Università degli Studi di Cagliari
DOTTORATO DI RICERCA IN ECONOMIA
XXIV Ciclo

Productivity, Markups, and Trade Liberalization: Evidence from Mexican Manufacturing Industries

Settore scientifico-disciplinare di afferenza SECS-P/06

Dottoranda:	Daniela Puggioni
Coordinatore Dottorato:	Prof. Romano Piras
Tutor/Relatore:	Prof. Fabiano Schivardi

Esame finale anno accademico 2011-2012

Productivity, Markups, and Trade Liberalization: Evidence from Mexican Manufacturing Industries*

Daniela Puggioni†

Abstract

In order to evaluate the effects of any policy or answer economic relevant questions it is of primary importance to accurately quantify the variables and the parameters that may be involved with the policy or the questions. Since production functions are a fundamental components of all economics, in most of the cases answering those questions requires to estimate production function parameters. Nonetheless, production function estimation is challenging since optimal input choices are correlated with firm-specific unobserved productivity shocks. In this contribution I apply a structural framework for estimating production function coefficients explicitly controlling for productivity through an observable proxy (investment or intermediate inputs) using a rich panel dataset of Mexican manufacturing firms between 1984 and 1990. I further derive firm-level markup estimates and use these estimates to evaluate the impact of the dramatic trade liberalization that took place in Mexico during the sample period on the profitability of domestic firms and exporters. My findings emphasize the importance of obtaining consistent estimates in order to correctly assess differences in technologies, productivity, and market power among the Mexican firms.

Settore Scientifico-Disciplinare: SECS-P/06.

JEL Classification: D22, D24, F14, L11, L60.

Keywords: Production function estimation, productivity, markups, trade liberalization, Mexican manufacturing industries.

* *Acknowledgment:* I would like to thank Professor James Tybout for providing me with the data and valuable suggestions for this project. I also thank Professor Bee-Yan Roberts and Professor Paul Grieco for thoughtful remarks. A special thanks goes to Professor Spiro Stefanou for his support and to Professor Fabiano Schivardi for his guidance. Furthermore, I am particularly grateful to Professor Stefano Usai and Professor Emanuela Marrocu for encouragement and valuable comments, and to Alexandros Fakos for helpful discussions and very useful computational tips.

† Università di Cagliari and The Pennsylvania State University, puggioni.daniela@gmail.com

Contents

1 Introduction

In order to evaluate the effects of any policy or answer economic relevant questions it is of primary importance to accurately quantify the variables and the parameters that may be involved with the policy or the questions. Since production functions are a fundamental component of all economics, oftentimes it is hard even to formulate a question appropriately without considering production functions and embedding them in the framework. This is because much of economic theory provides testable implications that are directly related to technology and optimizing behavior. Production functions relate productive inputs to outputs and applied economists started to worry since the early 1940s about the issues confronting their estimation because of the potential correlation between optimal input choices and unobserved firm-specific determinants of production. The rationale behind this concern is intuitive. Firms that experience higher productivity shocks are likely to respond increasing their input usage, therefore classical estimation methods as, for example, ordinary least squares (OLS) will yield biased coefficient estimates and biased estimates of productivity. Consequently, any further analysis or evaluation based on those biased estimates will be necessarily unreliable.

In the literature many alternatives to OLS have been proposed, from relatively simple instrumental variables and fixed effects solutions to more complex and sophisticated techniques like dynamic panel data estimators and structural empirical models. In this contribution I rely on the original insight of Olley and Pakes (1996) and the successive extension by Levinson and Petrin (2003) and attempt to correctly estimate production function parameters and productivity with a structural procedure using an observable proxy, either investment or intermediate inputs, to control for the correlation between input levels and the unobserved productivity shock. The essential assumption for successfully applying this methodology is that productivity and investment (or intermediate inputs) are linked through a unique monotonic relation so that observed investment (or intermediate inputs) choices contain valuable information about the productivity shock and can be used to consistently estimate production function coefficients. I take this empirical framework to a rich panel dataset including information on production and trade characteristics for over 2,000 Mexican manufacturing firms between 1984 and 1990.

With the unbiased production function estimates in hand, I further derive firm-level price-cost margins relying on a structural approach in which markups are given by the wedge between the cost share of factors of production and their revenue share. This approach has the advantage of being very general and flexible as it does not impose any strong restrictions on the underlying production function and it does not require to specify how firms compete in the market. I then compare these plant-level markup estimates with industry-level markups obtained through a simpler dual approach in order to verify the extent to which using micro-level information and directly controlling for unobserved firm-level productivity is important in correctly evaluating market power.

During the period covered in the data the Mexican economy tried to find its way out of a deep recession undergoing major structural reforms such as reduction in government expenditure, privatization of state-owned companies, elimination of subsidies, deregulation of financial markets, liberalization of foreign investment, and a dramatic re-orientation of trade policy. The trade policy reforms were perhaps the most striking leading Mexico to become one of the most open economy in the world in less than a decade. Therefore, the Mexican economic environment in those years is particularly suitable to analyze the effects of trade exposure on the Mexican manufacturing firms. More specifically, in order to investigate whether the outward looking trade reforms lowered the profitability of the domestic firms by boosting competition, I test the relation between markups and measures of import liberalization in a regression framework. In addition, I combine the markups and the productivity estimates to verify the prediction of several recent international trade models that exporters are more productive and thus able to charge higher markups.

The main findings of my contribution can be summarized as follows. First, controlling for unobserved productivity with the investment proxy successfully corrects the simultaneity bias in the production function parameter estimates. Second, the markups estimated at the firm level are more reasonable and significantly higher than the ones estimated at the industry level demonstrating that exploiting micro-level data and taking into account differences in productivity is important to assess the extent of market power. Third, the industry-level analysis on the impact of trade liberalization on the profitability of the Mexican manufacturing industries provides some evidence of import discipline but this result is not confirmed at the plant level. Lastly, the markup premium for exporters is significant only for "intensive" exporters, i.e. firms exporting a high percentage of their output.

The reminder of the work is organized as follows. Section 2 provides a review of the main issues and contributions in the literature regarding production function estimation. Section 3 includes the details of the empirical methodology used to estimate the production function parameters as well as the markups. Section 4 briefly characterizes the main features of the Mexican trade liberalization and illustrates some simple models suitable to relate markups and trade exposure. The data and the sample selection criteria are described in Section 5. Sections 6 and 7 present the results on the production function estimation and price-cost margins analysis, respectively. Section 8 concludes.

2 Literature Overview on Production Functions Estimation

2.1 Basic endogeneity issues

Production functions are an essential component in both theoretical and empirical economic models and their estimation has a long history in applied economics, starting in 1800. However, researchers are actually interested in estimating production functions because, in most cases, it is a tool for answering other questions, only partially related to the production function itself. Oftentimes it is hard even to formulate a question appropriately without considering production functions and embedding them in the framework. For example, a researcher may be interested in the presence of economies of scale in production, in whether productivity differences depend upon differences in the quality of labor or differences in R&D, in whether the marginal product of factors are equal to factor prices, in what is the market structure in different industries and how this is related to the profitability of the firms. All these questions require reliable estimates of cost or production functions and are so important and interesting in economics that it is worth trying to answer them, even though the estimation framework used for these purposes may be quite problematic.

Econometric production functions, as we know them today, essentially relate productive inputs (e.g. capital and labor) to outputs and have their roots in the work of Cobb and Douglas (1928) who proposed production function estimation as a tool for testing hypotheses on marginal productivity and competitiveness in labor markets. Criticism to their approach came very soon as Mendershausen (1938) argued that the data used by Douglas were too multi-collinear to allow for a credible determination of the production function coefficients. Marshack and Andrews (1944) were the first to explicitly state one of the main reasons why production function estimation is problematic.

"Can the economist measure the effect of changing amounts of labor and capital on the firm's output - the "production function" - in the same way in which the agricultural research worker measures the effect of changing amounts of fertilizers on the plot's yield? He cannot because the manpower and capital used by each firm is determined by the firm, not by the economist. This determination is expressed by a system of functional relationships; the production function, in which the economist happens to be interested, is but one of them."

To illustrate the issue consider the Cobb-Douglas production function technology

$$Y_j = A_j K_j^{\beta_k} L_j^{\beta_l}$$

with one output Y_j and two inputs: capital K_j and labor L_j. A_j is the Hicks-neutral efficiency level of firm j, that is unobservable by the econometrician. Taking natural logs the previous relation becomes linear

$$y_j = \beta_0 + \beta_k k_j + \beta_l l_j + \varepsilon_j \tag{2.1}$$

where lowercase letters express natural logarithms of the variables, (e.g. $\ln(K_j) = k_j$) and $\ln(A_j) = \beta_0 + \varepsilon_j$. The constant term β_0 can be view as the mean efficiency level across firms, while ε_j is the deviation from that mean for each firm j. ε_j represents all other disturbances, left out the factors, such as firm-specific technology, efficiency, or management differences, functional form discrepancies, measurement errors in output, or unobserved sources of variation in output. The observation made by Marshack and Andrews is that, since the right-hand-side variables are chosen by the firm is some optimal or behavioral fashion, they cannot really be treated as independent. In fact, if the firm knows its ε_j, or some part of it, when making input choices, these choices will likely be correlated with ε_j. One could argue that capital can be considered a fixed input, as it is usually predetermined for the duration of the relevant observation period, and it is therefore orthogonal with respect to the disturbance term. The same argument, however, will not apply to labor, even if we are willing to make the quite strong assumption that firms operate in perfectly competitive input and output markets and treat capital as a fixed input. If firms perfectly or imperfectly observe ε_j before choosing the optimal amount of labor to utilize in production, their choice will necessarily depend on ε_j and the usual exogeneity assumptions that are required for unbiasedness and consistency of OLS are unlikely to hold. Empirical results have actually shown that both capital and labor are usually correlated with the error term but most often the bias in the labor coefficient is larger than the bias on the capital coefficient. This is consistent with the view that labor is more easily adjustable than capital, this more variable, and therefore more highly correlated with ϵ_j.

Marshack and Andrews introduced a simultaneous equations methodology to production function estimation that can be exposed using a simple profit maximizing model of the joint determination of output and labor, given capital, output price (assuming, for simplicity that the price of output is the same across firms and it is normalized to 1) and input prices. In this context the marginal productivity condition, which is also the variable input demand function, is given by:

$$y_j = l_j + w_j - \ln(\beta_l) + \nu_j \tag{2.2}$$

where w is the natural logarithm of the price of labor and ν_j is a term representing all the deviations from the assumed conditions of perfect competition, absence of risk aversion and uncertainty, and possible measurement errors in y_j, l_j, and w_j. Equations (2.1) and (2.2) constitute a system of two structural equations whose reduced form is given by:

$$l_j = \left(\frac{1}{1-\beta_l}\right)[\beta_0 + \ln(\beta_l) + \beta_k k_j - (w_j + \nu_j) + \varepsilon_j] \tag{2.3}$$

$$y_j = \left(\frac{1}{1-\beta_l}\right)[\beta_0 + \beta_l(\beta_o + \ln(\beta_l)) + \beta_k k_j - \beta_l(w_j + \nu_j) + \varepsilon_j] \tag{2.4}$$

Thus the simple message of the Marshack and Andrews' contribution is that if labor is chosen even approxmately optimally, the production function disturbance is "transmitted" to the decision equation and l_j is a function of it. Simple OLS estimates of the production function coefficients will be biased and will not have the desired structural interpretation.

There is a second problem, perhaps less emphasized and documented in the literature, embedded in the OLS estimation of (2.1). Firm-level dataset are usually characterized by a significant level of attrition, i.e. firms entering and exiting but, obviously, researches have only data on firms prior to exiting. Assume firms can observe ε_j, then decide whether to exit or not, and choose labor and level of production optimally if they decided not to exit. Abstracting from dynamics implications, assume also that firms deciding to exit receive a non-negative remuneration equal to their sell-off value, thus firms will exit if the variable profits are lower than the sell-off value. The problem here is that this exit condition will generate correlation between ε_j and K_j, conditional on continuing to be in the dataset, i.e. continuing to produce. This is because, if firms know their ε_j when they have to decide whether to exit or stay, firms continuing to produce will have ε_j drawn from a selected sample and the selection will be partially dependent on the fixed input K_j. In other words, as firms with higher fixed capital are able to afford lower ε_j without having to exit, the sample selection of the firms remaining in business will generate negative correlation between ε_j and K_j. Once again, the orthogonality conditions for OLS estimation would be violated.

2.2 Traditional solutions: instrumental variables and fixed effects

The earliest responses to the concerns about the necessity of considering the endogeneity issues in production functions estimation came through the increasing availability of panel data and developed, traditionally, along two main directions: instrumental variables and fixed effects.

2.2.1 Fixed effects

Hoch and Mundlak were the pioneers in introducing the fixed effects methodology in economics in the context of production functions estimation. To understand the essence of this approach consider a modified formulation of (2.1)

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_j + \eta_{jt} \tag{2.5}$$

where η_{jt} is not observed by the firm before any production decision (input choice or exit) so that this term is not correlated with the firm's optimal choices. Conversely firms have knowledge of ω_j when they make input and exit choices. Intuitively, ω_j can represent entrepreneurial ability, labor quality, or any other factor affecting the production that firm can observe or predict and it is usually defined as the firms' unobserved (by the econometrician) productivity. η_{jt}, on the other hand, represents deviations from the expected values of these factors and can be also thought as the conventional measurement error in y_{jt} that is uncorrelated with input and exit decisions. Clearly the endogeneity issues concern only ω_j and not η_{jt}. The fact that ω_j is assumed to be constant over time, or at least over the length of the available panel, is the basic premise behind fixed effects estimation and allows for consistent estimation of production function coefficients using differencing, or least squares dummy variables estimation techniques. In general this implies that (2.5) can be consistently estimated via OLS specifying

$$(y_{jt} - y_{j\bar{t}}) = \beta_k(k_{jt} - k_{j\bar{t}}) + \beta_l(l_{jt} - l_{j\bar{t}}) + (\eta_{jt} - \eta_{j\bar{t}}) \tag{2.6}$$

where the notation $(x_{jt} - x_{j\bar{t}})$ represents averaging over the time dimension for each individual firm[1].

This approach is first stated briefly in Hoch (1955) and fully developed in Hoch (1962). In this latter contribution, Hoch makes use of combined time-series and cross-section data in the estimation of production function parameters for a sample of 63 Minnesota farms over a six-year period from 1946 to 1951. The main goal of the study is to estimate the elasticity of output with respect to inputs in order to draw inferences regarding the allocation of resources by the economic units of the sample. A Cobb-Douglas specification is used to derive a condition stating that firms equate the value of the marginal product of each input to its price multiplied by some constant. This constant represents the elasticity of output with respect to that input and can be interpreted as returns to scale. If firms are in fact profit maximizers, the value of the constant should be one as optimality requires the value of the marginal product to be the same as the price of the factor. Hoch argues that rationalizing the use of single equation estimates of the production function parameters is possible if one is willing to assume that firms maximize by differentiating anticipated output with respect to current input so that the observed input choices are not correlated with the disturbance term. The extent to which this assumption can be supported depends on the characteristics of the industry where the firms operates. In the case of agriculture, for example, it seems reasonable to believe that the term ε_j includes the effects of weather variability which do not affect the optimal choice of inputs. In this context a single equation estimation is justifiable. There are, however, other differences between firms, such as difference in technical efficiency, that will influence both output and inputs. Hoch points out that if there are differences in technical efficiency between firms, i.e. ω_j in (2.5) varies substantially across firms, firms that are more efficient will be able to produce more output for a given level of inputs and, by profit maximization, they will tend to have higher levels on inputs, thus the optimal choice of factors will depend on ω_j. A similar problem arises if productivity increases over time. As a way out of this difficulty, Hoch uses the analysis of covariance exploiting the time-series and cross-sectional dimensions of his data and estimating a system of equations similar to (2.3) and (2.4) including firm-specific and time-specific fixed effects in the production function equation. Since differences between firms and time periods affecting both output and input choices are accounted for these fixed effects, he argues that his model does not suffer from simultaneous equations bias.

[1] In the case of first-differencing (2.6) would be $(y_{jt} - y_{jt-1}) = \beta_k(k_{jt} - k_{jt-1}) + \beta_l(l_{jt} - l_{jt-1}) + (\eta_{jt} - \eta_{jt-1})$.

Despite the innovative approach, Hoch's results are not very encouraging. Moving from time correction estimates, where only time effects are included, to analysis of covariance estimates where also farm effects are considered, there is a significant drop in the estimated sum of elasticities, from almost 1 to approximately 0.75 which, in turn, generates unreasonably low estimated marginal returns (around 0.20) to labor. These figures force him to (questionably) interpret the shortfall as reflecting the fact that efficiency may increase with scale and that there may be returns to the unmeasured, fixed entrepreneurial factor.

Mundlak further exploits the fixed effect approach in his 1961 contribution with the scope of obtaining unbiased production function estimates in the presence of unobserved managerial ability. He notes that, instead of trying to rationalize the concept and the meaning of managerial capacity in order to include some index of management in the production function, one should assume that, whatever management is, it does not change substantially over time and, for at least a two year period, it can assumed to remain constant. Mundlak assumes a Cobb-Douglas specification very similar to (2.5) apart from the fact that the management variable is included among the inputs and has its own (constant) coefficient to be estimated. However, since management is not directly observable, the specification taken to the data is exactly the same as (2.5) with $\omega_j = cm_j$ being firm's j fixed effect, m_j being management, and c being the constant multiplicative term associated with it. If the production function is fully specified and the assumption of the classical regression model hold, unbiased and efficient estimates can be obtained using the analysis of covariance. Moreover, imposing the additional restrictions that management is the only fixed input for all firms, for at least a two year period, and that there are constant returns to scale, unbiased estimates of c and M_j (where $M_j = \exp(m_j)$) can also be recovered. Mundlak's results, obtained using a sample of 66 family farms in Israel from 1954 to 1958, show that management is positively correlated with most of the inputs and that the firm fixed effects are significantly different from zero, suggesting that the estimates obtained adopting a specification that does not include them are likely to be biased. However, the elasticity of output is fairly close to one when only time effects are considered, but drops to 0.87 when only firm fixed effects are present, and to 0.79 when both year and firm effects are included. Moreover, the "unbiased" model (where unbiased is the model including both year and firm effects which are found to be significant) delivers, again, an unrealistically low elasticity labor of 0.11 demonstrating that controlling explicitly for management bias does not necessarily improves on the credibility of the results.

