Introduction of new varieties of goods in the Chinese manufacturing sector

Irene Brambilla[†]
Yale University

September 2004

Abstract

This paper studies the introduction of new varieties of goods in the Chinese manufacturing sector during the period 1998-2000. I first develop a monopolistic competition model with multiproduct firms in which firms are heterogeneous in two dimensions: the fixed cost of introducing varieties into the market and the variable cost of production. Using firm-level data for the Chinese manufacturing sector, I compare the performance of foreign and domestic firms in innovation, measured by the number of new varieties that they introduce, and, as described in the model, I estimate whether the number of new varieties can be explained by differences in the cost of innovation and variable productivity.

Once observable firm characteristics are accounted for, I find that firms with more than 50 percent of foreign ownership introduce on average twice as many more new goods as private domestic firms; fully foreign-owned firms (100 percent of foreign ownership) introduce three times more new varieties than private domestic firms. Foreign firms in selected industries have advantages in productivity and in the cost of introducing new varieties. Advantages in productivity account for approximately 40 percent of the difference in the number of varieties, while advantages in the cost of innovation account for 13 to 25 percent of the same difference.

[†]I would like to thank G. Grossman, B. Honoré, E. Tamer and participants of the May 2004 Industrial Development Conference for helpful comments, and P. Keefer for access to the data. This paper has also benefited from early discussion with J. Altonji, S. Berry, P. Goldberg, R. Pande and G. Porto. email: irene.brambilla@yale.edu

1 Introduction

This paper studies empirically the introduction of new varieties of goods in the Chinese manufacturing sector by foreign and domestic firms. The explored hypothesis is that foreign firms are technically more efficient than domestic firms and have advantages in the cost of innovation; and that, as a result, they introduce a larger number of new goods.

To motivate the empirical exercise, I develop a model of horizontal differentiation and monopolistic competition with firms that can choose to produce more than one variety. Firms are heterogeneous in two dimensions, they face different fixed costs to develop and set up production of a particular variety (cost of innovation), and they bear different marginal costs of production (technical efficiency or productivity). A firm finds that introducing a particular variety is profitable when the cost of innovation is lower than variable profits. This is more likely to happen when a firm is technically more efficient and when it faces lower costs of innovation. As a result, firms that are efficient in innovation and/or in production will introduce a larger number or varieties than their less advantaged counterparts. Additionally, the expected number of varieties introduced by one particular firm can be approximated by a Poisson distribution.

The model builds on Melitz's (2003) extension of Dixit and Stiglitz (1997) model of monopolistic competition to incorporate heterogenous firms. Melitz models heterogeneity as different unit input requirement, which is equivalent to differences in a constant marginal cost of production. I add heterogeneity in the fixed costs of innovation (that could be interpreted alternatively as a fixed cost of production) and I allow firms to choose the number of varieties that they introduce.

I then apply the arguments described in the model to the study of the number of new varieties introduced by foreign and domestic manufacturing firms in China during the period 1998-2000. The presumption is that, due to their condition as multinational corporations with headquarters and subsidiaries in other countries, foreign firms enjoy economies of scale and/or advantages gained from learning by doing in the development of goods and production and organization techniques across the countries where they have production facilities. For these reasons, they are expected to be more efficient both in innovation and production, and therefore to introduce a larger number of new varieties into the Chinese market.

Once firm characteristics such as size, market share and age are accounted for, I find that firms with more than 50 percent of foreign ownership introduce on average twice as many new goods as private domestic firms; fully foreign-owned firms (100 percent of foreign ownership) introduce three times as many new varieties as private domestic firms.

I use total factor productivity (TFP) and expenditure on Research and Development and purchases of technology from outside sources as proxies for efficiency in production and in innovation, respectively (higher expenditures on technology per new variety introduced indicate a higher cost of innovation). To compute TFP, I estimate a production function for each industry using the method developed by Olley and Pakes (1996) and derive a measure of TFP as the difference between output and the part of it that can be explained by input usage.

I propose a method to explore possible selection bias arising from the fact that foreign firms may face higher costs of entry. Entry costs impose a truncation on the distribution of observed technical efficiencies (TFP), since only firms that are able to cover the fixed cost are observed. If foreign firms do indeed need to pay more than domestic firms to enter the Chinese market, they will be technically more efficient on average by construction, even if their underlying distributions of productivities are the same. It is possible to test for different truncation points in the distribution of TFPs across foreign and domestic firm, by estimating the truncation points with the lowest observed TFP levels.

Foreign firms in particular industries - Electronic Equipment, Household Appliances, and Motor Vehicles and Vehicle Parts - have a significant cost advantage both in production (TFP) and in innovation (expenditure on innovation per new product). In these same industries, advantages in productivity account for 13 to 31 percent of the difference in the number of varieties, while advantages in the cost of innovation account for only 0 to 2 percent of that same difference.

The paper is organized as follows: Section 2 presents the model; Section 3 discusses the application and the data; Section 4 presents evidence that foreign firms introduce more new varieties of goods than domestic firms; and Section 5 explores differences in productivity and innovation costs as possible explanations for the advantage in innovation enjoyed by foreign firms.

2 Model

In this section, I develop a model of monopolistic competition with heterogeneous firms that produce more than one differentiated good. It is based on the Dixit and Stiglitz (1977) model of monopolistic competition with horizontal differentiation and Melitz's (2003) addition of heterogeneous firms. In Dixit and Stiglitz's model, firms are homogeneous and produce differentiated goods at the same constant marginal cost. In equilibrium they all charge the same price and receive the same profits. There is a fixed cost of entry and firms enter until variable profits are equal to this fixed cost, so that in equilibrium net profits are zero and there are no incentives for entry or exit. There is only one decision (entry) and one fixed cost to pin that decision down. Melitz extends this model to incorporate heterogeneous firms. Firms are homogeneous before entry and, after paying an entry fee, they learn their marginal costs of production, which are different across firms. In addition to the entry fee, there is a fixed cost to set up production. Firms that learn that their marginal cost is so high that variable profits are lower than the fixed cost of production choose to exit. Ex-ante (i.e. before entry) firms are homogeneous and they decide to enter until expected profits net of the entry fee are zero. The fixed cost of production pins down which firms stay in the market and which firms exit. There are two decisions (entry and stay-exit) and two costs to pin down both decisions (the entry fee and the fixed cost of production).

In the model that I develop in this section, firms can choose to produce more than one variety. In addition to entry, they choose how many varieties to research and which of these to introduce into the market; there is a cost associated with each of these decisions. Firms are heterogeneous both in their efficiency in production (marginal cost) and in their ability to incorporate technology (fixed cost to develop a variety), and these differences determine which firms introduce more varieties.

The firm's decision-making process follows the following sequence: (1) firms pay a sunk fee to enter the market; after entry, they learn how efficient they are in developing and in producing different products (there are two parameters that capture these two different dimensions of efficiency). (2) Firms decide how many varieties they will start investigating; this means gathering information on development costs and profits associated with each variety. There is a cost associated with the information gathering that depends on the number of varieties investigated and it is common to all firms. A firm may choose to investigate zero varieties, which is equivalent to exiting the market immediately after entry. (3) Firms decide which of the investigated varieties to develop and to introduce into the market. They pay a fixed cost for each variety that they decide to introduce; furthermore, this fixed cost is different for each variety and determines which varieties are introduced and which are not. Given the assumptions on market structure and production technology, firms make a separate decision about each variety that depends solely on whether variable profits from sales net of fixed development and production costs are larger than zero. (4) After the firms have decided which varieties to introduce, they make price decisions subject to the residual demand faced for each variety.

Below, I describe this decision sequence backwards: I start by describing the price decision given the number of introduced varieties; then, I discuss which varieties are introduced given the total number of varieties investigated; thirdly, I explain how firms decide how many varieties to investigate. There is an additional step that involves the determination of the equilibrium number of firms and the price index.

Consumers are identical and their preferences are represented by a CES utility function over a continuum of differentiated goods in $[0, n^*]$. Let q_i be the quantity demanded of variety i, and $\sigma > 1$ the elasticity of substitution across the different varieties; the utility function can be written as,

$$U = \left[\int_0^{n^*} q(i)^{\frac{\sigma-1}{\sigma}} \, di \right]^{\frac{\sigma}{\sigma-1}}. \tag{1}$$

Given the exogenous income y and prices p_i, the resulting demand function for each variety takes the form

$$q_i = p_i^{-\sigma} P^{\sigma-1} y, \tag{2}$$

where P is the Dixit and Stiglitz price index defined by

$$P = \left[\int_0^{n^*} p(i)^{1-\sigma} \, di \right]^{\frac{1}{1-\sigma}}. \tag{3}$$

On the production side, there is a continuum of firms indexed by j over the interval $[0, J]$. Each firm produces a measure $n(j)$ of differentiated varieties. The aggregate measure of different varieties n^* satisfies

$$n^* = \int_0^J n(j)\, dj. \tag{4}$$

Hereafter, I will refer to J, $n(j)$ and n^* as the "number" of firms and varieties; it should be understood that each is a continuum.

The production technology is represented by a cost function for each variety i produced by firm j that takes the form

$$C^j(q_{ij}) = F_{ij} + c_j q_{ij}. \tag{5}$$

F_{ij} is the fixed cost to firm j of developing and setting up production of variety i, while c_j is the marginal cost of production; to simplify, I assume that a firm can produce all of its varieties at the same marginal cost.

The number of varieties introduced by each firm is small enough relative to the total number of varieties in the market, so that the effect of one firm in the aggregate price index is negligible and the index is taken as given in the profit maximization problem. Strategic effects across varieties produced by the same firm are disregarded as well, for the same reason. Under this assumption, firms act as monopolists over a residual demand with constant elasticity σ for each of their varieties.

Firms maximize variable profits given by $(p_{ij} - c_j)\, q(p_{ij})$, subject to the demand function in (2). Because the marginal cost is the same for all varieties produced by a given firm, and all goods enter the utility function symmetrically, firms choose the same price for all their own varieties and earn identical variable profits from each of them. Within a firm, varieties only differ in the fixed cost, which does not have any impact on the pricing decision. To emphasize this symmetry, I denote the price and variable profits for/from each variety produced by firm j by p_j and π_j, instead of indexing them as p_{ij} and π_{ij}. Under the CES utility assumption, prices are determined by a constant mark-up over the production cost that depends on the elasticity of substitution, $p_j = \frac{\sigma}{\sigma-1} c_j$. Variable profits take the form

$$\pi_j = k {c_j}^{1-\sigma} P^{\sigma-1} y, \tag{6}$$

with $k = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma}$.

The introduction of the different varieties involves two sequential processes and decisions. First, firms decide how many potential varieties to explore, N_j. In this exploration stage they gather information about the possible varieties, namely, the costs for developing and setting up production for each of them, and their associated profits. To perform this first overview, firms incur a cost represented by a function γ that is increasing and convex in the number of explored varieties, N_j. The function satisfies $\gamma(0) = 0$, $\gamma' > 0$,

$\gamma'' > 0$. A number n_j of the explored varieties will prove to be profitable and will be furthered developed and introduced into the market at a fixed cost that differs by variety, F_{ij}. For simplicity, this fixed cost includes both the cost of research and development and the fixed cost of production. The cost of exploration, $\gamma(N_j)$, is the cost of learning the fixed costs F_{ij} for a number N_j of varieties; this is realized by taking N_j draws of F_{ij} from a normal distribution with mean $\overline{F}^j$ and unit variance. The mean differs by firm to capture the varying capabilities to incorporate technology; methodologically, it introduces correlation in the fixed costs drawn by a same firm.

