Using firm-level data to study growth and dispersion in total factor productivity*

Dany Bahar

The Brookings Institution
Harvard Center for International Development

July 27, 2016

When looking at the evolution of total factor productivity (TFP) in the world in the last decade depicted in Figure 1, one notices two interesting stylized facts. First, the positive average TFP growth that the world experienced before the global financial crisis of 2008 has been reversed: since 2011, average TFP has continuously declined in the World. Second, the dispersion of TFP across countries has also increased since 2010. This deceleration and increased dispersion in aggregate figures must be a reflection of firm TFP dynamics. Using highly disaggregated firm-level data, this note identifies robust patterns regarding TFP levels, growth, dispersion and the relationship between them during the past decade.

To study firm behavior, we use data from the Orbis database from Bureau Van Dijk.[1] The database includes complete information to compute TFP for firms in about 30 countries (see Table 1) and 479 six digit NAICS codes, between years 2006 to 2014. The parameters of the production function per country and 3-digit NAICS code were estimated following the methodology put forward by Levinsohn and Petrin (2003) and Wooldridge (2009), and used to estimate revenue-based TFP for each firm and year in the sample. The results are robust to using other TFP estimation methodologies (Figure 2 compares TFP values in the sample across different methodologies).

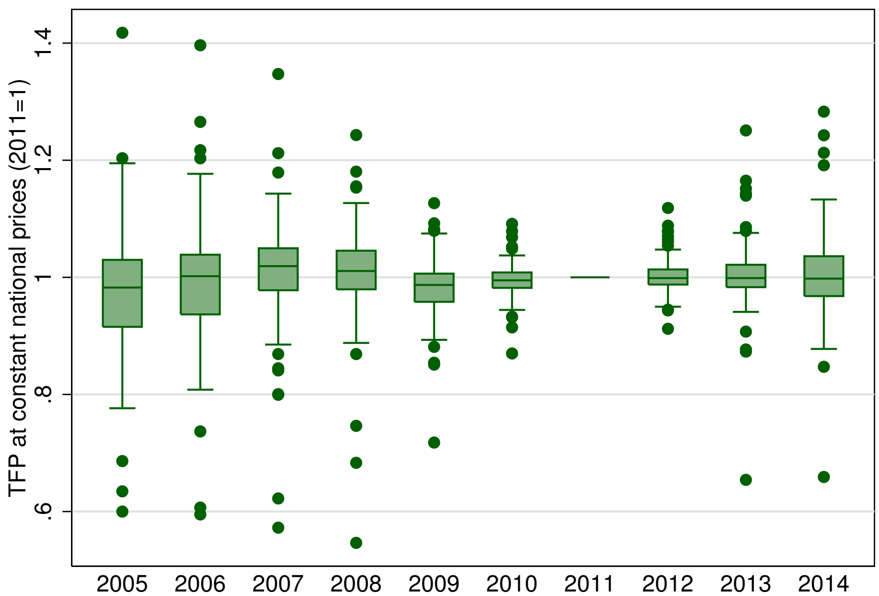
Figure 3 shows the distribution of log TFP for all firms in the sample, overall and by industry, both in 2008 and 2013. The figure reveals that median TFP has dropped across all sectors, while dispersion is persistent and has remained mostly unchanged across all cuts of the sample.

When digging into the data, there exists persistent dispersion within narrowly defined sectors within each country, consistent with the findings of Syverson (2004) and Hsieh and Klenow (2009), among others. Table 2 presents statistics on dispersion measures in the dataset. For example, on average, the TFP

*Thanks to Luis Omar Herrera for excellent research assistant. All errors are my own.

[1] This dataset has been employed also by a number of other researchers in the previous years (e.g. Bloom et al. 2012, Fons-Rosen et al. 2013).

