Cost and Product Advantages: Evidence from Chinese Manufacturing Firms

-Online Appendix-

Jordi Jaumandreu[*] Heng Yin[†]

This version: May 2018

[*]Boston University, Dep. of Econ., 270 Bay State Road, Boston, MA02215, USA. E-mail: jordij@bu.edu.

[†]Renmin University of China, National Academy of Development and Strategy, Haidian District, Beijing 100872, PRChina. E-mail: yheng@ruc.edu.cn.

A3. Demand specification.

Our exercise only needs well defined individual demands, not the adoption of an specific parametric model. The demands in equation (9) are expressions that, in logs, coincide with a first order approximation to any demand. Put aside the observed shifters for simplicity. The expression

$$q_j = \ln \alpha_0 - \eta(p_j - p) + \delta_j,$$

is the first order approximation in logs to demand $Q_j = Q(\frac{P_j}{P}, \delta_j)$ around the point $\alpha_0 = Q(1,0)$, where we drop the time subindex and denote the industry values by the absence of subindex. We use, without loss of generality, the restriction $\left.\frac{\partial \ln Q_j}{\partial \delta_j}\right|_{\alpha_0} = 1$. And we consider a common industry derivative $\left.\frac{\partial \ln Q_j}{\partial \ln(P_j/P)}\right|_{\alpha_0} = -\eta$. Notice that the intercept may be taken as the demand available to any firm selling at average price and with average demand advantages. The common elasticity assumption may be relaxed, as it could also be relaxed (at a higher cost) the independence of the elasticity from δ_j.

Our specification nests, as a particular case, the often used demands of the heterogeneous goods generalization of Dixit and Stiglitz (1977) CES system for a product differentiated industry. Let's see how. In Dixit and Stiglitz (1977) a representative consumer chooses the consumption level of the numeraire and the quantities of each variety of a differentiated good, aggregated in the index $Q = \left(\sum Q_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$. The demand for each variety turns out to be

$$Q_j = \frac{R}{P}\left(\frac{P_j}{P}\right)^{-\sigma} = Q\left(\frac{P_j}{P}\right)^{-\sigma},$$

where $R = \sum_j P_j Q_j = PQ$ is the income spent on the differentiated good and $P = \left(\sum P_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is the price of the aggregate quantity. The "elasticity of substitution" σ plays the role of the elasticity of demand η in our notation. Slightly departing from the original formulation, let's set the model in terms of average "equivalent" indices. Defining $\overline{Q} = \left(\frac{1}{N}\sum Q_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ and $\overline{P} = \left(\frac{1}{N}\sum P_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ we can write the demands as

$$Q_j = \overline{Q}\left(\frac{P_j}{\overline{P}}\right)^{-\sigma}.$$

where $\overline{Q}$ and $\overline{P}$ can be conveniently read as average values.

We can now allow the quantities Q_j to be the observed quantities of goods of varying degrees of attractiveness to the consumer by adjusting them by factors $\exp(\frac{\delta_j}{\sigma-1})$. The model treats the values $(\exp(\frac{\delta_j}{\sigma-1})Q_j)$ symmetrically in the quantity index and utility function (see Melitz 2000 for an specification of this kind). Now, at the same prices, the consumer buys more of the variety with a higher δ_j and the demands in terms of the observed quantities become

$$Q_j = \widetilde{Q}\left(\frac{P_j}{\widetilde{P}}\right)^{-\sigma}\exp(\delta_j),$$

where $\widetilde{Q} = \left(\frac{1}{N}\sum \exp(\frac{\delta_j}{\sigma})Q_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ and $\widetilde{P} = \left(\frac{1}{N}\sum \exp(\delta_j)P_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ stand for the average quantity and price consumer preference-adjusted indices. $\widetilde{Q}$ may also be written as $R/N\widetilde{P}$ (average deflated revenue spent on the goods). This coincides with the type of demand that we have specified. If the average quantity changes over time because of the resources allocated by the consumer to this industry or the number of firms producing actively, these changes can be incorporated as an aggregate change component of δ_j.

Although our models nests the CES framework we do not want to impose to the data the restrictions implied by this theoretical specification. For example, Hottman, Redding and Weinstein (2016) specify a model with symmetric preferences on the consumer valuations $(\varphi_j Q_j)$, with φ_j called consumer "appeal". This implies to specify the particular value $\delta_j = (\sigma-1)\ln\varphi_j$. An implication common to both our general model and the CES framework is the relationship

$$\frac{P_j Q_j}{P_s Q_s} = \left(\frac{P_j}{P_s}\right)^{-(\sigma-1)}\frac{\exp(\delta_j)}{\exp(\delta_s)},$$

that says that the relative sales of two varieties depend only on the relationship between prices and the relative unobserved advantages (notice here, in passing, how restrictive can be to forget about the observed shifters). With the Hottman, Redding and Weinstein (2016) specification this relation becomes

$$\frac{P_j Q_j}{P_s Q_s} = \left(\frac{P_j/\varphi_j}{P_s/\varphi_s}\right)^{-(\sigma-1)} \quad \text{or} \quad \frac{\varphi_j Q_j}{\varphi_s Q_s} = \left(\frac{P_j/\varphi_j}{P_s/\varphi_s}\right)^{-\sigma} \quad \text{for any two } j \text{ and } s.$$

These last two expressions tell us that, with "appeal"-adjusted prices, we will observe in equilibrium the same sales and the same total consumer valuations across varieties. As it is quite reasonable that prices reflect the different costs of "appeal" (for example, marginal cost may be proportional to the quality that determines the "appeal"), adjusted prices may be equal and then there is no other source of variation of sales across varieties. To see the consequence we follow up an example of the own authors. Assume that we observe the sales of a can of Coca-Cola and a water bottle of Perrier at adjusted prices. If the number of Perrier bottles is less than the number of Coke cans the model implies that consumers value more Perrier water and less the Coke can, exactly by the amounts that are needed to compensate the difference in quantities and get equal total valuations (and sales).[1]

A5. Data treatment.

A5.1. Linking the data.

Discontinuity of information for an existing firm occurs in the raw database for two reasons. First, and most importantly, a firm can be allocated a different ID (9 digit-code) during the period. Firms may receive a new ID if they are subject to some restructuring (change of name, ownership...), merger or acquisition. This creates numerous broken sequences and spurious exit and entry. Second, if a non-state owned firm falls below the sales threshold of RMB 5 million then it is not surveyed. If the firm re-enters the sample keeping its ID, we only get some missing observations in its time sequence. However, the firm can also not re-enter the sample. In this case we unfortunately have no strictly way to distinguish its disappearance from economic shutdown. The likelihood of these situations is however small.