The firm decision-making sequence can be solved backwards. In the second stage, firms take N_j draws of F_{ij}, which follows a distribution $N\left(\overline{F}^j, 1\right)$. A variety i is introduced if $\pi_j \geq F_{ij}$, which occurs with probability $\Phi\left(\pi_j - \overline{F}^j\right)$ (Φ denotes the cdf of the normal distribution). n_j is the number of successful draws, that is, the number of varieties for which $\pi_j \geq F_{ij}$.

For a given firm, each draw is an independent Bernoulli trial with the same probability of success, given by $\Phi\left(\pi_j - \overline{F}^j\right)$. Taking the number of trials (N_j) as given, the number of varieties introduced (n_j) follows a Binomial distribution. The expected number of varieties conditional on N_j is

$$E\left(n_j | N_j; \pi_j, \overline{F}^j\right) = N_j * \Phi\left(\pi_j - \overline{F}^j\right), \tag{7}$$

which is increasing in profitability and decreasing in development cost. If the number of trials is large enough and the probability of success is small, the distribution of n_j can be approximated by a Poisson with parameter $N_j * \Phi\left(\pi_j - \overline{F}^j\right)$. In the empirical section, I adopt a Poisson specification to explain the number of new varieties introduced by each firm.[1]

Prior to taking the draws of the fixed costs, the expected profit from a successful draw is $E\left(\pi_j - F_{ij} | \pi_j - F_{ij} \geq 0; \overline{F}^j\right)$. Firms choose N_j to maximize the expected profit from all successful draws, given by $E\left(\pi_j - F_{ij} | \pi_j - F_{ij} \geq 0; \overline{F}^j\right) * N_j * \Phi\left(\pi_j - \overline{F}^j\right)$, net of the exploration cost γ. Because the draws are normally distributed, the conditional expectation can be written in terms of the inverse Mill's ratio. The optimal number of draws solves

$$\max_{N_j \geq 0} \left[\left(\pi_j - \overline{F}^j\right) + \frac{\phi\left(\pi_j - \overline{F}^j\right)}{\Phi\left(\pi_j - \overline{F}^j\right)}\right] \Phi\left(\pi_j - \overline{F}^j\right) N_j - \gamma(N_j). \tag{8}$$

For simplicity, write the objective function as $A\left(P, c_j, \overline{F}^j\right) N_j - \gamma(N_j)$; the dependence on P and c_j comes

[1] To be able to describe the process generating the number of varieties by a Binomial or Poisson distribution, the trials need to be independent and the probability of success needs to be the same for all trials by a same firm. The assumption of $\pi_{ij} = \pi_j$ is not crucial for this result; varieties need to be identical ex-ante to taking the draws, but not necessarily ex-post (actually, if they were ex-post identical in every dimension either none or all varieties would be introduced). It is possible to relax the same variable profit assumption by introducing random costs of production and different demand functions. In this last case, goods enter the utility function asymmetrically, $U = \left[\int_0^{n^*} \beta(i)\, q(i)^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$. These richer assumptions do not add any significant insight into the development stage and come at the expense of complicating the pricing stage (in particular, the construction of the price index).

through π_j. Let $Lim_{N \to 0} \gamma'(N) = \gamma^* \geq 0$. Firms such that $A\left(P, c_j, \overline{F}^j\right) > \gamma^*$ choose to investigate a strictly positive number of varieties given by the solution to the first order conditions, $A\left(P, c_j, \overline{F}^j\right) = \gamma'(N_j)$; On the other hand, if $A\left(P, c_j, \overline{F}^j\right) \leq \gamma^*$, then $N_j = 0$, which is equivalent to exiting the market. The number of explored varieties is increasing in A, the expected profit from introducing one variety, and ultimately depends positively on the firm's profitability and negatively on the development costs.

The unconditional expectation - with respect to the number of draws - of the number of varieties for one firm is

$$E\left(n_j | \pi_j, \overline{F}^j\right) = N_j\left(P, c_j, \overline{F}^j\right) * \Phi\left(P, c_j, \overline{F}^j\right) \tag{9}$$

where $N_j\left(P, c_j, \overline{F}^j\right)$ solves (8). A firm that can obtain more profits from the varieties that it produces and that is more efficient in developing goods, chooses to investigate more varieties than a less advantaged firm; the probability of success for each variety is also higher. Consequently, the expected number of varieties that get introduced into the market is increasing in a firm's profits (represented here by efficiency in production) and decreasing in development costs.

In the next sections, I study the introduction of new varieties of goods by firms in the Chinese manufacturing sector and assess the roles played by advantages in productivity and innovation costs. In particular, I look at the different performances of foreign and domestic firms and examine whether there are systematic differences in productivity and in the cost of innovation between them. This exercise is not meant to be a test of the model. The model is as stylized as possible and provides basic guidelines for the empirical exercise, but does not describe some aspects of the exercise like ex-ante heterogeneity (foreign versus domestic firms) and dynamics in the introduction of varieties. In particular, I estimate the effects of productivity and innovation costs on the number of *new* varieties that firms introduce, rather than on the *total* number of varieties.

Equilibrium

In deriving the results above, the total number of firms and the price index were taken as exogenous, whereas these variables are in fact determined endogenously in equilibrium. Consider for a moment that the number of firms is fixed. The price index can be written as a function of the prices charged by each firm (which depend on the number of varieties introduced by each). The number of varieties depends in turn on the price index (through profits). These equations can be solved simultaneously for the equilibrium price index and number of varieties.

There is also the issue of how to determine the equilibrium number of firms. Firms can differ in two dimensions, the variable cost of production and the fixed costs of introducing varieties. The firm with expected profits from successful draws equal to the marginal increase in investigation costs evaluated at zero (γ^*) is the "cutoff" firm and is indifferent between exiting and investigating an infinitesimal number of varieties. Before entry, firms pay a sunk fee S and get a draw of $\left(c_j, \overline{F}^j\right)$ from a distribution common to all firms, afterwards they decide the number of varieties to investigate, N_j, the number of varieties to introduce, n_j, and the price for each of them, p_j. Firms enter until the expected profit net of the sunk fee is zero (that is, the expected profit before observing the realizations of c_j and $\overline{F}^j$, and the realizations of F_{ij} for all the varieties). The comparative statics results where more efficient firms (in terms of variable and fixed costs) introduce more varieties is left unchanged.

Consider the stage at which the firms choose the number of varieties to investigate to maximize expected profits, given by (8). The indirect expected profits function (optimized over N_j) is given by

$$A\left(P, c_j, \overline{F}^j\right) N_j\left(P, c_j, \overline{F}^j\right) - \gamma\left(P, c_j, \overline{F}^j\right). \tag{10}$$

This is the ex-post expected profit, after the realization of $\left(c_j, \overline{F}^j\right)$ and prior to the realizations of F_{ij}. In the fist step, firms choose to enter if the ex-ante expected profit, prior to the realization of $\left(c_j, \overline{F}^j\right)$, is higher than the entry fee S. Let G be the joint distribution of $\left(c_j, \overline{F}^j\right)$. The condition for free entry until expected profits are driven to zero can be written integrating over $\left(c_j, \overline{F}^j\right)$ as $F_{ij} \sim N\left(\overline{F}^j, 1\right)$

$$\int_{c_j}\int_{\overline{F}^j}\left[A\left(P, c_j, \overline{F}^j\right) N_j\left(P, c_j, \overline{F}^j\right) - \gamma\left(P, c_j, \overline{F}^j\right)\right] dG\left(c_j, \overline{F}^j\right) = S. \tag{11}$$

A definition of the price index in term of the primitives is needed to close the model. Plugging in the optimal prices for each variety, $p_i = \frac{\sigma}{\sigma-1}c_i$, into the definition of the price index in (3), the index becomes

$$P = \frac{\sigma}{\sigma-1}\left[\int_0^{n^*} c\left(i\right)^{1-\sigma} di\right]^{\frac{1}{1-\sigma}}. \tag{12}$$

To express the index in term of the model primitives, it is necessary to change variables and integrate over draws of $\left(c_j, \overline{F}^j\right)$. Consider the price decision of a firm that gets a draw c_j: all firms that get the same draw c_j charge the same price, and the density of firms that get c_j is $g_1\left(c_j\right) * J$, where J is the total number of firms that enter and g_1 is the marginal pdf of c_j. Consider the exploration decision of such firm: the expected number of varieties conditional on $\overline{F}^j$ is $N_j\left(P, c_j, \overline{F}^j\right) * \Phi\left(P, c_j, \overline{F}^j\right)$; the unconditional expectation is $\int_{\overline{F}^j} N_j\left(P, c_j, \overline{F}^j\right) * \Phi\left(P, c_j, \overline{F}^j\right) dG_2\left(\overline{F}^j\right)$, where G_2 is the marginal cdf of $\overline{F}^j$. The price index can then be written by integrating over c_j as well as

$$P^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\int_{c_j}\left(\int_{\overline{F}^j} N_j\left(P, c_j, \overline{F}^j\right)\Phi\left(P, c_j, \overline{F}^j\right) dG_2\left(\overline{F}^j\right)\right) c_j^{1-\sigma} g_1\left(c_j\right) J dc_j \tag{13}$$

or

$$P^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} J \int_{c_j}\int_{\overline{F}^j} c_j^{1-\sigma} N_j\left(c_j, P, \overline{F}^j\right)\Phi\left(c_j, P, \overline{F}^j\right) dG\left(c_j, \overline{F}^j\right) \tag{14}$$

Equations (8), (11) and (14) can be solved jointly for the equilibrium number of firms (J), the price index (P) and the rule of varieties to explore (N_j) as a function of the realization of the draws.

3 Overview and data

In the next two sections, I look at the new varieties of goods introduced by firms in 15 manufacturing industries in China during the period 1998-2000. I investigate whether foreign firms introduce more goods than domestic firms and if the difference can be attributed to the two factors described in the model: advantages in productive efficiency and/or in the cost of innovation. China is an interesting case-study because it has received a considerable amount of foreign direct investment in the last years and a large fraction of it has taken place in the form of "greenfield" FDI, as opposed to the acquisition of domestic plants by foreign firms. When a new production plant is set up from scratch, there is more room for heterogeneity between foreign and domestic firms.

The presumption is that, due to their condition as multinational corporations with headquarters and subsidiaries in other countries, foreign firms enjoy economies of scale and/or learning by doing in the development of goods and production and organization techniques across the countries where they have production facilities. In the case in which a foreign firm develops a new product to introduce in markets other than China, the marginal decision of whether to introduce it in China or not depends on setup costs and costs of production, but not on costs of R&D, which are incurred only once; if it is a joint decision across countries, still the Chinese subsidiary does not need to bear the full cost by itself. A multinational firm may also choose to develop some goods in China; it is likely that these goods are relatively similar to goods developed by the same firm elsewhere and that they can take advantage of economies of scope and learning-by-doing effects in R&D. Similar arguments apply to technology to introduce new products that is obtained from outside sources (by purchasing a license): if the license is valid to produce in many countries, then the license cost faced by the Chinese subsidiary is only a fraction of the total cost; if a license is needed to produce in each country or if goods produced in each country are not the same and each one requires the purchase of a different license, it is anyway reasonable to argue that a large multinational corporation acquiring many licenses enjoys better prices than a smaller firm purchasing only a few licenses for the Chinese market.

On the production side, if the same good has been already introduced in other markets, there is some experience about how to setup and carry out production; and, independently of producing the same goods in China and elsewhere, corporations develop experience even about how to organize production chains and the firm itself.

