms (e.g., reallocations as a propagation mechanism, as discussed in Basu and Fernald, 1997, or entry and exit), which are at work during business cycles.

Second, plant-level estimates are subject to biases, the direction of which is not always consistent. Overall, measurement errors in plant-level variables may cause returns to scale to be understated. Imperfect input measures may also overstate input changes over the cycle, causing a downward bias. For example, cyclical changes in inputs can be overstated when various types of overhead capitals (e.g., headquarters, intellectual property, organization capital) are omitted, or when cyclical changes in input quality and prices are not properly measured (e.g., lower quality of inputs or higher input prices during booms). In addition, as Klette and Griliches (1996) have pointed out, returns-to-scale estimates may be biased downward if firms sell outputs at different

---

[21] I follow Griliches and Hausman (1986) in choosing instruments. While using variables with longer lags as instruments helps to reduce the attenuation bias, it limits the sample to longer-lived survivors (e.g., 5-year survivors in Column 4), which tend to be bigger and may have higher returns to scale. While the IV estimate in Column 5 is not subject to this problem, it may react to industry-level variation, rather than plant-level variation, leading to an estimate that is closer to an industry-level estimate.

prices (i.e., imperfect competition), while firm-level outputs are deflated based on a common output deflator.[22] On the other hand, the estimates of returns will be biased upward if industry-level variation in inputs is correlated with technology changes. Failure to measure variation in utilization may also lead to an upward bias. While both industry- and plant-level estimates are subject to such biases, the size of a given bias may vary across different levels of aggregation.

4 Conclusion

In examining longitudinal plant-level data of U.S. manufacturing, I find that actual productivity may be more procyclical than observed aggregate productivity. While reallocations among producers over the business cycle create a countercyclical component in aggregate productivity, the effect of this reallocation is relatively weak for the manufacturing sector as a whole. However, the cyclical effects of reallocations between plants are statistically significant at the lower level of data aggregation and raise a caution in interpreting production function estimates based on disaggregate industry-level data. Reallocations may bias estimates of returns to scale based on industry-level data and lead to a different implication than would the true returns to scale of an average plant.

[22] In a recent study, Gorodnichenko (2006) argues that without a proper measure of firm-level price, production functions estimated on firm-level data estimates returns to scale in revenue function, not in production. With a markup less than unity, returns to scale in production can be larger than returns to scale in the revenue function.

REFERENCES

Baily, Martin N., , Eric J. Bartelsman, and John C. Haltiwanger, "Labor Productivity: Structural Change and Cyclical Dynamics," *Review of Economics and Statistics* 83:3 (August 2001), 420–433.

Baily, Martin N., Charles Hulten, and David Campbell, "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity: Microeconomics (1992),* 187–249.

Barlevy, Gadi, "The Sullying Effect of Recessions," *Review of Economic Studies* 69:1 (January 2002), 65–96

Bartelsman, Eric J., and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Micro Datasets," *Journal of Economic Literature* 38:3 (September 2000), 569–594.

Bartelsman, Eric J. and Gray, Wayne, "The NBER Manufacturing Productivity Database," National Bureau of Economic Research, Inc., NBER Technical Working Paper: No. 205 (1996).

Basu, Susanto and John G. Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy* 105:2 (April 1997), 249–283.

Basu, Susanto and John G. Fernald, "Aggregate Productivity and Aggregate Technology," *European Economic Review* 46:6 (June 2002), 963–991.

Basu, Susanto and John G. Fernald, and Miles Kimball, "Are Technology Improvements Contractionary?" National Bureau of Economic Research, Inc., NBER Working Paper: No. 10592 (2004).

Bils, Mark, "Real Wages over the Business Cycles: Evidence from Panel Data," *Journal of Political Economy* 93:4 (August 1985), 666–689.

Bils, Mark, and Yongsung Chang, "Understanding How Price Responds to Costs and Production," *Carnegie-Rochester Conference Series on Public Policy* 52 (2000), 33–78.

Caballero, R. J. and Richard K. Lyons, "External Effects in U.S. Procyclical Productivity", *Journal of Monetary Economics* 29:2 (April 1992), 209–225.

Caballero, Ricardo J. and Mohamad L. Hammour,"The Cleansing Effect of Recessions," *American Economic Review* 84:5 (December 1994), 1350–1368.

Chang, Yongsung, “Wages, Business Cycles, and Comparative Advantage,” *Journal of Monetary Economics* 46:1 (August 2000), 143–171.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, (Cambridge: MIT Press, 1996).

Dunne, Timothy, John C. Haltiwanger, and Kenneth Troske, "Technology and Jobs: Secular Changes and Cyclical Dynamics," *Carnegie-Rochester Conference Series on Public Policy* 46 (1997), 107–178.

