E S C O E

ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

ESCoE Research Seminar

Decomposing Differences in Productivity Distributions

Presented by Patrick Schneider, Bank of England

30 January 2018

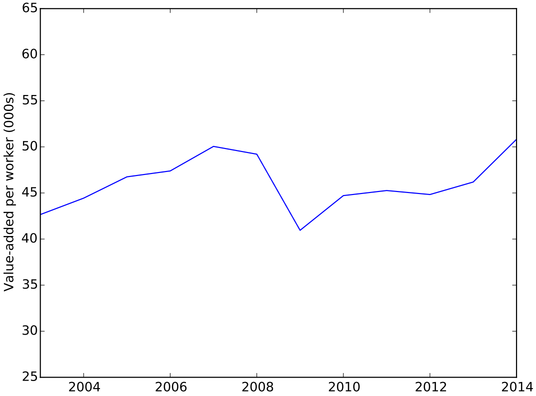
Decomposing differences in productivity distributions

Patrick Schneider
Bank of England

ESCoE Research Seminar, 30 January 2018

The views expressed in this presentation are those of the author, and not necessarily those of the Bank of England or its committees.

How can we use microdata to explain changes in productivity over time?



e.g. Barnett et al. (2014); Andrews et al. (2015); Riley and Bondibene (2016); Borio et al. (2016); Decker et al. (2017)

Bottom-up methods (Balk, 2016) are accounting decompositions

$$\Pi_t = \sum_i s_{it}\pi_{it} \Rightarrow \Delta\Pi_t = \sum_i s_{it}\pi_{it} - \sum_i s_{it-1}\pi_{it-1}$$

1. **Panel** methods track continuing firms (e.g. Griliches and Regev, 1995; Foster et al., 2001; Baily et al., 2001; Diewert and Fox, 2005)

$$\Delta\Pi_t = \underbrace{\sum_{i\in C} s_{it}\Delta\pi_{it}}_{within} + \underbrace{\sum_{i\in C} \Delta s_{it}\pi_{it-1}}_{between} + \underbrace{\sum_{i\notin C} \left(s_{it}\pi_{it} - s_{it-1}\pi_{it-1}\right)}_{net\ entry} \tag{1}$$

2. **Cross-section** methods track distribution moments (e.g. Olley and Pakes, 1996; Melitz and Polanec, 2015, add net entry)

$$\Delta\Pi_t = \underbrace{\Delta\bar{\pi}_t}_{mean} + \underbrace{\Delta cov\left[s_{it}, \pi_{it}\right]}_{efficiency} \tag{2}$$

Shortcomings

- Limited interpretation
- Reliance on tracking firms over time

This paper fits the researcher's question in a general decomposition framework (Fortin et al., 2011) that

1. tracks characteristics, not identities, so no need for panel data and different inferences
2. allows alternative methods for mean decompositions (e.g. nonlinear)
3. allows for decompositions of other distribution statistics (e.g. variance or quantiles)

Productivity is a weighted average

$$\Pi = \sum_i s_i \pi_i$$

...an unbiased estimator of the mean of productivity across workers, $Y \sim F_Y$

$$E[\Pi] = E[Y] = \int y \,\mathrm{d}F_Y(y)$$

...expanding F_Y to introduce the conditional effects of characteristics X (e.g. exporter)

$$F_Y = \int F_{Y|X}(y|x)\,\mathrm{d}F_X(x)$$

...so productivity is the interaction of **structure** ($F_{Y|X}$) and **allocation** (F_X).

$$E[Y] = \int y \,\mathrm{d}\underbrace{\left[\int F_{Y|X}(y|x)\,\mathrm{d}F_X(x)\right]}_{\textit{focus}}$$

Wish to explain difference in distribution

$$\Delta F_Y = F_Y - F'_Y$$

...construct counterfactual

$$F_Y^C = \int F_{Y|X}(y|x)\,\mathrm{d}F'_X(x)$$

...add, subtract and rearrange

$$\Delta F_Y = \underbrace{\int F_{Y|X}(y|x)\,\mathrm{d}\Delta F_X(x)}_{\textit{Allocation}} + \underbrace{\int \Delta F_{Y|X}(y|x)\,\mathrm{d}F'_X(x)}_{\textit{Structure}}$$

...and the same applies to functionals $v(F_Y)$ (mean, variance, quantiles, Gini...)

$$\Delta v_O = \underbrace{\Delta v_X}_{\textit{Allocation}} + \underbrace{\Delta v_S}_{\textit{Structure}}$$

