Cost and Product Advantages: A Firm-level Model for the Chinese Exports and Industry Growth*

Jordi Jaumandreu†
Boston University and CEPR

Heng Yin‡
Beijing Normal University

January 2013

(Preliminary and incomplete)

Abstract

This paper develops a simple model to estimate the demand and cost advantages of firms, both in their domestic and export markets, and applies it to characterize the spectacular growth of China exports and industry re-structuring, with data from 1998 to 2008. Using a structural model, we can separately estimate an idiosyncratic "product advantage," or degree of product atractiveness for the product of each firm in each market. Similarly, we can estimate the "cost advantage," or relative productivity that drives the evolution over time of the ability of the firm to produce at a different cost than its competitors. We first want to consider exogenously evolving persistent advantages, and then the impact of the firm investments on R&D and human capital, as well as some public policies, on these advantages. We expect the description of the distributions to shed light on many relevant questions about industrial growth.

*We thank Li Shigang, Liu Di and Urbashee Paul for excellent research assistance.

†Department of Economics, 270 Bay State Road, Boston, MA02215, USA. E-mail: jordij@bu.edu.

‡Department of Economics. 19 Xinjiekouwai Dajie, Haidian, Beijing 100088, PRC. E-mail: yheng@bnu.edu.cn.

1. Introduction.

Firms, even for narrowly defined industries, sell products that are different and exhibit different production costs. In addition, they typically set different prices. There are "product advantages," or advantages that drive the sales of a firm larger relative to the sales of its competitors. And there are "cost advantages," or relative productivity that drives the cost of the firm down from the cost of its competitors. Product differentiation (vertical, horizontal or any combination) is the most important explanation for persistent demand advantages of firms. Other demand advantages are generated by sales effort and customers exposure to advertising. The differences in the level of cost among firms which compete in (more or less) differentiated markets is the other key dimension that determines firms' shares in domestic as well as export markets.[1] To the extent that advantages may be considered endogenous, they become not only a way to explain the world but a possibility to inform public policy.[2] As both type of advantages crucially matter, it is clear the importance of developing methods of measurement that can describe in practice their distributions across firms and their change under firm action and government policies.

The bulk of economists' effort has been dedicated to the measurement of cost advantages. The measurement of productivity and its determinants at the firm-level has a long tradition, at least since Griliches (1979). See Syverson (2011) for a review that stresses the amount of heterogeneity in productivity found empirically. Recently, a literature on the estimation of production functions in the presence of unobserved productivity allows to recover the distribution of productivity across firms and time. The basic idea is to replace unobserved

[1] Trade economists have seen international competition from this point of view at least since Helpman and Krugman (1985, 1989). Recent empirical contributions in the area try to measure aspects, for example, as the role of efficiency (Eaton, Kortum and Kramarz, 2011) or the impact of quality (Crozet, Head and Mayer, 2012) in the "anatomy" of trade. Sutton (2001, 2007) has developed a model in which industrial development is linked to specific (changing) combinations of cost and quality of the products sold by firms.

[2] See, for example, the model by Aw, Roberts and Xu, 2011, where firms invest simultaneously in R&D and export, with both decisions affecting future productivity.

productivity by a function of observable variables, specifically by inverting the demand for a factor. This is the exercise of the pionering paper by Olley and Pakes (1996). Contributions to this literature are Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2006) and Wooldridge (2009). This way of estimating productivity constitutes today one of the most applied methods and there is a huge list of applications. Doraszelski and Jaumandreu (2013) propose a variant to measure productivity impacted by R&D or "endogenous productivity".

Only a few papers have tried however to estimate unobservable heterogeneous demand "advantages" outside very specific sectors. The main reason is probably the lack of suitable data and methods. Demand estimation since Berry (1994) and Berry, Levinsohn and Pakes (1995) has richly used the discrete choice framework to explain the product shares in specific markets. The price, together with the set of observable characteristics plus the unobservable term, fully explain the observed shares and, in this sense, the demand advantage of each product has been completely described. But this is hardly applicable when we have firms for which we only observe their sales in a broad industry, no product quantities are directly comparable, and only an approximate price index is available in the best of the cases and no price at all in many others. Part of the problem is in fact the multiproduct nature of the firms, which requires complex indices to transform the value of sales into quantity indices.[3]