Figure 1: Total Factor Productivity by year (world distribution)



Source: Penn World Tables 9.0

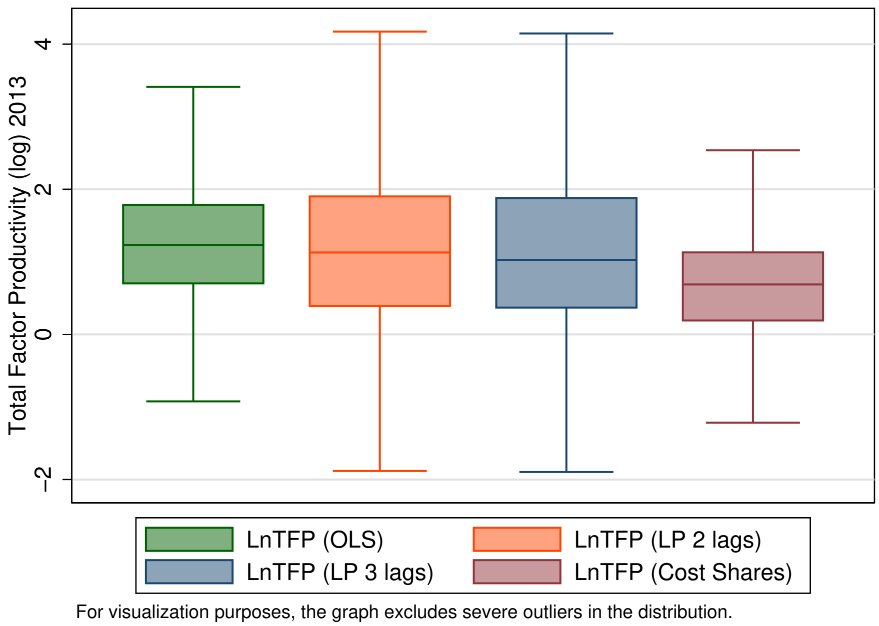
This figure visualizes the distribution of log TFP for all countries between 2005 and 2014. The horizontal line inside the box represents the median value. The edges of the box represents the values in between the 25th and 75th percentile, while the whiskers' edges represent adjacent values. Dots outside the box and its whiskers represent outlier values. The plot is constructed with values from Penn World Tables.

Table 1: Countries in Sample

Country	# Firm-Years	Cum Share
Spain	2,976,636	0.219
Italy	1,740,781	0.348
Romania	1,605,494	0.466
Portugal	1,552,508	0.580
France	1,523,388	0.693
Ukraine	597,507	0.737
Bulgaria	492,874	0.773
Czech Republic	447,763	0.806
Korea, Republic of	400,374	0.835
Croatia	347,537	0.861
Slovakia	288,979	0.882
Serbia	253,834	0.901
Finland	241,107	0.919
Norway	210,584	0.934
Germany	188,042	0.948
Estonia	141,497	0.959
Sweden	138,683	0.969
United Kingdom	130,669	0.979
Poland	87,146	0.985
Slovenia	74,873	0.990
Belgium	46,412	0.994
Bosnia and Herzegowina	38,687	0.997
Hungary	37,175	0.999
Latvia	2,811	1.000
Taiwan	1,980	1.000
Ireland	1,535	1.000
Austria	553	1.000

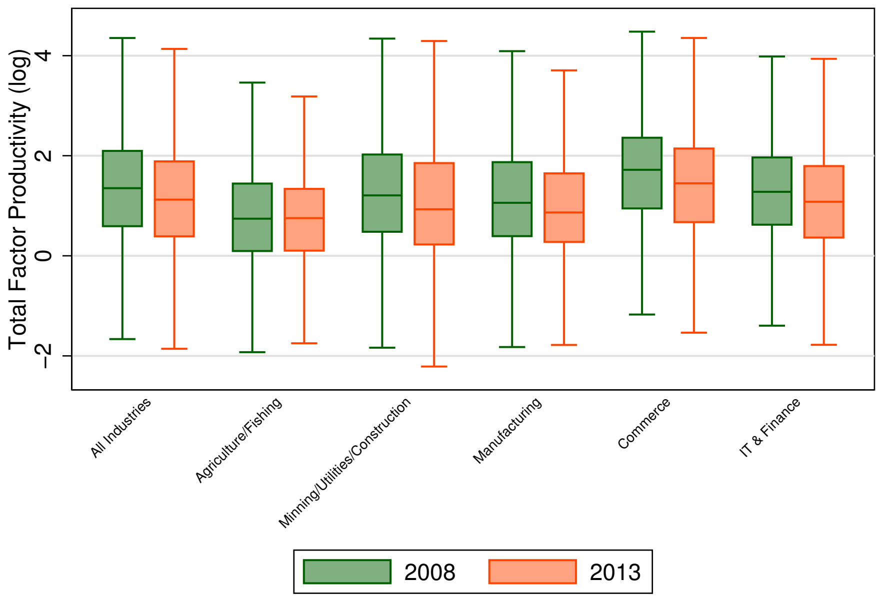
The firms in the sample are active in 485 different six-digit NAICS codes