Foreign firms can also be expected to be more productive because of selection. Less productive firms may find it profitable to produce in the domestic market, but they may not "make it" into foreign markets. Melitz (2003) formalizes this idea for exporting firms versus firms that only sell in the domestic market by including different fixed costs of entry. Helpman, Melitz and Yeaple (2004) extend the argument to include firms investing abroad. Fixed costs determine a cutoff below which firms are not profitable. If the fixed cost is larger for foreign firms than for domestic firms, the cutoff is higher for foreign firms and they result to be on average more productive even if the distribution of productivities is the same across both types of firms.

The hypothesis of economies of scale in R&D and learning by doing in R&D and production cannot be tested directly in the present context. What I do instead is to look at the number of new varieties introduced by firms over a three-year period, and at differences in productivity and differences in the costs of innovation using estimates of total factor productivity (TFP) and expenditure in R&D and purchases of technology from outside sources as proxies. Afterwards, I look at whether TFP and costs of innovation can explain the difference in the number of varieties introduced by each firm. I also look for evidence of selection based on different productivity cutoffs.

To compute TFP, I estimate a production function for each industry using the method developed by Olley and Pakes (1996) and compute TFP as the difference between output and the part of it that can be explained by input usage. This method has been used extensively in the trade and productivity literature. Among many examples, Pavcnik (2000) and Fernandes (2003) estimate changes in productivity at the firm level due to major trade liberalizations in Chile and Colombia, respectively; Bernard and Jensen (2004) study the causal relation between changes in productivity and changes in exporting status of U.S. firms; and Javorcik (2004) and Keller and Yeaple (2003), focus on whether there are productivity spillovers from foreign firms to domestic firms in Lithuania and the U.S., respectively.

I use firm-level data from the World Bank's 2001 Investment Climate Survey. This survey was run in collaboration with the Chinese National Bureau of Statistics and is part of the World Bank's larger project to study the business environment at the firm-level in Africa, Latin America, and South and East Asia. A total of 1,500 firms were interviewed in 2001 in five Chinese cities - Beijing, Tianjin, Shanghai, Guangzhou and Chengdu - by members of the Enterprise Survey Organization, of the Chinese National Bureau of Statistics. The surveyed unit is the main production facility of a firm. The General Managers of the main facilities responded to a questionnaire on ownership structure, relations with competitors, clients and suppliers, innovation, and market environment and investment climate. A second questionnaire about accounting data and characteristics of the labor force was answered by the Accountant or the Personnel Manager.

One thousand of these firms correspond to 27 different 3-digit and 4-digit level industries in the Manufacturing sector, while the other 500 correspond to Services. The original 27 industries were selected non-randomly with the purpose of focusing on the main sectors of the industry and on those with high growth and innovation rates. They can be categorized into 5 big groups: Apparel and Textiles, Household Appliances, Vehicles and Vehicle Parts, Electronic Equipment, and Electronic Components. Two hundred firms were surveyed in each of these groups. Within these groups, firms were chosen randomly and their composition is therefore representative of the population. I work with the 1,000 firms in the Manufacturing sector and I regroup the 27 industries into 15; otherwise, there are some industries with too few observations and it is not possible to estimate production functions consistently. A list of the final 15 industries and their industrial codes can be found in Table 1.

The data span the 1998-2000 period, however, firms were interviewed only once, in 2001. As a result, some questions are answered at the year level, while other answers involve information for the entire 3-year period in question. The accounting information on sales and input usage is yearly and most firms include the information for all three years; therefore, for the purpose of estimating a production function, the data are equivalent to a 3-year panel with no entry and exit of firms.

The questions on the introduction of new varieties of goods are answered for the entire 3-year period, that is, there is information on how many new varieties firms have introduced from the beginning of 1998 to the end of 2000, but not on how many per year. A new variety is not the same as a new production line; a new variety is defined as a good classified by the firm as different from all other previously existing goods that the same firm produces. To minimize problems of subjectivity, when a new good is similar to an old good that is being replaced, it is considered a new variety only if the price difference with respect to the old good is greater than 10 percent.[2]

Table 2 provides some descriptive statistics. After discarding plants that were established after 1998, those with missing accounting information and some outliers for which the reported number of new varieties is more than one hundred,[3] the sample includes 859 firms - 589 domestic firms and 270 foreign firms. Within domestic firms, 47 percent are privately owned, 29 percent are state-owned and 23 percent are cooperatives or collectively owned by workers. I consider a firm to be foreign when there is some degree of foreign participation in its capital. Of the 270 firms with some degree of foreign ownership, 70 percent are majority foreign-owned firms (15 percent are fully foreign-owned and 55 percent have foreign participation between 50 and 99 percent) while the remaining 30 percent have less than 50 percent of foreign ownership.

Columns 2 to 4 display the median number of workers in 1998, the average output per worker in 1998 and the average number of new varieties introduced during the period 1998 - 2000, each by ownership type. The overall median number of workers is 269, and there are substantial differences across ownership types. State-owned firms are the largest (in terms of workers) and private domestic and fully foreign-owned firms are the smallest. The performance of foreign firms in terms of output per worker is better than the performance of domestic firms. Within foreign firms, those with more than 50 percent of foreign ownership outperform firms with less than 50 percent of foreign ownership; within domestic firms, private ones do better than state-owned firms and cooperatives. Firms introduce on average 3 new varieties of goods; foreign firms are

[2] The definition of what is considered a new variety is to some degree subjective and different firms possibly follow different criteria. As long as the definitions of new varieties are uncorrelated to the ownership structure of the firms, the differences in criteria do not introduce a bias.

[3] Discarded observations include both foreign and domestic firms and do not follow a particular pattern. More than 97 percent of firms report introducing less than 20 new goods, and only 10 firms report introducing more than a hundred new goods. I do not include the latter to avoid these outlyers to drive results.

above this average (3.6 new varieties for firms with less than 50 percent of foreign ownership and 4.5 and 4.8 new varieties for the other foreign firms), while domestic firms are below this average (private firms introduce 2.7 new varieties on average, state-owned firms introduce 2.8 varieties and cooperatives 1.1 varieties).

4 Foreign firms introduce more goods

I assume that the number of new varieties introduced by firm i, n_i, follows a Poisson distribution with parameter λ_i. Under this assumption, the conditional probability that firm i introduces n_i new varieties is $P(n_i|\lambda_i) = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}$, and both the mean and the variance of the number of new varieties is given by λ_i. The parameter λ_i is specific to each firm and it is a deterministic function of observable firm characteristics, x, and the ownership structure of the firm, represented by the dummy variables FOR_1, FOR_2 and $NONPRIV$.

In particular, I adopt a log-linear specification, so that

$$\lambda_i = \exp\left(x_i'\beta_0 + \beta_1 FOR_{1i} + \beta_2 FOR_{2i} + \beta_3 NONPRIV_i\right) \tag{15}$$

Observable firm characteristics included in x are size - measured by the logarithm of the number of workers,- self-reported market share, the logarithm of the age of the firm and dummy variables for firms that started operating in 1997 and 1998. FOR_1 and FOR_2 are indicator variables for firms with some degree of foreign ownership; firms with less than 50 percent of foreign participation in capital are included in FOR_1 (hereafter, minority foreign-owned firms), while FOR_2 comprises firms with foreign capital ranging from 50 to 100 percent (majority foreign-owned firms). $NONPRIV$ is equal to one for non-private domestic firms, that is, state-owned firms and collectively-owned firms. Private domestic firms are the baseline category. I also include city and industry effects that capture differences in the mean number of new varieties introduced across these two dimensions. In addition, these fixed effects control for potential selection of foreign firms into cities or industries where more varieties are introduced.

Different forms of the Poisson distribution are usually adopted when the dependent variable is a non-negative integer. In the innovation literature, many authors model the number of patent applications as a Poisson process: Hausman, Hall and Griliches (1984) and Crepon and Duguet (1997) study the effect of expenditure on R&D on patents; Blundell, Griffith and Van Reenen (1995) investigate the relation between patents and market power and "stock of knowledge." Other distributional assumptions are explored and results prove to be robust.

The log-linear specification is convenient because it guarantees that the expected number of new varieties is positive. Compared to least squares, the Poisson specification handles count data more naturally. If OLS is used (where the expected number of varieties is linear in the explanatory variables instead of log-linear) the estimated expected number of new varieties is not necessarily non-negative. This problem does not arise when using non-linear least squares with a log-linear specification, however, there is the problem of how to treat the observations in which the number of new varieties is zero. The Poisson distribution models the probability of observing each non-negative integer, including zero, and therefore by-passes this problem.

The first column of Table 3 displays the results of the Poisson regression (15) of the number of new varieties on firm ownership. I find that firms with more than 50 percent of foreign participation introduce more new varieties of goods than private domestic firms, and that this difference is statistically significant. Firms with foreign capital lying between 1 percent and 50 percent, on the other hand, do not introduce more varieties than private domestic firms; as a matter of fact, the coefficient for this group of firms is negative, although not significant.

Because the regression function is not linear, the interpretation of the coefficients $\boldsymbol{\beta}$ is not straightforward. To provide a more intuitive interpretation of the results, the lower panel of Table 3 shows the incidence ratios of the explanatory variables, $\exp(\beta_k \Delta x_k)$. For an indicator variable, the incidence ratio is the ratio of the expected number of new varieties introduced by firms that belong to that particular category and by firms in the baseline category (private domestic firms). For example, the ratio of the expected number of varieties between majority foreign-owned firms and domestic firms is $\frac{\exp\left(x_i'\beta_0+\beta_2\right)}{\exp\left(x_i'\beta_0\right)} = \exp(\beta_2) = 2.02$, which means that these firms introduce roughly *twice* as many new varieties as private domestic firms - other things equal. The incidence ratios for minority foreign-owned firms and for non-private domestic firms are lower than one, implying that the underlying coefficients are negative; the number of new varieties introduced by non-private firms is 64 percent of the number of varieties introduced by private domestic firms. For continuous regressors, the incidence ratio is computed for a change of 10 percent. For example, a firm with 10 percent more workers introduces $\frac{\exp\left(x_i'\beta_0+10*\beta_{workers}\right)}{\exp\left(x_i'\beta_0\right)} = \exp(10 * \beta_{workers}) = 3.6$ times more new varieties than a firm of similar characteristics. The number of new varieties is increasing in market share, the coefficient is statistically significant but small in magnitude. The age of a firm is not significant and neither are the dummy variables for firms that initiated operations in 1997 and 1998 (not shown in the table). Columns [2] and [3] present the results from similar regressions excluding city and industry effects. Results do not differ substantially from the original specification in column [1].

A priori, there may be a reverse causality problem with the market share: firms that have larger market shares have more incentives to do R&D and to introduce more goods; on the other hand, firms that introduce newer varieties gain a larger fraction of the market from its rivals. In column [4], I utilize the self-reported number of competitors as an instrument for market share. In a Poisson regression, the probability of observing each number of counts is specified, but there is no underlying distribution for the continuous error term. In a case in which there is a distribution for the error term, instruments can be introduced into the estimation by specifying an equation for the endogenous regressor and estimating the original equation plus the equation for the endogenous regressor jointly by FIML or LIML (depending on whether the last equation is structural or reduced form). In the Poisson case, however, it is not possible to specify the joint distribution of the two error terms for the reason stated above, and joint MLE is not viable.

I use a 2 step *control function* approach as in Blundell and Powell (2003 and 2004) instead of FIML or LIML. In the first step, I run an OLS regression of the market share on the number of competitors (the instrument) plus the other explanatory variables in regression (15) and compute the residuals.[4] In the second step, I estimate the Poisson regression including the residual from the first step as a regressor; these residuals control for the endogeneity of the market share. It is analogous to including the inverse Mills-ratio in Heckman's 2-Step estimator to control for selection.