Fortin et al. (2011)

$$\Delta v_O = \underbrace{\Delta v_X}_{Allocation} + \underbrace{\Delta v_S}_{Structure}$$

Identifying assumptions

1. 'Simple counterfactual': no general equilibrium effects
2. 'Overlapping support': characteristics X describe both groups
3. 'Ignorability': X captures all relevant characteristics differentiating F_Y from F'_Y

$$\Delta v_O = \underbrace{\Delta v_X}_{Allocation} + \underbrace{\Delta v_S}_{Structure}$$

Many ways to implement that differ by

1. Choice of $v(\cdot)$: mean, variance, quantile, Gini coefficient etc
2. Counterfactual construction: often many options, depends on assumptions

1. What drove the change in aggregate productivity between 2003 and 2014?
2. What is behind the difference in productivity between London and the rest of the UK?

$$v(F_Y) = E[Y]$$

Linear model

- Measurement ▸ proof

$$E[Y] = E\left[E[Y|X]\right] = E[X]\beta$$

- Decomposition (Oaxaca, 1973; Blinder, 1973)

$$\Delta E[Y] = \underbrace{\Delta E[X]\beta}_{Allocation} + \underbrace{E[X]'\Delta\beta}_{Structure}$$

note equivalence to within/between for continuing firms where X is a set of firm dummies.

Quantile approximation

- Measurement ▸ proof

$$E[Y] \approx \frac{1}{Q}\sum_{i=1}^{Q} q_i(F_Y)$$

- Decomposition (Chernozhukov et al., 2013)

$$\Delta E[Y] \approx \underbrace{\frac{1}{Q}\sum_{i=1}^{Q} \left[q_i(F_Y) - q_i(F'_Y)\right]}_{Useful!}$$

$$= \underbrace{\frac{1}{Q}\sum_{i=1}^{Q} \left[q_i(F_Y) - q_i(F^C_Y)\right]}_{Allocation} + \underbrace{\frac{1}{Q}\sum_{i=1}^{Q} \left[q_i(F^C_Y) - q_i(F'_Y)\right]}_{Structure}$$

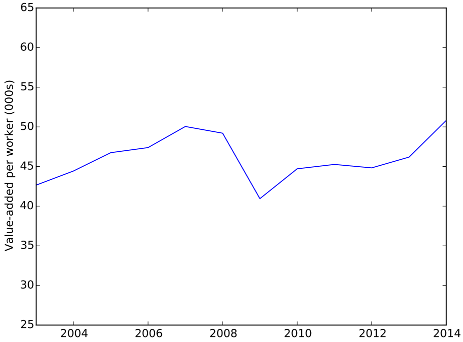
Source

- Annual Respondents Database X ONS, 2003–2014; 35,000–47,000 obs of reporting units per year
- Dropped Finance and Insurance Activities (SIC07 64–66), Agriculture, Forestry and Fishing (SIC07 01–03) and Public Administration and Defence (SIC07 84)
- Mining and Quarrying (SIC07 05–09) and Accommodation and Food Services Activities (SIC07 55–56) only included for London v. rest of country

Variables

- Productivity is real value-added per worker (SIC07 2-digit deflators)
- Characteristics X is SIC07 'division', region and foreign-owned dummy for 2003–2014 comparison
- Characteristics X is SIC07 'division', and exporter, import and foreign-owned dummies for London v. rest of country comparison

Figure: Summary statistics

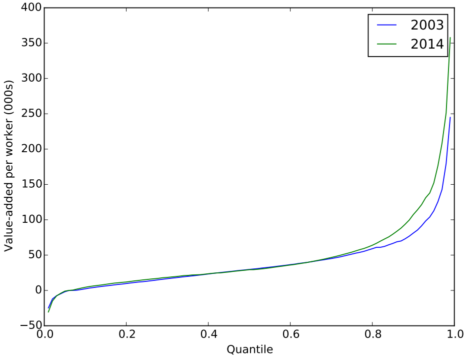


(a) Aggregate productivity over time

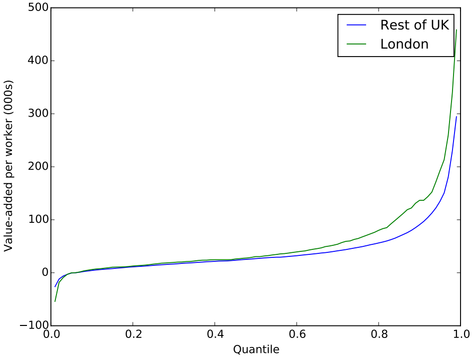
Region	Productivity (£000s)
London	64.6
Rest of UK	44.1
South East	51.3
West Midlands	45.5
East of England	45.0
Scotland	44.7
South West	43.5
North West	41.5
East Midlands	38.2
Wales	37.9
Yorkshire & Humberside	37.8