Figure 2: TFP comparison using different methods



This figure visualizes the distribution of log TFP for all countries and industries in 2013 using different methodologies. The horizontal line inside the box represents the median value. The edges of the box represents the values in between the 25th and 75th percentile, while the whiskers' edges represent adjacent values.

Figure 3: Firm Total Factor Productivity Distribution, 2008 and 2013



For visualization purposes, the graph excludes severe outliers in the distribution.

This figure visualizes the distribution of log TFP for all firms in different sectors both for 2008 and 2013. The horizontal line inside the box represents the average value. The edges of the box represents the values in between the 25th and 75th percentile, while the whiskers' edges represent adjacent values. Dots outside the box and its whiskers represent outlier values.

Table 2: TFP Dispersion Statistics

Measure	**Mean**	**25th**	**Median**	**75th**	**Std. Dev.**
IQ ratio	2.12	1.59	1.89	2.35	1.09
99th-Median ratio	6.99	2.83	3.97	6.42	30.22
90th-10th ratio	5.58	2.66	3.80	5.95	9.75
95th-5th ratio	13.98	4.01	6.56	12.29	63.93

interquartile (IQ) ratio in the sample is 2.12, implying that the firm in the 75th percentile in the TFP distribution is more than twice as productive as the firm in the 25th percentile. Similarly, the firm in the 99th percentile of the TFP distribution is about seven times more productive, on average, than the median firm of the distribution. Similarly, the figures for the 95th to 5th percentile is 14. It is worth mentioning that these ratios are reduced significantly, almost by half, when limiting the sample to manufacturing industries only, implying that productivity dispersion among service firms is larger.

What causes TFP dispersion? Naturally, given that firms are heterogenous agents, we expect them to have different levels of productivity, based on idiosyncratic characteristics of the firms themselves. It is worth mentioning that since our TFP is revenue-based, within industry price differences (i.e. driven by quality, for example) would also explain some TFP dispersion. Yet, if knowledge diffuses easily, one would expect a convergence effect: firms with low baseline TFP levels will tend to grow faster than firms in the upper end of the TFP distribution. In the absence of this linear relationship, we would expect that differential patterns of TFP growth could explain divergence, too. To shed light on this argument, I present the results of transition matrix for TFP using data for 2008 and 2013 in Table 3. Each cell in the table represents the probability of a firm being in quintile q_{2008} of the TFP distribution in year 2008 (in each row) transitioning to quintile q_{2013} of the TFP distribution in year 2013 (in each column). For instance, the first row of the table implies that, of all the firms that in 2008 were in the bottom quintile of the TFP distribution, about 55% of them stayed in the same quintile in 2013, 25% transitioned to the second quintile, 13% to the third quintile, 9% to the fourth quintile and 8% to the fifth quintile. Naturally, the diagonal elements of the matrix are high, implying that firms in the sample tend to remain in the same quintile of the TFP distribution, particularly for those firms in the bottom and in the top of the distribution (with a probability of 0.55 and 0.45, respectively). In short, the table suggests that mobility along the TFP distribution over time is limited.

To control for possible confounding factors related to across firms prices, country-level or industry-level variables, I run a number of TFP growth regressions which control for country-year and product-year fixed effects (products are defined as six digit NAICS codes). The dependent variable in this setting would be the growth rate in TFP for the same firm between two points in time in the sample, and thus, if pricing and quality remains stable within the firm, then growth rates would reflect only changes in productivity. In order to understand

Table 3: TFP Transition Matrix

2008 \ 2013	1	2	3	4	5
1	0.55	0.25	0.13	0.09	0.08
2	0.23	0.36	0.25	0.15	0.09
3	0.10	0.22	0.31	0.25	0.14
4	0.06	0.11	0.21	0.32	0.24
5	0.05	0.06	0.09	0.19	0.45