A few recent works have addressed demand heterogeneity measurement in broad samples. Foster, Haltiwanger and Syverson (2008) proceed to a selection of US industries with quasi-homogeneous products and show how demand and firm's efficiency have a separate impact on the pricing, quantity sold and market participation of the firms. Jaumandreu and Mairesse (2010) specify, for a sample of Spanish firms for which price index data are available, firm-level demand and cost functions compatible with product differentiation. Process innovations affect the cost advantages and advertising and product innovations affect the relative demand advantages. They find relevant impacts. Aw, Roberts and Xu (2011) use

[3] Katayama, Lu and Tybout (2009) and De Loecker (2011) are papers that gloss the difficulty to measure meaningfully productivity in the absence of "physical" measures of ouput. De Loecker paper models then the unobserved price as the result of an inverted demand known up to an iid shock across producers and over time.

a firm-level demand relationship and cost function, both subject to idiosyncratic shocks, to assess the impact of R&D investments and exporting on the dynamics of productivity. The cost side heterogeneity is modeled as endogenous with respect to R&D activities. Roberts, Xu, Fan and Zhang (2012) have later developed a model for the exports of Chinese footwear producers that draws on the diversity of destinations, prices set and quantities sold in each destination, by each producer, to infer firm demand and cost advantages.

Pozzi and Schivardi (2012) write a model with productivity and demand unobservable heterogenous terms and compare its predictions with the results of estimates for a sample of Italian firms. They label the demand shocks "market appeal," and find that both type of shocks have different impacts on firms adjustments. Petrin and Warzynski (2012) use a product level data set for Denmark to apply the discrete choice approach to estimate unobserved demand heterogeneity, that they label as "quality". Given the product level of the data, technical efficiency must be computed from multiproduct production functions. They find both product quality and multiproduct technical efficiency to be sensitive to R&D expenditures. So, the results of all these works stress the importance of demand heterogeneity and the endogenous character of the advantages.

This paper develops a method to estimate the idiosyncratic demand and cost advantages of firms, both in their domestic and export markets. The method will be applied to characterize the spectacular recent growth of China exports and simultaneous domestic industry re-structuring, with a firm-level panel data from 1998 to 2008 that includes more than 400,000 manufacturing firms. Using a structural model, we separate for each firm the idiosyncratic advantage that drives the evolution over time of the sales relative to the sales of its competitors once controlled all available observable factors. Similarly, we estimate the relative productivity that drives the evolution over time of the ability of the firm to produce at a different cost than its competitors.

We have no quantities and only rough approximations to relevant prices. We start by stressing the fundamental identification difficulty in this context. But we show that, at least in one case, non-restricted Markov processes for the cost and demand advantages can be separately identified without using prices. For this identification result we need firms that sell both in the domestic and export markets an identical product (that they hence produce at a unique marginal cost). This is the setting that Aw, Roberts and Xu (2011) employ. Intuitively, despite that the cost advantage is the same it produces a different revenue in each market because of the different demand elasticities. The estimations should then allow to recover the whole conditional distribution of cost and demand advantages for firms that participate in both markets.

Of course firms that participate in both markets are likely to be autoselected into this situation, so to make the estimates unconditional and compare them to the estimates for other firms we need to apply a standard correction procedure for selection. For the firms that only serve one market (domestic or exports) we are working on the development of other forms of possible identification. We have some individual price indices for a subset of years, so we can in any case compare the results with identification obtained by using these individual prices.

Firstly we are going to focus on "exogenous" advantages, that evolve with Markovian processes in which the current advantage depends only on its previous level and random uncorrelated shocks. Then, we want to explore the model of "endogenous advantages," that allows the firms' investments and public policy to impact the expected value of the advantages. The resulting distributions of "unobservable" advantages should allow for a streamlined description of the factors driving exports and industry structure over time, both intraindustries and across industries.

Firms are heterogeneous, so we expect the change in the distributions over time to point to the reasons behind the expansion or contraction of a particular industry. For example, if the increasing importance of machinery and electronic exports has been lead by the development of demand advantages (may be by "processing exports" firms, say), we expect two things: an increase over time of the demand advantages of some leading firms over the rest and the increase of the average demand advantages in these industries against the more traditional export industries. We are ready to analyze and test these changes as well as to detail them by subsets of firms (entrants, exitors and incumbents, SOE's and private firms, firms with different kinds of public support, firms with R&D activities...). Comparing formally the distributions amounts to perform tests of stochastic dominance (see, for example, Linton, Maasoumi and Whang, 2005).