The unsatisfactory results - low and often insignificant capital coefficients and unreasonably low returns to scale - obtained in the literature prove that the fixed effects framework, valid in theory, is not particularly successful in solving the endogeneity problem in practice. There are a number of reasons why this is the case. First, in order for the fixed effects methodology to be applicable, one needs to rely on the rather strong assumption that the unobserved productivity term ω_j is constant over time. This assumption is becoming less justifiable now that longer panel datasets are more easily available. Moreover, researchers are usually interested in studying major changes in the economic environment and, since significant changes are likely to affect different firms' productivity differently, firms are likely to adjust their optimal decision accordingly. If this is the case ω_j will obviously not be constant over time anymore. Second, when there is measurement error in inputs, the within transformation of the data through differencing may actually aggravate this problem and the estimates obtained with fixed effects are actually even less reliable than the OLS estimates. Nonetheless we can see the fixed effects approach as a useful and simple reduced-form way of exploring the data by decomposing the firms' heterogeneity into within and between effects.

2.2.2 Instrumental variables

The second classical solution to the endogeneity issue proposed in the literature is the use of instrumental variables. Consider a slight modification of (2.5)

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt} \tag{2.7}$$

where now the term ω_{jt} is allowed to change by firm and over time. Valid instruments would be variables that are correlated with the endogenous explanatory variables, in this case inputs, but do not enter the production function explicitly and are not correlated with the production function residuals. The theory of production provides some indication regarding natural candidates to be valid instruments: input prices. Input prices certainly influence input choices, as they are part of the input demand functions, but do not directly enter the production function. Moreover, input prices need to be uncorrelated with ω_j and this will depend on the nature of competition in the input market. Specifically, if input markets are perfectly competitive, firms take input prices as given, thus input prices are appropriate instruments. Other possible instruments would be output prices, once again under the condition that output markets are competitive, or any other variable that shifts either the demand for output or the supply of inputs. Nevertheless input prices are usually the most popular instruments because perfect competition is a more plausible assumption in input markets than in output markets.

In the wake of the "duality" revolution in production function theory, Nerlove (1963) is one of the very few successful contributions in the literature making use of input prices as instruments. His investigation on the returns to scale in electricity supply relays on several characteristics that render the U.S. electric power industry unique. Power cannot be stored in large quantities and must be supplied on demand; revenues from the sale of power by private companies depend primarily on rates set by public regulatory bodies; the input markets in this industry can be reasonably assumed to be competitive since fuel used in power production is purchased under long-term contracts at set prices, the industry is heavily unionized, and capital markets for utility companies are highly competitive. These features describe an industry where the output of a firm and the prices it pays for the production factors can be regarded as exogenous, even if the industry does not operate in perfectly competitive markets. Thus, the problem of the individual firm appears to be that of minimizing the total cost of production of output, subject to the given production function technology and factor prices. Specifically, if the production function is assumed to be a generalized Cobb-Douglas of the form:

$$Y_j = A_j K_j^{\beta_k} L_j^{\beta_l} F_j^{\beta_f} U_j \tag{2.8}$$

where capital K, labor L, and fuel F are the inputs of production and U is a residual expressing neutral variations in efficiency among firms, the problem of each firm j consists of minimizing the cost of production

$$\min_{K,L,F} \; C_j = p_{kj}K_j + p_{lj}L_j + p_{fj}F_j \tag{2.9}$$

subject to (2.8). The marginal productivity conditions associated with this problem are given by:

$$\frac{p_{kj}K_j}{\beta_k} = \frac{p_{lj}L_j}{\beta_l} = \frac{p_{fj}F_j}{\beta_f} \tag{2.10}$$

If the efficiency among firms varies neutrally, as indicated by the error term in (2.8), and the factor prices vary across firms, the input choices are not independent, but determined jointly by firm's efficiency, level of output and factor prices and the system of structural relations (2.8) and (2.10) suffers from simultaneity bias. However, if factor price data are available and factor prices do not move proportionally, it is possible to express the cost function as a reduced form of the system of equations (2.8) and (2.10)

$$c_j = k_j + \frac{1}{\gamma} y_j + \frac{\beta_k}{\gamma} \ln(p_{kj}) + \frac{\beta_l}{\gamma} \ln(p_{lj}) + \frac{\beta_f}{\gamma} \ln(p_{fj}) + v_j \tag{2.11}$$

where lowercase letters denote natural logarithms and $\gamma_j = \beta_k + \beta_l + \beta_f$. Assuming constant returns to scale, (2.11) can be rewritten as:

$$c_j - \ln(p_{fj}) = k_j + \frac{1}{\gamma} y_j + \frac{\beta_k}{\gamma} [\ln(p_{kj}) - \ln(p_{fj})] + \frac{\beta_l}{\gamma} [\ln(p_{lj}) - \ln(p_{fj})] + v_j \tag{2.12}$$

The fundamental duality between cost and production function, demonstrated by Shephard (1953), guarantees that the relation between the cost function empirically estimated and the underlying production function is unique. In other words, under the cost minimization assumption, cost functions and production functions are simply two different but equivalent ways of looking at the same concept.

Nerlove estimates (2.12) using data on 145 privately owned electrical utilities in 1955 and his main findings can be summarized as follows. There is substantial evidence of increasing returns to scale, but the degree of returns to scale varies inversely with output, especially for larger firms. The scale of operation affects the degree of returns to scale, but is does not significantly affect the marginal rate of substitution between factors of productions for given factor ratios. The elasticities of output with respect to labor and fuel are positive and of a plausible magnitude, while the elasticity of output with respect to capital is often very small and in some cases even negative.

Despite being a remarkable and innovative contribution, the peculiar environment and data used in Nerlove's study demonstrates why the instrumental variables approach, even if theoretically sound, may be challenging to apply in practice. First, firms do not usually report input prices and when they do, especially in the case of labor costs, they tend to report average wage per worker or per hour of labor. Ideally, the cost of labor should measure exogenous differences in labor market conditions, but it often capture also some component of unmeasured worker quality. It is very possible, for example, that firms employing higher quality workers will pay higher average wages. In this case the cost of labor will be correlated with the production function residuals and its validity as an instrument will be compromised. Second, the use of input prices an instruments requires that these variables have sufficient variation to identify production function coefficients. While input prices clearly change over time, they usually do not vary significantly across firms as inputs market conditions tend to be fairly national in scope. If input prices do not differ enough across firms in the data, or if the observed differences reflect unobserved input quality and not exogenous input market characteristics, the instrumental variable approach is not applicable. A third issue arises because the instrumental variables framework relies on the strong assumption that the term ω_{jt} in (2.7) evolves independently from input choices over time, thus firms cannot affect the evolution of ω_{jt} through input decisions. If this assumption does not hold, i.e. if the evolution in ω_{jt} is correlated with some inputs, finding valid instruments would require to identify variables that affect only those input choices without simultaneously affecting other input choices. Since individual input choices most likely depend on the prices of all inputs of production, the task to select valid instruments in such a context appears extremely challenging. Finally, the instrumental variables approach only addresses the endogeneity of input choice, not the endogeneity of firms' exit. If exit is endogenous, it will possibly depend, in part, on input prices so that firms facing higher input prices will be more likely to exit. This generates correlation between input prices, used as instruments, and the residuals in the production function rendering the instruments invalid.

2.3 Structural solutions

In the last twenty years, the increasing availability of firm-level data opened the door to more structural approaches to identifying production function coefficients controlling for simultaneity and selection problems.

2.3.1 The Olley and Pakes approach

Olley and Pakes, henceforth OP, in their 1996 contribution propose an innovative empirical framework with the goal of quantifying the impact of deregulation on measures of plant-level productivity in the U.S. telecommunication equipment industry between 1974 and 1987. Considering firms operating through discrete time, making production decisions to maximize the present discounted value of current and future profits, OP make use of the following assumptions. First, the production function is given by:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt} \tag{2.13}$$

where a_{jt} is the age (in years) of the plant expressed in natural logarithms. The motivation for introducing plant's age as an additional input is to analyze the impact of age on productivity.

Second, the unobserved productivity ω_{jt} evolves exogenously following a first-order Markov process of the form:

$$p(\omega_{jt+1}|\{\omega_{i\tau}\}_{\tau=0}^{t}\}, I_{jt}) = p(\omega_{jt+1}|\omega_{jt}) \tag{2.14}$$

where I_{jt} is firm j's information set at time t, and current and past realization of ω, i.e. $\{\omega_{jt}, \cdots \omega_{j0}\}$ are assumed to be part of I_{jt}. This is simultaneously an econometric assumption on the statistical properties of the unobservable term ω_{jt} and economic assumption on the way firms form their expectations on the evolution of their productivity over time. Specifically, at time $t+1$ these expectations depend only on the realization occurred at time t. Moreover the first-order Markov process is assumed to be stochastically increasing over time, i.e. a firm with a higher ω_{jt} today expects to have a better distribution of ω_{jt+1} tomorrow.

Third, capital is accumulated by firms through a deterministic dynamic investment process specified as:

$$k_{jt} = (1-\delta)k_{jt-1} + i_{jt-1} \tag{2.15}$$

This formulation implies that the firm's capital stock at period t was actually decided, through investment, at period $t-1$. Finally, the per-period profit function is given by:

$$\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - c(i_{jt}, \Delta_t) \tag{2.16}$$

Note that labor l_{jt} does not explicitly enter the profit function as it is considered a variable, non-dynamic input. Labor is variable in the sense that it is chosen and utilized in production in the same period and it is non-dynamic, unlike capital, because current labor decision do not impact future profits, i.e. labor is not a state variable. Therefore $\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ can be thought as a conditional profit function, where the conditioning is on the optimal static choice of labor input. Note also that both $\pi(\cdot)$ and $c(\cdot)$ depend on Δ_t which represents the economic environment where firms operate in a specific period. Δ is allowed to change overtime but, in a given time period, is considered to be constant across firms.

The firm maximization problem can be described by the following Bellman equation:

$$V_t(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) = \max \ \{\Phi, \sup_{i_{jt}\geq 0} \pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - c(i_{jt}, \Delta_t)$$
$$+\beta E[V_{t+1}(k_{jt+1}, a_{jt+1}, \omega_{jt+1}, \Delta_{t+1})|I_{jt}]\} \tag{2.17}$$

where Φ represents the sell-off value of the firm, and I_{jt} is, once again, the information available to the firm at time t, i.e. $(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t, i_{jt})$. The Bellman equation specifies that each firm compares its sell-off value and the expected discounted returns of staying in business. If the current state variables $(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ indicate that continuing in operation is not profitable, the firm will exit, while, in the opposite case it will choose an optimal, positive, investment level. Under the appropriate assumptions that an equilibrium exists and the difference in profits between continuing and exiting is increasing in ω_{jt}, i.e. firms with higher ω_{jt} are more likely to realize higher profits and thus decide to stay in business, the solution to the control problem in (2.17) generates an exit rule and an investment demand function. Defining χ_{jt} as the indicator function that takes the value of zero when the firm decides to exit we have that the exit decision rule and the investment demand function are written, respectively, as:

$$\chi_{jt} = \begin{cases} 1 & \text{if } \omega_{jt} \geq \bar{\omega}_{jt}(k_{jt}, a_{jt}, \Delta_t) = \bar{\omega}_t(k_{jt}, a_{jt}) \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

and

$$i_{jt} = i(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) = i_t(k_{jt}, a_{jt}, \omega_{jt}) \tag{2.19}$$

Investment is assumed to be strictly monotonic in ω as, conditional on k_{jt} and a_{jt}, firms with higher ω_{jt} will optimally invest more.

As long investment as is positive, since (2.19) is strictly monotonic in ω_{jt}, it is possible to invert it and generate

$$\omega_{jt} = h_t(k_{jt}, a_{jt}, i_{jt}) \tag{2.20}$$

which simply implies that, given a firm's levels of k_{jt} and a_{jt}, the investment demand i_{jt} provides sufficient information about ω_{jt}. This is because OP makes a scalar unobservable assumption, i.e. they assume that ω_{jt} is the only unobservable in the investment demand and there are no other unobservable (by the econometrician) variables that affect investment but not production.

Substituting (2.20) into (2.13) yields

$$y_{jt} = \beta_l l_{jk} + \phi_t(k_{jt}, a_{jt}, i_{jt}) + \eta_{jt} \tag{2.21}$$

where $\phi_t(k_{jt}, a_{jt}, i_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + h_t(k_{jt}, a_{jt}, i_{jt})$. Equation (2.21) is taken to the data in a first stage regression to recover an estimate of the labor coefficient. This is possible because the monotonicity and scalar unobservable assumption allows for "observing" the unobservable ω through investment eliminating the endogeneity problem for the labor coefficient. In this first stage of the estimation, however, the coefficients on capital and age are not identified because in (2.20) it is not possible to separate the effect of capital and age on the investment decision from their effect on output.

Rewriting (2.13) taking the term $\beta_l l_{jt}$ to the left-hand-side and taking expectation of both sides results in

$$\begin{aligned} E[y_{jt} - \beta_l l_{jt} | I_{jt-1}, \chi_{jt} = 1] &= E[\beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt} + \eta_{jt}] \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + E[\omega_{jt} | I_{jt-1}, \chi_{jt} = 1] \end{aligned} \tag{2.22}$$

The second line comes from the fact that k_{jt} and a_{jt} are known at time $t-1$ and η_{jt} is, by definition, uncorrelated with I_{jt-1} and exit. The last term of (2.22) can be expanded as:

$$\begin{aligned} E[\omega_{jt} | I_{jt-1}, \chi_{jt} = 1] &= E[\omega_{jt} | I_{jt-1}, \omega_{jt} \geq \bar{\omega}_t(k_{jt}, a_{jt}) \\ &= \int_{\bar{\omega}_t(k_{jt}, a_{jt})}^{\infty} \omega_{jt} \frac{p(\omega_{jt} | \omega_{jt-1})}{\int_{\bar{\omega}_t(k_{jt}, a_{jt})}^{\infty} p(\omega_{jt} | \omega_{jt-1})} d\omega_{jt} \\ &= g(\omega_{jt-1}, \bar{\omega}_t(k_{jt}, a_{jt})) \end{aligned} \tag{2.23}$$

where the first equality depends on the exit rule expressed in (2.18) and the last two lines from the exogenous first-order Markos process assumption on ω_{jt}.
While it is possible to estimate ω_{jt-1} since, from (2.21), for a given set of parameters $(\beta_0, \beta_k, \beta_a)$, $\hat{\omega}_{jt-1}(\beta_0, \beta_k, \beta_a) = \hat{\phi}_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}$, there is not direct knowledge of $\bar{\omega}_t(k_{jt}, a_{jt})$. OP try to control for $\bar{\omega}$ using data on observed exit. In fact, the probability of continuing operating at period t, conditional on the information available in the previous

period is

$$
\begin{aligned}
\Pr(\chi_{jt} = 1|I_{t-1}) &= \Pr(\omega_{jt} \geq \bar{\omega}_t(k_{jt}, a_{jt})|I_{jt-1}) \\
&= \Pr(\chi_{jt} = 1|\omega_{jt}, \bar{\omega}_t(k_{jt}, a_{jt})) \\
&= \tilde{\varphi}_t(\omega_{jt-1}, k_{jt}, a_{jt}) \\
&= \varphi_t(i_{jt-1}, k_{jt-1}, a_{jt-1}) \\
&= P_{jt}
\end{aligned}
\tag{2.24}
$$

where the first equality comes, again, form the exit rule (2.18) and the remaining equalities come from (2.20) and the fact that k_{jt} and a_{jt} are deterministic functions of i_{jt-1}, k_{jt-1}, and a_{jt-1}. OP obtain an estimate of $\hat{P}_{jt}$, i.e. the probability of firm j surviving to period t, through non-parametric methods.

Equation (2.24) also implies that $\bar{\omega}_t(k_{jt}, a_{jt})$ is a function of ω_{jt-1} and P_{jt}. Thus (2.22) becomes

$$
\begin{aligned}
E[y_{jt} - \beta_l l_{jt}|I_{jt-1}, \chi_{jt} = 1] &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\omega_{jt-1}, f(\omega_{jt-1}, P_{jt})) \\
&= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\omega_{jt-1}, P_{jt}) \\
&= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt})
\end{aligned}
\tag{2.25}
$$

The second stage in OP estimation requires to take to the data the following expression:

$$
y_{jt} - \beta_l l_{jt} = \beta_k k_{jt} + \beta_a a_{jt} + \tilde{g}'(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) + \xi_{jt} + \eta_{jt} \tag{2.26}
$$

where the function $\tilde{g}'$ includes the constant term β_0, and ξ_{jt} represents the innovation in productivity with $\xi_{jt} = \omega_{jt} - E[\omega_{jt}|\omega_{jt-1}, \chi_{jt} = 1]$. Substituting $\hat{P}_{jt}$, $\hat{\phi}_{jt}$, and $\hat{\beta}_l$, (2.26) can be estimated approximating the $\tilde{g}'$ function with polynomial or kernel methods. The coefficients associated with capital and age, β_k and β_a, can be identified in (2.26) because, given the information structure, the innovation in productivity is uncorrelated with k_{jt} and a_{jt} since these two variables are only function of the information at $t-1$, so that the orthogonality condition $E[\xi_{jt} + \eta_{jt}|I_{jt-1}, \chi_{jt} = 1] = 0$ holds. On the other hand, the labor input at time t is plausibly correlated with ξ_{jt} since it is free to adjust to shocks in productivity, thus the first stage of the estimation is needed to identify β_l. Finally, in (2.26) β_k and β_a are identified making use of the cross-sectional variation in k_{jt} and a_{jt} for firms with the same ω_{jt-1} and P_{jt} and the time variation in input usage across firms that have the same ω_{jt-1} and P_{jt}.