With regards to the case of the changing IDs, we have done intensive work (in the style of Brandt, Van Biesebroeck and Zhang, 2012) to link the data of the firms that presumably

[1] Hottman, Redding and Weinstein (2016) nest another CES utility over the different products of the multiproduct firm rather than considering single products. However, this doesn't change the basic propoerties of symmetry of the model, that shows the commented behavior for the whole set of products of the multiproduct firm and some additional properties on the number of products induced by the CES specification.

had the ID changed. This process has used extensive information such as: the firm's name, corporate representative, 6-digit district code, post code, address, telephone number, industry code, year of birth. We first check the neighbor years two by two, then the longer panel sequences with the following/previous years. After linking the data, we treat all disappearances from the sample as economic shutdowns assuming that the errors will be small. We test this by checking whether the exit rates make sense.

The results are quite satisfactory. Focusing on manufacturing, considering firm time sequences with a minimum of two years, we have a total of 445,397 firms and 2,253,388 firm-year data points. After our linking, firms stay in the sample an average of 5 years. We have time sequences of 5 or more years for more than half of the firms, and more than 80% of these sequences have no missing observations.

The linked data is summarized in Table 0a. Column (1) shows that the single observations discarded after the process are a small percentage, except for the starting and final years, at which the process of linking is more difficult. Columns (2) and (3) document the growth of the sample over time, particularly significant in the Census year of 2004. Entry and exit are reported in columns (4) and (5). Entry is defined as the set of firms that are newly included in the sample and born the same year or either of the two previous years. Its average rate is 9.4%. Exit is defined as the set of firms that are last seen in the sample the previous year and it shows an average rate of 7.9%. Exit is indirectly induced by our linking and can include failures in the linking process as well as firms in a process of drastic downsizing.[2]

Both entry and exit show sensible values and explain a significant part of the increase of the sample. The increase in newly born firms in the Census years of 2004 and 2008 is particularly high, probably reflecting the effort of administrative authorities in being exhaustive. The resulting net entry rate (entry minus exit), reported in column (6), is positive starting in 2003.

Column (7) documents the increases in the sample which are not related to entry and exit. Part of this improvement can be attributed to the increase of the number of firms

[2]It also includes the 22 SOEs below 5 M RMB in 2006 that are not going to be surveyed the following years. See the text.

with a size above the threshold. A detailed analysis of the size of these additions shows that between 60% and 80% are firms with sales between RMB 5 and 25 millions, the rest are firms with mostly bigger sales. Therefore, it is likely continuous statistical improvement of the Annual Census so that it covers more firms, which increases the size of the sample. Column (8) computes the ratio of the aggregate industry value added plus value added taxes to the industry GDP from the China Statistical Yearbook. Coverage is improving over time, reaching near 90% in the final years.[3] The degree of response of the sample firms, considered year to year, tends to exceed 96% (see column 9).

A5.2. Data cleaning.

We clean the linked data according to the conditions reflected in Table 0b. We set observations to a missing value if there are some particularly small values in revenue, capital, wage bill and the cost of materials; some abnormal values in other variables (details in the table); or some consistency problems (revenue is less than exports or less than sales effort); or variable cost (wage bill plus the cost of materials) greater than revenue; or financial capital is less than the sum of the reported components. This enlarges the number of data points without real information. We then use the longest time subsequence (adjacent years) with complete information for each firm, provided that is greater than one year. The cleaned sample retains 84% of the firms and 74% of observations.

A6. Results.

Table 3b explores the context of our a and b functions estimation. It reports the result of regressing the dependent variable for all firms (exporters and non-exporters, in total more than a million and a half observations) on a constant and a dummy which takes value one

[3] A comparison of industry aggregates with the aggregate numbers of the 2004 Census leads Brandt, Van Biesebroeck and Zhang (2012) to conclude that they match very well and account for about 90% of Chinese industrial output. Our numbers match well with their numbers, except for differences in the content of calculated variables. However, we prefer the comparison over time of the data aggregates with the global GDP estimates.

if the firm exports. We can read the result as telling us what is the mean price-average cost margin for non-exporters and how it differs for exporters. The mean margins for non-exporters look reasonable, ranging across industries from 14 to 20 percent. When compared with domestic margins in Table 3 it turns out that exporters tend to have slightly greater domestic margins. But the important conclusion of this regression is that price-average cost margins of the exporters are, if anything, slightly lower than the price-average cost of margins of the firms that sell only in the domestic market. In five industries price-average cost margins of exporters are lower by an amount that ranges from 1 to 2 percentage points, in two more industries there is no significant difference, and only in three industries the margins for exporters are up to 1 percentage point above the margin of non-exporters.[4] Our target is, however, a different object in a different sample: the price-marginal cost margins or markups of the exporters who also sell domestically, and in what follows we are going to focus on them.

A6.1 Robustness checks on estimating functions a and b.

Table 3c reports the results of several robustness checks on estimating functions a and b.

Margins computed with variable costs can be problematic because labor is affected by costs of adjustment. This is why some authors prefer to use the margin over material costs, although material costs are likely to be affected by other rigidities (such the ones created by outsourcing).[5] Columns (1) and (2) repeat the main regression with the log of the ratio of revenue over material costs as dependent variable. This measure seems a little noisier and in two industries the difference of margins changes sign. However, we do not see evidence

[4] In fact our regression can be compared with the regressions in de Loecker and Warzynski (2012), who obtain higher markups for the Slovenian exporters when compared with the markups of the domestic sellers. Their "Specification I" uses as dependent variable the log of revenue over the wage bill (instead of the whole variable costs) corrected by a constant representing the Cobb-Douglas elasticity of labor and an estimated error term. They run the regression controling for capital and labor. When we introduce these controls, two more industries show significantly smaller margins for exporters and differences tend in general to increase.

[5] See Doraszelski and Jaumandreu (2016) for an analysis of the effects of outsourcing on the ratio materials-labor.

of a systematic bias using variable cost.

Subsidies have been argued as a source of distortions in Chinese manufacturing. To check if this could be affecting the estimated margins, we again run the baseline regression with the parameter representing the function b interacted with the ratio subsidy over sales. Results are reported in columns (3) to (5). We find in 6 industries relatively better margins of the subsidized exporters but by negligible amounts (to see this multiply the coefficients by the mean subsidies in Table 2).

When domestic demand experiences a negative shock, firms can turn to the export market to try to sell their excess of domestic production. This can, in principle, generate a negative correlation between the disturbance of the equation and export intensity. This potential correlation brings a plausible explanation for a downward biased estimate of the function b (too big markup differences). To test for this possibility, we run the baseline specification by GMM using lagged export intensity as instrument. The results are reported in columns (6) and (7). The estimated difference between margins still increases in all industries. The results indicate, if anything, a slightly positive correlation between export intensity and the disturbance of the equation. This suggests that higher export intensity may be associated with better times. However the changes are not dramatic so we do not consider it necessary to change the baseline specification.