China extended recent period of industry re-structuring and exports growth provides a unique context to apply the method. The data cover just a crucial part of this period, 1998-2008, and we are working to add more recent years. At least three things converge to make this period extraordinary: the spectacular growth of the exports, in particular after China adhered WTO in 2001; the deep re-structuring of the SOE's and the spread change of the forms of ownership; the intensity of liberalization and competitive pressures, sharply inducing reallocation of activity and entry and exit.

Brandt, Van Biesebroeck and Zhang (2012) is the first comprehensive study of productivity growth at the firm-level for this period, using a data base drawn from the same source than ours. They document an average growth of gross output total factor productivity of 2.8% a year with an intense entry of high productivity firms, death of less efficient firms, and output reallocations among the surviving.

Regarding exports, Tang and Zhang (2012) report, with a transactions data base covering 2000-2006, the following facts: 25% average growth per year that has converted China in the wold top exporter, an increase in the number of exporters and average exports, number of exported products and number of countries, as well as important rates of entry and exit in the export activity. There has also been an important re-structuring of trade by ownership, and a sharp increase of the so-called "processing exports," in which foreign parts and components are imported to be assembled and then exported.

Many outcomes are affected by conscious actions of firms and government. Increasing wages have resulted from competition among firms for the best available workers. Many firms have invested heavily in R&D and workers training. Economic authorities try to heavily influence industrialization through instruments as the Economic Technical Development Zones (ETDZ) and Science and Technology Industrial Parks (STIP), as well as the privileges accorded to the firms located there (exemption of duties, workforce training...) in addition to the operation of geographic spillovers.

As a result of all these facts, the subperiod provides an intense variation of exogenous factors ideal to estimate the productivity and demand advantages of firms, their time evolution, as well as to test the effectiveness of the actions of firms and government measures.

This paper is an ongoing project. Section 2 develops the basic model and some possible modifications that will have to be taken into account in taking it to the data. Section 3 describes the data. We draw on our own ongoing work on the firm-level data on Chinese manufacturing that have been used elsewhere.

2. A model to indentify cost and product advantages.

In what follows we assume, as in Aw, Roberts and Xu (2011), that firms produce a single output (have a unique marginal cost) that sell in a domestic and an export maket (we discuss later the decision to operate in only one market). Both the domestic (D) and export (X) markets are monopolistically competitive.

Let's specifically suppose that demands for firm j in the export and domestic markets are Dixit-Stiglitz type

$$\frac{Q_{jt}^X}{Q_t^X} = \left(\frac{P_{jt}^X}{P_t^X}\right)^{-\eta_X} \exp(z_{jt}^X \alpha_X + \lambda \delta_{jt})$$

and

$$\frac{Q_{jt}^D}{Q_t^D} = \left(\frac{P_{jt}^D}{P_t^D}\right)^{-\eta_D} \exp(z_{jt}^D \alpha_D + \delta_{jt}),$$

where Q_t^X, P_t^X, Q_t^D and P_t^D are industry aggregates; η_X and η_D industry-level elasticities, and $Q_{jt}^X, P_{jt}^X, Q_{jt}^D$ and P_{jt}^D the quantities and prices determined by the firm. The terms in the exponents, $z_{jt}^X \alpha_X + \lambda \delta_{jt}$ and $z_{jt}^D \alpha_D + \delta_{jt}$, measure the factors by which the product of firm j is valued differently by consumers. According to the value of these factors, consumers buy more of the product of the firm when the price is the same that the price of a rival with these terms equal to zero. Alternatively, these terms scaled by the corresponding η may be read as describing how much consumers are willing to pay for the same quantity.

These "demand advantages" of the firms have two components. The first component consists of the impact of a vector of observables z, presumably different in each market, as for example being located in an area with high density of customers and the expenditure in sales promotion currently carried out by the firm. The second component consists of an idiosyncratic unobservable variable representing the level of "attractiveness" of the product. This term picks up the advantages that we are not able to observe and measure (quality, design, adequacy to the contractor...) and is likely to present a strong persistence. Importantly, the impact of product attractiveness may be different in the export and domestic market, and we normalize the impact to unity in the domestic market.