Olley and Pakes' findings demonstrate how important the bias created by not controlling for productivity and endogenous exit can be. Comparing the results obtained with their alternative method and the more classical OLS and fixed effect approaches, for both a balanced panel and the full sample (constructed be including exiting and entering firms), they find remarkable differences. Specifically, under OLS and fixed effects the coefficient on labor is overestimated and the coefficient on capital is heavily underestimated with respect to the OP approach and the differences are even larger when considering the balanced panel instead of the full sample. Qualitatively, the results show that changes in the regulatory structure of the telecommunication industry were followed by an increase in industry productivity generated, mainly, by a reallocation of capital and a shift in production towards more productive plants. Significant entry and exit appear to have facilitated this reallocation process.

2.3.2 The Levinson and Petrin approach

Levinson and Petrin (2003), LP henceforth, take a similar approach to Olley and Pakes for conditioning out serially correlated unobserved shocks in production function estimation. The key difference in their contribution is that they use an intermediate input demand function as a proxy for productivity instead of the investment demand function. The rationale behind this choice is that the OP procedure requires the investment function to be strictly monotonic in ω_{jt} in order to be inverted. Formally, the inversion can be done also in the presence of zero or lumpy investment levels, but zero or lumpy investment levels cast doubt on the strict monotonicity assumption on investment. On the other hand, restricting the sample to the sole observations for which $i_{jt} > 0$ could create a significant loss in efficiency. Specifically, LP observe that in the Chilean manufacturing dataset from 1979 to 1986 they use in their study more than fifty percent of the plant-year observations have zero investment level. Discarding these observations would imply loosing more than half of the sample with an obvious efficiency loss. In addition, LP note that investment is a control on a state variable that, by definition, may be costly to adjust. If investment is subject to non-convex adjustment costs, the investment function may present kinks that affect the reaction of investment to the transmitted productivity shock. In this case, the error term η_{jt} in (2.21) will be correlated with l_{jt} and the identification assumption on β_l would not hold.

To avoid the issues related to potentially large efficiency loss and adjustment cost non-convexities while using investment, LP suggest using intermediate inputs choices (energy or materials) to proxy for ω_{jt} as these variables are rarely zero and do not suffer from significant adjustment cost, thus the strict monotonicity assumption is more easily satisfied. They consider the production function

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt} + \eta_{jt} \tag{2.27}$$

where m_{jt} (intermediate input) is an additional input in production that, like labor, is assumed to be variable and non-dynamic. The intermediate input demand equation is specified as:

$$m_{jt} = m_t(k_{jt}, \omega_{jt}) \tag{2.28}$$

with the subscript t indicating that factors like input prices, market structure, or demand condition that can influence the demand for materials, are allowed to vary across time but not across firms. Note that (2.28) implies specific timing assumptions regarding the choice of m_{jt}. First, the intermediate input in period t is a function of ω_{jt}, i.e. it is chosen at the time production takes place. Second, labor does not enter (2.28) meaning that labor is chosen at the same time as intermediate inputs and, therefore, l_{jt} has no impact on the optimal choice of m_{jt}.
Assuming monotonicity of the intermediate input demand in ω_{jt}, analogously to OP, (2.28) is inverted to generate

$$\omega_{jt} = h_t(k_{jt}, m_{jt}) \tag{2.29}$$

Then, substituting (2.29) into (2.27) yields

$$y_{jt} = \beta_l l_{jt} + \phi_t(k_{jt}, m_{jt}) + h_t(k_{jt}, m_{jt}) + \eta_{jt} \tag{2.30}$$

The first stage of the LP procedure involves obtaining an estimate for β_l and $\phi_t(k_{jt}, m_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_m m_{jt} + h_t(k_{jt}, m_{jt})$ treating h_t non-parametrically. Note that β_k and β_m are not separately identified from the non-parametric function in this first stage. The second stage of LP consists of estimating

$$y_{jt} - \beta_l l_{jt} = \beta_k k_{jt} + \beta_m m_{jt} + \tilde{g}'(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_m m_{jt-1}) + \xi_{jt} + \eta_{jt} \tag{2.31}$$

Since k_{jt} is assumed to be decided at period $t-1$ it is orthogonal to the residual $\xi_{jt} + \eta_{jt}$. However, since m_{jt} is a variable input, it is certainly not orthogonal to the innovation component of productivity, ξ_{jt}, as ω_{jt} is observed at the time the intermediate input is chosen. Thus LP use its lag, m_{jt-1}, as an instrument for m_{jt}, such that the orthogonality condition $E[\xi_{jt} + \eta_{jt}|I_{jt-1}] = 0$ is satisfied for both k_{jt-1} and m_{jt-1}.

Levinson and Petrin find that using materials or electricity as a proxy for the unobserved productivity yields statistically significant estimates of the production function parameters for the Chilean manufacturing industry. Moreover, in line with Olley and Pakes, they also observe that, comparing their estimates with estimates obtained using OLS, the coefficient on labor is consistently upward biased and the opposite is true for the coefficient on capital. A final comparison between estimates obtained using the investment as proxy (OP method) and estimates using the intermediate input demand (LP method) delivers also higher coefficients on labor and lower coefficients on capital under the OP method suggesting that, at least in the case of the Chilean plants, the intermediate input seems to respond more fully to the productivity shock than investment.

2.3.3 The Ackerberg, Caves, and Frazer approach

Both OP and LP procedures rely on a crucial assumption regarding the nature of the labor decision, i.e. labor is not a state variable and, therefore, does not have implication in the firm's dynamic optimization problem. Nonetheless, Ackerberg, Caves, and Frazer, henceforth ACF, (2006) note that, if there are significant hiring or firing costs, or if a firm is highly unionized so that labor contracts are long term, current labor choices have dynamic implications and labor becomes a state variable.

In this case (2.20) and (2.29) become $\omega_{jt} = h_t(k_{jt}, l_{jt}, a_{jt}, i_{jt})$ and $\omega_{jt} = h_t(k_{jt}, l_{jt}, m_{jt})$, respectively, and the labor coefficient β_l cannot be identified from the first stage in neither OP or LP estimation because it is not possible to separate the impact of labor on the production from its impact on the non-parametric $h_t(\cdot)$ function. In other words, if the optimal labor choice is determined according to $l_{jt} = f_t(\omega_{jt}, k_{jt}) = f_t(h_t(\cdot), k_{jt})$, it is not feasible to simultaneously estimate a fully non-parametric, time-varying function of (k_{jt}, ω_{jt}) along with a coefficient associated with a variable, l_{jt}, that is merely a time-varying function of those same variables (k_{jt}, ω_{jt}). ACF further argue that the collinearity problem that prevents identification of β_l in the first stage is more serious and less easy to overcome in the LP approach than in the OP one. This is because the former method uses a proxy for productivity, the intermediate input demand, that is chosen in period t simultaneously with labor and production level after observing ω_{jt}, while the latter method relies on investment as a proxy that, by definition, is not directly linked to period t outcomes.

ACF suggest an alternative estimation procedure that avoids the collinearity problems arising in the estimation of the labor coefficient adopting a mild modification on the timing assumption for input choices. The main difference between their approach and OP and LP is that in the first stage no coefficient is estimated, instead the first stage serves the purpose of netting out the error from the production function η_{jt}.
More specifically, in the case of the intermediate input demand, with a production function specified as in (2.7), ACF assume that labor l_{jt} is chosen by firms at time $t-b$, with $(0 < b < 1)$, after capital k_{jt} was chosen at or before $t-1$ but prior to m_{jt} being chosen at t. The productivity process is assumed to evolve according to a first-order Markov process between the sub-periods $t-1$ and $t-b$, i.e.

$$p(\omega_{jt}|I_{jt-b}) = p(\omega_{jt}|\omega_{jt-b}) \quad \text{and} \quad p(\omega_{jt-b}|I_{jt-1}) = p(\omega_{jt-b}|\omega_{jt-1}) \tag{2.32}$$

(2.32) simply implies that labor and intermediates are both variable inputs with labor being "less variable" than intermediates. Also labor is not a function of ω_{jt}, but of ω_{jt-b}. With these timing assumption the demand for intermediate input and labor, respectively, are given by:

$$m_{jt} = m_t(k_{jt}, l_{jt}, \omega_{jt}) \tag{2.33}$$

$$l_{jt} = f_t(k_{jt}, \omega_{jt-b}) \tag{2.34}$$

and the collinearity problem has been solved as m_{jt} is now a function of $(k_{jt}, \omega_{jt-b}, \omega_{jt})$. Substituting into the production function, the first stage estimating equation in the ACF is given by:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + h_t(k_{jt}, l_{jt}, m_{jt}) + \eta_{jt} \tag{2.35}$$

where $h_t(\cdot)$ is, once again, the inverse of (2.33) used as proxi for ω_{jt}. As mentioned before, ACF run the first stage just to obtain an estimate of $\phi_t(k_{jt}, l_{jt}, m_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} +$

$h_t(k_{jt}, l_{jt}, m_{jt})$ in order to isolate η_{jt} and proceed to the second stage to estimate both β_k and β_l. This now requires two independent moment conditions for identification. The Markov process assumption on ω_{jt} implies that $\omega_{jt} = E[\omega_{jt}|I_{jt-1}] + \xi_{jt} = E[w_{jt}|\omega_{jt-1}] + \xi_{jt}$ so that ξ_{jt} is mean-independent of all the information known at $t-1$. Since k_{jt} is decided at $t-1$ and l_{t-1} is decided at $t-b-1$, both k_{jt} and l_{t-1} are included in the information set I_{t-1} so the orthogonality conditions required to identify β_k and β_l are

$$E\left[\xi_{jt} \middle| \begin{array}{c} k_{jt} \\ l_{jt-1} \end{array}\right] = 0 \tag{2.36}$$

For a given set of parameters (β_k, β_l), ACF compute the implied value of $\hat{\omega}_{jt}(\beta_k, \beta_l) = \hat{\phi}_{jt} - \beta_k k_{jt} - \beta_l l_{jt}$, then non-parametrically regress $\hat{\omega}_{jt}$ on $\hat{\omega}_{jt-1}$ to obtain $\hat{xi}_{jt}(\beta_k, \beta_l)$, and finally form the sample analogue

$$\frac{1}{T}\frac{1}{N}\sum_t \sum_n \xi_{jt}(\beta_k, \beta_l)\begin{pmatrix} k_{jt} \\ l_{jt-1} \end{pmatrix} \tag{2.37}$$

estimating (β_k, β_l) by minimizing (2.37).

In the case of the investment demand used as a proxy for ω_{jt} the ACF procedure is analogous as the procedure just described with the exception of (2.33) becoming

$$i_{jt} = h_t(k_{jt}, l_{jt}, \omega_{jt}) \tag{2.38}$$

and (2.35) becoming

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + h_t(k_{jt}, l_{jt}, i_{jt}) + \eta_{jt} \tag{2.39}$$

Once again $h_t(\cdot)$ is the inverse of (2.38) and $\phi_t(k_{jt}, l_{jt}, i_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + h_t(k_{jt}, l_{jt}, i_{jt})$.

ACF assert that their framework is completely consistent with labor choices having dynamic implications and with other unobservables (eg. input prices shocks or dynamic adjustment costs) impacting firm's choices of capital and labor. The results, obtained using the same Chilean manufacturing data as Levinson and Petrin, show how the estimates, obtained with their alternative procedure, differs significantly from both the classical OLS and LP methods. Specifically, the returns to scale estimated under OLS are higher than under ACF and this is mainly due to the fact that, as expected, the coefficient on labor is upward biased when not controlling for productivity shocks. Comparing ACF with LP estimates, they find that the coefficients are generally different in magnitude, with LP estimates of the labor coefficient being more often smaller than their ACF counterparts, suggesting that $\hat{\beta}_l$, which comes from the first stage in LP, may be downward biased because of the discussed collinearity issue.

2.4 Alternative solutions: dynamic panel models and endogenous productivity

2.4.1 Dynamic panel data methods

An alternative response to the simultaneity issues in production function estimation came from the dynamic panel data literature starting with Chamberlain (1982). Dynamic panel methods essentially extend the fixed effect framework allowing for a more sophisticated error structure and combine it with instrumental variables to control for collinearity. To see how this approach is developed consider the production function

$$
\begin{aligned}
y_{jt} &= \beta_k k_{jt} + \beta_l l_{jt} + (\alpha_j + \omega_{jt} + \eta_{jt}) && (2.40)\\
&= \beta_k k_{jt} + \beta_l l_{jt} + \psi_{jt} && (2.41)
\end{aligned}
$$

where the composite error term ψ_{jt} is the sum of all three error components, i.e. the unobserved, time-invariant firm-specific effect α_j, the productivity shock ω_{jt}, and the serially uncorrelated residual term η_{jt}.

The dynamic panel methodology relies on specific assumptions regarding the evolution of the error components α_j, ω_{jt}, and η_{jt}, and the correlation structure between these error components and the explanatory variables k_{jt} and l_{jt}. Given these assumptions, the estimation procedure requires finding functions of the aggregated error term ψ_{jt} that are uncorrelated with past, present, or future values of the explanatory variables. Commonly the imposed assumptions are as follows. First, the time invariant error component α_j may be correlated with capital and labor. Second, the term η_{jt} is i.i.d. over time and uncorrelated with capital and labor in every period. Third, the productivity process is usually modeled as a first-order linear autoregressive process of the form $\omega_{jt} = \rho\omega_{jt-1} + \xi_{jt}$. Lastly, while ω_{jt} is likely to be correlated with k_{jt} and l_{jt}, the innovation on ω_{jt} between $t-1$ and t, ξ_{jt}, is uncorrelated with all the input choices prior to period t. This is because the innovation in ω_{jt} is observed by the firm after period $t-1$ so that ξ_{jt} is uncorrelated with input chosen at $t-1$ or earlier. Note that the rationale behind this last assumption is similar to that behind the second stage identification conditions in OP, LP, and ACF. Given that ω_{jt} is AR(1), (2.40) has the dynamic common factor representation

$$\begin{aligned} y_{jt} = \beta_k k_{jt} - \rho\beta_k k_{jt-1} + \beta_l l_{jt} - \rho\beta_l l_{jt-1} \\ + \rho y_{jt-1} + ((1-\rho)\alpha_j + \xi_{jt} + \eta_{jt} - \rho\eta_{jt-1}) \end{aligned} \tag{2.42}$$

and given the definitions of ψ_{jt} and ω_{jt} the following function of ψ_{jt} can be specified:

$$(\psi_{jt} - \rho\psi_{jt-1}) - (\psi_{jt-1} - \rho\psi_{jt-2}) = \xi_{jt} - \xi_{jt-1} + (\eta_{jt} - \rho\eta_{jt-1}) - (\eta_{jt-1} - \rho\eta_{jt-2}) \tag{2.43}$$

Note that (2.43) contains only η_{jt} and the innovation in productivity, as the terms containing α_j have been differenced out. Moreover, since ξ_{jt} and ξ_{jt-1} have been assumed to be uncorrelated with the input choices prior to $t-1$ and η_{jt} is always uncorrelated with the the input choices, an appropriate moment condition for estimating β_k, β_l, and ρ would be

$$E\left[(\psi_{jt} - \rho\psi_{jt-1}) - (\psi_{jt-1} - \rho\psi_{jt-2}) \middle| \left\{ \begin{array}{c} k_{j\tau} \\ l_{j\tau} \end{array} \right\}_{\tau=1}^{t-2} \right] = 0 \tag{2.44}$$

(2.44) can be easily used to construct a sample analogue since (2.41) implies that, for given values of the parameter (β_k, β_l) any ψ_{jt} is observable. If the assumption that η_{jt} is uncorrelated with k_{jt} and l_{jt} in all time periods appears too strong, it can be substituted with the weaker assumption that η_{jt} is sequentially exogenous, i.e. uncorrelated with all inputs chosen prior to t. In this case, (2.44) does not hold anymore but can be substituted with

$$E\left[(\psi_{jt} - \rho\psi_{jt-1}) - (\psi_{jt-1} - \rho\psi_{jt-2}) \middle| \left\{ \begin{array}{c} k_{j\tau} \\ l_{j\tau} \end{array} \right\}_{\tau=1}^{t-3} \right] = 0 \tag{2.45}$$

because, while η_{t-2} is potentially correlated with k_{jt-2} and l_{jt-2}, it is uncorrelated with capital and labor decision prior to $t-2$.

(2.44) and (2.45) summarize the essence of the dynamic panel methodology, i.e. GMM estimators, which take first differences to eliminate firm-specific effects (α_j) and use lagged instruments to correct for simultaneity in the first-differenced equations, can be applied to estimate production function coefficients.

Blundell and Bond (2000), however, comment that the methodology described above tends to produce unsatisfactory results in the context of production function estimation. They mainly attribute this poor performance to the weak correlations that exist between the current growth rates of capital and labor and the lagged levels of these variables which result in weak instruments in the first-differenced GMM estimation procedure.
Using a panel of R&D performing US manufacturing firms between 1982 and 1989, Blundell and Bond propose to estimate (2.41) making use of (2.45) together with the following additional moment condition:

$$E\left[\left\{\begin{array}{c} k_{j\tau}-k_{j\tau-1} \\ l_{j\tau}-l_{j\tau-1} \end{array}\right\}_{\tau=2}^{t-3}\middle|\,\psi_{jt}\right]=0 \tag{2.46}$$

The moment condition in (2.45) allow the use of appropriately lagged levels of the variables as instruments in the first-differenced production function equation, while the moment condition in (2.46) allows the use of appropriately lagged first differences of the variables as instruments for the production function equation in levels. Both sets of moment conditions can be exploited as a linear GMM estimator in a system containing both first-differenced and levels equations. Their coefficient estimates obtained via the system GMM estimator are much more reasonable than the ones obtained under simple OLS, fixed effects, or the first-differenced GMM estimator with only the moment condition in (2.45). More precisely, the results show that the capital coefficient is higher and strongly significant, the additional instruments used in the system GMM are valid, and the hypothesis of constant returns to scale cannot be rejected in the data.

2.4.2 Estimating endogenous productivity

In all the frameworks presented so far the productivity process has been considered constant over a given period of time (fixed effects), or exogenous (structural approaches and dynamic panel methods), in the sense that firms' optimal decision do not affect the evolution in productivity. This modeling choice is mainly driven by the fact that endogenizing this process is problematic in the context of standard estimation procedures.
Nonetheless, it seems very reasonable to assume that firms can optimally choose to undertake activities to increment their productivity. A straightforward example is investment in R&D which generates knowledge-based assets accumulation, just like investment if physical capital, changing the firm's relative position with respect to other firms.