Exports can generate higher transportation costs for firms that are not located in the coastal (East) area of China (between 6 and 24 percent of the firms in our sample, depending on the industry, are located in the Middle and West areas). This cost varies with the amount of exports and for these firms it affects the average and marginal costs of exports. It is possible that this cost is not included or only partially included in our variable cost measure. We run the baseline regression interacting the slope with a dummy of Middle-West location. Results are in columns (8) to (10). We find that in 8 industries the exporting margins of these firms are not as poor as predicted by the general relationship. We conclude that their exporting prices are likely to be higher just to cover for this extra unobserved cost. As the proportion of affected firms is small, and we do not see a significant change in the rest of margins, we conclude that this is not likely to significantly distort our estimation of the elasticities.

Columns (11) and (12) report the result of estimating a and b over time. We recover the implicit time-series of price-average cost margins and we compute their standard errors over time. Variation over time is small without dramatic changes.

A6.2. Robustness checks on the system for exports and domestic sales.

We first check the result of allowing elasticities to vary across firms and over time. Equation (13) and our estimation procedure offer a nice framework to generalize this aspect. Grouping the firms in the industry into sets of firms that have different elasticities, it is possible to test for this variation (and its effects) by means of heterogeneous a and b functions to be brought to the system estimation. The elasticities of the firms in our sample tend to change with the size of the firms, quality of products (measured through workforce skills), and foreign participation, although differences are not dramatic. They do not change, instead, with either location and age of the firm.

Table 4b reports the results of estimating the system with elasticities that vary with size of the firms, quality of the product, and foreign participation. We define three dummies that take the value one when the firm has a value that is greater than the sample mean. Columns (2) to (4) and (6) to (8) report the result of interacting these dummies with the constant and the slope of the equation. This gives functions a and b that vary with the value of these dummies and that we brought to the estimation of the system. Bigger firms, firms with a higher quality product and with foreign participation tend to show greater market power domestically. We now individually estimate varying elasticities in the system, and columns (12) and (13) report averages across individuals. Nothing changes dramatically with respect to the baseline estimation. Additionally we check the estimated productivity and demand advantages (compared with the estimates that we report in Section 6 for the main specification), they change very little.

To consider arbitrary forms of heterogeneity, we reestimate the system for each industry including subindustry dummies at the four digit level. This requires the use of 392 total dummies (see Appendix E of the text). The results are reported in Table 4c. The new specification induces very small changes in the estimates of the coefficients, productivity and demand advantages. We conclude that we could successfully estimate our specification at a higher level of disaggregation with similar results.

Our sample only considers firms that simultaneously sell both in the domestic and export markets. As explained in Section 4, this opens the possibility that the system needs to be corrected for sample selection bias. In fact, we have found that the effect of this correction was small but significant in the equation that we use to estimate the a and b functions. We now test the need for correction in the system. If selection operates, as explained in the text, the expectations of the Markov processes become a function of an unobserved threshold. $\overline{\tau}_{jt}$

$$\begin{aligned} E[(\eta_{.}-1)\omega_{jt}+\delta_{jt}|I_{t-1},\tau(\omega_{jt},\delta_{jt}) &> \overline{\tau}_{jt}] = \\ \int_{\omega(\delta,\overline{\tau})}^{\infty}\int_{\delta(\omega,\overline{\tau})}^{\infty}[(\eta_{.}-1)\omega+\delta]\frac{P(\omega,\delta|\omega_{jt-1},\delta_{jt-1})d\omega d\delta}{\int_{\omega(\delta,\overline{\tau})}^{\infty}\int_{\delta(\omega,\overline{\tau})}^{\infty}P(\omega,\delta|\omega_{jt}-1,\delta_{jt-1})d\omega d\delta} &= \\ (\eta_{.}-1)\widetilde{g}(\omega_{jt-1},\delta_{jt-1},\overline{\tau}_{jt})+\widetilde{h}(\omega_{jt-1},\delta_{jt-1},\overline{\tau}_{jt}) &= (\eta_{.}-1)g(\cdot)+h(\cdot) \\ &\quad +\varphi(\omega_{jt-1},\delta_{jt-1},\overline{\tau}_{jt}), \end{aligned}$$

where I_{t-1} represents information available at $t-1$.

Because the probability of exporting is related to the unobservable threshold we can, in the tradition of Olley and Pakes (1996), invert this relationship and include the estimated probability in the Markov processes for productivity and demand advantages. We use the estimated probit equation for the probability of exporting based on the observations of all firms. In practice, we introduce in both equations a second order polynomial in the estimated probability, interactions of the probability with ω_{jt-1} and δ_{jt-1}, and the product of ω_{jt-1} and δ_{jt-1}.

Unfortunately, one of the effects of this introduction is that parameter ν tends to become very difficult to estimate and the routine crashes.[6] The model only converges normally for

[6] For a given a estimate, if ν becomes too small the elasticity has a discontinuity and may become negative.

Machinery and *Electronics.* We attribute this effect to the often observed multicollinearity problems introduced by the selection corrections. We correct this by fixing the parameter ν at its previously estimated value, allowing its components to be freely estimated. Table 4d reports in columns (1) to (5) the result of estimating the system. The results do not reveal any particular pattern that can be linked to selection. We might expect the coefficient of capital to increase once the negative correlation with the shocks conditional in exporting is controlled for[7]. The coefficient on capital tends instead to become smaller, which we attribute to the multicollinearity problem. We conclude that there is no reason to be worried about possible biases due to selection.

The same product characteristics can have a different impact in the domestic and export markets. We allow for this possibility by estimating an additional parameter λ as coefficient of δ_{jt} in the domestic market. The introduction and estimation of λ is not easy because it tends to pick up any unbalance between the two equations. Therefore, it tends to break the convergence of the model. Identification is probably very weak. Scaling the processes of both equations by a common factor, depending of the industry price index $(-[(\eta_X - 1) + (\eta_D - 1)]p_t)$, the model converges (except in one industry, *Textile*) and we get reasonable results. Table 4d reports the result of estimating the system specifying δ_{jt} as affected by a coefficient λ in the domestic market in columns (6) to (11). Parameter λ is estimated close to unity in at least half of the industries. The results indicate that in half of the industries the different impact of the advantages is not an important issue. However, in the other four estimable cases we have sundry values of λ that seem to indicate a different impact of the advantages. Interestingly, the elasticities and estimated productivities do not change dramatically in these industries.

A6.3. Several details on the estimated ω_{jt} and δ_{jt}.

The economics of cost and demand advantages.

[7] This is what happens in Olley and Pakes (1996) because bigger capital allows the firm to overcome the worst shocks.

