Sensitivity Analysis for "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *By Lucia Foster, John Haltiwanger, and Chad Syverson*

In the appendix of the published paper, we noted:

We also exclude establishments whose data appear to be imputed or suffer from reporting or recording errors. The Census Bureau imputes physical quantities when product-level data are not fully reported. Unfortunately, imputed data are not explicitly identified. To distinguish and remove imputed product-level data from the sample, we use techniques similar to those employed by Roberts and Supina (1996, 2000). To minimize the influence of reporting and recording errors, we also remove a small number of plants reporting physical quantities that imply prices greater than ten times or less than one-tenth the median price in a given year. In order to maintain the same sample over all exercises, we delete observations that are missing any one of the main regression variables. We also delete observations when the plant's labor or materials cost share is less than one-tenth of the corresponding industry's average cost share for that year, or when the cost share is more than one. Finally, we still find a relatively small number of obvious outliers in physical quantity measures, so we trim the one-percent tails of the physical productivity (TFPQ) distribution.

One method used by Roberts and Supina and others that we did not use in our final estimates was to identify possible imputed cases using modal price. However, we had conducted sensitivity and robustness analysis to excluding observations based on modal price. For the latter, the approach has been to remove the observations at the modal price within product*year cells. This has been interpreted as a possible indicator of imputed physical quantity data since price is computed as the ratio of PV (product value) to physical quantity (PQS). This document reports the results of sensitivity analysis to removing modal price cases. For this purpose, we focus on Table 1 (basic summary statistics) and Table 6 (marginal effects of indicators of productivity, prices and demand on selection).

In using this modal approach, one issue is what to do about rounding. This issue arises for the following reasons. If the imputation process is based upon a ratio method (e.g., $\hat{y}_{et} = r_{it}x_{et}$ where $r_{it} = \frac{y_{it}}{x_{it}}$ and e is establishment, i is industry, x is observed at establishment level but y is not for imputed cases), then it may be that the imputed $\hat{y}_{et}$ has been rounded in the stored data. In such cases, reverse engineering the modal imputed cases is complicated by such rounding. One way to overcome this is to first round the micro ratios before computing modes. So in the case of prices this implies rounding the computed prices (PV/PQS) before computing modes.[1] This is an imperfect process since rounding can also yield a greater likelihood of including non-imputed cases in the rounded mode.

We note that part of the imperfection of the process here is we don't know how the process was actually done. We note that in the micro data PQS values are all integer values suggesting that rounding was done to integer values after any imputation process was done. We also note an

[1] Of course in this case y/x=PQS/PV which is the inverse of the price but the modal and rounding issues are the same whether one identifies the modes with the price or inverse price.

apparent exception to this occurs in our datasets for ready mix concrete in the years 1987 and 1992 where PQS values are non-integer values. However, this feature was created by an adjustment made by our data processing since there was a unit change in the reported value of PQS in 1987. We adjusted the 1987 and 1992 data to be consistent with the prior unit values and we note we did not round the adjusted PQS for those years. In particular, we divided the PQS values in the original data by 1000 in 1987 and 1992 for ready-mix concrete. As we note below this has some relevance in considering reasonable rounding values.

An open question is what value of rounding is reasonable for identifying the modal values. Again this is an imperfect process but there are some rough rules of thumb that will depend on the scale of y at the micro level.[2] This is intuitive since if, for example, the scale of y is in the 1000's then rounding to integers yields less of a rounding problem then if y is in the say the 100's. A rough rule of thumb is that the rounding should be approximately equal to 1/Scale of y. So if y is typically around 500, the rounding factor should be around 0.002. In what follows, we show sensitivity of results to choosing a common rounding factor for all products and then choosing a rounding factor based on the median value of y in the micro data in each product by year cell. For the latter, we still use a common rounding factor within each product*year cell.[3]

In what follows, we report the following sensitivity analysis. First, we report Table 1 and Table 6 without removing any modal price cases. Followed by this are Table 1a and Table 6a which is based on the samples with modal price observations removed from product and year cells but without any rounding. Then, Table 1b and Table 6b reports the results from removing modal price observations where the price is first rounded to the nearest 0.002 before the mode is computed. Finally, Table 1c and Table 6c using rounding based upon the median of the product by year cell of PQS (so the rounding factor is 1/Median(PQS)). In applying this latter methodology, we note that in 1987 and 1992 for ready-mix concrete we use the original scaled values of PQS for this purpose. Since the original scaled values of PQS are 1000 times larger than the values in our datasets this mitigates the impact of rounding for this product in this year.[4] We note that in all of these sensitivity exercises we continue to use the other methods described in the paragraph above from the appendix to our earlier paper for identifying reporting errors and outliers. We have found that those additional steps are important to incorporate regardless of the modal price issue.