Doraszelski and Jaumandreu, DJ henceforth, in their very recent 2013 contribution develop a dynamic model of endogenous productivity change where firms carry out two types of investment, one if physical capital and another in knowledge through R&D expenditure. Firms operate through discrete time and make production decisions with the goal of maximizing the present discounted value of current and future profits. Each firm face a Cobb-Douglas production function of the form:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt} \tag{2.47}$$

Capital accumulation follows the investment process described in (2.15) and the usual i.i.d. across time and across firms assumption on η_{jt} holds. However, the key difference in this model is that the evolution of firm-level productivity over time is endogenized and productivity is assumed to be governed by a controlled first-order Markov process

$$p(\omega_{jt}|\omega_{jt-1}, r_{jt-1}) \tag{2.48}$$

with r_{jt} being R&D expenditure expressed in natural logarithms. The firm's dynamic maximization problem is summarized by the Bellman equation

$$V_t(k_{jt}\omega_{jt}) = \max_{i_{jt}, r_{jt}} \pi(k_{jt}\omega_{jt}) - C_i(i_{jt}) - C_r(r_{jt}) + \beta E[V_{t+1}(k_{jt+1}, \omega_{jt+1})|I_{jt}] \tag{2.49}$$

where $\pi(\cdot)$ denotes per-period profits, $C_i(\cdot)$ and $C_r(\cdot)$ are cost functions for investment and R&D, respectively, and I_{jt} represent the information set at time t that includes $(k_{jt}, \omega_{jt}, i_{jt}, r_{jt})$. The solution to (2.49) generates two policy function $i_t(k_{jt}, \omega_{jt})$ and $r_t(k_{jt}, \omega_{jt})$ for investment in physical capital and knowledge. The firms anticipates the effect of R&D on ω_t when making decisions about investment in knowledge in period $t-1$, thus the Markov process assumption yields

$$\omega_{jt} = E[\omega_{jt}|\omega_{jt-1}, r_{jt-1}] + \xi_{jt} = g(\omega_{jt-1}, r_{jt-1}) + \xi_{jt} \tag{2.50}$$

(2.50) shows that the productivity at period t, ω_{jt} can be decomposed into expected productivity $g(\cdot)$ and a random shock ξ_{jt}. While the expected productivity depends on R&D expenditure, the innovation in productivity ξ_{jt} does not. The term ξ_{jt} generally captures factors that persistently influence productivity such as absorption of techniques, modification of production precesses, fluctuations due to changes in labor composition and man-

agerial ability but, when firms engage in R&D the innovation in productivity also captures the uncertainties inherent in the knowledge accumulation process such as the chance of discovery, or the success in implementation. The timing structure behind firms' behavior is also emphasized in (2.50), i.e. in period $t-1$, when the optimal level of investment in knowledge r_{jt-1} is decided, the firm can only form an expectation regarding the impact of R&D on ω_{jt}, but the actual impact depend on the realization of ξ_{jt} that occurs only after the investment has been completely carried out.
The estimation of an endogenous productivity process is challenging when data on R&D are not available because the estimation procedures analyzed so far do not apply for the following reasons. First, input prices are invalid instruments because, when the productivity process is not exogenous, the transitions from current to future productivity are affected by the choice of the additional unobserved R&D input, whose optimal choice depends on the prices of all the other inputs. Second, the scalar unobservable assumption necessary for a structural estimation like OP, LP, or ACF is violated in this context because r_{jt} and ω_{jt} are both unobservable and recovering the productivity process using capital, investment, or intermediate input demand may not be possible. Furthermore, even when data on R&D are available, there may still be problems with identification as noted by Buettener (2005).

DJ estimation procedure relies on R&D data and builds on the LP insight that, since static inputs like labor and material are chosen once the current productivity realization is known, they contain useful information about it. More specifically, given the production function in (2.48), assuming that labor is a static input implies that the optimal demand for labor is given by:

$$l_{jt} = \frac{1}{1-\beta_l}(\beta_0 + \ln(\beta_l) + \mu + \beta_k k_{jt} + \omega_{jt} - (w_{jt} - p_{jt})) \tag{2.51}$$

where $\mu = \ln E[\exp(\eta_{jt})]$ and $(w_{jt} - p_{jt})$ is the real wage. Solving for ω_{jt}, the inverse labor demand function is

$$\omega_{jt} = h_t(k_{jt}, l_{jt}, (w_{jt} - p_{jt})) = -\beta_0 - \ln(\beta_l) - \mu + (1-\beta_l)l_{jt} - \beta_k k_{jt} + (w_{jt} - p_{jt}) \tag{2.52}$$

Lagging (2.52) yields $\omega_{jt-1} = h_{t-1}(k_{jt-1}, l_{jt-1}, (w_{jt-1} - p_{jt-1}))$ thus, substituting h_{t-1} for ω_{jt-1} into (2.50) and using it in the production function, JD take to the data the following estimating equation:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + g(h_{t-1}(k_{jt-1}, l_{jt-1}, (w_{jt-1} - p_{jt-1}), r_{jt-1}) + \xi_{jt} + \eta_{jt} \tag{2.53}$$

Note that in the above expression k_{jt}, whose value is determined in period $t-1$ through i_{jt-1} and r_{jt-1} is uncorrelated with ξ_{jt}, but the same is not true for l_{jt}. However, non-linear functions of the other variables, as well as lagged values of l_{jt} can be used as instruments for labor.

Applying a non-linear GMM technique, Doraszelski and Jaumandreu estimate (2.53) using an unbalanced panel of Spanish manufacturing firms during the 1990s. Their findings confirm the well established empirical pattern that, compated to OLS estimates, the labor coefficient labor decreases and the capital coefficient considerably increases when controlling for exogenous productivity. This result is even stronger when the productivity process is endogenized through R&D. Moreover, there is evidence of complementarities and increasing return to R&D, but the R&D process shows significant non-linearity and uncertainty. The returns to R&D are considerably higher than the returns to physical capital and the expected productivity of firms that engage in R&D is systematically more favorable compared to that of firms not performing R&D. Finally, R&D expenditures are found to be the primary source of productivity growth in the Spanish manufacturing sector, as firms investing in R&D contribute 65 to 85 percent to the productivity growth in the industries with intermediate or highly innovative activities.

2.5 Concluding remark

This survey has demonstrated that, even if production function estimation is challenging because of the possibility of simultaneity and selection bias, to obtain realistic and reliable estimates of production function coefficients, is the first step to answering more complex and interesting economic questions.

Since firms' responses to changes in the operating environment typically depend on how these changes affect their productivity, to separate the evolution in productivity from the variation in input choices, which also react to changes in the environment, requires an explicit model describing how firms' optimal choices are made. The appropriateness of different models and assumptions remains an empirical issue that needs to be addressed in each specific case, given the environment and the available data.

The literature has suggested a considerable variety of alternatives, nonetheless the common message emphasized in all the proposed approaches is that productivity studies must explicitly take into account the fact that changes in productivity are the main determinant of firms' response to the changes being analyzed, therefore, changes in productivity cannot be ignored in any estimation procedure.

3 Empirical Methodology

3.1 A structural framework to estimate production function coefficients

My first empirical goal is to obtain estimates of production function parameters. However, as well documented in the literature, this is not a simple task since, while firms' productivity is not directly observable, optimal input decisions are based on it. The fact that differences in productivity are known to firms when they choose their inputs and, for a given firm, productivity is usually highly correlated over time generates the classic simultaneity bias.

3.1.1 The empirical model

To address the simultaneity problem I rely on the insight of Olley and Pakes (1996), who propose to include directly in the estimation a proxy for productivity. This proxy is derived from a structural dynamic model of firm behavior that allows for firm-specific productivity differences, characterized by idiosyncratic changes over time, and specifies the information available to the firms when input decisions are made. Specifically, consider a firm j in industry i at time t (to simplify notation the industry subscript is omitted) producing output Q_{jt} according to the production function technology

$$Q_{jt} = F(X_{jt}, K_{jt}, \beta) \exp(\omega_{jt}) \tag{3.1}$$

where X_{jt} is a set of variable inputs, K_{jt} is capital stock, and β is a common set of technology parameters that governs the transformation of inputs to units of output in industry i. ω_{jt} is a firm-specific, Hicks neutral productivity shock.

Define value added as $Y_{jt} = Q_{jt} - M_{jt}$, with M_{jt} being intermediate inputs such as material and energy. Allowing for measurement error and for unanticipated shocks to production, the observed value added is given by $Y_{jt}\eta_{jt}$ and the value added industry-specific production function is

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{it} k_{jt} + \omega_{jt} + \eta_{jt} \tag{3.2}$$

where lower cases denote natural logarithms of the variables. Capital is a state variable accumulated accordingly to the deterministic dynamic investment process $k_{jt} = (1-\delta)k_{jt-1} + i_{jt-1}$. Note that this particular formulation of the capital accumulation process implies that period t capital stock was actually determined at time $t-1$. On the other hand, labor is assumed to be a perfectly variable input decided either at time t, when production takes place, or at time $t-b$, after capital but before production decisions occur. The importance of these assumption regarding the timing of input choices is related to identification and will become clear shortly. The error in (3.2) is assumed to be additively separable in the transmitted productivity component ω_{jt} and in the i.i.d. component η_{jt}. The main difference between these two components is that the former is assumed to be known by the firm when making optimal input choices while the latter is not so that η_{jt} simply represents a random optimization error. Note also that (3.2) is a translog production function but it easily allows to recover the Cobb-Douglas specification by dropping the higher order terms $(\beta_{ll}l_{jt}^2, \beta_{kk}k_{jt}^2)$ and the interaction term $(\beta_{lk}l_{jt}k_{jt})$.

In order to obtain consistent estimates of the production function coefficients, I directly control for unobserved productivity shocks, which are potentially correlated with labor and capital choices, adopting, again, the approach proposed by Olley and Pakes. Specifically, I use the investment function to proxy for productivity under the assumption that a firm's optimal investment demand, $i_{jt} = h_t(k_{jt}, \omega_{jt})$, is a strictly increasing function of its current productivity. The investment demand function contains all current state variables for the optimizing firm, i.e. its current level of capital and its current productivity. Conversely, labor does not enter the state space because it is a non-dynamic input and values of ω_{jt} prior to period t do not enter the state space either because the evolution of ω_{jt} is assumed to be governed by a first-order Markov process of the form $p(\omega_{jt}|\omega_{jt-1})$. Furthermore, the h function is only indexed by t (and not jt) since variables such as input prices and demand shifters, which may be also part of the state space, are allowed to vary only across time but not across firms as it is plausible to assume that firms operate in the same inputs market and under the same demand conditions. Given that the investment function is strictly monotonic in ω_{jt}, it can be inverted to obtain

$$\omega_{jt} = h_t^{-1}(k_{jt}, i_{jt}) \tag{3.3}$$

Following the same reasoning and maintaining the same assumptions on the evolution of the productivity process and the static/dynamic nature of the inputs, I also use the approach suggested by Levinson-Petrin (2003). They observe that investment levels are, in many cases, zero or very lumpy and propose to control for unobserved productivity using the intermediate input demand function as a proxy. In this case, if the optimal expenditure level in intermediates, $m_{jt} = f_t(k_{jt}, \omega_{jt})$, is assumed to be a strictly increasing function of the current productivity, it can be inverted to generate

$$\omega_{jt} = f_t^{-1}(k_{jt}, m_{jt}) \tag{3.4}$$

3.1.2 Estimation procedure

Equations (3.3) and (3.4) show that investment or, alternatively, intermediate input demand can be substituted into the production function as a proxy for the unobserved productivity term ω_{jt}, so that the estimating equation in (3.2) becomes

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{jt} k_{jt} + h_t^{-1}(k_{jt}, i_{jt}) + \eta_{jt} \tag{3.5}$$

or

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{jt} k_{jt} + f_t^{-1}(k_{jt}, m_{jt}) + \eta_{jt} \tag{3.6}$$

The estimation of (3.5) or (3.6) consists of two stages. The first stage serves the purpose of obtaining an estimate of the expected value added ϕ_{jt} and an estimate of η_{jt} alternatively running the following regressions:

$$y_{jt} = \phi_t(l_{jt}, k_{jt}, i_{jt}) + \eta_{jt} \tag{3.7}$$

or

$$y_{jt} = \phi_t(l_{jt}, k_{jt}, m_{jt}) + \eta_{jt} \tag{3.8}$$

where in (3.7) $\phi_{jt} = \phi_t(l_{jt}, k_{jt}, i_{jt}) = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{jt} k_{jt} + h_t^{-1}(k_{jt}, i_{jt})$, while in (3.8) $\phi_{jt} = \phi_t(l_{jt}, k_{jt}, m_{jt}) = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{jt} k_{jt} + f_t^{-1}(k_{jt}, m_{jt})$. In addition, the functions h_t^{-1} in (3.5) and f_t^{-1} in (3.6), are given by:

$$h_t^{-1}(k_{jt}, i_{jt}) = \bar{\beta}_k k_{jt} + \beta_i i_{jt} + \bar{\beta}_{kk} k_{jt}^2 + \beta_{ii} i_{jt}^2 + \beta_{ki} k_{jt} i_{jt} \tag{3.9}$$

and

$$f_t^{-1}(k_{jt}, m_{jt}) = \bar{\beta}_k k_{jt} + \beta_m m_{jt} + \bar{\beta}_{kk} k_{jt}^2 + \beta_{mm} m_{jt}^2 + \beta_{km} k_{jt} m_{jt} \tag{3.10}$$

Note that, due to the specification of (3.9) and (3.10), in the first stage the coefficients associated with capital and capital squared in (3.5) and (3.6), respectively, are not identified. These coefficients will be identified only in the second stage of the estimation using an appropriate set of moment conditions.

Moreover, under the Cobb-Douglas specification, i.e. $y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + h_t^{-1}(k_{jt}, i_{jt}) + \eta_{jt}$

with the investment demand, or $y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + f_t^{-1}(k_{jt}, m_{jt}) + \eta_{jt}$ with the intermediate input demand, the coefficient associated with labor, β_l, can be identified and estimated in the first stage as well.

In the second stage the (remaining) production function coefficients can be obtained relying on the Markov process assumption and the law of motion for productivity. More specifically, I model the productivity process non parametrically as a third degree polynomial of lagged productivity in the following way:

$$\omega_{jt} = \gamma_0 + \gamma_1 \omega_{jt-1} + \gamma_2 \omega_{jt-1}^2 + \gamma_3 \omega_{jt-1}^3 + \xi_{jt} \tag{3.11}$$

Using the estimated $\hat{\phi}_{jt}$ from the first stage, the value of productivity for any given vector of β, where $\beta = (\beta_l, \beta_k, \beta_{ll}, \beta_{kk}, \beta_{lk})$, can be computed as:

$$\omega_{jt}(\beta) = \hat{\phi}_{jt} - \beta_l l_{jt} - \beta_k k_{jt} - \beta_{ll} l_{jt}^2 - \beta_{kk} k_{jt}^2 - \beta_{lk} l_{jt} k_{jt} \tag{3.12}$$

By regressing $\omega_{jt}(\beta)$ on its lag $\omega_{jt-1}(\beta)$ it is possible to recover the innovation in productivity given by $\xi_{jt}(\beta)$. Specifically, denote $\beta Z_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_{ll} l_{jt}^2 + \beta_{kk} k_{jt}^2 + \beta_{lk} l_{jt} k_{jt}$, then the productivity process in (3.14) can simply be rewritten as $\omega_{jt}(\beta) = \hat{\phi}_{jt} - \beta Z_{jt}$ and the term ξ_{jt} in (3.13) is given by:

$$\begin{aligned} \xi_{jt}(\beta) =& \hat{\phi}_{jt} - \beta Z_{jt} - \gamma_0 - \gamma_1(\hat{\phi}_{jt-1} - \beta Z_{jt-1}) - \\ & \gamma_2(\hat{\phi}_{jt-1} - \beta Z_{jt-1})^2 - \gamma_3(\hat{\phi}_{jt-1} - \beta Z_{jt-1})^3 \end{aligned} \tag{3.13}$$

Equation (3.31) allows for calculating a $\xi_{jt}(\beta)$ term for every firm and every period which can be used in a GMM context to form appropriate moments in order to finally obtain estimates of the production function parameters.
More precisely, for the Cobb-Douglas specification, I carry on the estimate for β_l from the first stage and identify β_k using the moment condition on current capital

$$E[\xi_{jt}(\beta) k_{jt}] = 0 \tag{3.14}$$

The rationale behind the validity of this moment comes from the assumptions on the timing of input choices discussed above. Assuming that the optimal level of l_{jt} is chosen at time t, when also the innovation in productivity is known to the firm, implies that l_{jt} is correlated with ξ_{jt} and the coefficient on labor β_l needs to be identified in the first stage. Conversely,

the optimal level of k_{jt} is assumed to be chosen at time $t-1$, thus k_{jt} is not correlated with ξ_{jt} and the moment condition in (3.14) identifies the coefficient on capital β_k in the second stage of the estimation.
Regarding the translog production function I estimate the whole set of coefficients in the second stage relying on the moment conditions

$$E\left[\xi_{jt}(\beta)\begin{bmatrix} l_{jt-1} \\ k_{jt} \\ l_{jt-1}^2 \\ kjt^2 \\ l_{jt-1}k_{jt} \end{bmatrix}\right] = 0 \tag{3.15}$$

These moments exploit the following assumptions on the timing of input choices. Once again, current capital is assumed to be decided one period ahead therefore, at time t, k_{jt} is not correlated with the innovation in productivity ξ_{jt}. Lagged labor is used to identify the coefficient on labor if current labor, l_{jt}, is expected to react to shocks to productivity and hence $E[\xi_{jt}(\beta)l_{jt}]$ is expected to be different from zero. Thus, the moment conditions in (3.15) identify the whole set of coefficients $(\beta_l, \beta_k, \beta_{ll}, \beta_{kk}, \beta_{lk})$ in the translog production function. The standard errors of the estimated coefficients are obtained by block-bootstrap which is a special bootstrap technique designed to maintain the structure of the panel. Specifically, I bootstrap along the firm dimension, i.e. I randomly sample with replacement a number of firms equal to the number of firms present in each industry 400 times.