Foster, Lucia, John C. Haltiwanger, and C. J. Krizan, “Aggregate Productivity Growth: Lessons from Microeconomic Evidence,” in Hulten, Charles R., Dean, Edwin R., and Harper, Michael J., eds., *New Developments in Productivity Analysis*, (Chicago: University of Chicago Press, 2001), 303–363.

Foster, Lucia, John C. Haltiwanger, and Chad Syverson, “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” University of Chicago Working Paper (2006).

Goolsbee, Austan, “The Importance of Measurement Error in the Cost of Capital,” *National Tax Journal* 53:2 (2000), 215–228.

Gorodnichenko, Yuriy, "Using firm optimization to evaluate and estimate returns to scale", Manuscript, University of Michigan (2006).

Griliches, Zvi, and Jerry Hausman, “Errors in Variables in Panel Data,” *Journal of Econometrics* 31:1 (February 1986), 93–118.

Klette, Tor J. and Zvi Griliches, “The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous,” *Journal of Applied Econometrics* 11:4 (July-Aug 1996), 343–361.

Marschak, Jacob, and William H. Andrews, “Random Simultaneous Equations and the Theory of Production,” *Econometrica* 12:3&4 (July-October 1944), 143–205.

Rotemberg, Julio J., and Michael Woodford, “Dynamic General Equilibrium Models with

Imperfectly Competitive Product Markets," in Cooley, Thomas ed. *Frontiers of Business Cycle Research*, (Princeton: Princeton University Press, 1995).

Solon, Gary, Robert Barsky, and Jonathan Parker, "Measuring the Cyclicality of Real Wages: How Important is Compositional Bias?" *Quarterly Journal of Economics* 109:1 (February 1994), 587–616.

Stockman, Alan, "Aggregation Bias and the Cyclical Behavior of Real Wages," Manuscript, University of Rochester (1983).

Westbrook, Daniel, and James Tybout, "Estimating Returns to Scale with Large Imperfect Panels: An Application to Chilean Manufacturing Industries," *World Bank Economic Review* 7:1 (1993), 85–112.

Wilson, Daniel J., "Estimating Returns to Scale: Lo, Still No Balance," *Journal of Macroeconomics* 22:2 (April 2000), 285–314.

Table 1: Shares and Relative TFP of Entrants and Exiting Plants

	Shares in Total Number of Plants (Entry/Exit Rate)		Employment Shares		Output Shares		Relative TFP	
	Entering Plants	Exiting Plants	Entering Plants	Exiting Plants	Entering Plants	Exiting Plants	Entering Plants	Exiting Plants
All sample years	.074	.096	.029	.023	.024	.031	1.094	.977
Boom	.077	.108	.025	.019	.021	.034	.981	.935
Recession	.054	.075	.026	.026	.020	.026	1.072	.974

Note: Booms are years in which log change of real GDP > 4% (1972, 1973, 1976, 1977, 1978, 1983, 1985, and 1988). Recessions are years in which log change of real GDP <1% (1975, 1980, 1982, and 1991).

Table 2: Countercyclical Effect of Reallocations

– regression of GDP growth on reallocation terms

Dependent variable:	[1] *Reallocations* (Between-plant and Net Entry)	[2] $\sum_{j\in Conti}\Delta s_{jt}(\overline{\ln tfp_j}-\overline{\ln TFP})$ (Between-plant)	[3] $\sum_{j\in Entry}s_{jt}(\ln tfp_{jt}-\overline{\ln TFP}) - \sum_{j\in Exit}s_{j,t-1}(\ln tfp_{j,t-1}-\overline{\ln TFP})$ (Net Entry)
Manufacturing	-0.123	-0.097	-0.027
	(0.164)	(0.145)	(0.059)
Num. of obs.	20	20	20
Nondurables	-0.076	-0.084	0.008
	(0.255)	(0.239)	(0.056)
Num. of obs.	20	20	20
Durables	-0.497*	-0.442*	-0.055
	(0.148)	(0.145)	(0.062)
Num. of obs.	20	20	20

Note: The dependent variable is a subset of the "*Reallocations*" term in Equation (4), stated in the column heading. The independent variable is the log change in real GDP. The sample period is 1972–97, excluding the years 1974, 1979, 1984, 1989, and 1994. * Significant at 1% level.