Removal of the modal observations without rounding of price has only a modest impact on the number of observations and very little impact on the results in Table 1 and Table 6. Removal of the modal observations with rounded modal price (0.002) has a non-trivial impact on the number of observations but only a modest impact on the results in Tables 1 and 6. We observe modest increases in dispersion in prices and revenue based measures of productivity. The marginal effects for the determinants of productivity are about the same in Table 6b. Finally removing

---

[22] Note that the scale of x is not relevant as long as any rounding of x occurred before the imputation of y. So for example if x is includes imputed values but the imputation of x was done prior to the imputation of y then there is no problem (and as long as the values in the stored data are the data used in the imputation of y).

[3] So, for example, if the median of y is 1000 in a product year cell then the common rounding factor for that product year cell is 0.001.

[4] Note that for the 0.002 rounding case we did not make this adjustment to return PQS to its original scale in 1987/92 in ready mix concrete. In that respect the 0.002 is identifying too many cases for this product in those years.

the modal observations with rounded modal price (1/Median(PQS) as the rounding factor) removes substantially fewer observations. The reason is that the scale of PQS is typically much larger than 500 which yields the 0.002 rounding factor. The large scale of PQS mitigates the impact of this rounding issue in practice. This suggests that the results for 0.002 are conservative estimates of the need for rounding (especially since in the 0.002 case we did not undo the adjustment in scale for PQS that we made in 1987 and 1992 for ready mix concrete). We also note that the magnitudes and statistical significance are quite similar between Tables 1 and 1c and between 6 and 6c. Finally, we note that we considered alternative common rounding factors (i.e, 0.005) and obtained similar results.

Note that similar issues apply to using modal ratios to identify variables for other imputes. Sensitivity analysis to these rounding issues for other possible imputed variables yields similar results.

Table 1. Summary Statistics for Output, Price, and Productivity Measures (Original Specification)

Correlations								
Variables	Trad'l. Output	Revenue Output	Physical Output	Price	Trad'l. TFP	Revenue TFP	Physical TFP	Capital
Traditional Output	1.00							
Revenue Output	0.99	1.00						
Physical Output	0.98	0.99	1.00					
Price	-0.03	-0.03	-0.19	1.00				
Traditional TFP	0.19	0.18	0.15	0.13	1.00			
Revenue TFP	0.17	0.21	0.18	0.16	0.86	1.00		
Physical TFP	0.17	0.20	0.28	-0.54	0.64	0.75	1.00	
Capital	0.86	0.85	0.84	-0.04	0.00	-0.00	0.03	1.00
Standard Deviations								
	1.03	1.03	1.05	0.18	0.21	0.22	0.26	1.14

Note: This table shows correlations and standard deviations for plant-level variables for our pooled sample of 17,669 plant-year observations. We remove product-year fixed effects from each variable before computing the statistics. All variables are in logs. See the text for definitions of the variables.

Table 6. Selection on Productivity or Profitability?