Two remarks regarding the estimation procedure are needed. First, I do not explicitly model entry and exit. This is because the panel I use is essentially closed given that, when a firm exited the sample, it was replaced by a similar firm and this new firm was assigned the same identifier as the exiting one. Consequently, it is not possible to keep track of entry and exit patterns and focus on selection issues. However, as Griliches and Mairesse (1998) and Levinson and Petrin (2003) note, the selection correction seems to make little difference once the simultaneity correction is in place. Second, I observe revenue instead of physical output, hence I actually estimate "revenue" production function parameters deflating the sales with an industry-wide price index. This is an imperfect solution since, if the unobserved firm-specific output price index substantially differs from the industry price index, I am actually introducing a price error. Furthermore, if input decisions are correlated with the price error, the estimated coefficients of the production function may be biased downward because, as mentioned in the original contribution by Klette and Griliches (1996), more inputs will lead to higher output and decrease prices, *ceteris paribus*. Nonetheless, this imperfect solution appears to be the best possible solution, given the limitations in the available data

3.2 A structural approach to derive firm-level markups

My second empirical goal is to derive markup estimates at the firm-level. To achieve this goal I follow the approach proposed by De Loecker and Warzynski (2012), which has the advantage of not depending on the availability of very detailed data. The data requirements, indeed, are limited to total expenditure on variable inputs (labor and materials), capital, investment and output at the firm-level. This approach is fairly direct from an economic theory perspective, since it relies on standard optimal input demand conditions that can be obtained from standard cost minimization. Moreover, it is straightforward to implement empirically, since the estimation is simply based on the insight that the cost share of factors of production are not equal to their output revenue share when markets are not perfectly competitive, so that the estimated markups can be interpreted as a measure of market power. Finally, this approach is flexible as it can be applied to a wide range of production functions and it is able to correct the markup bias by directly controlling for the firm-specific unobserved productivity.

To derive an expression for markups consider, once again, a firm j in industry i at time t (the industry subscript is again omitted for simplicity) producing output Q_{jt} using variable inputs $(X_{jt}^1, \dots, X_{jt}^V)$, which may include labor, materials, and energy, and capital K_{jt} as factors of production, and with productivity level ω_{jt}. This firm aims to minimize its cost of production by solving the problem

$$\begin{aligned} &\min_{X_{jt}, K_{jt}} \sum_{v=1}^{V} P_{jt}^{X^v} X_{jt}^v + r_{jt} K_{jt} \\ &\text{s.t. } Q_{jt} = Q_{jt}(X_{jt}^1, \dots, X_{jt}^V, K_{jt}, \omega_{jt}) \end{aligned} \tag{3.16}$$

where $P_{jt}^{X^v}$ denotes the price of any variable input and r_{jt} denotes the price of capital. The technology constraint takes a very general form and the only restriction imposed on $Q_{jt}(\cdot)$ is that it is continuous and twice differentiable with respect to its argument. The

Lagrangian associated with the minimization problem in (3.16) is given by:

$$\mathcal{L}(X_{jt}, K_{jt}, \lambda_{jt}) = \sum_{v=1}^{V} P_{jt}^{X^v} X_{jt}^v + r_{jt}K_{jt} + \lambda_{jt}(Q_{jt} - Q_{jt}(\cdot)) \tag{3.17}$$

with the first order condition with respect to each variable input being

$$\frac{\partial \mathcal{L}}{\partial X_{jt}^v} = P_{jt}^{X^v} - \lambda_{jt}\frac{\partial Q_{jt}(\cdot)}{\partial X_{jt}^v} = 0 \tag{3.18}$$

The Lagrange multiplier λ_{jt}, in this context, measures the marginal cost of production since $\frac{\partial \mathcal{L}}{\partial Q_{jt}} = \lambda_{jt}$. This is because, formally, λ_{jt} represents the shadow value of the constraint, i.e. the increase in cost generated by a marginal expansion in output.
Rearranging terms, multiplying both sides of (3.18) by X_{jt}/Q_{jt}, and dividing by λ_{jt} yields

$$\frac{\partial Q_{jt}(\cdot)}{\partial X_{jt}^v}\frac{X_{jt}^v}{Q_{jt}} = \frac{1}{\lambda_{jt}}\frac{P_{jt}^{X^v} X_{jt}^v}{Q_{jt}} \tag{3.19}$$

(3.19) simply states that cost minimization requires the optimal input demand being satisfied when a firm equalizes the output elasticity of input X_{jt}^v to $\frac{1}{\lambda_{jt}}\frac{P_{jt}^{X^v} X_{jt}^v}{Q_{jt}}$. Note that, in the special case of constant marginal cost, given the interpretation of the Lagrange multiplier, (3.19) implies that, at the optimum a firm equalizes the output elasticity of any variable input to its cost share.
Defining μ_{jt} as the the markup, implies that $\mu_{jt} = \frac{P_{jt}}{\lambda_{jt}}$ or, in a more compact way,

$$\mu_{jt} = \frac{\theta_{X_{jt}^v}}{\alpha_{X_{jt}^v}} \tag{3.20}$$

where $\theta_{X_{jt}^v}$ is the output elasticity with respect to the variable input X_{jt}^v and $\alpha_{X_{jt}^v}$ is the share of X_{jt}^v's expenditure in total revenue.

As mentioned before, the technology constraint in (3.16) is very general and can easily encompass different specifications. Assuming that the technology takes the form of the value added production function in (3.2) (where labor is the only variable input) and estimating the production function parameters following the procedure illustrated in the previous section, the estimated output elasticity of labor is given by $\hat{\theta}_{L_{jt}} = \hat{\beta}_l + 2\hat{\beta}_{ll}l_{it} + \hat{\beta}_{lk}k_{it}$ under the translog, and by $\hat{\theta}_{L_{jt}} = \hat{\beta}_l$ under the Cobb-Douglas specification. Additionally, the expenditure share is $\alpha_{L_{jt}} = \exp(\hat{\eta}_{jt}) \frac{P_{L_{jt}} L_{jt}}{P_{jt} \tilde{Y}_{jt}}$, where $\tilde{Y}_{jt}$ is observed value added, given by $Y_{jt} + \exp(\eta_{jt})$, allowing for measurement error. Note that the correction with the error term η_{jt} is important to eliminate any variation in the expenditure share that comes from variation in output not correlated with factor of production choices. Finally, with $\hat{\theta}_{L_{jt}}$ and $\alpha_{L_{jt}}$, the expression for the estimated markup for each firm in each period is derived as:

$$\hat{\mu}_{jt} = \frac{\hat{\theta}_{L_{jt}}}{\alpha_{L_{jt}}} \tag{3.21}$$

(3.21) remarks the rationale behind this approach for estimating markups, i.e. market power can be detected when the output elasticity of labor does not equalize the labor expenditure share.

3.3 A simpler dual approach to derive industry-level markups

The computational intensive methodology illustrated above allows for estimating firm-specific markups using disaggregated micro-level data. Here, I present a more parsimonious approach, that can be applied to more aggregated data, and is suitable to estimate industry-specific markups, i.e. all firms belonging to the same industry are assumed to share the same price-cost margin. The purpose of exploring this alternative method is to have a standard of comparison between a simpler and less demanding (in terms of data requirements and computational burden) approach and a more structural and onerous one.

The basic idea behind this unsophisticated approach is that, under certain assumptions, total factor productivity can be calculated either as the residual of the production function or, alternatively, as the residual of the dual cost function. However, the correlation between these theoretically equivalent measures is hard to verify empirically. Roeger (1995) argues that this lack of correlation can be explained by the presence of a positive markup of prices over marginal costs. In fact, with imperfect competition, the difference in the growth rate of output and a weighted average of the factor inputs cannot be entirely attributed to technical change. This is because, if price exceeds marginal cost, the input shares per unit of output do not sum up to one and are lower, instead, because of the presence of a positive markup.

Formally, consider an industry i characterized by a linearly homogeneous production function. The value added for this industry at time t is given by $Y_{it} = \Theta_{it} F(L_{it}, K_{it})$ where L_{it} is labor, K_{it} is capital and Θ_{it} is an industry- and period-specific shock in production.

Note that the productivity term Θ_{it} can be thought of as including an unanticipated and random element as well as an element that can be foreseen by all the firms in the industry. Carrying on the insight of Hall (1988), the decomposition of the Solow residual (SR) into a pure technology component and a markup component can be formulated as:

$$SR_{it} = \hat{Y}_{it} - \alpha_{it}\hat{L}_{it} - (1-\alpha_{it})\hat{K}_{it} = \beta_{it}\left(\hat{Y}_{it} - \hat{K}_{it}\right) + (1-\beta_{it})\hat{\Theta}_{it} \tag{3.22}$$

where the hat represents growth rates, P_{it}^{L} and P_{it}^{Y} are price of labor (i.e. wage) and price of output, respectively, and $\alpha_{it} = \frac{P_{it}^{L}L_{it}}{P_{it}^{Y}Y_{it}}$ is the labor expenditure share in total value added expressed in growth rates. In this context, market power can be recovered from the Lerner index $\beta_{it} = \frac{P_{it}^{Y}-c_{it}}{P_{it}^{Y}} = 1 - \frac{1}{\mu_{it}}$, with c_{it} denoting the marginal cost, and $\mu_{it} = \frac{P_{it}^{Y}}{c_{it}}$ the price-cost markup. While value added, input factor usage, and input shares can be easily observed in the data, the Lerner index and the productivity shock cannot. Hence, the estimation of β_{it} in (3.22) is problematic because $(\hat{Y}_{it} - \hat{K}_{it})$ and $\hat{\Theta}_{it}$ are positively correlated since optimal input decisions (in this case concerning capital) are made taking into account the partially known productivity shock. A possible solution to this problem would require identifying appropriate instruments that are correlated with output, but are neither a consequence nor a cause of technological innovation yet, as well documented in the literature, finding such instruments is a difficult task.
To deal with this issue, Roeger derives the dual price-based Solow residual (SPR):

$$SPR_{it} = \alpha_{it}\hat{P}_{it}^{L} + (1-\alpha_{it})\hat{R}_{it} - \hat{P}_{it}^{Y} = -\beta_{it}\left(\hat{P}_{it}^{Y} - \hat{R}_{it}\right) + (1-\beta_{it})\hat{\Theta}_{it} \tag{3.23}$$

where R_{it} represents the rental rate of capital. Subtracting (3.22) from (3.23), the *net* Solow residual is given by:

$$\begin{aligned} SR_{it} - SPR_{it} &= \left(\hat{Y}_{it} + \hat{P}_{it}^{Y}\right) - \alpha_{it}\left(\hat{L}_{it} + \hat{P}_{it}^{L}\right) - (1-\alpha_{it})\left(\hat{K}_{it} + \hat{R}_{it}\right) \\ &= \beta_{it}\left[\left(\hat{Y}_{it} + \hat{P}_{it}^{Y}\right) - \left(\hat{K}_{it} + \hat{R}_{it}\right)\right] \end{aligned} \tag{3.24}$$

(3.24) can be further rewritten to obtain a direct measure of the price-cost markup, i.e.

$$\left(\hat{Y}_{it} + \hat{P}_{it}^{Y}\right) - \left(\hat{K}_{it} + \hat{R}_{it}\right) = \mu_{it}\left[\alpha_{it}\left(\left(\hat{L}_{it} + \hat{P}_{it}^{L}\right) - \left(\hat{K}_{it} + \hat{R}_{it}\right)\right)\right] \tag{3.25}$$

Note that (3.25) simply states that the markup captures the difference between the net change in nominal value added and the net change in nominal labor payments weighted by the labor share in value added, where net means that the change in nominal capital has been netted out from both variables. Also note that the term $(1 - \beta_{it})\hat{\Theta}_{it}$, causing the endogeneity issue in (3.22) and, potentially, in (3.23), does not appear in (3.25) so that this equation can be consistently estimated without using instrumental variables. Moreover, (3.25) provides a way of estimating markups indirectly controlling for (i.e. netting out) productivity.

From (3.25) it is also clear that markup estimates at the industry-level can be easily obtained using only aggregated data on the nominal value added (calculated as the nominal value of sales minus the nominal value of materials), the total labor remuneration in nominal terms, and the nominal value of capital (calculated as the product between the real capital stock and the nominal interest rate).

4 Trade Exposure and Price-Cost Margins

4.1 The Mexican case

From the early 1950s until the early 1980s Mexico[2], like many other developing countries, adopted a growth strategy based on import substitution. Relying on protection measures against world competition and government intervention in the domestic economy, this strategy encouraged investment in industry, suppressed agricultural prices and expanded government enterprises. Between 1960 and 1981 Mexico experienced an average increase of real GDP of 7 percent per year; even accounting for the high rate of population growth in Mexico over the that period, this translated into an average increase of GDP per capita of 4 percent per year. During the 1970-1982 period, however, the import substitution policy began to be less effective as policies of deficit spending and monetary expansion financed by public sector borrowing from international banks were implemented. As a result, Mexico experienced rising inflation, which together with a fixed nominal exchange rate, led to substantial real exchange rate appreciation and growing current account deficit. Despite the substantial economic imbalances, the Mexican economy continued to expand on an average growth rate of real GDP of 6.2 percent over 1970-1982.

In 1982 the import substitution policy, and the Mexican economy with it, fell apart. Faced with a massive public debt owned by foreign banks, sharply rising international interest rates, and falling oil prices due to the worldwide recession, Mexico could not meet its debt obligations. The peso collapsed, the government nationalized banks and implemented strict exchange rate controls and the economy entered a deep recession.
In late 1982, under newly elected President Miguel de la Madrid Hurtado, Mexico embarked on its first steps on the long road to recovery. During the 1983-85 period, with the financial support of the International Monetary Fund, the new administration implemented a series of policies designed to cut the public sector deficit and turn the large trade deficit into a surplus. These policies included reduction in government expenditures, increases in taxes and in the prices of public services, elimination of many subsidies and closure of some public enterprises, enforcement of license requirements for all imports and the abolition of the exchange controls. Although this program was successful in creating a trade surplus and in partially lowering inflation, it was not enough to prevent another crises. In late 1985 fiscal discipline began to waver, IMF funding ended, an earthquake in Mexico City caused

[2]The following data on the Mexican economy as well as the main features of the trade reform are taken from Kehoe (1995).

disruption and imposed significant costs, and the oil prices started on a steep decline that continued until 1987.

The 1985-87 period was characterized by falling output and accelerating inflation. It was during this period, however, that Mexico began some of the policy reforms that were crucial in bringing deficit and inflation rate to acceptable levels and restoring economic prosperity during the 1987-93 period. Major initiative included privatization of state-owned companies, deregulation of financial markets, liberalization of foreign investment regulations and a dramatic re-orientation of trade policy.

The trade policy reforms were perhaps the most striking. In 1985 the process of *apertura*, openness to foreign trade and investment, began and between 1982 and 1994 Mexico went from being a relatively closed economy, even for developing countries' standards, to be one of the most open in the world. In 1982 tariffs were as high as 100 percent and there was substantial dispersion in tariff rates. Licenses were required for importing any good and, as a general rule, foreigners were restricted to no more than 49 percent ownership of Mexican enterprises. In 1982 import licenses, not tariffs, were Mexico's most significant trade barriers. Starting in late 1983 quantitative restrictions were replaced with tariffs. The portion of tariff items subject to license requirements fell from 100 percent in 1983 to 65 percent in 1984 and reached 10 percent in 1985. By 1992 it was just 2 percent. Even so, the portion of the value of imports subject to license requirements fell more slowly: from 100 percent in 1983 to 83 percent in 1984, to 35 percent in 1985, to 11 percent in 1992. As import licenses were replaced by tariffs as the major tool of trade policy, average tariffs initially rose and then fell. The average tariff went from 23.2 percent in 1983 to 25.4 percent in 1985, to 13.1 in 1992. The trade-weighted average tariff went from 8.0 percent in 1983 to 13.3 percent in 1985, to 11.1 percent in 1992.
Equally effective with the change in average tariff rates was the simplification of the tariff schedule. These measures were major steps in making the Mexican trade policy less protective and more transparent. They were accompanied by a number of other supporting policies: in 1986 Mexico acceded the GATT adopting the Harmonized Commodity Description and Coding System, the Foreign Trade Law and the GATT Anti-Dumping Code. In short, in about five years Mexico dramatically liberalized its trade regime. The liberalization process was almost complete by the end of 1987, although the impact on the flow of imports was softened by real devaluations. The reforms helped to promote exports. In terms of both import penetration and export rates, the manufacturing sector was substantially open as a consequence.

4.2 Relating markups and measures of trade liberalization

Prior to the consistent liberalization started in 1983, trade accounted for a small share of manufacturing production in most Mexican industries. Both the ratio of imports over domestic consumption and the ratio of exports over domestic production were below 10 percent. Nonetheless, as a consequence of the rapid and dramatic process of foreign trade and investment liberalization, in merely a decade Mexico went from being a relatively closed economy, even for developing countries' standards, to be one of the most open in world. In order to investigate whether this outward-looking reform generated import discipline I relay on two simple models that allow for quantifying the impact of trade liberalization on the price-cost margins. The first model is a variant of Domowitz, Hubbard and Peterson (1986) and it is suitable for a industry-level analysis on the markups. Consider the relation:

$$\mu_{it} = f\left(H_{it}, TRADE_{it}, H_{it} * TRADE_{it}, KQ_{it}, I_i, T_t\right) \tag{4.1}$$

where the explanatory variables include a measure of industry structure, the Herfindahl index H_{it}, a measure of trade liberalization, the industry-level capital-output ratio KQ_{it}, as well as industry I_i and time T_t controls. Regarding the measure of trade exposure I alternatively use the share of total industrial output falling into commodity categories subject to import licenses ($QUOTA_{it}$), and the production-weighted official tariff rate ($TARIFF_{it}$). When the industry dummies are not included, most of the variation occurs across industries and the Herfindahl index and the capital-output ratio should identify variations in technology and the degree of competition among domestic producers. If a pro-competitive effect of trade exposure exists, it should manifest as a negative correlation between measures of trade liberalization and markups. Moreover, if highly-concentrated industries do not operate under perfect competition, they should be relatively more sensitive to foreign competition, therefore the interaction term between the Herfindahl index and the trade indicator should reflect the same negative relation between trade openness and price-cost margins.