The first step regression is a reduced form regression of market share on all explanatory variables plus the number of competitors. The assumption is not simply that the market share and number of competitors are correlated and that the number of competitors is uncorrelated to the error term - as would be the case in the instrumental variables case - but rather that the first step regression function correctly specifies the econometric model that determines the market share - as would be the case if using LIML. In addition, the error term in the first step regression is assumed to be uncorrelated to the error term in the second step regression.

Results from the control function regression are displayed in column [4] of Table 3. The coefficient on majority foreign-owned firms is larger than before but not substantially (foreign firms introduce 2.2 as many new goods as private domestic firms, instead of 2 as many new goods as in column [1]). The coefficient on market share is not statistically significant.

In equilibrium, the number of workers is determined jointly with the number of varieties. First of all, firms that introduce more varieties need to hire more workers to increase production; to minimize this reverse causality problem, I use the number of workers in 1998, the first year of data. Second, the number of new varieties and the number of workers may depend on unobserved firm characteristics not included in the regression that affect scale. The number of workers is included as a scale variable and its coefficient should not be interpreted as the change in the number of new varieties predicted by a change in the number of workers of a particular firm; the coefficient has a reduced form interpretation that relates the number of workers and the number of varieties in equilibrium without predictive value. In column [5], I exclude the number of workers from the regression. The coefficient for majority foreign-owned firms is larger than before since these firms are on average larger than domestic firms, but the difference is not substantial.

Columns [6], [7] and [8] test the sensitivity of the results to different distributional assumptions. The Poisson assumption implies that both the mean and variance of the dependent variable are equal to the parameter λ. In many applications with count data, however, it is argued that the variance of the dependent variable can be higher than the mean. This phenomenon is referred to as *overdispersion* and is usually attributed to two factors: unobserved heterogeneity and excess zeros (more zeros than the Poisson distribution can account for).[5]

[4] More general non-parametric methods can be used instead (Blundell and Powell, 2004).

[5] Gourieroux, Monfort and Trognon (1984a, 1984b) show that, provided the true distribution belongs to the linear exponential family - as is the case of the Poisson - and provided the conditional mean is correctly specified, the estimators for the coefficients

To account for unobserved heterogeneity, I estimate a Poisson-Gamma mixture in which the Poisson parameter λ is now a stochastic function of observed characteristics x and depends also on a scalar representing unobserved characteristics, ξ, such that $\exp(\xi)$ follows a Gamma distribution; the resulting distribution for the number of varieties is Negative Binomial.[6] To correct for excess zeros, I adopt a process that is usually referred to as *Zero-Inflated Poisson* and is given by a Probit-Poisson mixture. There are two types of firms, those that do not participate in the introduction of new goods and those that *consider* introducing new goods (but may end up introducing zero new goods). A firm does not participate with probability $\Phi(z_i'\delta)$, where Φ is the Normal cdf and z are observed firm characteristics including the ownership dummies. The number of goods introduced by firms that participate is distributed Poisson with parameter λ. For more details on Poisson mixtures see Cameron and Trivedi (1998).

Column [6] corresponds to the Poisson-Gamma mixture (Negative Binomial) and column [7] corresponds to the Zero-Inflated Poisson specification. The coefficients and incidence ratios can be directly compared to the previous columns in Table 3. There is no substantial difference in the results, majority foreign-owned firms introduce over twice and slightly less than twice as many new varieties as private domestic firms in the negative binomial and zero-inflated Poisson specifications, respectively.

The last column of Table 3 displays the results from an OLS regression including the same variables as before. In this case the expectation is a linear function of the regressors and the coefficients can be interpreted directly. Majority foreign-owned firms introduce 2.6 more new varieties than private domestic firms, minority foreign-owned firms and non-private firms introduce less varieties than private firms and the coefficients are not significant. Other things equal, an increase on 10 percent in the number of workers translates into 4 more new varieties. Qualitative results are almost identical to the original Poisson specification (with the exception of the coefficient on non-private firms, which is not statistically significant in the linear case). The magnitude of the coefficient fore majority foreign-owned firms (2.6) can be compared to the difference in the expected values of new varieties between majority foreign-owned firms and private domestic firms in the Poisson specification. For a given firm, this difference is given by $\exp(x_i'\beta_0 + \beta_2) - \exp(x_i'\beta_0)$. I compute two aggregate measures, the difference evaluated at the mean level of x, and the average of the difference for each firm. The first difference is 1.88, while the second is 2.81. The OLS coefficient of 2.6 lies between these two values.

In Table 4, I explore three additional categories of foreign ownership. Column [1] shows that fully foreign-owned firms (100 percent of foreign share in ownership, labeled in the table as *Foreign 100%*) do not introduce more varieties of goods than firms with foreign ownership between 50 percent and 99 percent: they both introduce twice the number of varieties as private domestic firms (the coefficients are almost identical to the coefficient for majority foreign-owned firms in Table 3 (previous table)). In the second

are consistent even if higher order moments of the distribution are not correctly specified; standard errors need to be corrected.

[6] $\lambda_i = \exp\left(x_i'\beta_0 + \beta_1 FOR_{1i} + \beta_2 FOR_{2i} + \beta_3 NONPRIV_i + \xi_i\right)$

column I disaggregate fully foreign-owned firms into two categories: those that are owned by a foreign firm or multinational corporation (labeled in the table as *Foreign 100%)* and those who are owned by foreign individuals or foreign investors (labeled as *Foreign 100% - Investor*). The former (firms fully-owned by a foreign corporation) turn out to introduce more than 50 percent more new varieties than foreign firms with foreign ownership lying between 50 and 99 percent, and over *three* times more new varieties than private domestic firms. The latter (firms fully owned by foreign individuals or investors) do not introduce more new varieties than private domestic firms (the coefficient is actually negative, but not significant).[7]

This result indicates that what matters for the introduction of new products is not the foreign origin of capital per se, but rather that foreign firms be owned by multinationals. Fully foreign-owned firms that are owned by multinationals can easily introduce new goods in China that they have developed elsewhere; whereas fully foreign-owned firms that are owned by an investor do not operate in other countries and do not have advantages in innovation capabilities over domestic firms.

In columns [3] and [4] of Table 4, I reestimate the two previous regressions using the number of competitors as an instrument for market share (including the first stage control function in the Poisson regression as described before). In this case, the coefficient of fully foreign-owned firms is larger than the coefficient on majority foreign-owned firm up to 99 percent of foreign capital even when firms owned by multinationals and firms owned by investors are grouped together. However, when the two categories of fully foreign-owned firms are included separately, the coefficient on fully foreign-owned firms owned by investors is negative, which is consistent with column [2] and its implications about the advantage of having production facilities in other countries enjoyed by multinationals.

Summarizing these findings, controlling for size, market share, age, industry and location, majority-owned foreign firms (50 percent to 100 percent of foreign capital) and firms owned fully by a foreign corporation introduce two and three times as many varieties as domestic firms, respectively; collectively-owned domestic firms (state-owned and cooperatives), minority foreign-owned firms (1 percent to 50 percent foreign) and firms owned fully by a foreign individual or investor introduce fewer varieties than private domestic firms, but results are not significant in the last two cases.

The cutoff for the definition of minority and majority foreign-owned firms is not strict. The regression results are not dramatically different when the groups are defined taking 40 percent or 60 percent as cutoffs. I have also experimented with a continuous variable indicating the degree of foreign participation in a firm's capital obtaining a positive and significant coefficient (not shown). The dummy specification is preferred given that the relation appears to be non-linear.

[7] There are only 35 fully foreign-owned firms in the sample - 15 are owned by foreign firms and 20 by foreign investors.

5 Profit and cost

The model in Section 2 describes two dimensions of firm-level heterogeneity that explain why firms choose to introduce a different number of varieties: efficiency in production and efficiency in research and development. In this section, I investigate whether there are systematic differences in these two dimensions of heterogeneity between domestic and foreign firms and assess the impact that both dimensions have on the number of varieties introduced by a firm. I decompose the total difference in the number of varieties introduced by domestic and foreign firms into the fractions explained by differences in productive efficiency and efficiency in innovation.

To measure efficiency in production, I estimate a Cobb-Douglas production function for each of the fifteen manufacturing industries in the sample and compute total factor productivity (TFP) for each firm. TFP is a measure of Hicks-neutral productive efficiency since it enters the production function in a multiplicative manner. Throughout the estimation, all firms in a given industry are assumed to be equally efficient in the use of inputs.

I approximate efficiency in innovation by the expenditure in R&D and purchases of technology from outside sources per new variety. For foreign firms, I consider expenditure by Chinese subsidiaries only, as the intention is to capture the cost of introducing new varieties in China, not worldwide. So, if a foreign firm has developed a new variety elsewhere, the R&D cost for this variety is zero in China; this captures the returns to scale advantage of foreign firms that do not need to duplicate R&D costs. I also compare the preferred ways of getting access to new varieties across firms.

5.1 Are foreign firms more productive?

I answer this question in two steps. I first estimate total factor productivity (TFP) for each firm and later I regress the estimated TFP on firm characteristics, including the ownership structure of the firms.

In the first step, I estimate a Cobb-Douglas production function where all firms in the same industry have the same time-invariant labor and capital coefficients but differ in a Hicks-neutral efficiency parameter, or TFP. TFP is exogenous and can vary over time. The estimating equation takes the form,

$$\log Y_{ijt} = \alpha_{1j} \log L_{ijt} + \alpha_{2j} \log K_{ijt} + \omega_{ijt} + \varepsilon_{ijt} \tag{16}$$

where i, j and t index firm, industry and year, respectively; Y is output, and L and K are labor and capital. There are two Hicks-neutral technology shocks, ω and ε. The first component, ω, includes characteristics of the firm that are unobserved by the researcher but observed by the firm prior to making input decisions, such as managerial ability. It is independent across firms but serially correlated across time for a given firm. The second component, ε, is an independent and identically distributed (across firms and time) zero-mean shock, that is only observed by the firm ex-post. It includes, for example, unforeseeable waste or spoilage of inputs and loss of output. Total factor productivity is given by $\exp(\omega_{ijt} + \varepsilon_{ijt})$.

As first pointed out by Marschak and Andrews (1944) and Griliches (1957), labor and capital choices are endogenous to the firm and depend on the observed component of TFP, ω, since more efficient firms employ more inputs. Hence, in an OLS regression where $(\omega_{ijt} + \varepsilon_{ijt})$ is treated as the error term, labor and capital are correlated with the error and the resulting residuals are inconsistent estimators for TFP. Generally plausible instruments are not available in this case; the panel is not long enough to use lagged input values as instruments; in addition, there is no data on differences in input quality, ruling out the possibility of using input prices as instruments (as prices are likely to reflect differences in quality and not just differences in the cost of access to inputs). Another possible solution to the endogeneity problem is to treat ω as a time-invariant firm-level unobserved efficiency term and to estimate the labor and capital coefficients by exploiting the variation in input usage and output within a same firm (Mundlak, 1963); however, I argue below that, in the present case, productivity may vary over time because of the introduction of new varieties.