Table 3: Composition Bias in Returns-to-Scale Estimates

Independent variable:	[1] *Reallocations* (Between-plant and Net Entry)	[2] $\sum_{j \in Conti} \Delta s_{jt} (\overline{\ln tfp_j} - \overline{\ln TFP})$ (Between-plant)	[3] $\sum_{j \in Entry} s_{jt} (\ln tfp_{jt} - \overline{\ln TFP}) - \sum_{j \in Exit} s_{j,t-1} (\ln tfp_{j,t-1} - \overline{\ln TFP})$ (Net Entry)
Manufacturing			
δ	–.043	–.037	–.006
(Std. Err)	(.096)	(.085)	(.035)
Num. of obs.	20	20	20
Nondurables			
δ	–.096	–.032	–.027
(Std. Err)	(.298)	(.174)	(.040)
Num. of obs.	20	20	20
Durables			
δ	–.118	–.107	–.011
(Std. Err)	(.075)	(.071)	(.027)
Num. of obs.	20	20	20

Note: The independent variable is a subset of the "*Reallocations*" term in Equation (4), stated in the column heading. The dependent variable is the cost-share-weighted change in inputs (dx). The sample period is 1972–97, excluding the years 1974, 1979, 1984, 1989, and 1994. The coefficients of the constant terms are not reported.

Table 4: Returns-to-Scale Estimates at Different Levels of Aggregation, Two-digit SIC

A. Nondurables

SIC code	Industry		[1] Plant level (pooled) OLS	[2] Plant level (pooled) IV	[3] Aggregate (all plants) OLS
20	Food	γ	.688	**.741**	.632
		(Std. Err)	(.021)	(.041)	(.154)
		Num. of obs.	118,839	79,337	20
21	Tobacco	γ	.682	**.659**	.413
		(Std. Err)	(.078)	(.082)	(.214)
		Num. of obs.	1,389	1,030	20
22	Textiles	γ	.876	**.868**	1.022
		(Std. Err)	(.030)	(.036)	(.132)
		Num. of obs.	37,572	25,598	20
23	Apparel	γ	.837	**.869**	.833
		(Std. Err)	(.017)	(.029)	(.110)
		Num. of obs.	62,086	33,192	20
26	Paper	γ	.904	.971	1.053
		(Std. Err)	(.028)	(.040)	(.212)
		Num. of obs.	45,584	32,434	20
27	Printing	γ	.729	**.806**	.584
		(Std. Err)	(.020)	(.054)	(.238)
		Num. of obs.	75,493	38,712	20
28	Chemicals	γ	.864	.972	.465
		(Std. Err)	(.034)	(.096)	(.303)
		Num. of obs.	68,289	45,441	20
29	Petroleum	γ	.874	**.983**	.425
		(Std. Err)	(.056)	(.062)	(.169)
		Num. of obs.	15,455	10,200	20
30	Rubber	γ	.855	.996	1.271
		(Std. Err)	(.020)	(.031)	(.098)
		Num. of obs.	55,865	34,389	20
31	Leather	γ	.897	**1.060**	.836
		(Std. Err)	(.059)	(.082)	(.136)
		Num. of obs.	10,066	6,537	20

B. Durables

SIC code	Industry		[1] Plant level (pooled) OLS	[2] Plant level (pooled) IV	[3] Aggregate (all plants) OLS
24	Lumber	γ	.745	**.803**	.885
		(Std. Err)	(.017)	(.048)	(.127)
		Num. of obs.	67,711	35,644	20
25	Furniture	γ	.919	**1.000**	1.296
		(Std. Err)	(.022)	(.049)	(.105)
		Num. of obs.	28,549	16,290	20
32	Stone, Clay, & Glass	γ	.990	**1.074**	1.109
		(Std. Err)	(.021)	(.067)	(.219)
		Num. of obs.	28,549	29,721	20
33	Primary Metals	γ	.792	.914	1.212
		(Std. Err)	(.037)	(.032)	(.099)
		Num. of obs.	40,854	27,782	20
34	Fabricated Metals	γ	.840	**.952**	1.293
		(Std. Err)	(.016)	(.032)	(.202)
		Num. of obs.	112,483	66,116	20
35	Nonelectrical Machinery	γ	.872	**.943**	1.608
		(Std. Err)	(.016)	(.034)	(.181)
		Num. of obs.	118,588	68,887	20
36	Electrical Machinery	γ	.927	**1.029**	1.621
		(Std. Err)	(.018)	(.035)	(.264)
		Num. of obs.	69,047	45,032	20
37	Transportation Equipment	γ	.889	**.997**	1.207
		(Std. Err)	(.023)	(.066)	(.077)
		Num. of obs.	39,173	25,331	20
38	Instruments	γ	.855	**.845**	.524
		(Std. Err)	(.030)	(.050)	(.122)
		Num. of obs.	31,455	19,733	20
39	Miscellaneous Durables	γ	.905	**.978**	.887
		(Std. Err)	(.043)	(.056)	(.175)
		Num. of obs.	26,491	13,944	20

Note: ASM sample weight is used. The sample period is 1972–97, excluding the years 1974, 1979, 1984, 1989, and 1994. Plant-level IV estimates appear in bold if the Hausman test rejects the consistency of the corresponding OLS estimates at the 5% level of significance.