Assume for simplicity that all firms are equal in observed costs and demand shifters, and therefore heterogeneity is exclusively driven by the unobserved terms ω_{jt} and $\delta_{jt}/(\eta_D - 1)$. It is easy to check that total profit (the sum of profits obtained in each market) is increasing for ω_{jt} and δ_{jt}. In addition, cost and demand advantages are "complements", in the sense that the amount of each advantage increases the marginal profitability of the other advantage. Isoprofit curves in the plane $(\omega, \delta/(\eta_D - 1))$ have a negative slope steeper than -1 and are concave. This is the result of the export market having a higher elasticity of demand, which makes it desirable for the firm that has greater efficiency to sell more in this market. For the sake of the argument, suppose that firms can transform one advantage into the other given heterogeneous total endowments of amount $[\omega^2 + (\delta/(\eta_D - 1))^2]^{\frac{1}{2}}$. That is, firms have perfectly balanced transformation curves at different distances of the origin (represented by the relevant portion of circumferences of different ratios). With profit maximizing firms, our observations on the pairs $(\omega_{jt}, \delta_{jt}/(\eta_D - 1))$ would lie on a positively sloped curve with slope less than one.

Decomposition of growth of demand advantages.

We decompose the demand advantages growth into a gross component and the effect of entry. The demand advantage of an individual firm in a particular market includes an average component that is common with the rest of firms (see Section A3). This component will shrink if the demand has to be shared with an increased number of firms. We can hence write

$$\begin{aligned}
& \Delta\delta_{jt}/(\eta - 1) \\
= \quad & -[S_{jt}^{D}(\Delta \ln N_t^{DE} - \Delta \ln N_t^{DX})/(\eta_D - 1) + S_{jt}^{X}(\Delta \ln N_t^{XE} - \Delta \ln N_t^{XX})/(\eta_X - 1)] + \Delta\delta_{jt}^{*}/(\eta - 1),
\end{aligned}$$

where the term in brackets is the weighted sum of net entry faced by the firm[8] and the second term is the gross growth of the firm's demand advantage.

We estimate rough rates of entry and exit in each industry and in the corresponding exports market using the linked data and following the same methodology as in Section 5 to separate economic entry from additions. Net entry estimates from 1998 to 2008 (entry

[8]Entry in the export market is different because it includes starts and stops.

minus exit) for the domestic and exports market are reported in columns (1) and (2) of Table 6b. They amount to an average net entry of 13% in the domestic market and 8% in the export market.

Weighted means decomposition.

Calling x_{jt} the variable of interest (ω_{jt} or $\delta_{jt}/(\eta-1)$) and w_{jt} the revenue weights, the Melitz and Polanec (2015) decomposition is

$$\sum w_{j2}x_{j1} - \sum w_{j1}x_{j1} = (x_2^S - x_1^S) + w_2^E(x_2^E - x_2^S) + w_1^X(x_1^S - x_1^X),$$

where $x_t^I = (\sum_{j \subset I} w_{jt}x_{jt})/\sum_{j \subset I} w_{jt}$ and $w_t^I = \sum_{j \subset I} w_{jt}/\sum_{j \subset I} w_{jt}$. The growth of survivors can be further split as follows (Olley and Pakes, 1996)

$$x_2^S - x_1^S = (\overline{x}_2^S - \overline{x}_1^S) + [\sum w_{j2}(x_{j2} - \overline{x}_2^S) - \sum w_{j1}(x_{j1} - \overline{x}_1^S)],$$

where $\overline{x}_1^S$ and $\overline{x}_2^S$ are simple means and the terms involved in brackets are covariances multiplied by N_t.

We report the three terms of the first formula in columns (5) to (12) of Table 6 and we detail the split of the survivors term under the names "Shift" and "Covariance" in columns (3) to (6) of Table 6b. We observe that the shares of entrants and exitors are all very significant and the estimates should be very robust statistically. Shares are reported in columns (7) to (10) of Table 6b.

References

Brandt, L., J. Van Biesebroeck and Y. Zhang (2012), "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97, 339-351.

Dixit, A. and J. Stiglitz (1977), "Monopolistic Competition and Optimal Product Diversity, *American Economic Review*, 67, 297-308.

Doraszelski, U. and J Jaumandreu (2017), "Measuring the Bias of Technological Change," forthcoming *Journal of Political Economy.*

Hottman, C., S. Redding and D. Weinstein (2016), "Quantifying the Sources of Firm Heterogeneity," *Quarterly Journal of Economics*, 1291-1364.

Melitz, M. (2000), "Firm Productivity Estimation in Differentiated Product Industries," mimeo, Harvard University.

Melitz, M. and S. Polanec (2015), "Dynamic Olley-Pakes decomposition with Entry and Exit," forthcoming *Rand Journal of Economics,* 46, 362-375.

Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263-1298.

Table 0a: Manufacturing linked data[a]

Years	Discarded single obs.[b]	No. of firms[c]	Sample growth[d]	Entry rate[e]	Exit rate[f]	Net entry rate[g]	Additions[h]	Aggreg. output /Industry GDP[i]	Response rate[j]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1998	0.153	129,671	-	0.142	-	-	-	0.557	1.000
1999	0.026	145,949	0.112	0.044	-	-	-	0.578	0.971
2000	0.025	149,371	0.023	0.050	0.093	-0.043	0.066	0.608	0.955
2001	0.021	159,471	0.063	0.081	0.110	-0.029	0.092	0.605	0.950
2002	0.018	170,979	0.067	0.070	0.075	-0.005	0.072	0.638	0.946
2003	0.030	184,537	0.073	0.084	0.080	0.004	0.069	0.626	0.943
2004	0.067	247,854	0.255	0.176	0.099	0.077	0.178	0.741	0.966
2005	0.009	263,681	0.060	0.069	0.046	0.023	0.037	0.760	0.939
2006	0.010	288,433	0.086	0.088	0.055	0.033	0.053	0.813	0.953
2007	0.021	315,769	0.087	0.086	0.057	0.029	0.058	0.881	0.966
2008[k]	0.167	333,330	0.053	0.145	0.092	0.053	0.000	0.870	1.000
1998-2008		445,397							0.963

[a] We only retain firms which stay two and more years.
[b] As proportion of the remaining number of firms.
[c] There are 2,253,388 firm-year observations.
[d] New firms as proportion of number of firms at t. $Sample\ growth = Entry\ rate - Exit\ rate + Additions$.
[e] Newly included firms born in t, $t-1$ or $t-2$, as proportion of number of firms at t.
[f] Firms last seen at $t-1$ as proportion of number of firms at t. Not defined for 1998 and 1999.
[g] Entry rate - exit rate.
[h] Sample growth - net entry.
[i] $(Revenue - Cost\ of\ Materials + VA\ tax)/Industry\ GDP\ from\ China\ Statistical\ Yearbook$.
[j] Proportion of firms in sample at year t which report information.
[k] 2008 entrants, 48,369 firms, treated (in this row) as if they were to stay two or more years.

Table 0b: Filters used to clean the linked data

Values are set to missing in the following cases:

Small values:
- Less than 8 workers or 30,000 RMBs in *Revenue*, *Capital*, *Wage bill*, *Cost of materials*.

Abnormal values:
- Negative value in *Exports* or *Sales effort*.
- Zero or less in finacial capital, negative value in a financial component.
- Born before 1949 or after 2008.