With panel data it is possible to further control for persistent differences across industries in technology and market structure by including industry dummy variables. In this case, the estimated coefficient reflect only temporal variation in the data and, since measures of trade policy change through time, price-cost margins regressions including industry dummies are better suited to capture the disciplining effect of trade liberalization.

The second model, proposed by Schmalensee (1985), aims to explain the extent of competition within a given industry by studying firm-level margins. The rationale behind this second exercise is to detect whether cross-firm variations are due to industry-wide effects or to firm-specific market shares. In general, more efficient firms should be larger and have higher profits, therefore a positive relation between market shares and price-cost margins is usually expected and it is not necessarily an indication of market power, as emphasized by Demsetz (1973) in his famous critique. However, if the degree of market power varies across industries, industry dummies should capture this source of difference in firm-level profitability. If industry dummies are not significantly different across industries, the evidence suggests absence of heterogeneity in market power. For the purpose of verifying the effect of trade liberalization on profitability at the firm level consider the following specification:

$$\mu_{jit} = f\left(S_{jit}, S^2_{jit}, TRADE_{it}, S_{jit} * TRADE_{it}, KQ_{jit}, KQ^2_{jit}, I_i, T_t\right) \tag{4.2}$$

where the price-cost margin μ_{jit} of firm j in industry i in year t depends on its share of output in total domestic manufacturing production, S_{jit} and S^2_{jit}, on the capital-output ratio KQ_{jit}, on an industry-specific measure of trade exposure, $TRADE_{it}$, as well as industry and year dummy variables.

4.3 Relating markups and export status

The two previous models relate markups with trade exposure using trade indicators that mainly capture the extent of import liberalization. In fact, both the quota coverage and the average tariff rate measure protective restrictions on imports. A number of recent models of international trade, however, emphasize the implications of trade openness, productivity and profitability for exporters. More specifically, these models generate the result that more productive firms are more profitable because they can charge higher markups, the higher profitability allows those firms to pay an export entry cost and become exporters, thus exporters have usually higher markups.

In the literature two main reasons for this positive relation between markups and firm's export status have been identified. Bernard, Eaton, Jensen, and Kortum (2003), as well as Melitz and Ottaviano (2008) attribute the source of the markup premium for exporters to differences in productivity. Both contributions essentially predict that exporters will charge higher markups because they are more productive than their domestic rivals and this productivity wedge allows them to be more profitable and more competitive. On the other hand, Kugler and Verhoogen (2011) and Hallak and Sivadasan (2009) explore the role of quality differences between exporters and non exporters assuming that if exporters produce higher quality goods while using higher quality inputs, they can charge higher markups, all things equal.

Given that with the structural approach I can estimate firm-level markups, I can easily relate a firm's markup to its export status in a regression framework as follows:

$$\ln(\mu_{jt}) = \psi_0 + \psi_1 E_{jt} + z_{jt}\rho + \varepsilon_{jt} \tag{4.3}$$

where μ_{jt} is the markup for firm j at time t and E_{jt} is a dummy variable that takes the value of one when firm j is an exporter. Thus, the coefficient associated with this dummy, ψ_1, measures the percentage markup premium for exporters. In addition, z_{jt} is a set of variables including capital and labor use that control for differences in size and factor intensity, as well as year- and industry-specific dummy variables that control for aggregate trends in markups. The vector ρ collects the coefficients associated with the whole set of controls.

After obtaining an estimate for ψ_1 it is possible to recover the level markup difference, denoted as μ_E, by calculating the percentage difference with respect to the constant term ψ_0, which captures the markup average for domestic firms. Specifically, $\mu_E = \psi_1 \exp(\psi_0)$. A positive and significant μ_E would imply that there is in fact a markup premium for exporters with respect to the domestic producers.

5 Data

My entire analysis is conducted using plant-level panel data collected through the Mexico's Annual Industrial Survey INEGI. These data were made available by Mexico's Secretary of Commerce and Industrial Development SECOFI, (now Secretariat of Economy) and includes a sample of active Mexican manufacturing plants during the period 1984-1990. For a typical industry, the sample is representative of about 80 percent of the total output in that industry therefore, even if the smallest plants were excluded, the sample can be considered fairly representative. Note that, because of the way the data are reported, it is not possible to identify which plants belong to the same firm. Therefore, even if there are certainly multi-plant firms in the sample, I formally treat a plant as a firm and do not try to capture the extent to which multi-plant firms may have a different strategic production behavior due to their multi-plant nature. For this reason the words "firm" and "plant" are used interchangeably as, in this dataset, they are essentially the same. Furthermore, as mentioned before, when a firm exited the sample, it was replaced with a firm with similar production characteristics and the new firm was assigned the same identifier as the exiting one. Thus the panel can be considered essentially closed as it is not possible to keep track of entry and exit patterns. The panel is however unbalanced since a (marginally) decreasing number of firms is included in the sample over the years.

For each plant in each year it is possible to observe data on value of production, revenue, input expenditure, labor remunerations,value of fixed capital, investment, inventories, and input costs. Each plant can be traced and uniquely identified over time using a combination of industry (RAMA), class (CLASE) and plant (FOLIO) identity codes. The dataset also contains price indices at the industry level for output and intermediate inputs, and sector-wide deflators for machinery and equipments, buildings and land. Moreover, the dataset contains detailed information about imports, exports, and commercial policy features like coverage of import license and tariff rates at the industry level. This information is particularly useful to describe the Mexican trade liberalization process and to verify its effects on the price-cost margins.

5.1 Data preparation: relevant variables and sample selection criteria

The original sample consisted of 22,526 observations on 3,218 plants during the period 1984-1990. All the variables used in the analysis are reported in Table 1, the monetary variables were converted to millions of 1980 Mexican pesos using specific deflators.

Table 1: Variables from the original dataset used in the analysis

Variable	Description
Labor force	
TOTREMUN[1]	total labor remunerations
TOHHOM[1]	total hours worked
TOPEOC	total employment
Inputs costs	
TOTMASUM[5]	total material cost
GTRENALQ[1]	rent and leasing costs
VAENELCN[6]	value of electricity consumed
GASTMAQU[1]	cost of subcontractors
Value of output	
VALPROEL[2]	value of output
Revenue	
INSERMAQ[1]	income from subcontracting
Fixed capital	
V68[4]	machinery and equipment valued at replacement cost
V92[4]	machinery and equipment produced for own use
V69[3]	construction and installation valued at replacement cost
V93[3]	construction and install assets produced for own use
V70[1]	land valued at replacement cost
V71[7]	transportation equipment valued at replacement cost
V94[7]	transportation equipment assets produced for own use
V72[1]	other assets valued at replacement cost
V95[4]	other asset produced for own use
Trade indicators	
TAI630	average tariff on input (June 30)
TAI1230	average tariff on input (Dec. 30)
TAQ630	average tariff on output (June 30)
TAQ1230	average tariff on output (Dec. 30)
LCI630	license coverage on input (June 30)
LCI1230	license coverage on input (Dec. 30)
LCQ630	license coverage on output (June 30)
LCQ1230	license coverage on output (Dec. 30)
Price indices	
[1]PM	wholesale price index
[2]PPP	producer price index
[3]PKE	construction price index
[4]PK	machinery price index
[5]PMP1	raw materials price index
[6]PEMP	electricity price index
[7]PKT	transportation price index

In addition, I use some of the original variables present in the dataset to construct new variables useful for the analysis. These variables, their description and the calculation details are reported in Table 2.

First, the expenditure in intermediates was calculated without taking into account inventories. This choice was dictated by the fact that in 1985 the variables characterizing inventories presented many missing observation thus, following one of the sample selection criteria (described in detail below) of withdrawing incomplete series, to consider inventories would have caused the elimination of half of the plants from the analysis.

Second, total capital stock for each plant was calculated as the sum of replacement cost of capital and the capitalized value of the rents with a 10 percent discount rate.

Third, the variables involved in the calculation of the value added and the value added itself were corrected in order to account for the measurement error in the intermediate inputs expenditure for the *maquiladoras*[3]. Specifically, the value added was corrected adding the income from subcontracting and subtracting the cost of subcontractors. The gross value of output, which suffers from the same bias, was corrected under the assumption that the ratio between value added and output and between primary materials and total inputs are constant through time and among plants running the following regression[4]: CORGVO=GVO+(b((INSERMAQ/PM)-(GASTMAQU/PM))). The value of the b[5] parameter used in the correction was estimated at a two-digit national accounts classification level (RAMA) using only the plants that did not conduct *maquila* activities. Finally, the corrected value of expenditure in intermediates was simply calculated by subtracting the corrected value added from the corrected value of gross output.

Around 20 percent of the original observations were eliminated discarding negative, zero, and missing values of the following variables: total employment, total hours worked, capital stock, gross value of output, corrected gross value of output, value added, corrected value added, intermediates, corrected intermediates, labor remunerations. This process ended up with the elimination of 4,234 observations. Among the 18,292 observations left, other 4,924 were eliminated dropping the incomplete series, i.e. plants that were discarded

[3]The *maquiladora* is a firm-concept very diffused in Mexico.*Maquiladoras* are manufacturing firms that are allowed to import materials and equipment on a duty-free and tariff-free basis for assembling of manufacturing and then sell the assembled or manufactures products to the domestic firm which commissioned the *maquila* service or re-export the products outside the Mexican border. The *maquiladoras* generate measurement error because the Mexican accounting system attributes to the firm that order the subcontracting service the expenditure in intermediates actually used by the subcontractor.

[4]See Table 2 for a detailed description of the variables involved in this regression.

[5]The average value of b is 1.47 with standard deviation 0.12.

Table 2: Variables constructed

Variable	Description	Calculation
GVO	Gross value of output	GVO=(VALPROEL/PPP)+(V92/PKM) +(V93/PKE)+(V94/PKT)+(V95/PKM)
INT	Intermediates	INT=(TOTMASUM/PMP1)+(VAENELCN/PEMP)
VA	Value added	VA=GVO-INT
CORVA	Corrected value added	CORVA=VA+(INSERMAQ/PM) -(GASTMAQU/PM)
CORGVO	Corrected gross value of output	CORGVO=GVO+(b*((INSERMAQ/PM) -(GASTMAQU/PM)))
CORINT	Corrected intermediates	CORINT=CORGVO-CORVA
TRCK	Total replacement cost of capital	TRCK=(V68/PKM)+(V69/PKE)+(V70/PM) +(V71/PKT)+(V72/PM)
KSTOCK	Capital stock	KSTOCK=TRCK+((GTRENALQ/PM)/0.10)
INVEST	Investment	INVEST=$\text{KSTOCK}_t - 0.9{*}\text{KSTOCK}_{t-1}$
TLPM	Deflated total labor remunerations	TLPM=TOTREMUN/PM
TLPMPH	Labor remunerations per hour	TLPMPH=TLPM/TOHHOM

for at least one year because of one or more of the above reasons were completely eliminated from the sample. The final sample used in the analysis contained 13,368 observations on 2,088 plants. Moreover, in order to carry on the structural production function estimation using the investment as a proxy for productivity, 2,092 observations were further dropped in order to create the investment series. Finally, two sectors, Tobacco and Nonferrous metals, were dropped because the extremely low number of plants left after the sample selection was not adequate to perform a meaningful empirical analysis in those two industries.

5.2 Sample characteristics

Table 3 reports in detail the industrial classification codes aggregated into each sector, the average number of plants in each sector as well as some other characteristics that describe the relative importance of each sector in the total manufacturing output and the openness to trade. As shown in Table 3 there is substantial heterogeneity in all these characteristics among the Mexican manufacturing industries.

Table 4 summarizes the data by presenting the number of plants and various indicators of plant size pooling all the manufacturing plants during the period 1985-1990. Except for 1986, average plant growth is positive whether measured by gross output, value added, or total employment and it is particularly high in the last 2 years included in the sample. Average capital stock per plant decreases from 1985 to 1986 probably as a consequence of the physical destruction caused by the earthquake of 1985 and the low level of net

Table 3: Industry-specific indicators

Rama	Sector	Industry	# of plants	Share of output	Share of imports	Share of exports	% of plants exporting
11-19	1	Food	226	11.66	7.24	3.09	0.20
20-22	2	Beverages	108	8.42	2.59	2.64	0.22
23	3	Tobacco	6	1.35	2.58	0.22	0.10
24-26	4	Textiles	103	2.25	10.00	8.17	0.24
27	5	Clothing and Apparel	81	0.82	19.22	2.29	0.10
28	6	Leather and Footwear	19	0.18	1.62	58.00	0.38
29-30	7	Wood and Furniture	61	0.64	3.42	5.25	0.13
31-32	8	Pulp and Paper	117	4.80	15.22	1.82	0.13
33-40	9	Chemicals	277	15.72	12.54	11.27	0.42
41-42	10	Plastic and Rubber	159	3.21	14.54	4.51	0.21
43	11	Glass	22	3.07	6.90	15.73	0.62
44	12	Cement	27	2.71	2.05	9.06	0.39
45	13	Nonmetal Minerals	95	1.22	9.68	3.01	0.13
46	14	Iron and Steel	73	10.58	3.10	4.88	0.24
47	15	Nonferrous Metals	6	3.68	2.77	48.68	0.51
48-50	16	Metal Products	106	2.87	14.34	8.80	0.32
51	17	Nonelectrical Machinery	116	1.85	28.80	26.14	0.29
52 -55	18	Electrical Machinery	109	5.49	20.52	8.60	0.41
56-58	19	Transport Equipment	116	19.06	19.54	39.94	0.44
59	20	Other Manufacturing	46	0.54	16.92	4.14	0.20

Note: The share of output is reported as average over the sample period. The shares of imports and exports are calculated as shares of total sectoral imports and exports, respectively, over sectoral output and are reported as averages over the sample period.

investment during the recession of 1986. Its upward trend after 1987 is consistent with the recovery of the economy and the exit from the sample of small firms, which occurs mainly in 1989-1990. Capital productivity is characterized by ups and downs during the entire period and this may reflect underutilization of capacity and delays in replacing old equipment. Finally, investment follows also a very irregular pattern with sharp drops in 1986 and 1988 which are likely picking up, again, the adverse effects of the earthquake and the recession. Additional variables that are used in the regression models and further help to characterize the Mexican manufacturing environment are reported in Table 6 at the end of the section.

Table 4: Production characteristics

Variable	1985	1986	1987	1988	1989	1990
Numbers of Plants	1,949	1,972	1,953	1,919	1,873	1,614
Gross Value of Output[a]	465.01	439.07	453.21	473.80	554.19	619.88
Value Added[b]	415.56	243.88	247.99	259.75	304.36	361.79
Δ Gross Value of Output[c]		7.81	6.66	4.76	21.27	15.09
Δ Value Added[d]		-2.20	53.45	98.89	61.86	42.65
Capital Stock[e]	3.14	2.92	2.99	2.94	3.05	3.22
Capital Productivity[f]	4.59	4.70	4.74	4.35	4.56	4.56
Investment[g]	24.94	13.38	37.03	19.17	36.64	36.53
Total Employment	369.29	348.28	343.33	350.20	802.02	409.50
Δ Total Employment [h]		7.64	1.62	2.40	17.82	44.55

Note: [a]In millions of 1980 pesos; [b]In millions of 1980 pesos; [c]Percentage; [d]Percentage; [e]In millions of 1980 pesos; [f]Average plant-level gross value of output/capital stock; [g]In millions of 1980 pesos; [h]Percentage.

5.3 Trade statistics

The trade data on imports and exports, used to calculate the statistics at the industry level reported in the last three columns of Table 3, came from the Commodity Trade database of the United Nations Statistical Office, which provides information at the four-digit level ISIC classification and categorized products by end of use. These data were merged with the Mexico's Annual Industrial Survey, which on the other hand categorized products by production technology, trying to achieve a reasonable match relying on detailed product codes available in the industrial survey. Also since the trade data are reported in dollars, they were first converted into 1980 dollars and then into pesos using the 1980 exchange rate in order to render the figures comparable removing the exchange rate fluctuations.

In addition, the data on commercial policies were provided by the SECOFI and were already harmonized with the classification scheme of the industrial census. These data, summarized by industry and time sub-periods in Table 5, clearly demonstrate that most of the changes in commercial policy took place between 1985 and 1988.

Table 5: Average annual change in trade protection

	Δ *Import coverage*		Δ *Average tariff rate*	
	1985-1988	1988-1990	1985-1988	1988-1990
Food	-24.34	-11.46	-22.97	-5.97
Beverages	-33.09	-29.63	-34.73	-7.33
Tobacco	3.17	-0.17	-26.57	-8.46
Textiles	-49.56	-25.70	-24.89	-12.37
Clothing and Apparel	-41.60	-47.83	-25.16	-11.93
Leather and Footwear	-56.32	-40.18	-25.44	-11.01
Wood and Furniture	-71.98	-31.37	-25.50	-12.99
Pulp and Paper	-53.11	-45.68	-33.80	-10.98
Chemicals	-53.08	-23.34	-21.31	-7.23
Plastic and Rubber	-70.13	-30.21	-21.95	-12.12
Glass	-48.72	-4.71	-34.00	-11.14
Cement	-25.36	-0.81	-20.27	-6.93
Nonmetal Minerals	-50.98	-8.57	-24.12	-10.22
Iron and Steel	-42.95	-4.92	-18.71	0.15
Nonferrous Metals	-51.64	-6.68	-23.51	-8.04
Metal Products	-66.38	-19.88	-27.73	-9.74
Nonelectrical Machinery	-44.07	-14.83	-18.70	-3.20
Electrical Machinery	-72.64	-40.66	-22.54	-10.03
Transport Equipment	-44.31	-36.09	-24.39	-7.16
Other Manufacturing	-47.78	-17.44	-23.73	-9.60

Note: The change is expressed in percentage.