Since I cannot safely use instrumental variables or fixed effects to account for unobserved efficiency, I use the method developed by Olley and Pakes (1996). The innovation introduced by Olley and Pakes consists of using investment to control for unobserved productivity in the regression function. They develop a dynamic model in which labor and investment - the choice variables - are a function of observed productivity and capital stock - the state variables; observed productivity follows a Markov process and investment is incorporated into the capital stock in the following period. They show that the investment function is strictly increasing and can be inverted to express productivity as a function of investment and the capital stock. Regression (16) can be estimated by introducing a non-parametric function of investment and the stock of capital to control for the unobserved shock ω, yielding a consistent estimator for the labor coefficient. The coefficient on the capital stock cannot be separately identified from such the non-parametric function and its estimation requires a second step that involves a non-parametric approximation to the Markov process that determines the evolution of productivity.[8]

The Investment Climate survey includes information on inputs and output for 1998, 1999 and 2000. Since the questionnaire is answered only once, in 2000, there is no entry and exit of firms in the panel. Labor is measured as labor expenditure deflated by the average wage in the region. This measure is preferred to the plain number of workers because differences in wages across firms within the same region may capture differences in workers' skill levels that affect the level of output. Capital is the book value of the stock of capital in real terms. Output is measured as value added: the cost of intermediate inputs in real terms is subtracted from total sales, which are deflated using 2-digit level price indices that differ by region.

I estimate separate regressions for each of the 15 industries. The number of firms with non-missing and

[8] Ackerberg and Caves (2003) have pointed out that if the decision on labor and investment is taken jointly and depends on the same variables - capital and the realization of observed productivity - then the labor coefficient is not separately identified from the non-parametric function on investment and capital in the first step regression. They propose a different decision making sequence in which labor is chosen prior to the investment decision and depends on a *partial* realization of observed productivity. In practice, this timing assumption introduces an error term in the choice of labor and solves the indeterminacy of the labor coefficient in the first step.

non-zero information on investment, inputs and output for the three-year period ranges from 21 to 96 by industry. The coefficients on labor and capital are different for each industry but are assumed constant over time. In principle, different coefficients can be estimated by year and region, but a large number of firms is needed in each industry. I use fourth order polynomials to approximate the non-parametric functions in the first and second step of the Olley and Pakes' methodology, with full interactions in investment and capital in the first step. Results are not significantly different when a third order polynomial is used instead. The first-step polynomial is derived from a reduced form demand function for investment and reflect the market conditions in a particular industry, time and region. To control for changes in these market conditions, I estimate one polynomial for each industry and include year and region dummies.[9] I also estimate different polynomials in the second step to account for differences in the Markov processes for productivity across industries, but these are fixed over time.

Notice that in order to be able to include a given firm in the estimation, that firm needs to have a strictly positive level of investment; otherwise, the investment function is not strictly increasing in productivity and cannot be inverted to express productivity as a function of investment and capital in the first step. In practice, it is not unusual to find that the majority of firms report no investment for several years, which can reduce the sample size dramatically.[10] Because the Investment Climate Survey has purposively targeted high-growth sectors, a large fraction of the sample reports strictly positive investment. Between 70 and 80 percent of the firms producing apparel and leather goods, and vehicles and vehicles parts report non-zero investment; this fraction lies between 80 and 90 percent for the remaining sectors - electronic equipment, electronic components, and household appliances.

Results of production function coefficients by industry are displayed in Table 5. Industries are sorted by intensity in the use of labor. The most labor-intensive industries are the manufacturing of apparel and leather goods, where the ratio of the labor to capital coefficients, α_1/α_2, is above 2.4. Capital-intensive industries include the manufacturing of communications equipment, electron tubes, household appliances, and knitted textiles; the ratio α_1/α_2 ranges from 0.54 to 1.1 in these four cases. In all other industries - motor vehicles, vehicle parts, motorcycles, computers, audio and video, small appliances and other electronic components, α_1/α_2 is roughly between 1.4 and 2.

Productivity and ownership structure

Figure 1 shows the empirical density functions of the logarithm of the estimated TFP for private domestic firms and majority foreign-owned firms (50 percent to 100 percent), averaged over the period 1998-2000, for 12 of the 15 industries (the excluded industries are the ones with the fewer number of observations). This graphs provide a first visual inspection of the distribution of productivities of foreign and domestic firms.

[9] When a large number of firms is observed, full interaction terms between year and region dummies and all first order to fourth order terms in capital and investment can be included - which is equivalent to estimating different polynomials for each time period and region.

[10] Levinsohn and Petrin (2003) propose using intermediate inputs instead of investment to avoid this problem.

The first three graphs correspond to Apparel, Textiles and Leather; there is no evidence that foreign firms have a productivity advantage in these industries. In most remaining sectors (with the exceptions of the manufacture of electron tubes, of capacitors and resistors and of large household appliances), the pdf of majority foreign-owned firms first order stochastically dominates the pdf of private domestic firms, which implies that foreign firms are on average more productive.

As a more formal test for evidence of the existence (or absence) of productivity advantages in foreign firms, I regress estimated TFP on ownership structure and firm characteristics. The regression function takes the following form,

$$\log TFP_{ij} = x'_{ij}\delta_0 + \delta_1 FOR^1_{ij} + \delta_2 FOR^2_{ij} + \delta_3 NONPRIV_{ij} + \mu_{ij} \qquad (17)$$

The ownership structure dummies are the same as in the Poisson regression: FOR^1, FOR^2 are dummies for minority and majority foreign-owned firms, respectively, and $NONPRIV$ indicates that the firm is domestic and state or collectively owned. Other control variables are age, dummies for firms that started operating in 1998 and 1997 and industry and city effects. Since ownership variables are time invariant during the sample, there is no advantage from incorporating different time periods and the regression reduces to a cross-section of firms.

I run two separate regressions, one for industries in Textiles, Apparel and Leather (the first three industries in Figure 1) and a different regression for industries in Electronics, Machinery and Appliances. Hereafter, I refer to these two groupings of industries as Group 1 and Group 2, respectively. Table 6 displays the results from regression (17) with average TFP over the period 1998 to 2000 as the dependent variable; results are similar if TFP in 1998 is used instead. Column [1] shows the results for Group 1 (Textiles, Apparel and Leather) and column [3] the results for Group 2 (Electronics, Machinery and Appliances). Majority foreign-owned firms are on average 23 percent more productive than private domestic firms in industries in Group 2, while they have no advantage in industries in Group 1 (the coefficient is negative and not statistically significant).

The same pattern repeats when more categories of foreign ownership are included in the previous regression. In columns [2] and [4], I include two categories of fully foreign-owned firms (100 percent of foreign ownership): those owned by firms and those owned by investors. In these two regressions, the variable *Foreign 50%-100%* includes firms of up to 99 percent of foreign ownership. None of the categories of foreign firms have a productivity advantage in firms in Group 1 (coefficients are not significant - see column 2). Whereas in Group 2, firms with 50 to 99 percent of foreign capital are 22 percent more productive than private domestic firms; and fully foreign-owned firms that are owned by multinationals are 31 percent more productive than private domestic firms (column 4). On the other hand, fully foreign-owned firms that are owned by investors are not significantly more productive. This last finding supports the hypothesis of economies of scale and learning by doing, since firms owned by multinationals can take advantage of the experience that they have developed in other countries, while firms owned by investors cannot.

Minority foreign-owned firms (1 to 50 percent of foreign capital) are not more productive than domestic firms in either of the groups and specifications, furthermore, they are significantly less productive than domestic firms in Textiles, apparel and leather (30 percent less productive - columns 1 and 2). Non-private domestic firms are less productive than private domestic firms in Group 2, but the difference is not significant. In Group 1, newer firms are more productive. Regarding city effects, there is no differences in location for firms in Group 1; while in Group 2, firms located in Shanghai are the most productive, Beijing and Guangzhou follow, and firms in Tianjin and Chengdu are the least productive (not shown in table).[11]

Selection

Possible selection of foreign firms to most productive industries and cities is controlled for with industry and city effects. There is still a potential selection problem if firms of different origin face systematic differences in the costs of entry, as in Melitz (2003) and Helpman, Melitz and Yeaple (2004).[12] Suppose all foreign firms need to pay the same fixed cost of entry which is higher than the cost of entry paid by domestic firms. A firm finds it profitable to enter and/or to stay in the market (depending on when the uncertainty about profits unravels) if variable profits can cover the fixed costs, which translates into a cutoff for variable profits. If differences in variable profits depend solely on differences in TFP, the profitability condition can be expressed as a cutoff in the level of TFP necessary to find it profitable to stay in the market: firms with TFP above the cutoff can cover the fixed costs of entry while firms below the cutoff cannot and exit the market. Since the fixed cost of entry is higher for foreign firms than for domestic firms, the required TFP level is also higher for foreign firms. Thus, the observed TFP levels are on average higher for foreign firms even in the case that the distribution of TFP is the same for both types of firms.

I address this question by estimating truncations in the distribution of TFP for foreign and domestic firms in each industry. Consider the distribution of TFP levels within a particular firm type (for example foreign) in one industry, and assume there is a cutoff, so that observed firms have TFP levels which are all above the cutoff (firms below the cutoff exit and are not captured in the data). As the number of firms goes to infinity, the lowest realized TFP converges to the cutoff level of TFP. Following this argument, I estimate the cutoffs for majority foreign-owned firms and private domestic firms in each industry by choosing the firms with the lowest estimates for TFP in each of the two ownership categories and industry. Let η_{Fj} and η_{Dj} be the estimators of the TFP cutoffs for foreign and domestic firms in industry j. They are defined by

$$\begin{aligned} \eta_{Fj} &= \min_i \{TFP_{ij} \mid i \text{ is a foreign firm in industry } j\} \\ \eta_{Dj} &= \min_i \{TFP_{ij} \mid i \text{ is a domestic firm in industry } j\}. \end{aligned} \tag{18}$$

[11] Chengdu is an interior city, historically less developed than the other four cities in the sample.

[12] Even another selection problem arises in the case of mergers and acquisitions, since foreign firms may be inclined to choose the most productive domestic firms.

Table 7 displays the difference in cutoffs $(\eta_{Fj} - \eta_{Dj})$ for each industries. The difference is positive and significant only for two industries: Communications equipment and Motor vehicles.

Ten percent confidence intervals are constructed by bootstrapping firms and recomputing the cutoffs for each bootstrap sample. I take a thousand bootstrap samples of firms. For each sample, I reestimate the production function and measured TFP, and I find $\left(\eta_{Fj} - \eta_{D_j}\right)$ - they do not necessarily correspond to the same firm in each sample. With these results, I construct the 10 percent confidence interval for the difference in minimum measured TFP's by computing the 5 and 95 percentiles.[13]

Note that this is not strictly a test for the existence of cutoffs in TFP. When the difference is statistically significant, it means that the lower bounds of the supports of the distribution of TFP levels of firms that stay in the market are different for foreign and domestic firms; this may be due to the decision of firms to exit the market (cutoffs) or just to the fact that the supports of the full distributions (without the truncation resulting from selection) do not coincide. On the contrary, the *absence* of a statistically significant difference, is evidence that unequivocally contradicts the existence of cutoffs.

In the case of cutoffs, the selection bias arises because domestic firms with TFP levels between η_{Fj} and η_{Dj} stay in the market and are included in the sample in the estimation of (17), while foreign firms with TFP in this range are not observed. In Table 6, column [5], I reestimate the regression of TFP on ownership of column [3] excluding the domestic firms with TFP between the two cutoff levels in the two industries with significant cutoffs (Communications equipment and Motor vehicles). That is, I take a uniform truncation point, given by η_{Fj} and include only firms that are above η_{Fj}, no matter their origin. In this case, majority foreign-owned firms are 20 percent more productive than their private domestic counterparts. As noted above, this method would not be correct if the difference in η_{Fj} and η_{Dj} is due to differences in the lower bounds of the support rather than on cutoffs. The regression is just a robustness check: majority foreign-owned firms are substantially more productive than domestic firms, even when the less productive domestic firms are not included in the comparison.