Consistency:
- Revenue less than *Exports*, *Sales effort* or *Variable cost* (*Wage bill*+ *Cost of materials*).
- Financial capital is less than the sum of its finacial components.

A missing value is an interruption of the firm time sequence. We only use the firm's longest time subsequence provided that is longer than one year.

The cleaned sample retains 83.7% of the firms and 73.6% of the observations.

Table 0c: Data and sample.

			Average levels				Firm's category (proportions)[e]				Average size (labor) by category[e]			
		Firms	Revenue[a]	Capital[b]	Labor[c]	Materials[d]	Surv. (and exp.)	Add.	Entrants (/Starts)	Exitors (/Stops)	Surv. (and exp.)	Add. (/Starts)	Entrants	Exitors (/Stops)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Total data	1998	75,949	43.732	28.001	333	32.529	0.246			0.754	407			307
	2008	215,402	120.521	27.661	211	79.706	0.087	0.137	0.776		499	238	204	
	Growth[f]		0.158	0.110	0.032	0.142								
Sample	1998	11,948	117.852	86.926	739	86.971	0.199			0.801	1002			674
	2008	34,749	301.975	72.834	451	193.722	0.068	0.127	0.805		1201	567	413	
	Growth[f]		0.130	0.092	0.041	0.111								

[a] Deflated by industry output price indices (Millions of RMBs).
[b] Deflated by an investment price index (Millions of RMBs).
[c] Number of workers.
[d] Deflated by industry materials price indices (Millions of RMBs).
[e] Categories in the sample are: Survivors and exporterss, Additions exporting, Entrants/Start exporting, Exitors/Stop exporting.
[f] Unweighted average of rates of growth.

Table 3b: Regressing $\ln \frac{R_{jt}}{C_{jt}}$ on export status, OLS.

	Number of obs.	Constant (s. e.)[b]	Exports dummy (s. e.)[b]
	(1)	(2)	(3)
1. Food, drink and tobacco	168,548	0.193	-0.009
		(0.001)	(0.002)
2. Textile, leather and shoes	243,120	0.138	-0.017
		(0.001)	(0.001)
3. Timber and furniture	57,228	0.167	-0.020
		(0.001)	(0.002)
4.Paper and printing products	85,732	0.170	0.002
		(0.001)	(0.003)
5. Chemical products	273,603	0.203	-0.012
		(0.001)	(0.002)
6. Non-metallic minerals	144,666	0.199	0.008
		(0.001)	(0.002)
7. Metals and metal products	162,002	0.138	0.006
		(0.001)	(0.001)
8. Machinery	205,014	0.178	0.012
		(0.001)	(0.001)
9.Transport equipment	78,987	0.173	0.001
		(0.001)	(0.002)
10. Electronics	176,791	0.187	-0.019
		(0.001)	(0.001)

[a] Standard errors are robust to heteroskedasticity and autocorrelation.

Table 3c.Robustness checks on estimating functions a and b.

	R over cost of materials		Effect of subsidies[a]			GMM estimation		Middle-West location[a]			Time varying margins	
	a (s. e.)[b]	b (s. e.)[b]	a (s. e.)[b]	b_0 (s. e.)[b]	b_1 (s. e.)[b]	a (s. e.)[b]	b (s. e.)[b]	a (s. e.)[b]	b_0 (s. e.)[b]	b_1 (s. e.)[b]	Dom. (s. d.)[c]	Exports (s. d.)[c]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. Food, drink and tobacco	0.332	0.114	0.234	0.106	0.009	0.193	0.147	0.173	0.112	-0.037	0.271	0.143
	(0.007)	(0.010)	(0.005)	(0.008)	(0.005)	0.012	0.007	(0.005)	(0.008)	(0.008)	(0.018)	(0.012)
2. Textile, leather and shoes	0.260	-0.025	0.144	0.037	-0.008	0.105	0.046	0.098	0.036	-0.006	0.158	0.117
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.014)	(0.009)
3. Timber and furniture	0.291	0.030	0.175	0.046	-0.007	0.141	0.062	0.131	0.047	-0.028	0.199	0.150
	(0.006)	(0.009)	(0.004)	(0.007)	(0.005)	(0.009)	(0.006)	(0.004)	(0.006)	(0.009)	(0.034)	(0.022)
4.Paper and printing products	0.303	0.015	0.193	0.056	-0.010	0.124	0.064	0.121	0.056	-0.094	0.218	0.142
	(0.007)	(0.012)	(0.005)	(0.009)	(0.013)	(0.011)	(0.006)	(0.005)	(0.008)	(0.034)	(0.018)	(0.026)
5. Chemical products	0.351	0.080	0.232	0.093	-0.023	0.181	0.115	0.169	0.093	-0.068	0.266	0.158
	(0.005)	(0.008)	(0.003)	(0.006)	(0.005)	(0.008)	(0.004)	(0.003)	(0.006)	(0.011)	(0.017)	(0.014)
6. Non-metallic minerals	0.356	-0.041	0.229	0.038	-0.020	0.162	0.046	0.155	0.038	-0.034	0.261	0.229
	(0.006)	(0.010)	(0.004)	(0.007)	(0.009)	(0.010)	(0.005)	(0.004)	(0.007)	(0.013)	(0.012)	(0.033)
7. Metals and metal products	0.275	0.017	0.163	0.034	-0.003	0.117	0.039	0.114	0.033	-0.004	0.182	0.141
	(0.004)	(0.006)	(0.003)	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)	(0.008)	(0.016)	(0.013)
8. Machinery	0.405	0.131	0.218	0.071	-0.014	0.162	0.079	0.157	0.069	-0.050	0.247	0.170
	(0.004)	(0.131)	(0.002)	(0.004)	(0.005)	(0.005)	(0.003)	(0.002)	(0.004)	(0.013)	(0.012)	(0.015)
9.Transport equipment	0.358	0.095	0.198	0.056	-0.012	0.140	0.059	0.136	0.054	-0.046	0.227	0.165
	(0.006)	(0.011)	(0.003)	(0.006)	(0.006)	(0.008)	(0.004)	(0.003)	(0.006)	(0.018)	(0.022)	(0.022)
10. Electronics	0.378	0.110	0.214	0.087	-0.018	0.168	0.097	0.162	0.084	-0.057	0.247	0.148
	(0.004)	(0.006)	(0.002)	(0.004)	(0.004)	(0.005)	(0.003)	(0.002)	(0.004)	(0.012)	(0.021)	(0.013)

[a] Parameter b_1 is the coefficient on the interaction term.
[b] Standard errors robust to heteroskedasticity and autocorrelation.
[c] Standard deviations over time of the estimated margins.

Table 4b: Specifying and estimating with elasticities which vary with firm size, product quality and foreign capital.