Representing all the industry-level terms by the variables Φ_t^X and Φ_t^D (their impact can be estimated by means of dummies), we can write

$$Q_{jt}^X = \Phi_t^X (P_{jt}^X)^{-\eta_X} \exp(z_{jt}^X \alpha_X + \lambda \delta_{jt}) \tag{1}$$

and

$$Q_{jt}^D = \Phi_t^D (P_{jt}^D)^{-\eta_D} \exp(z_{jt}^D \alpha_D + \delta_{jt}).$$

It is worthy of noticing that this specification of demand would allow for the estimation of the unobserved demand advantage in quite general conditions, had we information on prices and quantities and a shifter z whose level can be considered optimally determined by the firm.[4]

In the absence of price and quantity, however, we still can obtain identification adding the production side. To see this let's assume that the firm has a Cobb-Douglas production function

$$Q_{jt} = K_{jt}^{\beta_K} L_{jt}^{\beta_L} M_{jt}^{\beta_M} \exp(\omega_{jt})$$

where ω represents productivity. Assuming optimal choice of L_{jt} (alternatively we could

[4] Take as example the domestic demand with a unique z^D. Suppose that the firm sets optimally P_{jt} and z_{jt}^D incuring a cost of one money unit per unit of z. A Dorfman-Steiner type of condition for z gives

$$z_{jt}^D = \frac{1}{\alpha_D}(\ln \frac{\eta_D}{\alpha_D} + (\eta_D - 1) p_{jt}^D - \delta_{jt}),$$

where p_{jt}^D stands for the log of price. Assume that δ follows a first order Markov process $\delta_{jt} = g(\delta_{jt-1}) + \varepsilon_{jt}$. The condition for z can be written for moment $t-1$, solved for δ_{jt-1}, and the obtained expression used to substitute for δ_{jt-1} in the demand written in terms of the Markov process. This would raise an Olley and Pakes (1996) type of estimation in the version proposed by Doraszelski and Jaumandreu (2013).

use M_{jt}), it can be shown that marginal cost may be written as

$$MC_{jt} = \exp(\beta_{MC}) K_{jt}^{-\beta_K} W_{jt}^{(1-\beta_M)} P_{Mjt}^{\beta_M} L_{jt}^{(1-\nu)} \exp(-\omega_{jt}) = \overline{MC}_{jt} \exp(-\omega_{jt}),$$

where $\overline{MC}_{jt}$ represents the observable part (up to a set of parameters) of marginal cost.

Let's assume that the unobservables follow first order exogenous Markov processes: $\omega_{jt} = h(\omega_{jt-1}) + \xi_{jt}$ and $\delta_{jt} = g(\delta_{jt-1}) + \varepsilon_{jt}$, and let's now combine the demand and production sides.

First order conditions of profit maximization with a common production marginal cost are

$$\begin{aligned} P_{jt}^X (1 - \frac{1}{\eta_X}) &= \overline{MC}_{jt} \exp(-\omega_{jt}) \\ P_{jt}^D (1 - \frac{1}{\eta_D}) &= \overline{MC}_{jt} \exp(-\omega_{jt}). \end{aligned} \tag{2}$$

The easiest way to consider these equations is as part of a broader system that includes the first order conditions for the choice of any variable factor, labor and materials say, and hence determining simultaneously the marginal cost for a given capital (that we can assume, as is usual, previously chosen). Notice that this implies that both price and quantity will be different in each market but also correlated with both ω and δ.

Write (1) in terms of deflated revenues $R_{jt}^X = \frac{P_{jt}^X Q_{jt}^X}{P_t^X}$ and $R_{jt}^D = \frac{P_{jt}^D Q_{jt}^D}{P_t^D}$. Replace prices of the right hand side by their optimal choice according to (2). Taking logs (that we will represent by lowercase letters) we have

$$\begin{aligned} r_{jt}^X &= \varphi_t^X - (\eta_X - 1)\overline{mc}_{jt} + z_{jt}^X \alpha_X + (\eta_X - 1)\omega_{jt} + \lambda \delta_{jt} \\ r_{jt}^D &= \varphi_t^D - (\eta_D - 1)\overline{mc}_{jt} + z_{jt}^D \alpha_D + (\eta_D - 1)\omega_{jt} + \delta_{jt}, \end{aligned} \tag{3}$$

where $\overline{mc}_{jt} = \beta_{MC} - \beta_K k_{jt} + (1 - \beta_M) w_{jt} + \beta_M p_{Mjt} + (1 - \nu) l_{jt}$. These equations show how revenues in both markets depend on the observed part of marginal cost, the observed demand advantages, the unobserved cost advantage of the firm and the unobserved demand advantage.