Productivity versus profitability

Theoretical production functions explain quantities of output with quantities of inputs. However, when production functions are estimated, quantities of output and some inputs are replaced with values. The basic reason to use values instead of quantities is that even at a very detailed industrial disaggregation, products are heterogeneous and quantities cannot be directly compared, using values is arguably a good way to control for differences in quality. It can be argued that there is also a pure practical reason to use values: that firm-level data usually includes only values and not firm-level prices. This is actually a similar reason as

[13] The result is suggestive of the absence of differences in productivity cutoffs for foreign and domestic firms - except for firms that manufacture communication equipment and motor vehicles. The formal validity of the procedure depends on the validity of the bootstrapped standard errors, which is not guaranteed because the procedure involves the estimation of a boundary parameter.

the first: a particular firm usually produces more than one product and there is no such thing as a firm-level price, values are needed even to compare or aggregate goods produced by the same firm.

Klette and Griliches (1996) have pointed out that the estimators from a production function regressions using values of output are inconsistent. The problem comes from the fact that the value of output includes both prices and quantities. While quantities can be directly linked to inputs through the production function, prices are the result of the interaction of supply and demand. Therefore, price times quantity is not reflecting just the production side, it also includes demand and market structure since prices are an equilibrium outcome that depend on these forces in addition to technology. Klette and Griliches focus on estimators of returns to scale, but the same argument can be applied to the estimators of TFP. The resulting estimates of TFP can be considered efficiency in generating value of output rather than just efficiency in production. To address this problem, it is necessary to include information about the demand and market structure into the estimation.[14]

To sum up, it is likely that measured TFP is capturing profitability in a broader sense rather than strict technical efficiency. This issue is not particularly important in the context of the present application. In the model of section (2) , I show how the introduction of new varieties depends on profits. For expositional simplicity, profits differ by firm only because of differences in productive efficiency, but the model could be easily adapted to include demand parameters that vary per product. In this latter case, profits depend on both productive efficiency and the demand side (although the CES assumption does not allow for different demand elasticities and mark-ups are the same for all varieties even when they enter the utility function asymmetrically). Consequently, capturing a firm's profitability instead of productive efficiency would not bias the results in the present case, quite the contrary, it may be desirable to do so and give the correct interpretation of profitability rather than productivity to the TFP estimates.

One consequence of capturing profitability in addition to technical efficiency when measuring TFP in the present context is that measured TFP is presumably higher after firms introduce new varieties. The number of new varieties affects the value of output not only through quantities but also through prices; firms that introduce more new goods earn higher profits and are measured as more "productive.[15]" In Section 5.3, I include TFP as an explanatory variable in the Poisson regression of the number of new varieties. As I explain in that section, I use estimated TFP in 1998 instead of average TFP during 1998–2000 to avoid a reverse causality problem.

[14] Katayama, Lu and Tybout (2003) use a model of oligopoly with differentiated products and extend the argument to different data scenarios. They show that firms that face a more inelastic demand are able to charge higher prices and they appear to be more productive according to the TFP estimates.

[15] As a consequence, a fixed effect estimator for TFP would not be accurate as it would be assumed that TFP is constant over time.

5.2 Is the cost of innovation lower for foreign firms?

I first take a look at different ways in which firms introduce new products by running three different probit regressions among firms that did introduce a positive number of new varieties. The dependent variables in each of these regressions are whether firms have transferred at least one new variety from a company in the same corporate group, whether they have purchased at least one license for a new product from a foreign source, and whether they have developed at least one new variety in-house (for multinational firms, this means in the subsidiaries located in China). The explanatory variables are ownership structure, captured by the same categories as in the TFP analysis (majority foreign-owned, minority foreign-owned, private domestic, collectively owned), the logarithm of the number of workers, the logarithm of age, dummy variables for firms that started production in 1997 and 1998, industry and city fixed-effects.

Results are displayed in Table 8 and can be interpreted straightforwardly: majority foreign-owned firms are 17 percent more likely to transfer new products from a firm in a the same corporate group than are private domestic firms, they are 38 percent more likely to purchase at least one foreign license, and they are 19 percent *less* likely to develop products in-house. All results are significant. The difference in the probability of occurrence of any of these three events is not significant for minority foreign-owned firms and collectively-owned firms (the baseline category is private domestic firms).

The fact that foreign firms purchase more licenses and developed less products in the Chinese facilities than domestic firms do, does not imply that they have access to technology at a lower cost. It might well be the case that they need to purchase technology from outside sources because they are not efficient at in-house R&D. On the other hand, foreign multinationals transfer more goods from other firms in the same corporate groups, which is in fact an indicator of economies of scale in R&D production.

To address the question of whether foreign firms have a cost advantage in the introduction of new goods in a more direct way, I look at the expenditure on R&D plus expenditure on purchases of technology from outside sources over the period 1998-2000 per new variety successfully introduced, as a proxy for the average fixed cost of introducing new goods. For foreign multinationals, I only consider expenditure on R&D by the Chinese subsidiaries, to assess the marginal cost of introducing new varieties in China, not elsewhere.

Table 9 shows the results of a regression of this proxy for development costs on ownership structure categories (same as above). Controlling for total sales in 1998, self-reported market share, logarithm of age, dummies for firms that started operating in 1997 and 1998, and industry and city effects, majority foreign-owned firms have a cost advantage of 838 thousand dollars compared to private domestic firms (column [1]), although this estimated difference is not statistically significant. When instrumenting for market share with the self-reported number of competitors, the advantage for majority-owned foreign firms is 965 thousand dollars and significant at the 10 percent level. Fully foreign-owned firms are included separately in columns [3] and [4]. These firms have cost advantages of 1.6 and 1.2 million dollars per product in the least squares and instrumental variables regression, respectively. Minority foreign-owned firms are also more cost efficient than private domestic firms, on the four specifications, but the coefficients are imprecisely estimated.

5.3 How much does each factor explain?

Foreign firms in Group 2 (Electronics, Machinery and Appliances) appear to have a cost advantage over domestic firms, both in productivity and in the incorporation of new technology, as measured by TFP and expenditures on R&D and purchases of outside technology. In this section, I estimate the extent to which these differences affect the ability of firms to introduce new varieties of goods.

I reestimate the regressions that explain the number of new varieties (section (4)) including the estimates for TFP (hereafter, tfp denotes the estimate for the logarithm of TFP) and the expenditure of R&D and purchases of outside technology per new product introduced (hereafter, RD) over the sample of firms that introduce a positive number of new varieties. RD is a proxy for the average cost of innovation. The Poisson regression takes the form,

$$E\left(n_i|.\right) = \exp\left(x_i'\beta_0 + \beta_1 FOR_{1i} + \beta_2 FOR_{2i} + \beta_3 NONPRIV_i + \beta_4 tfp_i + \beta_5 RD_i\right) \tag{19}$$

where x are firm characteristics (number of workers, market share, age and initial year dummies), FOR_1, FOR_2 and $NONPRIV$ are the ownership dummies, and n_i is the number of varieties introduced during the period 1998-2000. tfp is estimated TFP during 1998. The reason to use estimates for 1998 instead of 1999, 2000 or the average is that these measures may suffer from a reverse causality problem. As discussed above, TFP estimates may not only reflect technical efficiency but also include demand factors that affect prices and consequently the value of output used to compute TFP. In principle, measured TFP can be affected by the number of products introduced. Using estimates of TFP for 1998 minimizes this potential problem.

Using an estimated value of TFP instead of the true value does not lead to an inconsistent estimator for the vector $\boldsymbol{\beta}$. Unlike the case of measurement error in survey data, the error in the estimation of TFP converges to zero asymptotically since tfp is a consistent estimator for $\log(TFP)$. Ignoring the fact that tfp is previously estimated, however, does lead to inconsistent standard errors for $\boldsymbol{\beta}$. To compute consistent standard errors I take 100 bootstrap samples of firms and re-estimate the production functions, TFP and regression (19) for each sample.

Results from Poisson, Negative Binomial and Least Squares regressions are reported in Table 10, with and without instrumenting for market share (the instrument is the number of competitors). The effect of TFP on the number of new varieties is positive and significant in five out of seven specifications. In the Poisson and Negative Binomial case an increase of 10 percent in TFP explains the introduction of approximately 50 percent more new varieties (see the incidence ratios in the lower panel). In the Least Squares specification, a 10 percent increase in TFP explains the introduction of 2.6 to 3.3 more new varieties. The coefficient of the cost of innovation (RD) has the expected negative sign (firms with higher cost of innovation introduce less new products) but it is not significant in any of the specifications.

Columns [3] and [7] include exports as a fraction of total sales as an explanatory variable. Firms that export have access to larger markets and presumably could derive larger profits from the introduction of a new variety than a firm that operates only in the Chinese domestic market; this effect would lead to the introduction of more new varieties than domestic firms. Empirically, however, there is no correlation between the number of new varieties and exports; coefficients are not significant (additionally, these are the two specifications for which TFP is not significant either).

To assess the impact of cost heterogeneity on the number of new varieties, I perform a decomposition of the predicted number of new varieties for two groups of firms - majority foreign-owned and private domestic - on the different explanatory factors.

Consider firstly the case of a linear specification, as the Least Squares regression of Table 10. The conditional expectation takes the form,

$$E\left[n_i | x_i, tfp_i, RD_i\right] = x_i'\beta_0 + \beta_4 tfp_i + \beta_5 RD_i. \tag{20}$$

For simplicity of notation, I include the ownership variables FOR_1, FOR_2 and $NONPRIV$ in x. Let $\widehat{n}_F$ be the predicted number of products for foreign firms, evaluated at the mean of x, tfp and RD over the sample of majority foreign-owned firms. That is, $\widehat{n}_F = \overline{x}_F{}'\widehat{\beta}_0 + \widehat{\beta}_4\overline{tfp}_F + \widehat{\beta}_5\overline{RD}_F$, where $\overline{x}_F$, $\overline{tfp}_F$ and $\overline{RD}_F$ are the means of x, tfp, and RD over this particular sample. Let $\widehat{n}_D$, $\overline{x}_D$, $\overline{tfp}_D$ and $\overline{RD}_D$ be the means computed over private domestic firms. By subtracting the predicted values $\widehat{n}_F$ and $\widehat{n}_D$ we obtain the following decomposition

$$(\widehat{n}_F - \widehat{n}_D) = (\overline{x}_F - \overline{x}_D)'\,\widehat{\beta}_0 + \widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right) + \widehat{\beta}_5\left(\overline{RD}_F - \overline{RD}_D\right). \tag{21}$$

The total contribution of differences in TFP to differences in the number of products between majority foreign-owned firms and private domestic firms is given by

$$\frac{\widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right)}{(\overline{x}_F - \overline{x}_D)'\,\widehat{\beta}_0 + \widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right) + \widehat{\beta}_5\left(\overline{RD}_F - \overline{RD}_D\right)}. \tag{22}$$

The definition for the contribution of RD is analogous.