Table 4: TFP Growth Regression

Dependent Variable: TFP Growth Rate (CAGR)				
	3 years		5 years	
	(1)	(2)	(3)	(4)
lnTFP	-0.1498	-0.3489	-0.0967	-0.1991
	(0.023)***	(0.047)***	(0.019)***	(0.038)***
lnTFP_sq		0.0527		0.0268
		(0.008)***		(0.006)***
N	1795333	1795333	774351	774351
r2	0.07	0.09	0.16	0.18

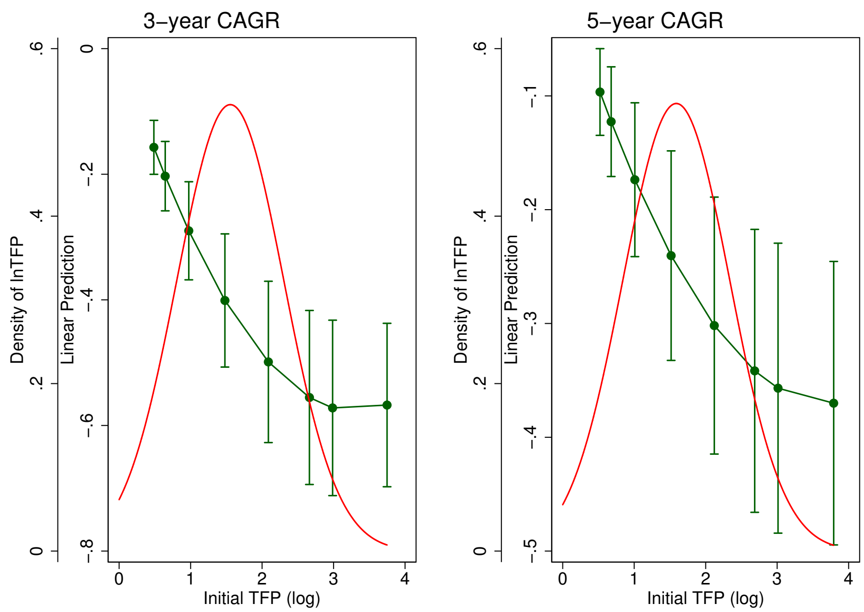
All columns include country-year and product-year fixed effects. Standard errors clustered at the country and product level are presented in parenthesis.
$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

whether there are non-linearities in the relationship I allow for a quadratic term in TFP baseline values. The results are presented in Table 4. Columns (1) and (3) present the results of a linear growth regression, which is consistent with the initial intuition: firms with higher initial levels of TFP would tend to grow slower (hence, the negative estimator). Columns (2) and (4) allow for a quadratic term, and interestingly enough, the estimator for such term is positive and statistically significant. This implies that there exists a non-linear, U-shaped, relationship that describes TFP convergence (see Figure 4). That is, firms in the upper end of the TFP distribution would grow faster than those in the middle. This result, by definition, implies quite the opposite: divergence. That is, firms in the lower end of the distribution are able to "catch-up" to firms with average TFP, but these average firms won't catch up with those in the upper end of the distribution, thus generating dispersion in TFP levels.

I study these average trends by dividing the sample into developed (OECD) and developing (non-OECD) countries in the sample. The results are presented in Table 5. Overall, the point estimates suggest that for developed nations the speed of convergence is somewhat slower while the non-linearity is more pronounced, implying that dispersion is more likely to occur.

When splitting the sample across all sectors (one digit NAICS codes), the

Figure 4: Expected TFP Growth based on initial TFP levels



This figure visualizes the expected TFP annual growth (CAGR) as a function of initial TFP level, using firm data for firms between the 1st and 99th percentile of the TFP distribution. The red line represents the distribution of firms based on their initial TFP level. The estimation controls for country-year and product-year fixed effects, where each product is a six digit NAICS code.