We are going to show that identification is possible if the elasticities that the firm faces in each market are different. Intuitively, there is identification just because each advantage has a different impact in each market. Assume from now on, with the only purpose to simplify notation, that there are not observed demand advantages z^X and z^D. Equations (3), inverted, give a system for ω_{jt} and δ_{jt}. This system of equations can be solved to get the unobservables in terms of observables. The solution gives

$$\omega_{jt} = \gamma_t^X + \frac{1}{d} r_{jt}^X - \frac{\lambda}{d} r_{jt}^D + \overline{mc}_{jt} \tag{4}$$

and

$$\delta_{jt} = \gamma_t^D + \frac{\eta_X - 1}{d} r_{jt}^D - \frac{\eta_D - 1}{d} r_{jt}^X$$

where $d = (\eta_X - 1) - \lambda(\eta_D - 1)$, $\gamma_t^X = \frac{\varphi_t^D - \varphi_t^X}{d}$ and $\gamma_t^D = \frac{(\eta_X - 1)\varphi_t^D - (\eta_D - 1)\varphi_t^X}{d}$. Notice that the coefficients on the observed variables identify η_X, η_D and λ.

The assumption that ω_{jt} and δ_{jt} follow Markov processes allows us to set an estimable system of equations by replacing ω_{jt} and δ_{jt} in (3) by the Markov processes. Substituting expressions (4), once lagged, for ω_{jt-1} and δ_{jt-1}, we have

$$\begin{aligned}
r_{jt}^X &= \varphi_t^X - (\eta_X - 1)\overline{mc}_{jt} + (\eta_X - 1) h(\gamma_{t-1}^X + \frac{1}{d} r_{jt-1}^X - \frac{\lambda}{d} r_{jt-1}^D + \overline{mc}_{jt-1}) + (\eta_X - 1)\xi_{jt} \\
&\quad + \lambda g(\gamma_{t-1}^D + \frac{\eta_X - 1}{d} r_{jt-1}^D - \frac{\eta_D - 1}{d} r_{jt-1}^X) + \lambda \varepsilon_{jt} \\
r_{jt}^D &= \varphi_t^D - (\eta_D - 1)\overline{mc}_{jt} + (\eta_D - 1) h(\gamma_{t-1}^X + \frac{1}{d} r_{jt-1}^X - \frac{\lambda}{d} r_{jt-1}^D + \overline{mc}_{jt-1}) + (\eta_D - 1)\xi_{jt} \\
&\quad + g(\gamma_{t-1}^D + \frac{\eta_X - 1}{d} r_{jt-1}^D - \frac{\eta_D - 1}{d} r_{jt-1}^X) + \varepsilon_{jt}.
\end{aligned}$$

where $\overline{mc}_{jt}$ has to be replaced by its corresponding expression.

In this system the variable factors l_{jt} and m_{jt} are endogenous: their level has been chosen according to the knowledge of ω_{jt} and δ_{jt} and hence they are correlated with ξ_{jt} and ε_{jt}. Their lags l_{jt-1} and m_{jt-1}, however, are correlated with ω_{jt} and δ_{jt} but uncorrelated with ξ_{jt} and ε_{jt} because these shocks were not known when l_{jt-1} and m_{jt-1} were chosen. For similar reasons r_{jt-1}^X and r_{jt-1}^D are uncorrelated with ξ_{jt} and ε_{jt} ($q_{t-1}^X, p_{t-1}^X, q_{t-1}^D$ and p_{t-1}^D were chosen before knowing ξ_{jt} and ε_{jt}). As in the marginal cost expression we only need one variable input, identification amounts to find an instrument for the included variable l_{jt} or m_{jt}. We can use as instrument the lag of the non-included variable. If labor should be considered costly to adjust, a sensible arrangement can be the use of materials in the marginal cost function and lagged labor as instrument.

Estimation may be carried out by nonlinear GMM once that the unknown functions $h(\cdot)$ and $g(\cdot)$ are replaced by polynomials in its functional arguments.

Notice that estimation only needs information on the revenues in each market and the variables that determine marginal cost (that is, capital, employment, materials, wages and price of materials). In principle, the elasticities in both markets and all parameters of the production function are identified. To see why notice that in the system (4) the elasticities are identified. With the elasticities identified we can identify the parameters in $\overline{mc}$.

It is straightforward to generalize the model to consider endogenous Markov processes, where functions $h(\cdot)$ and $g(\cdot)$ are shifted by actions of the firms (e.g. workers training, R&D, reception of government support...)