Industry	Estimating varying functions of the elasticities[a]								Estimating the system with varying elasticities: Input elasticity			Demand elas.[b]	
	a_0 (s. e.)[c]	a_{size} (s. e.)[c]	$a_{quality}$ (s. e.)[c]	$a_{foreign}$ (s. e.)[c]	b_0 (s. e.)[c]	b_{size} (s. e.)[c]	$b_{quality}$ (s. e.)[c]	$b_{foreign}$ (s. e.)[c]	k (s. e.)[c]	l (s. e.)[c]	m (s. e.)[c]	η_D	η_X
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. Food, drink and tobacco	0.160	0.085	0.073	0.064	0.020	0.131	0.095	0.051	0.051	0.121	0.829	6.7	13.6
	(0.004)	(0.013)	(0.008)	(0.012)	(0.007)	(0.024)	(0.013)	(0.018)	(0.013)	(0.015)	(0.007)		
2.Textile, leather and shoes	0.126	0.014	0.020	0.030	0.030	0.011	0.016	-0.001	0.020	0.209	0.703	21.1	133.2
	(0.002)	(0.004)	(0.003)	(0.006)	(0.003)	(0.006)	(0.004)	(0.007)	(0.005)	(0.012)	(0.012)		
3. Timber and furniture	0.154	0.019	0.021	0.041	0.019	0.031	0.029	0.034	0.019	0.110	0.812	11.2	20.1
	(0.005)	(0.011)	(0.007)	(0.014)	(0.008)	(0.015)	(0.011)	(0.019)	(0.007)	(0.015)	(0.015)		
4.Paper and printing products	0.161	0.031	0.028	0.040	0.034	0.041	0.018	-0.009	0.032	0.101	0.797	10.0	19.9
	(0.005)	(0.010)	(0.008)	(0.012)	(0.009)	(0.019)	(0.014)	(0.022)	(0.014)	(0.011)	(0.013)		
5. Chemical products	0.168	0.058	0.061	0.085	0.030	0.093	0.047	0.063	0.058	0.104	0.891	5.2	7.7
	(0.003)	(0.008)	(0.005)	(0.009)	(0.006)	(0.015)	(0.009)	(0.014)	(0.009)	(0.012)	(0.013)		
6. Non-metallic minerals	0.196	0.012	0.041	0.067	0.029	-0.007	0.021	0.031	0.005	0.157	0.721	11.2	18.9
	(0.004)	(0.008)	(0.007)	(0.011)	(0.008)	(0.016)	(0.013)	(0.019)	(0.016)	(0.015)	(0.015)		
7. Metals and metal products	0.136	0.004	0.035	0.044	0.017	0.010	0.029	0.006	0.045	0.130	0.780	17.3	32.8
	(0.003)	(0.006)	(0.005)	(0.007)	(0.005)	(0.009)	(0.007)	(0.011)	(0.006)	(0.008)	(0.009)		
8. Machinery	0.187	-0.014	0.048	0.051	0.064	-0.015	0.015	0.006	0.048	0.202	0.722	8.2	16.3
	(0.003)	(0.004)	(0.004)	(0.006)	(0.005)	(0.009)	(0.008)	(0.010)	(0.006)	(0.021)	(0.023)		
9.Transport equipment	0.165	0.001	0.037	0.051	0.032	0.025	0.019	0.023	0.068	0.097	0.819	10.5	20.2
	(0.004)	(0.007)	(0.005)	(0.007)	(0.006)	(0.015)	(0.011)	(0.013)	(0.007)	(0.012)	(0.013)		
10. Electronics	0.174	0.000	0.055	0.041	0.051	0.026	0.037	0.024	0.046	0.281	0.640	8.1	20.3
	(0.003)	(0.005)	(0.004)	(0.005)	(0.004)	(0.008)	(0.007)	(0.008)	(0.009)	(0.021)	(0.020)		

[a] a. and b. denote the coefficients of the interactions.
[b] Averages over the sample of the individual elasticities.
[c] Standard errors robust to heteroskedasticity and autocorrelation.

Table 4c: Estimating the system with subindustry dummies.

Industry	No of subind.	Input elasticity k (s. e.)[a]	l (s. e.)[a]	m (s. e.)[a]	Demand elas. η_D	η_X	$\Delta\omega$	$\Delta\delta/(\eta-1)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Food, drink and tobacco	49	0.048	0.217	0.693	6.7	17.7	0.188	0.068
		(0.010)	(0.029)	(0.033)				
2.Textile, leather and shoes	33	0.016	0.440	0.469	20.1	64.0	0.441	0.024
		(0.006)	(0.023)	(0.023)				
3. Timber and furniture	13	0.016	0.214	0.682	13.1	28.7	0.320	0.014
		(0.012)	(0.038)	(0.012)				
4.Paper and printing products	10	0.063	0.266	0.628	10.2	20.9	0.406	0.022
		(0.013)	(0.042)	(0.039)				
5. Chemical products	61	0.033	0.280	0.571	11.5	173.4	0.298	0.108
		(0.011)	(0.025)	(0.032)				
6. Non-metallic minerals	30	0.063	0.303	0.518	15.4	31.8	0.596	0.036
		(0.014)	(0.031)	(0.033)				
7. Metals and metal products	37	0.054	.230	0.656	15.3	27.4	0.427	-0.023
		(0.008)	(0.020)	(0.021)				
8. Machinery	73	0.075	0.193	0.692	8.8	17.9	0.304	0.368
		(0.006)	(0.020)	(0.021)				
9.Transport equipment	23	0.075	0.113	0.779	10.2	19.2	0.586	-0.046
		(0.009)	(0.016)	(0.018)				
10. Electronics	63	0.084	0.392	0.559	6.4	11.7	0.376	0.407
		(0.007)	(0.034)	(0.033)				

[a]Standard errors robust to heteroskedasticity and autocorrelation and corrected for two-step estimation.

Table 4d: Other robustness checks on the estimation of the system.