Table 5: Returns-to-Scale estimates for the Manufacturing Sector

	[1] OLS	[2] IV	[3] IV	[4] IV	[5] IV
γ	.828	.910	.937	1.049	1.077
(Std. Err)	(.006)	(.012)	(.062)	(.091)	(.128)
Num. of obs.	1,078,471	655,350	766,787	512,176	1,078,471
Hausman test		263.73	3.00	3.75	3.81
stat. (p-value)		(.000)	(.083)	(.053)	(.051)

Note: For the IV estimates, the following instrumental variables were used for each column.

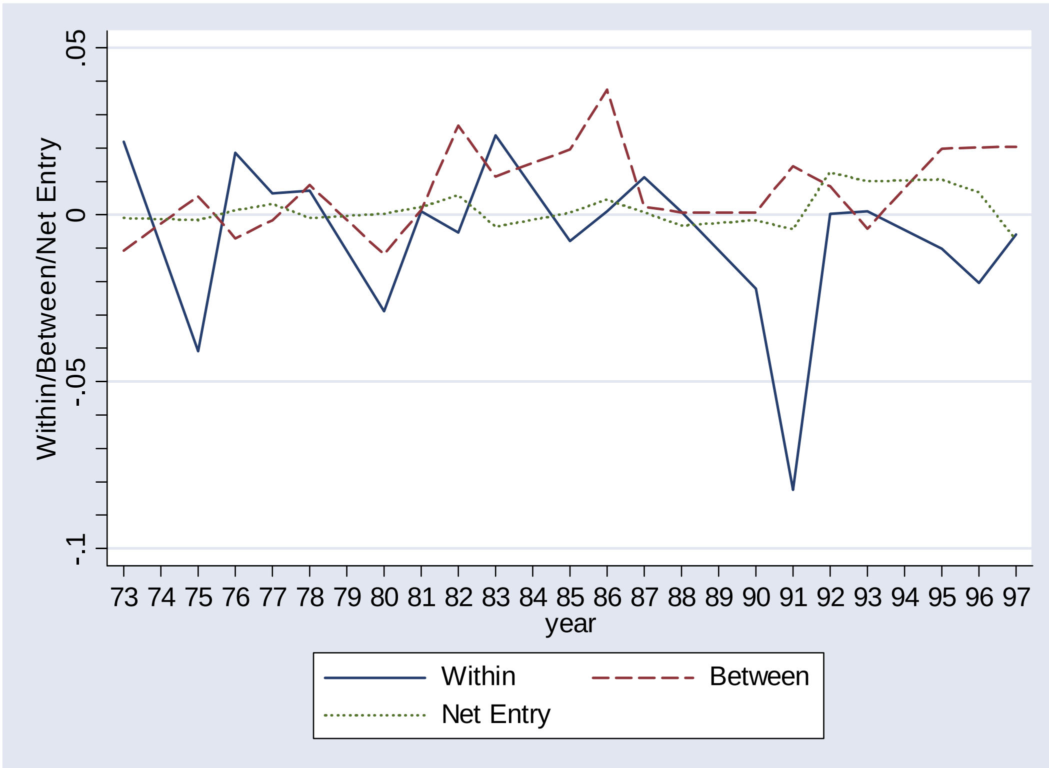
Column 2: plant-level changes in the cost-share weighted inputs (*dx*) between $t + 1$ and $t - 2$.

Column 3: 2-year and 3-year lags of cost-share weighted inputs (level).

Column 4: 4-year and 5-year lags of cost-share weighted inputs (level).

Column 5: 4-digit industry-level changes in the cost-share weighted inputs (*dx*) between t and $t - 1$.

Figure 1: Productivity Decompositions



Note: 1974, 1979, 1984, 1989, & 1994 interpolated in the graph.

Appendix

Table A6: Estimates of Returns to Scale over Different Time Horizons

– Plant-level pooled regressions with the same sample of continuing plants

	[1]	[2]	[3]	[4]
	t & $t-1$	t & $t-2$	t & $t-3$	t & $t-4$
γ	0.889	0.915	0.924	0.930
(Std. Err)	(0.009)	(0.007)	(0.006)	(0.005)
Num. of obs.	632,268	632,268	632,268	632,268

Note: The dependent variable is the log change in real output measured over the stated time period in the column heading. The independent variable is the cost-share-weighted change in inputs over the same stated time period. The regressions are run for the same sample of plants that have operated for at least four consecutive years, to exclude the effects of sample changes due to changes in the time period. ASM sample weight is used.