Specification:	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Traditional TFP	-0.073						
	0.015						
Revenue TFP		-0.063					
		0.014					
Physical TFP			-0.040			-0.062	-0.034
			0.012			*0.014*	*0.012*
Prices				-0.021		-0.069	
				0.018		*0.021*	
Demand Shock					-0.047		-0.047
					0.003		*0.003*
Controlling for Plant Capital Stock							
Traditional TFP	-0.069						
	0.015						
Revenue TFP		-0.061					
		0.013					
Physical TFP			-0.035			-0.059	-0.034
			0.012			*0.014*	*0.012*
Prices				-0.030		-0.076	
				0.018		*0.021*	
Demand Shock					-0.030		-0.029
					0.004		*0.004*
Capital Stock	-0.046	-0.046	-0.046	-0.046	-0.023	-0.046	-0.023
	0.003	*0.003*	*0.003*	*0.003*	*0.004*	*0.003*	*0.004*

Note: These results are from various probits of plant exit by the next census (shown by column) on plant-level productivity, price, demand, and capital stock measures (shown by row) as well as a full set of product-year fixed effects. The sample is our pooled sample of 17,314 plant-year observations (355 observations from the main sample are excluded because we cannot determine exiting plants in the 1997 CM, the final year of observation). Standard errors, clustered by plant, are in italics.

Table 1a. Summary Statistics for Output, Price, and Productivity Measures (Removing Modal Price Cases from Product*Year cells)

Correlations

Variables	Trad'l. Output	Revenue Output	Physical Output	Price	Trad'l. TFP	Revenue TFP	Physical TFP	Capital
Traditional Output	1.00							
Revenue Output	0.99	1.00						
Physical Output	0.98	0.99	1.00					
Price	-0.03	-0.03	-0.19	1.00				
Traditional TFP	0.19	0.18	0.15	0.13	1.00			
Revenue TFP	0.17	0.21	0.18	0.16	0.86	1.00		
Physical TFP	0.17	0.20	0.29	-0.54	0.64	0.74	1.00	
Capital	0.86	0.85	0.84	-0.04	0.00	0.00	0.03	1.00

Standard Deviations

	1.03	1.04	1.06	0.18	0.21	0.22	0.26	1.14

Note: This table shows correlations and standard deviations for plant-level variables for our pooled sample of about 17,500 plant-year observations. We remove product-year fixed effects from each variable before computing the statistics. All variables are in logs. See the text for definitions of the variables.

Table 6a. Selection on Productivity or Profitability? (Removing Modal Price Observations)

Specification:	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Traditional TFP	-0.072						
	0.015						
Revenue TFP		-0.064					
		0.014					
Physical TFP			-0.041			-0.064	-0.035
			0.012			*0.014*	*0.012*
Prices				-0.020		-0.069	
				0.018		*0.021*	
Demand Shock					-0.047		-0.047
					0.003		*0.003*
Controlling for Plant Capital Stock							
Traditional TFP	-0.068						
	0.015						
Revenue TFP		-0.062					
		0.014					
Physical TFP			-0.036			-0.060	-0.035
			0.012			*0.014*	*0.012*
Prices				-0.029		-0.076	
				0.018		*0.021*	
Demand Shock					-0.030		-0.030
					0.004		*0.004*
Capital Stock	-0.046	-0.046	-0.046	-0.046	-0.023	-0.046	-0.023
	0.003	*0.003*	*0.003*	*0.003*	*0.004*	*0.003*	*0.004*

Note: These results are from various probits of plant exit by the next census (shown by column) on plant-level productivity, price, demand, and capital stock measures (shown by row) as well as a full set of product-year fixed effects. The sample has about 17,000 observations (smaller than Table 1c because we cannot determine exiting plants in the 1997 CM, the final year of observation). Standard errors, clustered by plant, are in italics.

Table 1b Summary Statistics for Output, Price, and Productivity Measures (Removing Rounded price (0.002) Modal Cases)

Correlations								
Variables	Trad'l. Output	Revenue Output	Physical Output	Price	Trad'l. TFP	Revenue TFP	Physical TFP	Capital
Traditional Output	1.00							
Revenue Output	0.99	1.00						
Physical Output	0.97	0.98	1.00					
Price	-0.03	-0.03	-0.21	1.00				
Traditional TFP	0.19	0.18	0.15	0.15	1.00			
Revenue TFP	0.16	0.21	0.18	0.17	0.85	1.00		
Physical TFP	0.17	0.20	0.28	-0.57	0.61	0.70	1.00	
Capital	0.86	0.85	0.84	-0.04	0.00	-0.00	0.03	1.00
Standard Deviations								
	1.04	1.05	1.07	0.21	0.21	0.23	0.27	1.15

Note: This table shows correlations and standard deviations for plant-level variables for a sample of about 14,000 plant-year observations. We remove product-year fixed effects from each variable before computing the statistics. All variables are in logs. See the text for definitions of the variables.