Table 6: Trade and performance variables for selected years

	Average tariff rate			*Quota coverage*			*Herfindal index*			*Capital-output ratio*		
Industry	1985	1988	1990	1985	1988	1990	1985	1988	1990	1985	1988	1990
Food	0.233	0.099	0.121	0.660	0.280	0.222	0.011	0.012	0.016	0.502	0.661	0.559
Beverages	0.541	0.148	0.163	0.879	0.233	0.129	0.036	0.042	0.044	1.249	2.259	2.119
Tobacco	0.432	0.166	0.178	0.803	0.879	0.877	0.237	0.256	0.240	0.892	0.728	0.826
Textiles	0.351	0.137	0.142	0.429	0.029	0.019	0.029	0.025	0.027	0.728	0.807	0.730
Clothing and Apparel	0.429	0.170	0.172	0.748	0.008	0.005	0.061	0.067	0.074	0.387	0.636	0.480
Leather and Footwear	0.437	0.176	0.178	0.826	0.004	0.003	0.051	0.059	0.087	0.358	0.587	0.369
Wood and Furniture	0.399	0.156	0.156	0.701	0.014	0.011	0.035	0.045	0.056	0.761	0.775	0.698
Pulp and Paper	0.343	0.078	0.103	0.513	0.029	0.010	0.035	0.036	0.042	0.920	1.215	0.703
Chemicals	0.227	0.105	0.122	0.558	0.056	0.041	0.012	0.014	0.013	0.715	0.791	0.992
Plastic and Rubber	0.304	0.139	0.142	0.552	0.010	0.008	0.053	0.061	0.067	5.185	0.806	0.798
Glass	0.459	0.129	0.143	0.541	0.048	0.047	0.065	0.079	0.074	1.097	1.063	0.812
Cement	0.183	0.084	0.098	0.491	0.132	0.131	0.059	0.056	0.057	1.818	2.007	3.091
Nonmetal Minerals	0.335	0.140	0.145	0.452	0.026	0.024	0.037	0.037	0.039	0.851	1.223	0.904
Iron and Steel	0.153	0.077	0.106	0.440	0.009	0.008	0.100	0.085	0.101	0.910	1.267	1.060
Nonferrous Metals	0.262	0.110	0.126	0.446	0.007	0.007	0.415	0.770	0.938	0.819	1.650	1.767
Metal Products	0.308	0.112	0.124	0.483	0.011	0.010	0.039	0.052	0.059	0.716	0.719	0.585
Nonelectrical Machinery	0.230	0.118	0.142	0.447	0.049	0.032	0.036	0.058	0.074	1.329	1.502	1.160
Electrical Machinery	0.331	0.147	0.153	0.534	0.009	0.006	0.025	0.047	0.088	0.536	0.758	0.598
Transport Equipment	0.266	0.109	0.127	0.652	0.109	0.054	0.056	0.056	0.065	0.641	0.862	0.541
Other Manufacturing	0.335	0.140	0.154	0.527	0.050	0.042	0.068	0.063	0.080	0.437	0.467	0.443

6 Production Function Estimation Results

In this section I exploit the structural framework described in Section 3.1 to estimate production function parameters controlling for endogenous productivity for eighteen Mexican manufacturing sectors. I estimate several models under different production technology specifications (Cobb-Douglas and translog) with both the Olley and Pakes and Levinson and Petrin approaches. The estimation results suggest that the Cobb-Douglas specification with the investment function used as a proxy for productivity (Olley and Pakes method) is the most adequate to fit the data, therefore I provide all the main results on production function parameters and productivity adopting this specification. At the end of the section I report robustness check results obtained with alternative models.

6.1 Production function parameters

I begin by presenting the production function estimates for the whole sample comparing the structural estimation results with the ones obtained using more standard OLS and fixed effects estimation techniques. I then test whether there is statistically significant evidence that the production function coefficients change during the period considered.

6.1.1 Comparing different estimators

The last two columns of Table 7 report the results obtained estimating a Cobb-Douglas production function using the investment as a proxy for productivity (Olley and Pakes approach). Specifically, in the first stage the Cobb-Douglas version of equation (3.7), i.e.

$$\begin{aligned} y_{jt} &= \beta_l l_{jt} + \beta_k k_{jt} + h_t^{-1}(k_{jt}, i_{jt}) + \eta_{jt} \\ &= \phi_t(l_{jt}, k_{jt}, i_{jt}) + \eta_{jt} \end{aligned} \tag{6.1}$$

is estimated with OLS. The results of the first stage estimation, i.e. $\hat{\phi}_{jt}$ and $\hat{\eta}_{jt}$, are carried through the second stage where the residual ξ_{jt} of the productivity process from equation (3.11) is again obtained by OLS. Finally, equation (3.13) is estimated by GMM exploiting the moment condition on capital in (3.14). Note that, since the coefficient on labor β_l is identified and estimated in the first stage, I rely on one moment condition to identify the only remaining parameter, β_k, in the second stage. Thus, the system is just identified and the identity matrix is the optimal weighting matrix used in the GMM objective function.

In almost all sectors, with the exception of Food (1) and Glass (11), the coefficient associated with labor is significant and ranges from 0.15 in Chemicals (9) to 0.96 in Leather and Footwear (6). The coefficient on capital, on the other hand, is significant for only twelve sectors and ranges from 0.36 in Plastic and Rubber (10) to 0.74 in Nonelectrical machinery (17). As expected, there are significant differences in the production function parameters, thus in technology, across sectors. In particular, some sectors like Clothing and Apparel (5), and Plastic and Rubber (10) are more labor intensive, while other sectors like Pulp and Paper (8), Chemicals (9), and Transportation equipment (19) are more capital intensive.

Table 7: Estimates of production function coefficient
under different estimation methodologies

	OLS		FE		Structural	
Industry	β_l	β_k	β_l	β_k	β_l	β_k
1	0.7407***	0.2243***	0.5601***	0.3161***	0.0537	0.5749**
2	0.5873***	0.3654***	0.2402***	0.2492***	0.2751**	0.6385**
4	0.7115***	0.1243***	0.4717***	0.0845	0.4914**	0.0833
5	0.7949***	0.2067***	0.6128***	0.0364	0.5967**	0.2334
6	0.8408***	0.2353***	0.5624***	0.2017**	0.9606**	0.0086
7	0.7293***	0.2424***	0.5486***	0.3465***	0.4192**	0.0615
8	0.6923***	0.3273***	0.2629***	0.6271***	0.2320**	0.5934**
9	0.6454***	0.3251***	0.1343***	0.1500***	0.1545**	0.6202**
10	0.7943***	0.2525***	0.5239***	0.1128**	0.5108**	0.3617**
11	0.7291***	0.1906***	0.9015***	0.3534***	0.0475	0.6243**
12	0.8219***	0.1667***	0.4523***	-0.0043	0.4852**	0.8142
13	0.8143***	0.1804***	0.5834***	0.2179***	0.3992**	0.3970**
14	0.8039***	0.2285***	0.5298***	0.2009***	0.2622**	0.4180**
16	0.7238***	0.3396***	0.4642***	0.3298***	0.3432**	0.5092**
17	0.7454***	0.2758***	0.6412***	0.2909***	0.4699**	0.7388**
18	0.9131***	0.1723***	0.7183***	0.0440	0.2678**	0.6330**
19	0.6379***	0.4406***	0.3661***	0.3624***	0.3099**	0.7342**
20	0.9414***	0.0883***	0.4884***	0.0369	0.3188**	0.1178

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). For the structural estimation the standard errors are obtained by block-bootstrap.

The comparison between the results from the structural estimations and those obtained with a simple OLS regression yields a well established empirical evidence. First, the coefficients on labor and capital are highly statistically significant across all sectors. Second, focusing only on the coefficients that are significant in the structural estimation, the OLS coefficient on labor is always bigger while the coefficient on capital is always smaller than its structural counterpart. This pattern is well documented in the literature and is determined by the correlation structure between the transmitted productivity shock and the production inputs. More precisely, the variable input labor is supposed to be positively correlated with the unobserved productivity, thus the OLS coefficient on labor is likely to be biased upward. On the other hand, if current capital is not correlated with the current productivity shock, as it is decided one period ahead, or if capital is much less weakly correlated with productivity than labor, the OLS estimate on capital is likely to be biased downward.

Finally, looking at the estimates obtained using plant-level fixed effects (third and fourth column of Table 7), it is clear that, at least for labor, this approach partially mitigates the bias discussed above, i.e. the fixed effect coefficient on labor is always significant and smaller than the OLS one. However, the estimation of the capital parameter under fixed effects appears more problematic with some insignificant values and an unclear pattern with respect to the magnitude of the coefficient, which is higher than its OLS counterpart in some cases but smaller in some other cases. Nonetheless, the fixed effects estimates still remain higher for labor and lower for capital than those obtained with the structural approach. This is because the former is just an indirect way of controlling for unobserved productivity, whereas the latter fully accounts for the transitory productivity shock.

With Cobb-Douglas technology, the production function coefficients represent the elasticity of output with respect to the inputs and their sum can be interpreted as returns to scale. Table 8 reports the estimated returns to scale, i.e. $\beta_l + \beta_k$. With OLS in most of the industries the sum of the two coefficients is very close to one but the constant returns to scale hypothesis is statistically verified only for half of the industries. The within estimator (plant-level fixed effects) delivers returns to scale that are in general below one and overall lower than in the OLS case. However, for fourteen industries constant returns to scale are statistically verified. Finally, the returns to scale estimated with the structural procedure are mostly in between the OLS and FE results and, again, in thirteen out of eighteen industries the constant returns to scale hypothesis cannot be rejected.

Since the structural approach, and to some extent also the FE, should deliver more credible estimates as they control (directly or indirectly) for productivity shocks, the empirical evidence seems to support the presence of constant returns to scale in the majority of the Mexican manufacturing industries.

Table 8: Returns to scale
under different estimation methodologies

	OLS	FE	Structural
Industry	$\beta_l + \beta_k$	$\beta_l + \beta_k$	$\beta_l + \beta_k$
1	0.9650	0.8762	0.6286**
2	0.9527**	0.4894**	0.9136**
4	0.8358**	0.5562**	0.5747
5	1.0016	0.6492**	0.8301**
6	1.0761**	0.7641**	0.9692**
7	0.9717	0.8951	0.4807**
8	1.0196	0.8900**	0.8254
9	0.9705**	0.2843**	0.7747
10	1.0468**	0.6367**	0.8725
11	0.9197**	1.2549	0.6718**
12	0.9886	0.4566**	1.2994**
13	0.9947	0.8013**	0.7962**
14	1.0324	0.7307**	0.6802
16	1.0634**	0.7940**	0.8524**
17	1.0212	0.9321	1.2087**
18	1.0854**	0.7623**	0.9008**
19	1.0785**	0.7285**	1.0441**
20	1.0297	0.5253**	0.4366**

Note: The stars indicate that the constant returns to scale hypothesis $H_0 : \beta_l + \beta_k = 1$ cannot be rejected at a 5 percent significance level.

6.1.2 Testing for a structural change in the production function parameters

In order to verify whether the trade liberalization process generated factor reallocation phenomena across the Mexican manufacturing industries by modifying the factor intensity, I re-estimate the structural model dividing the sample into two sub-periods, the first from 1985 to 1987 and the second from 1988 to 1990. This choice is dictated by the fact that in the first three years (1985-1987) the most dramatic reforms took place, while the last three years (1988-1990) can be mainly considered a consolidation period. In order to carry on the test I modify (6.1) and (3.11) in the following way:

$$\begin{aligned} y_{jt} =& \beta_l l_{jt} + \tilde{\beta}_l D_t l_{jt} + \beta_k k_{jt} + \tilde{\beta}_k D_t k_{jt} + \bar{\beta}_k k_{jt} + \tilde{\bar{\beta}}_k D_t k_{jt} + \beta_i i_{jt} + \tilde{\beta}_i D_t i_{jt} \\ &+ \beta_{kk} k_{jt}^2 + \tilde{\beta}_{kk} D_t k_{jt}^2 + \beta_{ii} i_{jt}^2 + \tilde{\beta}_{ii} D_t i_{jt}^2 + \beta_{ki} k_{jt} i_{jt} + \tilde{\beta}_{ki} D_t k_{jt} i_{jt} + \eta_{jt} \qquad (6.2) \end{aligned}$$

$$\omega_{jt} = \gamma_0 + \bar{\gamma}_0 D_t + \gamma_1 \omega_{jt-1} + \gamma_2 \omega_{jt-1}^2 + \gamma_3 \omega_{jt-1}^3 + \xi_{jt} \qquad (6.3)$$

where D_t is a dummy variable taking the value of one from 1988 on and zero otherwise. Note that in (6.3) only the constant, i.e. the average productivity, is allowed to possibly change between the two sub-periods. The intuition behind (6.2) is simply that, if $\tilde{\beta}_l$ and $\tilde{\beta}_k$ are significantly different from zero there is evidence of a structural change in the production function parameters.

Table 9: Production function coefficients estimates for the sample divided in two sub-periods

	1985-1987		1988-1990	
Industry	β_l	β_k	$\beta_l + \tilde{\beta}_l$	$\beta_k + \tilde{\beta}_k$
1	0.0355	0.0002	0.0898	1.9850
2	0.1994**	0.6455	0.3573**	0.5397
4	0.4175**	0.2257	0.5968**	0.1530
5	0.5314**	0.0236	0.6527	0.3806
6	0.9343**	0.0000	0.9617	0.3746
7	0.4317**	0.2614	0.4083	0.5654
8	0.2109**	0.6516	0.2500	0.5409
9	0.0815**	0.6566**	0.2214**	0.5900
10	0.4031**	0.4216	0.6390**	0.3083
11	-0.0091	0.6738**	0.0415	2.0660
12	0.4614**	0.4091	0.5405	0.9828
13	0.3372**	0.5547	0.4672**	0.1800
14	0.2578**	0.5190	0.2643	0.1855
16	0.3263**	0.4570	0.3391	0.5750
17	0.4652**	0.2690	0.4961	1.2480
18	0.2274**	0.5754**	0.3053	0.6739
19	0.3630**	0.6670	0.2577	0.8071
20	0.2593**	1.3320**	0.4975	0.0154

Note: In the first two columns, the stars indicate that zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping the sample. In the last two columns, the stars indicate that the coefficient is significantly different between the first and the second sub-period.

Table 9 shows that, reasonably, in almost all the cases the estimated coefficients for the two sub-periods can be considered as an upper and lower bound for the coefficients estimated using the whole sample (reported in the last two columns of Table 7). However, the first two columns of Table 8 demonstrate that, especially for the capital coefficient, the division of the sample compromises the significance of the estimates. Moreover, regarding capital, the coefficient associated with dummy variable is never significant meaning that there is no evidence that the capital parameter changed in the second part of the sample. As for labor, a significant structural change occurs after 1987 for just five industries: Beverages (2), Textiles (4), Chemicals (9), Plastic and Rubber (10), and Nonmetal minerals (13), with the labor coefficient always increasing in the second sub-period. Nonetheless, since only in the Chemicals industry the capital coefficient is significant in the first sub-period and does not change between the two sub-periods, I conclude that, overall, the factor intensity remained fairly constant during the trade liberalization process for the majority of the industries with the exception of the Chemicals sector which became more labor-intensive. The coefficient associated with the dummy variable in (6.3), not reported here, is insignificant in every industry suggesting that the average productivity did not change from the first sub-period to the second.

6.2 Productivity analysis

The structural framework illustrated in Section 3.1 is suitable to obtain a characterization of the technology in each industry through the production function coefficients as well as an estimate of the productivity process for each firm in each year. Specifically, with a Cobb-Douglas technology, the productivity process can be recovered, after estimating β_l and β_k, as $\hat{\omega}_{jt} = \hat{\phi}_{jt} - \hat{\beta}_l l_{jt} - \hat{\beta}_k k_{jt}$. Furthermore, recall that the first-order Markov productivity process is modeled as a third degree polynomial in lagged productivity of the form: $\omega_{jt} = \gamma_0 + \gamma_1 \omega_{jt-1} + \gamma_2 \omega_{jt-1}^2 + \gamma_3 \omega_{jt-1}^3 + \xi_{jt}$.

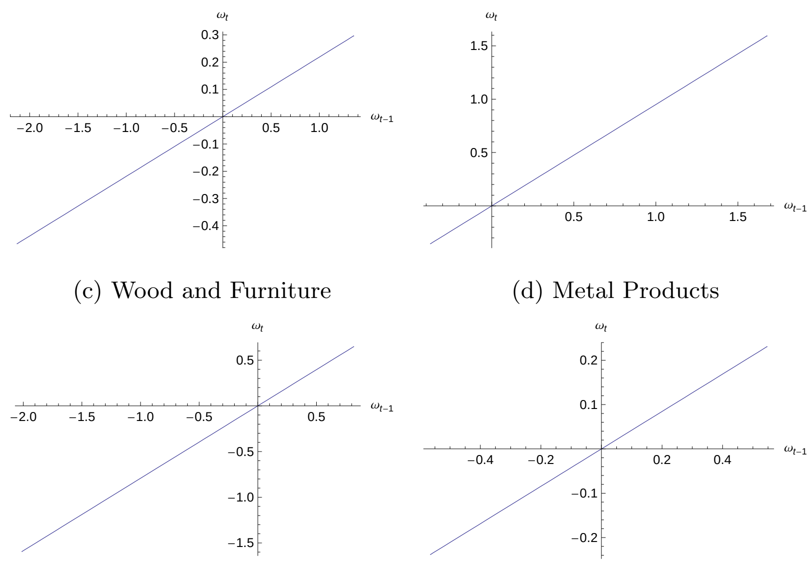
The empirical evidence suggests that, since for almost all the industries the γ_0, γ_2, and γ_3 coefficients are statistically insignificant, the productivity process can be actually approximated by the AR(1) process $\omega_{jt} = \gamma_1 \omega_{jt-1} + \xi_{jt}$. Therefore, the current productivity depends only linearly on the value of the previous productivity. Moreover, the γ_1 coefficient is estimated to be always below one (except for the Cement industry (12)) meaning that the productivity process is stationary. Figure 1 depicts the productivity process for four industries chosen for illustrative purposes.

For two of those four industries, Figure 2 shows the smoothed plots for capital and investment. Specifically, in each panel the vertical axis measures the estimated productivity shock, while the horizontal axis running left measures investment levels and the horizontal axis running right measures capital usage. The structural estimation procedure is based on a crucial monotonicity assumption regarding productivity and investment, i.e. conditioning on any observed levels of capital usage the investment level should increase in productivity. As demonstrated in Figure 2 this monotonicity condition seems to hold.