Industry	Estimating the system with selection					Different impacts of demand advantages					
	Input elasticity			Demand elas.		Input elasticity				Demand elas.	
	k (s. e.)[a]	l (s. e.)[a]	m (s. e.)[a]	η_D	η_X	k (s. e.)[a]	l (s. e.)[a]	m (s. e.)[a]	λ (s. e.)[a]	η_D	η_X
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1. Food, drink and tobacco	-0.004	0.237	0.685	6.3	14.5	0.043	0.276	0.648	1.039	6.2	14.3
	(0.012)	(0.044)	(0.044)			(0.014)	(0.043)	(0.015)	(0.116)		
2.Textile, leather and shoes[b]	0.012	0.211	0.697	20.3	66.6						
	(0.003)	(0.011)	(0.011)								
3. Timber and furniture	0.015	0.223	0.673	13.1	28.9	0.023	0.229	0.672	1.729	12.3	25.0
	(0.010)	(0.033)	(0.033)			(0.014)	(0.043)	(0.048)	(1.752)		
4.Paper and printing products	0.050	0.248	0.648	10.0	19.9	0.060	0.273	0.623	1.001	10.0	19.9
	(0.013)	(0.030)	(0.030)			(0.013)	(0.042)	(0.040)	(0.205)		
5. Chemical products	0.027	0.185	0.737	6.3	12.1	0.067	0.056	0.862	0.937	6.5	12.6
	(0.009)	(0.018)	(0.018)			(0.008)	(0.018)	(0.018)	(0.088)		
6. Non-metallic minerals	0.051	0.261	0.562	15.0	30.1	0.066	0.292	0.529	0.755	15.6	32.9
	(0.014)	(0.022)	(0.022)			(0.017)	(0.037)	(0.027)	(0.196)		
7. Metals and metal products	0.018	0.266	0.622	15.0	26.4	0.059	0.214	0.672	0.952	15.4	27.9
	(0.013)	(0.024)	(0.024)			(0.008)	(0.020)	(0.019)	(0.010)		
8. Machinery	-0.008	0.197	0.694	8.3	16.0	0.075	0.209	0.706	1.501	7.0	11.5
	(0.010)	(0.022)	(0.009)			(0.007)	(0.024)	(0.026)	(0.253)		
9.Transport equipment	0.025	0.141	0.754	9.9	18.3	0.046	0.180	0.685	0.445	14.4	45.2
	(0.008)	(0.025)	(0.025)			(0.010)	(0.040)	(0.042)	(0.292)		
10. Electronics	0.012	0.568	0.388	6.2	11.0	0.077	0.314	0.616	0.879	7.3	15.4
	(0.020)	(0.046)	(0.013)			(0.005)	(0.026)	(0.024)	(0.016)		

[a]Standard errors robust to heteroskedasticity and autocorrelation and corrected for two-step estimation.

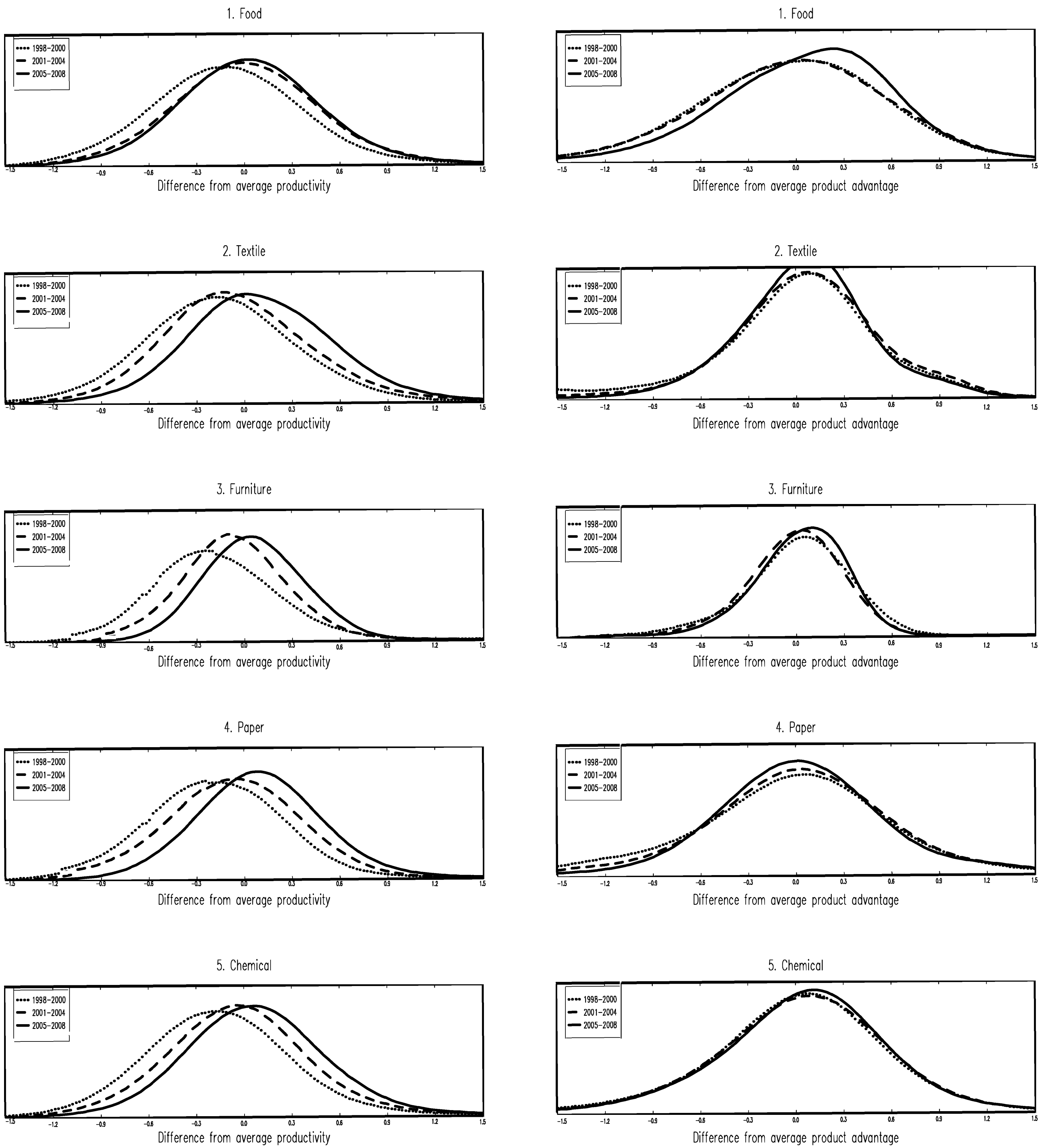
[b] We have been unable to compute the model with λ.

Table 6b: Rates of entry, decomposition of survivors' growth, and shares of survivors, entrants and exitors.

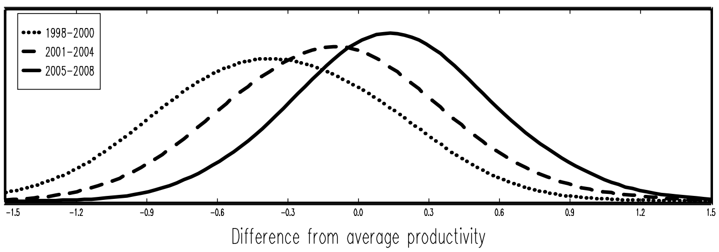
Industry	Market rates of net entry[a]		Decomposition of survivors' growth of ω		Decomposition of survivors' growth of $\frac{\delta}{(\eta-1)}$		Market shares[b] 2008		Market shares[b] 1998	
	Domestic	Exports	Shift	Covariance	Shift	Covariance	Survivors[c]	Entrants[d]	Survivors[c]	Exitors[e]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Food, drink and tobacco	-0.131	0.014	0.219	-0.093	0.072	-0.196	0.400	0.600	0.466	0.534
2.Textile, leather and shoes	0.233	0.073	0.369	0.136	0.046	-0.163	0.339	0.661	0.209	0.791
3. Timber and furniture	0.529	0.181	0.264	-0.010	0.024	0.068	0.230	0.770	0.270	0.730
4.Paper and printing products	-0.213	0.000	0.370	0.348	0.037	-0.643	0.302	0.698	0.297	0.703
5. Chemical products	0.152	0.052	0.321	0.119	0.039	-0.088	0.389	0.611	0.366	0.634
6. Non-metallic minerals	0.115	0.033	0.536	0.096	0.016	-0.010	0.302	0.698	0.326	0.674
7. Metals and metal products	0.226	0.072	0.371	0.334	0.026	-0.292	0.447	0.553	0.399	0.601
8. Machinery	0.133	0.087	0.277	0.208	0.396	-0.215	0.420	0.580	0.314	0.686
9.Transport equipment	0.039	0.124	0.538	0.205	0.031	-0.300	0.380	0.620	0.369	0.631
10. Electronics	0.205	0.168	0.387	0.235	0.458	-0.590	0.360	0.640	0.405	0.595