Table 6b. Selection on Productivity or Profitability? (Removing Modal Price Cases with 0.002 Rounding)

Specification:	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Traditional TFP	-0.073						
	0.017						
Revenue TFP		-0.062					
		0.015					
Physical TFP			-0.037			-0.061	-0.030
			0.013			*0.015*	*0.012*
Prices				-0.020		-0.067	
				0.018		*0.021*	
Demand Shock					-0.050		-0.049
					0.003		*0.003*
Controlling for Plant Capital Stock							
Traditional TFP	-0.067						
	0.016						
Revenue TFP		-0.058					
		0.014					
Physical TFP			-0.030			-0.056	-0.029
			0.013			*0.015*	*0.012*
Prices				-0.029		-0.072	
				0.018		*0.021*	
Demand Shock					-0.031		-0.031
					0.004		*0.004*
Capital Stock	-0.048	-0.048	-0.048	-0.049	-0.024	-0.048	-0.024
	0.003	*0.003*	*0.003*	*0.003*	*0.004*	*0.003*	*0.004*

Note: These results are from various probits of plant exit by the next census (shown by column) on plant-level productivity, price, demand, and capital stock measures (shown by row) as well as a full set of product-year fixed effects. The sample is about 13,700 (smaller than Table 1c since some observations are excluded because we cannot determine exiting plants in the 1997 CM, the final year of observation). Standard errors, clustered by plant, are in italics

Table 1c Summary Statistics for Output, Price, and Productivity Measures (Removing Rounded price (1/Median(PQS)) Modal Cases)

Correlations								
Variables	Trad'l. Output	Revenue Output	Physical Output	Price	Trad'l. TFP	Revenue TFP	Physical TFP	Capital
Traditional Output	1.00							
Revenue Output	0.99	1.00						
Physical Output	0.98	0.99	1.00					
Price	-0.04	-0.02	-0.19	1.00				
Traditional TFP	0.18	0.17	0.15	0.14	1.00			
Revenue TFP	0.17	0.21	0.18	0.17	0.86	1.00		
Physical TFP	0.16	0.20	0.29	-0.55	0.64	0.73	1.00	
Capital	0.86	0.85	0.84	-0.04	0.00	0.00	0.03	1.00
Standard Deviations								
	1.04	1.04	1.06	0.18	0.20	0.22	0.26	1.15

Note: This table shows correlations and standard deviations for plant-level variables for a sample of about 16,800 plant-year observations. We remove product-year fixed effects from each variable before computing the statistics. All variables are in logs. See the text for definitions of the variables.

Table 6c. Selection on Productivity or Profitability? (Removing Rounded Modal Price Cases (1/Median(PQS))).

Specification:	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Traditional TFP	-0.077						
	0.017						
Revenue TFP		-0.065					
		0.014					
Physical TFP			-0.042			-0.065	-0.036
			0.012			*0.014*	*0.012*
Prices				-0.020		-0.069	
				0.018		*0.021*	
Demand Shock					-0.047		-0.047
					0.003		*0.003*
Controlling for Plant Capital Stock							
Traditional TFP	-0.072						
	0.016						
Revenue TFP		-0.062					
		0.014					
Physical TFP			-0.036			-0.060	-0.035
			0.012			*0.014*	*0.012*
Prices				-0.028		-0.075	
				0.018		*0.021*	
Demand Shock					-0.029		-0.029
					0.004		*0.004*
Capital Stock	-0.046	-0.046	-0.046	-0.046	-0.024	-0.046	-0.024
	0.003	*0.003*	*0.003*	*0.003*	*0.004*	*0.003*	*0.004*

Note: These results are from various probits of plant exit by the next census (shown by column) on plant-level productivity, price, demand, and capital stock measures (shown by row) as well as a full set of product-year fixed effects. The sample is about 16,500 (smaller than Table 1c since some observations are excluded because we cannot determine exiting plants in the 1997 CM, the final year of observation). Standard errors, clustered by plant, are in italics