An important question is the possibility of identification of the advantages for the firms that operate in only one market, especially in the domestic market. Right now we have not a clear best option, we want to explore several possibilities. Firstly, we want to check the identification power of the decision of not to participate in the foreign market. A simple way to think about this decision is as the result of finding that participation is not profitable given its fixed costs. As markets are segmented, according to our specification

$$\pi_{jt}^X = \frac{1}{\eta_X} R_{jt}^X - F_{jt}^X$$

where F_{jt}^X are the fixed costs of acceding the foreign market. We may consider that there is a latent optimal $(R_{jt}^X)^*$ that is not observed because it doesn't produce positive profitability. Secondly, the revenue equation for the domestic market produces a composite unobserved advantage $(\eta_D - 1)\omega_{jt} + \delta_{jt}$. Assuming that this composite advantage follows a first order Markov process it can be estimated and compared with the composite advantages of the firms that participate in both markets. Thirdly, it is always possible to define the relevant relationships in terms of $p_{jt}^D + \omega_{jt}$ and $-(\eta_D - 1)p_{jt}^D + \delta_{jt}$ or "value" of the advantages. If one is willing to assume a Markov process on these values, the relative importance of the advantages can be assessed.

3. Data.

The source is the Annual Census of Industrial Production, a firm-level survey conducted by the National Bureau of Statistics (NBS) of China, covering the period 1998-2008. We are right now working to complete some data for 2008 and to include the data corresponding to 2009 and 2010. This annual census includes all industrial non-state firms with more than 5 million RMB (about $600,000) in annual sales plus all industrial state-owned firms (SOEs). Our source is then the same used in Brandt, Van Biesebroeck, and Zhang (2012). A comparison with the (complete) Census of 2004 led them to conclude that the aggregates correspond extremely well and that the sample accounts for somewhat more than 90% of Chinese industrial output.

Our sample hence consists basically of larger firms and some smaller SOEs. The available information includes firm demographics such as region, industry code, the date of birth and ownership together with output, sales, number of employees, capital and materials, some non-financial details and financial statements. In order to use it as a panel, we have had to address two important related questions: the problem of discontinuity of information and the detection of economic entry and exit as different from time improvements of information.

Discontinuity of information (other than incidental and attrition) for an existing firm can happen for two reasons: a firm is non-state owned and falls below the sales threshold of RMB 5 million and, more importantly, a firm is allocated a different ID (9 digit-code). Firms occasionally receive a new ID if they are subject to some restructuring (change of name, ownership...), merger or acquisition. In the first case, if the firm re-enters the sample keeping its ID, we only get some missing observations in the time sequence of the firm. But when the firm doesn't re-enter sample we have strictly no way to distinguish its disappearance from economic shutdown.

With regards to the case of the IDs, we have done an intensive work (in the style of Brandt, Van Biesebroeck, and Zhang, 2012) to link over time the data of the firm that presumably had the ID changed. This process has used extensively information such the firm's name, corporate representative, 6-digit district code, post code, address, telephone number, industry code, year of birth, and has been implemented in several steps: first checking on neighbor years two by two, then longer panel sequences with the following/previous years.

The results seem very satisfactory. Our focus is going to be manufacturing, a sector for which considering sequences of a minimum of two years we have a total of 445,397 firms and 2,340,676 firm/years. So, after our linking, firms stay in the sample by an average of 5.3 years. We have time sequences of 5 or more years for more than half of the firms and, on average, more than 80% of these sequences have information all years.

The work done allows us to assess economic entry and exit. As the information includes the age of birth of the firm, we can distinguish economic start ups or entry from inclusion in the sample due to information improvement. Table 1 reports in column (1) the manufacturing sample size year to tear. Column (2) shows that the sample size grows over time except in 2008 (data can still change), 34% in the Census year of 2004, and at an average pace of near 8% the rest of years. In column (3) we report Entry defined for each year as firms included for the first time in the sample that are born the same year or any of the two previous years. Column (4) reports Exit as the proportion of firms included in the previous year that disappear from the sample. We can consider this a quite reliable estimate of economic shut downs although, as mentioned before, it can include some firms in a process of drastic downsizing and some attrition.

Sample growth should equal approximately the rate of net entry (difference between the rates of entry and exit as defined) plus the rate of growth because sample information improvements. So, according to the table, net entry went from negative to positive around 2003 and information has improved all years, 30% in the Census year 2004. The degree of response of the sample firms considered year to year tends to be higher than 95%.

Table 2 also provides some basic statistical information on key variables included in the panel. They are unweighted averages ignoring all firm/year observations for which there are some abnormal values affecting the years t and/or t-1. Columns (5) to (7), show a doubling over the period of the real output per firm, stability of the real capital and an important fall in the average number of workers. Columns (8) to (9), show a intense average growth of output, closely followed by capital, and a positive growth for employment after 2001.