Consider now the Poisson case,

$$E\left[n_i | x_i, tfp_i, RD_i\right] = \exp\left(x_i'\beta_0 + \beta_4 tfp_i + \beta_5 RD_i\right) \tag{23}$$

The predicted number of products for foreign firms is given by the specification of the conditional mean: $\widehat{n}_F =$

$\exp\left(\overline{x}_F{}'\widehat{\beta}_0 + \widehat{\beta}_4\overline{tfp}_F + \widehat{\beta}_5\overline{RD}_F\right)$, the definition is analogous for domestic firms. Because this expression is non-linear in the explanatory variables, it is convenient to consider incidence ratios. The predicted number of new varieties is $\frac{\widehat{n}_F}{\widehat{n}_D}$ times larger for foreign firms, and this ratio can be written as,

$$\frac{\widehat{n}_F}{\widehat{n}_D} = \exp\left[(\overline{x}_F - \overline{x}_D)'\,\widehat{\beta}_0 + \widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right) + \widehat{\beta}_5\left(\overline{RD}_F - \overline{RD}_D\right)\right]. \tag{24}$$

In this non-linear case, there are different ways to define the contribution of a particular variable (for example tfp) to the difference in the number of products between foreign and domestic firms. I take the contribution to explain the index, given by

$$\frac{\widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right)}{(\overline{x}_F - \overline{x}_D)'\,\widehat{\beta}_0 + \widehat{\beta}_4\left(\overline{tfp}_F - \overline{tfp}_D\right) + \widehat{\beta}_5\left(\overline{RD}_F - \overline{RD}_D\right)}, \tag{25}$$

which is exactly the same as in the Least Squares case.

In Table 11, I construct the ratios in (22) and (25) for TFP, cost of innovation and exports using the estimates in Tables 10. Differences in TFP between foreign and domestic firms explain between 13 percent and 31 percent of the predicted advantage of foreign firms in the introduction of new varieties, while differences in the cost of R&D and purchases of new technology explain between 0 percent and 2 percent of this advantage. The difference attributable to export share of sales lies between 5 and 7 percent. In the Least Squares case, the predicted difference in the number of new varieties is by construction equal to the observed difference. The total explanatory power of both factors of cost heterogeneity lies between 13 and 33 percent; when exports are included in the regression, the three factors together explain betwee 24 and 28 percent.

6 Conclusion

This paper describes some aspects of the introduction of new goods in the Chinese manufacturing sector by domestic and foreign firms, and provides empirical evidence on the effects of advantages in TFP and the cost of innovation of foreign firms.

Firms with more than 50 percent of foreign capital introduce twice the number of new products as privately-owned domestic firms; and firms fully owned by multinationals introduce three times as many new products. Differences in TFP and the cost of innovation are explored as possible explanations for the difference in the number of new varieties. The hypothesis is that foreign firms have already introduced the same or similar varieties in foreign markets and have already incurred development costs and gained experience in production; as a result, both the cost of innovation and of production are lower in China when compared to the cost of domestic firms (that operate only in the Chinese market) that have to set up the production of new varieties from scratch.

I find that majority foreign-owned firms are more efficient than domestic firms in the manufacturing of Electronics, Machinery and Appliances, but not in the production of Apparel, Textiles and Leather goods. They have an advantage in costs of innovation (measured as expenditure on R&D plus purchases of outside technology by Chinese subsidiaries per new variety introduced into the Chinese market), although it is not statistically significant. The advantage in TFP explains between 13 and 31 percent of the predicted difference in the number of new varieties between foreign and domestic firms; the lower cost of innovation explains only up to 2 percent of the difference.

References

Ackerberg, D. and K. Caves (2003): "Structural Identification of Production Functions," *Department of Economics, UCLA*

Bernard, A. and J.B. Jensen (2004): "Exporting and Productivity," *Tuck School of Business, Dartmouth College*

Blundell, R., R. Griffith and J. Van Reenen (1995): "Dynamic Count Data Models of Technological Innovation," *Economic Journal,* 105, 333-344

Blundell, R. and J. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," *Advances in Economics and Econometrics: Theory and Applications*, ed. by M. Dewatripont, L. P. Hansen and S. Turnovsky, Vol. 2, Econometric Society Monograph #36, Cambridge University Press, Cambridge.

Blundell, R. and J. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," forthcoming *Review of Economic Studies*

Cameron, A.C. and P. Trivedi (1998): *Regression Analysis of Count Data,* Cambridge University Press, Cambridge

Crepon, B. and E. Duguet (1997): "Research and Development, Competition and Innovation. Pseudo-Maximum Likelihood and Simulated Maximum Likelihood Methods Applied to Count Data Models with Heterogeneity," *Journal of Econometrics,* 79, 355-378

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

Fernandes, A. (2003): "Trade Policy, Trade Volumes and Plant-Level Productivity in Colombian Manufacturing Industries," The World Bank, Working Paper No. 3064

Gourieroux, C., A. Monfort and A. Trognon (1984a): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681-700

Gourieroux, C., A. Monfort and A. Trognon (1984b): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica,* 52, 701-720

Griliches, Z. (1957): "Specification Bias in Estimates of Production Functions," *Journal of Farm Economics,* 39, 8-20

Hausman, J., B. Hall Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D relationship," *Econometrica,* 52, 909-938

Helpman, E., M. Melitz and S. Yeaple (2004): "Export versus FDI with Heterogeneous Firms," *American Economic Review,* 94, 300-316

Javorcik, B. (2004): "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages," forthcoming *American Economic Review*

Katayama, H., S. Lu and J.R. Tybout (2003): "Why Plant-Level Productivity Studies are Often Misleading, and an Alternative Approach to Inference," *Department of Economics, Pennsylvania State University*

Keller, W. and S. Yeaple (2003): "Multinational Enterprises, International Trade, and Productivity Growth: Firm Level Evidence from the United States," *NBER*, Working Paper 9504

Klette, T.J. and Z. Griliches (1996): "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics,* 11, 343-361

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

Marschak, J. and W. H. Andrews (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica,* 12, 143-205

Melitz, M. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica,* 71, 1695-1725

Mundlak, Y. (1963): "Estimation of Production and Behavioral Functions from a Combination of Cross-Sections and Time Series Data," *Measurement in Economics: Studies in Mathematical Economics and Econometrics in memory of Yehuda Grunfeld,* ed. by C. Christ et al. Stanford University Press, Stanford

Olley, G.S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica,* Vol. 64, No. 6, 1263-1297

Pavcnik, N. (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies,* Vol. 69, No. 1

TABLE 1. Industrial sectors included in the survey

	INDUSTRY CODES
APPAREL AND LEATHER GOODS	
Textiles	178
Apparel manufacturing	181, 1923, 1952
Leather tanning and finishing; product manufacturing	191, 192
ELECTRONIC EQUIPMENT	
Communications equipment manufacturing	411
Audio and video equipment manufacturing	413
Computer and peripheral equipment manufacturing	414, 4171,4172
ELECTRONIC COMPONENTS	
Electron tube manufacturing	4151
Printed circuit manufacturing and assemble	4155
Semiconductor and related device manufacturing	4153, 3617
Electronic capacitors, resistors, coils, transformers, connectors	416, 4021, 4023, 4024
CONSUMER PRODUCTS	
Household cooking, refrigerating and laundry appliances	4066, 4063, 4061
Small electrical appliance manufacturing	4069
VEHICLES AND VECHICLE PARTS	
Motor vehicle manufacturing, including body and trailer	3721, 3725, 3726
Motor vehicle accessories	3727, 3512, 4222, 3727
Motorcycle, bicycle and parts manufacturing	373, 374

TABLE 2. Summary statistics by ownership type

Type of firm	Number of firms	Median number of workers	Average Output per worker	Average number of new products
	[1]	[2]	[3]	[4]
Private domestic	279	172	207	2.7
State-owned	171	602	70	2.8
Cooperative	139	286	97	1.1
Foreign 1%-50%	81	223	254	3.6
Foreign 50%-99%	148	306	483	4.5
Foreign 100%	41	165	470	4.8
All firms	859	269	226	3.0

TABLE 3. Regression of number of new varieties on foreign ownership

	Poisson			2-Step Poisson		Negative Binomial	Zero-Inflated Poisson	OLS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Foreign 50%-100%	0.707***	0.607**	0.619**	0.797**	0.823***	0.801***	0.667***	2.617**
	0.273	*0.269*	*0.285*	*0.325*	*0.313*	*0.259*	*0.238*	*1.251*
Foreign 1%-50%	-0.181	-0.213	-0.167	-0.117	-0.150	-0.135	-0.240	-0.212
	0.257	*0.262*	*0.244*	*0.267*	*0.261*	*0.208*	*0.289*	*0.630*
Non-Private	-0.447**	-0.494**	-0.442*	-0.300	-0.350	-0.528***	-0.429*	-0.840
	0.242	*0.245*	*0.247*	*0.258*	*0.237*	*0.194*	*0.239*	*0.622*
logWorkers	0.130**	0.152**	0.111**	0.156**	-	0.245***	0.097	0.413**
	0.064	*0.062*	*0.050*	*0.065*		*0.061*	*0.069*	*0.174*
Market share	0.012***	0.011***	0.012***	0.007	0.009**	0.011***	0.004	0.044***
	0.003	*0.003*	*0.003*	*0.004*	*0.004*	*0.003*	*0.003*	*0.017*
logAge	0.133	0.175	0.123	0.083	0.185	0.141	0.132	0.324
	0.123	*0.119*	*0.121*	*0.135*	*0.130*	*0.110*	*0.135*	*0.371*
Initial Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City effects	Yes	-	-	Yes	Yes	Yes	Yes	Yes
Industry effects	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes
INCIDENCE RATIOS								
Foreign	2.03	1.83	1.86	2.22	2.28	2.23	1.95	
Foreign 1%-50%	0.83	0.81	0.85	0.89	0.86	0.87	0.79	
Non-Private	0.64	0.61	0.64	0.74	0.70	0.59	0.65	
logWorkers	3.67	4.57	3.03	4.76	-	11.59	2.63	
Market share	1.13	1.12	1.13	1.07	1.09	1.12	1.04	
logAge	3.77	5.73	3.44	2.30	6.36	4.10	3.74	

Dependent variable: number of new varieties. Other controls: setup up year dummies. Robust standard errors in italics. ***, ** and * indicate significancy at the 1%, 5% and 10% level. Incidence ratios are computed for a discrete change in categorical variables and a 10% change in continuous variables.

TABLE 4. Fully foreign-owned firms

	Poisson		2-Step Poisson	
	[1]	[2]	[3]	[4]
Foreign 1%-50%	-0.182	-0.184	-0.137	-0.139
	0.255	*0.253*	*0.257*	*0.257*
Foreign 50%-99%	0.705**	0.702**	0.728**	0.726**
	0.307	*0.304*	*0.327*	*0.325*
Foreign 100%	0.712*	1.203**	1.025*	1.460**
	0.411	*0.488*	*0.597*	*0.723*
Foreign 100% - Investor	-	-0.252	-	-0.176
		0.429		*1.539*
INCIDENCE RATIOS				
Foreign 1%-50%	0.83	0.83	0.87	0.87
Foreign 50%-99%	2.02	2.02	2.07	2.07
Foreign 100%	2.04	3.33	2.79	4.31
Foreign 100% - Investor	-	0.78	-	0.84

Poisson regression of the number of new varieties. [1]: Foreign 100% includes all firms with no domestic participation. [2]: Foreign 100% includes firms with no domestic participation that are owned by a foreign firm, while Foreign 100% - Investor indicates that the firm is owned by a foreign investor, not a firm. Other controls: Non-Private, logWorkers, Market share, logAge, setup year dummies, industry effects, city effects. Robust standard errors in italics. ***, ** and * indicate significancy at the 1%, 5% and 10% level.