Table 5: TFP Growth Regression, OECD vs non-OECD

Dependent Variable: TFP Growth Rate (CAGR)				
	3 years		5 years	
	OECD	Not OECD	OECD	Not OECD
lnTFP	-0.2871	-0.4855	-0.1507	-0.2818
	(0.025)***	(0.040)***	(0.018)***	(0.036)***
lnTFP_sq	0.0415	0.0687	0.0193	0.0328
	(0.004)***	(0.010)***	(0.003)***	(0.006)***
N	1367798	427531	587642	186709
r2	0.06	0.15	0.14	0.29

All columns include country-year and product-year fixed effects. Standard errors clustered at the country and product level are presented in parenthesis.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 6: TFP Growth Regression, by sector

Dependent Variable: TFP 5-year Growth Rate (CAGR)					
	Agriculture	Mining,Utilities,Constr	Mnftr	Commerce	IT & Finance
lnTFP	-0.2749	-0.2953	-0.1704	-0.1527	-0.2467
	(0.028)***	(0.045)***	(0.029)***	(0.030)***	(0.038)***
lnTFP_sq	0.0391	0.0410	0.0186	0.0195	0.0342
	(0.007)***	(0.008)***	(0.005)***	(0.004)***	(0.007)***
N	26172	111141	160659	252328	224051
r2	0.27	0.21	0.21	0.15	0.22

All columns include country-year and product-year fixed effects. Standard errors clustered at the country and product level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

non-linearity is present, as can be seen in Table 6.

Does dispersion matter?

Productivity dispersion within narrowly defined sectors is a sign that technology or knowledge is not being diffused across different firms coexisting in the same country. This could be a result of the difficulties of transferring knowledge, which would be intensified in economies with poor labor mobility or lack of competition. Yet, does TFP dispersion matter?

The particular causes of TFP dispersion are out of the scope of this note; yet, I explore whether dispersion has any robust relationship to aggregate TFP growth. In order to study this, I compute a number of TFP inequality measures based on firms within each six digit NAICS codes for each country-year combination available in the dataset.[2]

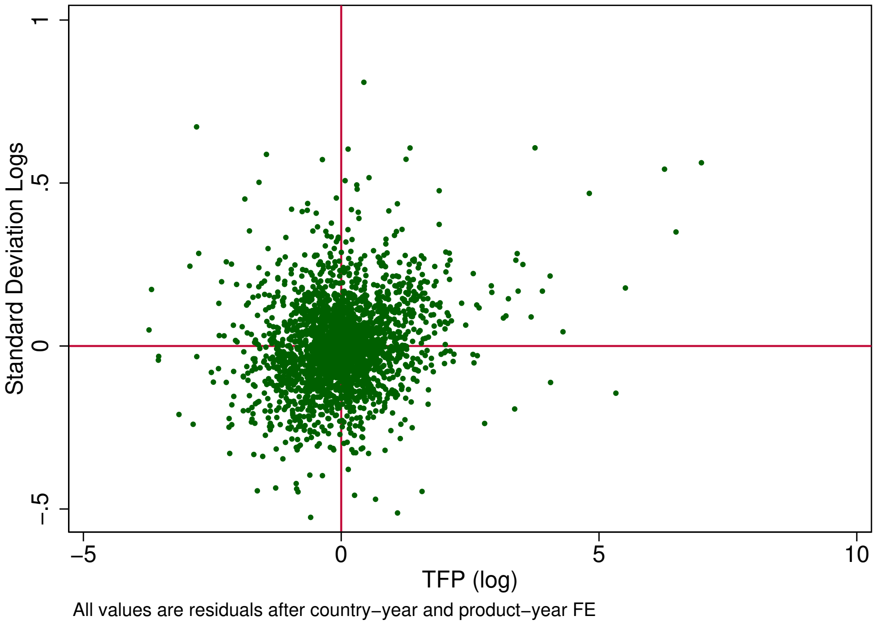
These measures of dispersion have no clear relationship between the correspondent average TFP level of a six-digit industry and country combination. This is depicted in Figure 5 that plots one of the measures of TFP dispersion (on the vertical axis) against the average TFP for every country-product combination in the dataset in the year 2010 (the figures are residuals after controlling for country-year and product-year fixed effects). There is no clear correlation in the data: there are country-industry combinations both above and below mean TFP levels with both high and low levels of dispersion.

Does dispersion matter for future growth? With this data in hand, it is possible to study the relationship between the average growth rate of a country-industry combination and its baseline level of TFP dispersion, using regression analysis. The data reveals that, almost across all measures, TFP dispersion is negatively correlated to its future growth, after controlling for baseline TFP levels, as can be seen in Table 7.[3]

[2] The measures are the relative mean deviation, the coefficient of variation, the standard deviation of logs, the Gini coefficient, the Theil index and the mean log deviation.