Taken together, these average levels and rates of growth of capital and employment imply a strong reallocation of capital and employment from the biggest to the small firms. Column (11) shows the growth of TFP, especially intense after 2001. The numbers are compatible with Brandt, Van Biesebroeck, and Zhang (2012) estimates. In Table 2, column (1), we also report TFP computed by industries.

Let's focus on some facts relevant for our exercise. Table 2 reports by industries, in columns (2), (3) and (4), the average proportion of firms that operate in both markets, the foreign market only and the domestic market only, respectively. The proportion of firms that both export and sell home ranges from 10% (Paper) to 32% (Apparel) and, as reported in column (6), has increased over time in all sectors but Apparel, with very important shifts in Transportation equipment and Electronics. The absolute number of these firms is reported for 2008 in column (5), and the proportion which account of industry sales in column (7). This latest number is never less than 20% and is more than 40% in 6 sectors out of 10. The export intensity of these firms, or proportion of sales which go to the foreign markets, can be seen to range from 30% to 60%in column (8). This set of facts make this sample highly interesting to identify cost and demand advantages without the need to resort to price information.

Appendix

Main Variables Construction:

Output

Firms report the value of their output in both nominal and real prices from the 1998-2003, using a set of "reference" prices provided by NBS. Collecting price indices across industries and provinces from several statistical sources we construct deflators for the years 2004-2008.

Capital stock

Capital is defined as the real value of fixed assets at the end of the year. We construct this real capital series by estimating the net nominal value and net investment at the firm level, and deflating them by the price index for capital and investment goods. Price deflators for investment goods are taken from the China Fixed Asset Investment Year Book (various years) published by the NBS. Each firm reports information on the value of their fixed capital stock at original purchase prices, and their capital stock at original purchase prices less accumulated depreciation. As these book values are the sum of nominal values for different years, we convert them into real values that are comparable across time and across firms by using Brandt, Biesebroeck, and Zhang (2012) method.

Intermediate input (Materials)

According to the definition of the NBS, intermediate input must equal the sum of the value of total output less value added and the net value-added tax. We deflate this value by intermediate input deflators. Following Brandt, Biesebroeck, and Zhang (2012), we calculate input deflators using the output deflators and information from the National Input-Output table.

Employment and wage payments

Labor input is measured by total number of employees, which includes all the full-time production and nonproduction workers recorded for a firm, excluding part-time workers and casual workers. Firms report total annual employment and several components of employee compensation. The latter include wages, employee supplementary benefits, and unemployment insurance. Firms began to report retirement and health insurance in 2003 and housing benefits in 2004. Therefore, our wage compensation includes: (1) payable wage; (2) unemployment insurance premium; (3) pension and medical insurance premium; (4) housing mutual fund; (5) total welfare fees. We deflate wage compensation with consumer price indices by region.

Export revenues.

Firms report the yearly value of industrial export sales at current prices.

References

Ackerberg, D., K. Caves and G. Frazer (2006), "Structural Identification of Production Functions," Working paper, UCLA.

Aw, B.Y., M. Roberts and D.Y. Xu (2011), "R&D Investment, Exporting and Productivity dynamics," *American Economic Review*, 101, 1312-1344.

Berry, S. (1994), "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242-262.

Berry, S., J. Levinsohn and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Brandt, L., T.G. Rawski and J. Sutton (2008), "China's Industrial Development," in L. Brandt and T. Rawski (eds.), *China's Great Economic Transformation*, Cambridge University Press, 569-632.

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.

Crozet, M., K. Head and T. Mayer (2012), "Quality Sorting and Trade: Firm-level evidence for French Wine," *The Review of Economic Studies*, 79, 609-644.

De Loecker (2011), "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79, 1407-1451.

Eaton, J., S. Kortum and F. Kramarz (2011), "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 79, 1453-1499.

Doraszelski, U. and J Jaumandreu (2013), "R&D and Productivity: Estimating Endogenous Productivity," *The Review of Economic Studies,* forthcoming.

Foster, L., J. Haltiwanger and C. Syverson (2008), "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98, 394-495.

Griliches, Z. (1979), "Issues in Assessing the Contribution of R&D to Productivity Growth," *Bell Journal of Economics*, 10, 92-116

Helpman, E. and P. Krugman (1985), *Market Structure and International Trade,* MIT Press.

Helpman, E. and P. Krugman (1989), *Trade Policy and Market Structure,* MIT Press.