(b) 2014 labour-productivity across regions

Figure: Comparing distributions



(a) Productivity distributions over time

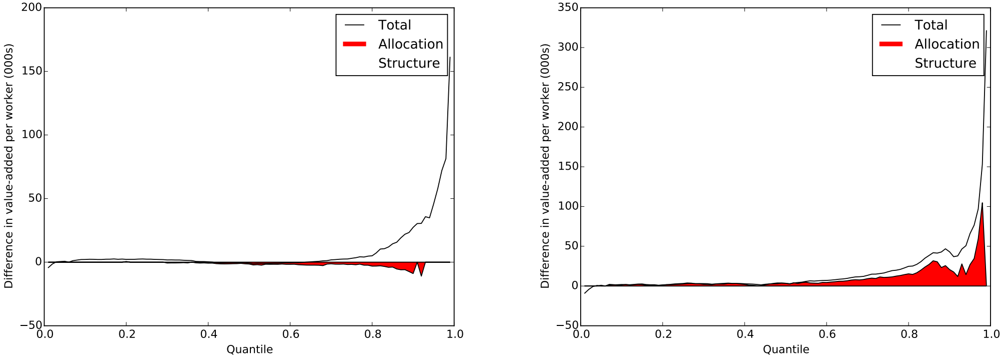


(b) Productivity distributions over space (2014)

Table: Summary of results (2011 £000s CVM)

	2003–14			**London gap**		
	Δ	Allocations	Structure	Δ	Allocations	Structure
Mean	8.2	0.1	8.1	20.5	9.6	10.9
Quantile approx.	8.0	-1.4	9.4	17.1	8.6	8.5
q1–q50	*0.6*	*-0.2*	*0.7*	*0.9*	*1.0*	*-0.0*
q51–q75	*0.1*	*-0.5*	*0.6*	*2.4*	*1.6*	*0.8*
q76–q99	*7.4*	*-0.7*	*8.1*	*13.8*	*6.0*	*7.7*

Figure: Contributions to differences in distributions, by quantile



(a) 2003 to 2014 difference

(b) Rest of UK to London difference

Take these results with a grain of salt.

- Model based, proper inference requires bootstrapping
- Very limited characteristic set (ignorability satisfied?)
- Index number issues from counterfactual treatment
- Use of sector level deflators does not identify firm level quantities
- Limited sector coverage

- Most productivity change decompositions require panel data
- But productivity analysis fits well within a general framework for decomposing distribution statistics
  - The Fortin et al. (2011) framework is generally useful if you're dealing with micro-data
- Recognising this:
  - Frees you from a need for panel data
  - Gives you more angles of attack even if you have it
  - Allows you to look at different relevant facts like quantiles or variance
- Productivity distributions are very skewed, so differences between most productive firms tend to explain differences in the mean

Define a matrix of indicators $\mathbf{I}$ that denote whether a firm belongs to one or other of J disjoint subsets (e.g. sectors) and suppose that membership of each subset causes a mean-shift in the conditional distribution the firm draws its productivity level from:

$$\pi_i = \mathbf{I}_i \beta + e_i; \; e_i \sim IID[0, \sigma^2]$$
$$\pi = \mathbf{I}\beta + e$$

The labour-share-weighted estimator for β is estimated by weighted-least-squares using the weight-matrix $\mathbf{S}$ with the labour share vector on the diagonal and zero off-diagonals (i.e. $s_i = L_i / \sum_i L_i$ is the i-th element of diag($\mathbf{S}$)). This yields

$$\hat{\beta} = \left(\mathbf{I}'\mathbf{S}\mathbf{I}\right)^{-1} \mathbf{I}'\mathbf{S}\pi$$

In the special case where $J = 1$, this coefficient estimate is equivalent to the standard aggregate productivity formula

$$\hat{\beta} = \left(\sum_i s_i\right)^{-1} \sum_i s_i \pi_i = \sum_i s_i \pi_i = \Pi$$