[a] Rates 1998-2008.
[b] 1% of observations at each tail of the distribution of δ have been trimmed for this exercise.
[c] Includes additions that are not new born or starts in the export market.
[d] Includes starts in the export market.
[e] Includes firms that stop exporting.

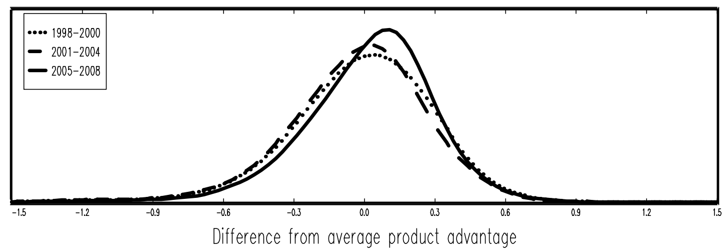
Figure A1: Density functions of ω and $\delta/(\eta_D\text{-}1)$



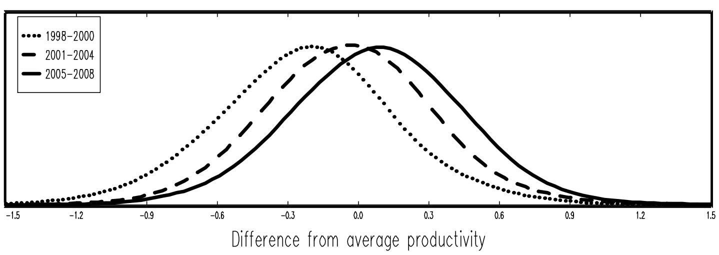
6. Non-met. minerals



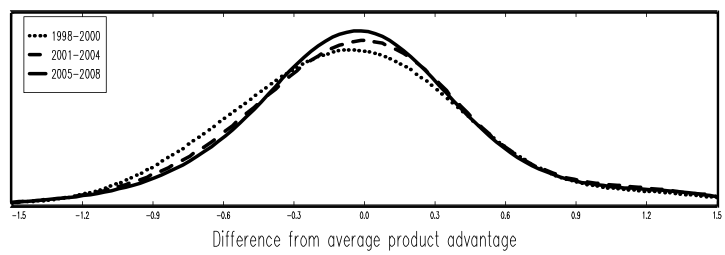
6. Non-met. minerals



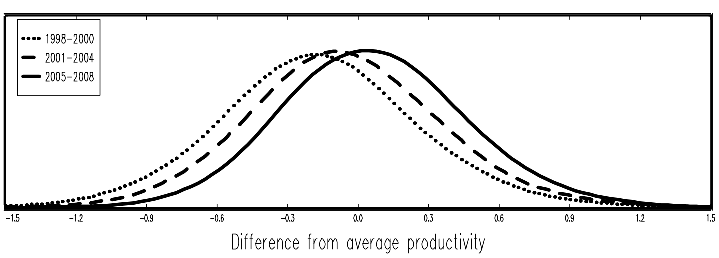
7. Metal products



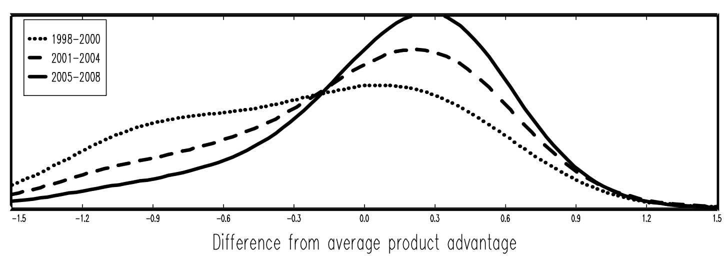
7. Metal products



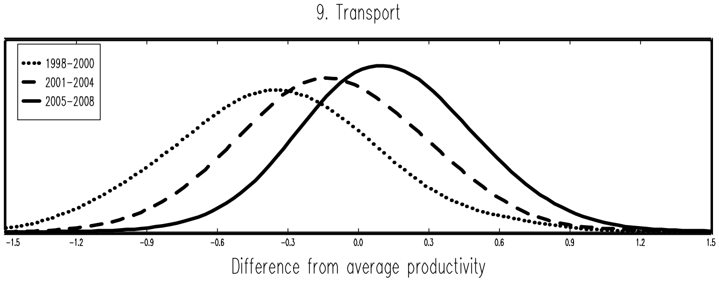
8. Machinery



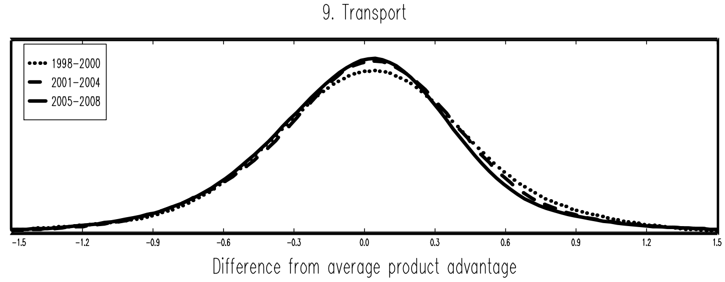
8. Machinery



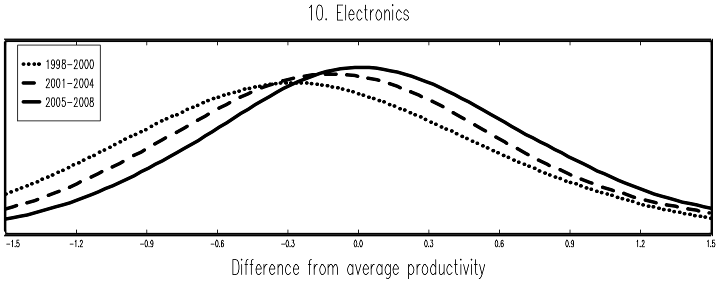
9. Transport



9. Transport



10. Electronics



10. Electronics