Jaumandreu, J. and J. Mairesse (2010), "Innovation and Welfare: Results from Joint Estimation of Production and Demand Functions," NBER Working Paper 16221.

Katayama, H., S. Lu and J. Tybaut (2009), "Firm-level Productivity Studies: Illusions and a Solution," *International journal of Industrial Organization*, 27, 403-413.

Levinsohn, J. and A. Petrin (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317-341.

Linton, O., E. Maasoumi and Y. Whang (2005), "Consistent Tests of Stochastic Dominance under General Sampling Schemes," *The Review of Economic Studies*, 72, 735-765.

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

Petrin, A., and F. Warzynski (2012), "The impact of research and Development on quality, Productivity and Welfare," Working Paper, University of Minnesota and Aarhus University.

Pozzi, A. and F. Schivardi (2012), "Demand or Productivity: What Determines Firm Growth?," Working Paper 9184, CEPR.

Roberts, M., D.Y. Xu, X. Fan and S. Zhang (2012), " A Structural Model of Demand, Cost, and Export Market Selection for Chinese Footwear Producers," NBER Working Paper 17725.

Sutton, J. (2001), "Rich Trades, Scarce Capabilities: Industrial Development Re-visited," Keynes Lecture, Proceedings of the British Academy.

Sutton, J. (2007), " Quality, Trade and the Moving Window: The globalization Process," *The Economic Journal*, 117, 469-498.

Syverson, C. (2009), "What Determines Productivity?," *Journal of Economic Literature*, 49, 326-365.

Tang, H. and Y. Zhang (2012), "Exchange Rates and the Margins of Trade: Evidence from Chinese Exporters," *CESifo Economic Studies*, 58, 671-702.

Wooldridge, J. (2009), "On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables," *Economic Letters*, 104, 112-114.

Table 1: Manufacturing sample breakdown by years and basic descriptive statistics

Years	No of firms	Sample growth	Entry	Exit	Average levels			Average growth rates			TFP growth
					Output	Capital	Employment	Output	Capital	Employment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1998	129,671		0.142								
1999	145,949	0.126	0.044	0.000	55.0	15.7	359	0.046	0.024	-0.020	0.001
2000	149,371	0.023	0.050	0.095	64.7	15.3	354	0.066	0.014	-0.009	0.018
2001	159,471	0.068	0.081	0.117	72.7	16.5	342	0.050	0.025	-0.009	0.003
2002	170,979	0.072	0.070	0.080	79.2	15.5	320	0.104	0.060	0.012	0.022
2003	184,537	0.079	0.084	0.086	92.0	15.4	313	0.136	0.093	0.032	0.025
2004	247,854	0.343	0.176	0.133	106.6	15.5	299	0.158	0.107	0.022	0.033
2005	263,681	0.064	0.069	0.049	99.4	13.2	254	0.211	0.169	0.058	0.033
2006	288,433	0.094	0.088	0.061	115.8	14.3	257	0.189	0.127	0.033	0.021
2007	315,769	0.095	0.086	0.062	129.0	14.9	246	0.213	0.126	0.040	0.033
2008	284,961			0.098	128.6	15.5	235	0.126	0.189	0.041	
1998-2008								0.146	0.114	0.027	0.024

Table 2: Productivity, types of firms, and firms with export and domestic markets.

	TFP growth average 1999-2008	Proportion of firms that: Export and sell home	Export only	Sell home only	Firms that export and sell home: Number 2008	Change prop. 1999-2008	Prop. ind. sales in 2008	Export intensity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Food, drink and tobacco	0.026	0.156	0.033	0.811	2,942	0.013	0.302	0.437
2. Textile,leather and shoes	0.024	0.319	0.172	0.509	9,729	-0.080	0.409	0.580
3. Timber and furniture	0.025	0.200	0.105	0.695	1,719	0.043	0.279	0.554
4. Paper and printing products	0.020	0.099	0.020	0.881	1,086	0.032	0.276	0.328
5. Chemical products	0.021	0.212	0.041	0.747	7,055	0.010	0.423	0.381
6. Non-metallic minerals	0.034	0.117	0.031	0.852	2,056	0.035	0.228	0.377
7. Metals and metal products	0.007	0.192	0.054	0.754	4,319	0.035	0.446	0.469
8. Machinery	0.027	0.229	0.028	0.744	6,631	0.039	0.469	0.351
9. Transport equipment	0.028	0.219	0.033	0.748	2,506	0.099	0.558	0.352
10. Electronics	0.035	0.312	0.093	0.595	7,958	0.074	0.601	0.485