TABLE 5. Production function coefficients

	Labor		Capital	
Leather	0.69	*0.09*	0.25	*0.44*
Apparel	0.67	*0.06*	0.27	*0.14*
Bare printed circuits	0.65	*0.38*	0.31	*0.43*
Semiconductor and related devices	0.56	*0.16*	0.28	*0.25*
Motor vehicles	0.68	*0.15*	0.34	*0.44*
Audio and video equipment	0.56	*0.21*	0.32	*0.11*
Computer equipment	0.49	*0.14*	0.28	*0.13*
Motor vehicles accessories	0.46	*0.07*	0.27	*0.14*
Motorcycles, bicycles and parts	0.51	*0.31*	0.32	*0.36*
Electronic capacitors, resistors and related	0.58	*0.15*	0.38	*0.19*
Small electrical appliances	0.54	*0.11*	0.39	*0.37*
Textiles	0.49	*0.24*	0.44	*0.24*
Large household appliances	0.55	*0.18*	0.53	*0.26*
Communications equipment	0.39	*0.15*	0.45	*0.21*
Electron tube manufacturing	0.32	*0.14*	0.59	*0.24*

Standard errors in italics

FIGURE 1. Distribution of TFP

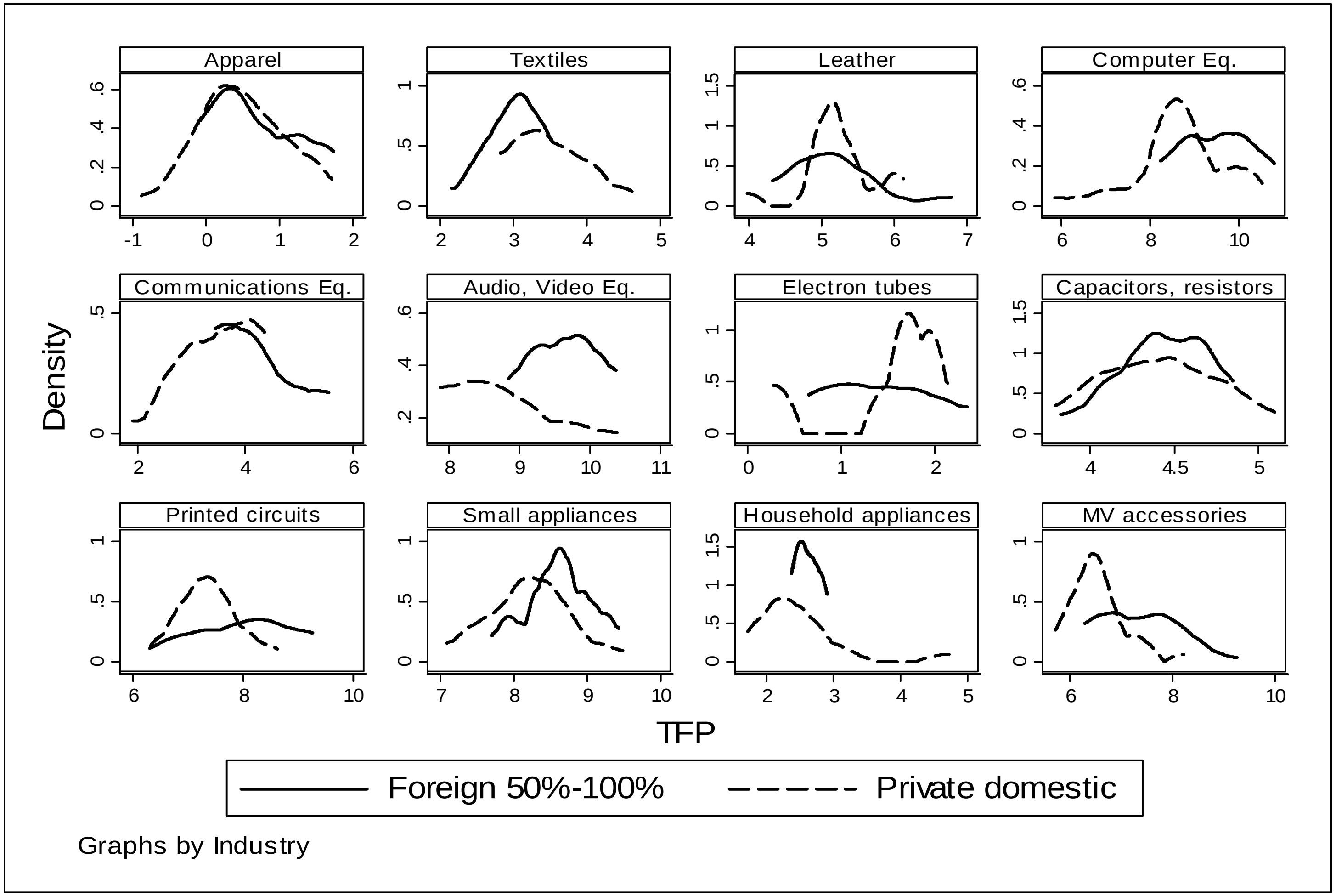

TABLE 6. Regression of average TFP on ownership structure

	Group 1		Group 2		Cutoff
	[1]	[2]	[3]	[4]	[5]
Foreign 50%-100%	-0.153	-0.181	0.226***	0.217***	0.196***
	0.133	*0.144*	*0.074*	*0.080*	*0.072*
Foreign 100% - Firms	-	0.045	-	0.314*	-
		0.156		*0.174*	
Foreign 100% - Investors	-	-0.074	-	0.221	-
		0.236		*0.178*	
Foreign 1%-50%	-0.305**	-0.305**	-0.012	-0.013	-0.009
	0.139	*0.141*	*0.087*	*0.087*	*0.086*
Non-Private	0.047	0.049	-0.084	-0.085	-0.099
	0.138	*0.139*	*0.075*	*0.075*	*0.075*
log Age	-0.301***	-0.302***	-0.004	-0.003	0.007
	0.072	*0.072*	*0.042*	*0.042*	*0.042*

Group 1: Textiles, Apparel and Leather. Group 2: Electronics, Machinery and Appliances. The definition of Foreign 50%-100% depends on the column; in columns [1], [3] and [5] it includes fully foreign-owned firms, while in columns [2] and [4], fully foreign-owned firms are included in two separate categories (Foreign100% - Firm, and Foreign 100% - Investors). Column [5] does not include firms below the estimated producitvity cutoff. Other controls: setup year dummies, industry and city effects. Standard errors in italics; *, ** and *** indicates significance at the 10%, 5% and 1% level.

TABLE 7. Difference in minimum estimated TFP levels between majority foreign-owned firms and private domestic firms

Apparel	0.69
Textiles	-0.67
Leather	0.35
Computer equipment	2.35
Communications equipment	1.68*
Audio and video equipment	1.14
Electron tube manufacturing	0.38
Semiconductors and related devices	0.35
Electronic capacitors, resistors and related	-0.06
Bare printed circuits	0.26
Small electrical appliances	0.67
Large household appliances	0.50
Motorcycle, bicycle and parts	0.84
Motor vehicles	2.80*
Motor vehicles accessories	0.60

* Indicates that it is significant at the 10% level. 90% confidence intervals constructed with 1,000 bootstrapped replications

TABLE 8. Ways in which firms introduce new varieties

	Transferred [1]	Licensed [2]	In-house [3]
Foreign 50%-100%	0.174***	0.379***	-0.189**
	0.073	*0.083*	*0.079*
Foreign 1%-50%	-0.014	-0.027	0.000
	0.065	*0.074*	*0.084*
Non-Private	-0.081	0.017	-0.064
	0.049	*0.073*	*0.062*

Probit regressions. Table displays incremental effect of regressors. Dependent variables: [1] Transferred at least one new good from company in same corporate group; [2] Purchased at least one license from a foreign source;[3] Developed a new good in house. Regressions only include firms that introduced new goods. Other controls: logWorkers, logAge, set-up year dummies, industry effects, city effects. Standard errors in italics. ***, ** and * indicate significancy at the 1%, 5% and 10% level.

TABLE 9. Expenditure on technology

	OLS [1]	IV [2]	OLS [3]	IV [4]
Foreign 1%-50%	-0.385	-0.622	-0.583	-0.542
	0.441	*0.445*	*0.450*	*0.493*
Foreign 50%-100%	-0.838	-0.965*	-0.816	-0.711
	0.535	*0.572*	*0.585*	*0.628*
Foreign 100%	-	-	-1.655**	-1.246**
			0.737	*0.534*
Non-Private	-0.549*	-0.621**	-0.598*	-0.480
	0.306	*0.315*	*0.316*	*0.299*
Sales	0.0025***	0.0025***	0.0025***	0.0057***
	0.0007	*0.0007*	*0.0007*	*0.0016*
Market share	0.0029	0.0016	0.0021	0.0031
	0.0061	*0.0187*	*0.0189*	*0.0062*
logAge	-0.147	-0.192	-0.185	-0.133
	0.198	*0.203*	*0.204*	*0.200*

Robust standard errors in italics. ***, ** and * indicate significancy at the 1%, 5% and 10% level. Dependent variable is expenditure in R&D including purchases of technology from outside sources, measured in 1998 millions of dollars. Regressions only include firms that introduced new goods. Other controls: setup year dummies, industry and city effects.

TABLE 10. Number of new varieties

	Poisson	2-Step Poisson	2-Step Poisson	2-Step Neg. Binomial	OLS	IV	IV
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Foreign 50%-100%	0.498	0.499	0.449	0.441	3.477	5.811*	3.191
	0.318	*0.331*	*0.749*	*0.290*	*2.446*	*3.346*	*6.474*
Foreign 1%-50%	-0.362	-0.332	-0.209	-0.315	-1.923	0.948	0.660
	0.256	*0.261*	*0.654*	*0.261*	*1.489*	*2.828*	*4.997*
Non-Private	-0.529*	-0.407	-0.721	-0.279	-3.189**	-2.742	-1.037
	0.277	*0.271*	*0.653*	*0.234*	*1.628*	*1.871*	*4.965*
TFP 1998	0.439**	0.430**	0.379	0.382**	2.628**	2.704**	3.342
	0.189	*0.188*	*0.339*	*0.156*	*1.176*	*1.367*	*3.006*
R&D and Licenses	-0.075	-0.072	-0.056	-0.029	-0.233	-0.123	-0.235
	0.053	*0.051*	*0.082*	*0.038*	*0.173*	*0.249*	*0.689*
logWorkers	0.086	0.063	0.141	0.090	0.624	-0.264	0.028
	0.079	*0.078*	*0.213*	*0.071*	*0.467*	*0.713*	*1.885*
Market share	0.011***	0.009*	0.000	0.005	0.075**	0.369**	0.337
	0.003	*0.005*	*0.042*	*0.005*	*0.027*	*0.183*	*0.393*
logAge	0.156	0.123	-0.041	0.071	1.041	2.303**	0.778
	0.118	*0.124*	*0.335*	*0.120*	*0.736*	*1.150*	*2.442*
Exports	-	-	0.324	-	-	-	5.160
			0.488				*4.202*
INCIDENCE RATIOS							
Foreign 50%-100%	1.645	1.648	1.567	1.555			
Foreign 1%-50%	0.696	0.718	0.811	0.730			
TFP 1998	1.551	1.537	1.460	1.465			
R&D and Licenses	0.928	0.930	0.945	0.972			
Exports	-	-	1.383	-			

Dependent variable: number of new goods. Incidence ratios computed for a change of one unit for continuous variables. Other controls: setup up year dummies, industry and city fixed effects. Standard errors in italics. ***, ** and * indicate significancy at the 1%, 5% and 10% level.

TABLE 11. Decomposition of the difference in the predicted number of new varieties between foreign and domestic firms

	Poisson [1]	2-Step Poisson [2]	2-Step Poisson [3]	2-Step Neg. Binomial [4]	OLS [5]	IV [6]	IV [7]
TFP 1998	23.4%	24.6%	21.7%	31.4%	19.4%	12.9%	16.0%
R&D and Licenses	2.0%	2.1%	1.7%	1.2%	0.9%	0.3%	0.6%
Exports	-	-	5.3%	-	-	-	7.0%