And where $J > 1$, each element j of $\hat{\beta}$ is the sub-sample productivity formula

$$\hat{\beta}_j = \left(\sum_{i \in j} s_i\right)^{-1} \sum_{i \in j} s_i \pi_i = \sum_{i \in j} \frac{s_i}{\sum_{i \in j} s_i} \pi_i = \sum_{i \in j} s_i^j \pi_i = \pi_j$$

Finally, if $J = N$ then $\mathbf{I}$ is the identity matrix I_N, indexing observations, and each element j of β is the j-th firm's calculated productivity

$$\hat{\beta}_j = \pi_i$$

The final case is inestimable (zero degrees of freedom) but the coefficients are observed in their own right.

back

The population mean is the integral over quantiles (i) of the unconditional distribution.

$$\begin{aligned} E[Y] &= \int_y y \; \mathrm{d}F_Y(y) \\ &= \int_y y \; \mathrm{d}\left[\int_0^1 F_{Y|i}(y|i) \; \mathrm{d}F(i)\right] \\ &= \int_0^1 \int_y y \; \mathrm{d}F_{Y|i}(y|i) \; \mathrm{d}F(i) \\ &= \int_0^1 q_i(F_Y) \; \mathrm{d}F(i) \end{aligned} \tag{3}$$

Where $q_i(F_Y) = \int_y y \; \mathrm{d}F_{Y|i}(y|i)$ is the i-th quantile for the distribution of Y and $F(i)$ is a uniform distribution over the support $[0, 1]$. This can be approximated by summing over a number Q of equally spaced quantiles

$$E[Y] \approx \frac{1}{Q} \sum_{i=1}^{Q} q_i(F_Y) \tag{4}$$

The approximation is not exact and will be biased if there is skew in the distribution of Y (in the opposite direction of the skew), but it becomes better and less biased as Q grows, such that $\lim_{Q\rightarrow\infty} \frac{1}{Q} \sum_{i=1}^{Q} q_i(F_Y) = E[Y]$, as in equation (3).

back

- Linear: oaxaca lprod {$characteristics} [aw=labour], by(group)
- Quantiles: cdeco lprod {$characteristics} [aw=labour], by(group) quantiles(0.01(0.01)0.99) method(logit)

Andrews, D., C. Criscuolo, P. Gal, et al. (2015). Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries. Technical report, OECD Publishing.

Baily, M. N., E. J. Bartelsman, and J. Haltiwanger (2001). Labor productivity: structural change and cyclical dynamics. *The Review of Economics and Statistics 83*(3), 420–433.

Balk, B. M. (2016). The dynamics of productivity change: A review of the bottom-up approach. In W. H. Greene, L. Khalaf, R. Sickles, M. Veall, and M.-C. Voia (Eds.), *Productivity and Efficiency Analysis*, Cham, pp. 15–49. Springer International Publishing.

Barnett, A., A. Chiu, J. Franklin, and M. Sebastiá-Barriel (2014). The productivity puzzle: a firm-level investigation into employment behaviour and resource allocation over the crisis. *Bank of England Working Paper* (495).

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.

Borio, C. E., E. Kharroubi, C. Upper, and F. Zampolli (2016). Labour reallocation and productivity dynamics: financial causes, real consequences. *BIS Workping papers* (534).

Chernozhukov, V., I. Fernndez-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2017, May). Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown. *American Economic Review 107*(5), 322–326.

Diewert, W. E. and K. A. Fox (2005). On measuring the contribution of entering and exiting firms to aggregate productivity growth. *Price and productivity measurement 6*.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of Labor Economics 4*, 1 – 102.

Foster, L., J. C. Haltiwanger, and C. J. Krizan (2001). Aggregate Productivity Growth: Lessons from Microeconomic Evidence. In *New Developments in Productivity Analysis*, NBER Chapters, pp. 303–372. National Bureau of Economic Research, Inc.

Griliches, Z. and H. Regev (1995). Firm productivity in israeli industry 1979-1988. *Journal of Econometrics 65*(1), 175–203.

Melitz, M. J. and S. Polanec (2015, 06). Dynamic Olley-Pakes productivity decomposition with entry and exit. *RAND Journal of Economics 46*(2), 362–375.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Olley, G. S. and A. Pakes (1996, November). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Riley, R. and C. R. Bondibene (2016). Sources of labour productivity growth at sector level in briatin, after 2007: a firm level analysis. *NESTA Working Paper* (16/01).