[3] The results for 5-year CAGR are still negative and statistically significant for most -though not all- dispersion measures.

Figure 5: TFP Dispersion



This figure plots for every country and six digit NAICS code the TFP dispersion (measured by the standard deviation of log TFP) against industry average TFP (computed as a weighted sum of firms' TFP, using operating revenue as the weight). The values presented are residuals after controlling for country-year and product-year fixed effects, where each product is a six digit NAICS code. The red lines represent the average dispersion and aggregate TFP in the sample.

Table 7: TFP Growth and Dispersion

Dependent Variable: TFP 5-year Growth Rate (CAGR)						
	Rel Mean Dev	Coef Var	Std Dev Log	Gini	Theil	Mean Log Dev
Dispersion	-0.2301	-0.0226	-0.0297	-0.1947	-0.0766	-0.0888
	(0.059)***	(0.003)***	(0.041)	(0.053)***	(0.014)***	(0.032)**
lnTFP	-0.0239	-0.0199	-0.0275	-0.0246	-0.0203	-0.0225
	(0.009)**	(0.008)**	(0.009)***	(0.009)**	(0.008)**	(0.008)***
N	1768	1768	1768	1768	1768	1768
r2	0.39	0.43	0.37	0.39	0.42	0.39

All columns include country-year and product-year fixed effects. Standard errors clustered at the country and product level are presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Concluding Remarks

The existence of important productivity dispersion within narrowly defined industries has been established in the literature in several settings. The reason behind large TFP dispersion could be a consequence of supply side variables such as distortions of factor prices (Hsieh and Klenow, 2009), differential technology adaption or managerial techniques, or demand side variables such as low within-industry product substitutability (Syverson, 2004).[4]

Establishing the causal link between TFP dispersion and growth at the aggregate level represents important empirical challenges. It could well be that as average TFP increases, dispersion tends to decrease (a result that would be consistent with decades long research on economic growth and income inequality). On the other hand, dispersion in TFP levels could play a role in hindering future TFP growth. For instance, if frontier firms grow faster than laggard firms (as evidenced in Tables 4 and 6, as well as in Figure 4), then the latter would be discouraged of investing in innovation and technology adoption. Given that these laggards represent a large mass of firms, then this behavior could hurt overall TFP growth in that industry. This result would be consistent with the U-shaped between innovation and competition documented by Aghion et al. (2005): in the presence of a TFP distribution that is highly skewed towards frontier firms, other firms in the lower and middle part of the distribution would have less incentives to innovate and invest in their own productivity.

All in all, the results in this note present an important input to understand the era of secular stagnation: few highly productive firms remain at the frontier of productivity, while there is no full TFP convergence dynamics. This generates large productivity dispersion that could be, in turn, slowing down overall aggregate growth.

References

Aghion, Phillipe, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. "Competition and innovation: an inverted-u relationship." *Quarterly Journal of Economics* 120, May: (2005) 701–728.

Bloom, Nicholas, Raffaella Sadun, and J. Van Reenen. "The Organization of Firms Across Countries." *The Quarterly Journal of Economics* 127, 4: (2012) 1663–1705.

Fons-Rosen, Christian, Sebnem Kalemli-Rosen, Bent E. Sorenson, Carolina Villegas-Sanchez, and Vadym Volosovych. "Quantifying Productivity Gains from Foreign Investment." *NBER Working Papers Series* 18920: (2013) p. 69, Cambridge MA.

Hsieh, CT, and P.J. Klenow. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124, 4: (2009) 1403–1448.

[4]See Syverson (2011) for a review of this literature.

Levinsohn, James, and Amil Petrin. "Estimating Production Inputs to Functions Using Control for Unobservables." *The Review of Economic Studies* 70, 2: (2003) 317–341.

Syverson, Chad. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86, 2: (2004) 534–550.

———. "What Determines Productivity?" *Journal of Economic Literature* 49, 2: (2011) 326–365.

Wooldridge, Jeffrey M. "On estimating firm-level production functions using proxy variables to control for unobservables." *Economics Letters* 104, 3: (2009) 112–114.