Figure 1: Productivity process

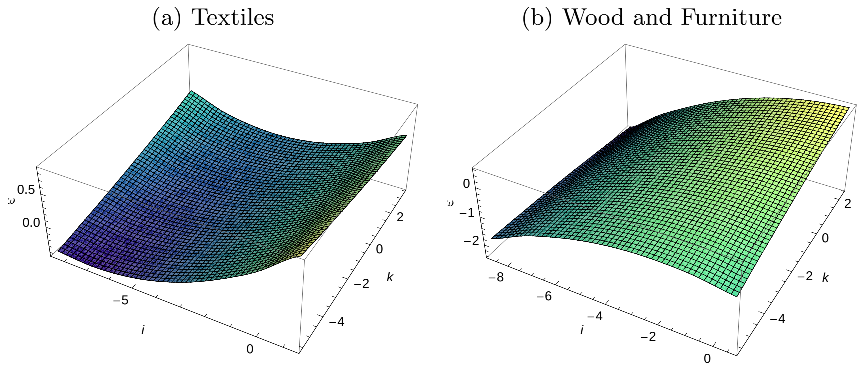
(a) Food (b) Textiles

(c) Wood and Furniture (d) Metal Products



The ability of obtaining a direct estimate of the productivity process allows me to analyze the growth in productivity for each firm from year to year. In fact, knowing ω_{jt}, the growth in productivity can be easily calculated as $\Delta\omega_{jt} = \omega_{jt} - \omega_{jt-1}$. Table 10 displays the mean and the standard deviation of the productivity growth between 1986 and 1990. The average annual growth in productivity is relatively small, below 0.1 percent, for the majority of the industries and for six of them (Wood and Furniture (7), Pulp and Paper (8), Chemicals (9), Cement (12), Nonmetal minerals (13), and Iron and Steel (14)) the average growth rate is negative. The standard deviation, however, is relatively high demonstrating that there are significant differences among the firms in each industry with respect to productivity growth performances. The last column of Table 10 shows the percentage of firms that have moved across the quartiles of the productivity growth distribution. The figures are always above 60 percent demonstrating that in each industry there is a lot of heterogeneity and reshuffling across firms. This results can be observed further in Figure 3 where the frequency and the kernel approximated distribution of productivity growth is depicted for three industries in 1986 (left panel) and 1990 (right panel). It is easy to see that, even if the distribution is always centered around zero, its shape considerably changes between the first and the last year in each of the three industries. Furthermore,

Figure 2: Productivity as a function of capital and investment



in the Beverage sector the distribution is fairly symmetric although very leptokurtic in both the first and last year. In the Chemicals sector the distribution is skewed to the right in 1986 but skewed to the left in 1990 and strongly leptokurtic in both years. In the Nonelectrical machinery sector the distribution is right skewed and leptokurtic in both 1986 and 1990 but skewness and kurtosis are lower in the last year. In conclusion in none of the sector the distribution of productivity growth seems normal.

6.3 Robustness checks with alternative models

In this section I compare the production function coefficient estimates obtained applying the structural approach to alternative models. In particular I estimate the following additional specifications:

I Value added with translog technology using the investment as a proxy for productivity (Olley and Pakes method) and estimating all the production function coefficients in the second stage relying on the moment conditions in (3.15).

II Value added with Cobb-Douglas technology using the demand for intermediate inputs as a proxy for productivity (Levinson and Petrin method), estimating the coefficient on labor in the first stage and the coefficient on capital in the second stage relying on the moment condition in (3.14).

III Value added with translog technology using the Levinson and Petrin method and estimating all the production function coefficients in the second stage relying on the moment conditions in (3.15).

Table 10: Annual productivity growth

Industry	Mean	Std. Dev.	% Firms moving
1	0.0078	0.2392	73.29
2	0.0048	0.1110	78.57
4	0.0005	0.1082	73.78
5	0.0022	0.2607	74.50
6	0.0042	0.1154	70.75
7	-0.0049	0.2390	72.12
8	-0.0027	0.1285	74.56
9	-0.0039	0.1231	76.46
10	0.0006	0.1212	75.49
11	0.0079	0.3033	85.06
12	-0.0121	0.2199	78.85
13	-0.0001	0.1292	77.17
14	-0.0084	0.1186	67.41
16	0.0004	0.0959	70.93
17	0.0045	0.1164	67.20
18	0.0065	0.2574	80.98
19	0.0002	0.1728	74.89
20	0.0221	0.2660	71.35

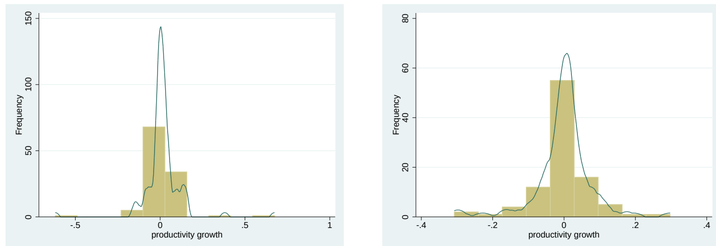
Note: The last column reports the percentage of firms that during the period 1986-1990 have moved across the quartiles of the productivity growth distribution.

IV Gross output with Cobb-Douglas technology including intermediate inputs in the production function, using the Levinson and Petrin method, and relying on the moment conditions of lagged labor, lagged intermediate inputs, and current capital to identify the coefficients associated with labor, intermediate inputs, and capital, respectively, in the second stage.

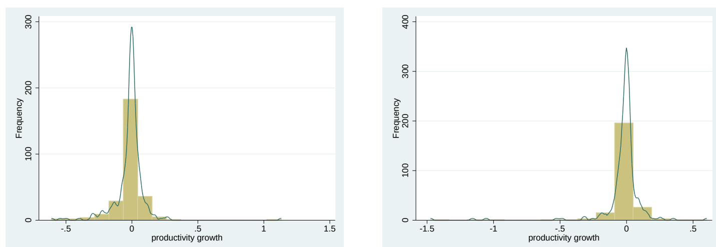
The results for the second and fourth specifications are omitted here because the estimation of these models was particularly problematic. The solving algorithm could not find a solution satisfying the optimization criteria and the resulting production function parameters were in many cases zero which implies that, since I imposed the restriction for the βs to be nonnegative, this constraint was often binding. The results for the first and third specifications are presented in Tables 11, 12, and 13. In particular, I estimate the third specification twice, first using the entire sample and then using a subsample excluding the *maquiladoras*. This is because the Mexican accounting system includes in the books of the firm that orders a subcontracting service the value of expenditure in intermediate inputs

Figure 3: Distribution of productivity growth

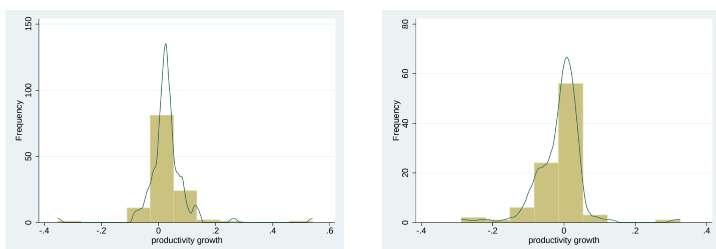
(a) Beverages



(b) Chemicals



(c) Nonelectrical machinery



used by the subcontractor, generating a measurement error problem with the intermediate inputs.

The results in Tables 11, 12, and 13 demonstrate, first, that even if the translog specification allows for a more flexible way of modeling technology, the overall significance of the estimates is much lower because the higher order and interaction terms potentially generate collinearity issues. Moreover, the interpretation of the coefficients and returns to scale is complicated with translog because the sign and the magnitudes of the coefficients do not have a straightforward meaning as they capture more elaborated and complex interactions between inputs.

Table 11: Estimates of production function coefficients,
translog technology, Olley and Pakes method

Industry	β_l	β_k	β_{ll}	β_{kk}	β_{lk}
1	-0.5721**	0.8281**	-0.2091**	0.1178	-0.0163
2	-0.1930	0.7812**	-0.1736**	0.0080	0.1688
4	0.2847	0.2217**	-0.0709	0.0658	-0.0519
5	-0.0629	0.4851**	-0.1372	0.0439	0.0179
6	1.1807**	-0.9369**	0.4271	0.2104**	-0.8233**
7	-0.1161	0.4346	-0.2170	-0.0726	0.1719
8	0.0109	0.4832	-0.1031	0.0496	-0.0376
9	0.0249	0.5997**	-0.0522**	0.0461	-0.0056
10	0.1905	0.6011**	-0.1205**	0.0214	0.0960
11	-0.7413	1.1676	-0.3956	-0.1522	0.5748
12	-0.6070	0.9394	0.0396	-0.1361	0.4806
13	0.0622	0.3478**	-0.1197**	0.0199	0.0065
14	-0.2046	0.7527**	-0.2260	-0.0090	0.2257
16	-0.0524	0.4558**	-0.1157	0.1322	-0.1350
17	0.2323	0.6427	-0.0173	0.0281	-0.0093
18	-0.8505	1.3917**	-0.5748**	-0.1901	0.7524
19	0.3172**	0.7545**	0.0107	0.0715	-0.0268
20	-0.6517**	0.9863**	-0.3648**	-0.0030	0.2928

Note: The stars indicate significance levels (**$p < 0.05$), i.e. zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping.

The lack of significance is exacerbated when the intermediate inputs demand is used as a proxy for productivity as confirmed by the results reported in Table 12 and 13. The cause of this problem is likely to be the measurement error in intermediate inputs, originated by the peculiar way of recording expenditure in intermediates for the *maquiladoras*, which does not appear to be resolved even when these firms are excluded. In fact, a crucial requirement for the Levinson and Petrin method to be successfully applied is the absence of measurement error in intermediate inputs expenditure.

Table 12: Estimates of production function coefficients, translog technology, Levinson and Petrin method (full sample)

Industry	β_l	β_k	β_{ll}	β_{kk}	β_{lk}
1	0.1802	0.2095	-0.0513	0.0472	-0.0954
2	0.0427	0.5830**	-0.1206	0.0164	0.0997
4	3.4026	1.2897	0.4224	0.1580	-0.1516
5	-0.0905	0.1157	-0.1374	0.0480	0.0742
6	1.0747	-0.7804	0.3546	0.1824	-0.6245**
7	-0.3855	0.1080	-0.2711	-0.0653	0.2937
8	0.4332	0.2107	0.0085	0.0867	-0.1770
9	0.2912	0.4189	-0.0070	0.0442	-0.0515
10	0.8374	0.0504	0.1267	0.1365	-0.2883
11	0.0808	0.6522	0.0868	0.0135	0.0125
12	-0.6883	1.5939**	0.4354	-0.2522	0.3858
13	-0.1825	0.0699	-0.1802	0.0058	0.1736
14	0.5714**	0.2521	-0.0752	0.0166	0.0229
16	0.0089	0.4300	-0.1181	0.1094	-0.1072
17	-0.2529	-0.7824	-0.1235	-0.0594	0.1201
18	-0.0418	0.5842	-0.2297	0.0530	0.1067
19	0.4602**	0.5693**	0.0380	0.0830**	-0.0936
20	-0.7006	0.4892	-0.2989	0.0049	0.2461

Note: The stars indicate significance levels (**$p < 0.05$), i.e. zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping.

Table 13: Estimates of production function coefficients,
translog technology, Levinson and Petrin method (no *maquiladoras*)

Industry	β_l	β_k	β_{ll}	β_{kk}	β_{lk}
1	0.3688	0.0799	0.0246	0.0713	-0.2004
2	0.3015	0.4871**	0.0186	0.0313	-0.0108
4	1.9368	0.8479	0.2181	0.0842	-0.1069
5	-0.0888	0.0604	-0.1406	0.0232	0.0899
6	1.1008	-1.3294	0.3612	0.0915	-0.6068
7	-0.4666	0.0064	-0.2858	-0.0727	0.2973
8	0.3385	0.2600	-0.0135	0.0961	-0.1688
9	0.2820	0.3859	0.0008	0.0477	-0.0735
10	0.3165	0.5362	-0.0904	0.0070	0.0880
11	-0.5346	0.9487	-0.2968	-0.0829	0.3910
12	0.6469	0.6338	0.0770	-0.0684	0.0222
13	-0.2048	0.0895	-0.1775	0.0130	0.1527
14	0.1499	0.5091**	-0.2016	0.0056	0.1357
16	-4.8433	1.9756	-1.5002	-0.1237	0.7065
17	-0.1554	-0.7857	-0.0907	-0.0421	0.0681
18	-0.0285	0.4356	-0.1706	0.1099	-0.0467
19	0.4329**	0.5857**	-0.0032	0.0735	-0.0279
20	-0.7159	0.4555	-0.3246	-0.0238	0.3011

Note: The stars indicate significance levels (**$p < 0.05$), i.e. zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping.

7 Price-Cost Margins Results

In this section I present the results on the industry-level markups estimated with the simple dual approach outlined in Section 3.3. I then compare these with the plant-level markups derived by plugging into equation (3.19) the production function parameters previously obtained within the structural framework. Finally, with the results of the markup estimations in hand, I investigate the relation between price-cost margins and trade openness.

7.1 Industry-level Markups

Table 14 reports the mean, median and standard deviation of the year- and industry-specific markups recovered by estimating (3.25) by OLS. Note that with this procedure it is only possible to estimate one markup for each industry in each year, therefore the variation is along the (firm) cross-sectional dimension. The average markup ranges between 0.24 for Chemicals and 2.69 for Cement and in almost all the industries the mean and the median are different, with the mean being usually higher than the median, implying that the distribution of the markups is not symmetric. Moreover, the standard deviation is high indicating that in every industry the markups vary consistently across years. In many industries the magnitude of the markups is quite unreasonable (significantly lower than one) and the significance of the estimates is fairly poor. Overall this simple estimation procedure delivers very imprecise and unreliable results and it does not appear to be a valid alternative to the structural approach.

7.2 Plant-level Markups

Table 15 summarizes the plant-level markups recovered combining the output elasticity with respect to labor $\hat{\beta}_l$, obtained estimating (6.1), i.e. a Cobb-Douglas production function with the investment as a proxy for productivity, and data on labor expenditure and value added as described in (3.21). For sixteen out of eighteen industries the average markup is significantly different than zero and above or very close to one. Once again the mean is higher than the median implying that the markups distribution is positively skewed in almost all the industries. This result is confirmed in Figure 4 where the distribution of the markups for some representative industries is plotted. The same figure shows also that the distribution of the markups is asymmetric, as expected. This is because markups are supposed to be bigger than or equal to one as they represent the ratio between price and

Table 14: Industry-level markup estimates

Industry	Mean	Med.	St.D.
1	1.00	0.47	1.36
2	2.54	1.63	3.82
4	0.99	0.59	1.02
5	0.79	0.80	0.47
6	1.45	1.11	1.37
7	1.50	0.93	1.75
8	1.28	1.26	1.16
9	0.24	0.22	0.31
10	0.89	1.10	0.73
11	2.58	0.77	4.17
12	2.69	1.42	3.53
13	0.90	0.54	1.00
14	0.70	0.48	0.74
16	0.70	0.62	0.72
17	0.92	0.43	1.46
18	0.65	0.37	0.83
19	0.30	0.06	0.50
20	0.73	0.56	0.63

Note: Mean, median, and standard deviation of the markups are calculated for each industry pooling all the years.

marginal cost, therefore their distribution should be truncated around one. The standard deviation is very high indicating a substantial variation in markups across firms in each manufacturing sector.

The comparison between the results from Table 14 and those in Table 15 clearly highlights that the plant-level markups obtained with the structural approach are usually higher than the ones estimated at the industry level using the simplified approach. This is because the industry-level markups are estimated in first differences which usually leads to a downward bias. At plant level the highest significant markup, 2.20, is estimated for the Cement industry while the lowest significant markup, 0.82, is estimated form the Chemical industry and this results is the same at the industry level. Nonetheless, the correlation between the industry-level average markups and the plant-level average markups is merely 0.10.

Table 16 shows the markups estimated at the plant level using different specification for the production functions, i.e. the alternative models I and III described before. The plant-level markups with translog technology are confirmed to be higher than the industry-level

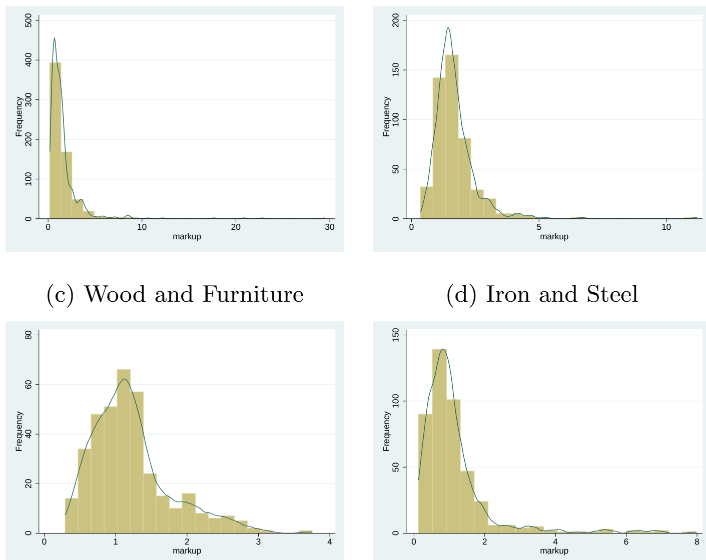
Table 15: Plant-level markup estimates,
Cobb-Douglas technology

	Olley&Pakes (full sample)		
Industry	Mean	Med.	St.D.
1	0.40	0.28	0.55
2	1.61**	1.12**	2.17
4	1.04**	0.92**	0.60
5	1.68**	1.52**	0.85
6	2.19**	2.30**	0.35
7	1.21**	1.13**	0.56
8	0.96**	0.70**	4.27
9	0.82**	0.51**	3.11
10	1.13**	1.02**	0.86
11	0.29	0.26	0.21
12	2.20**	2.21**	1.12
13	1.25**	1.12**	0.65
14	1.13**	0.89**	1.04
16	0.93**	0.83**	0.53
17	0.99**	0.91**	0.50
18	0.82**	0.70**	0.59
19	1.37**	0.79**	2.78
20	1.08**	0.96**	0.57

Note: The stars indicate significance level (**$p < 0.05$), i.e. zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping.

ones in almost all the sectors under any specification. However, the significance of these results is much lower than the significance of the plant-level markups obtained with Cobb-Douglas technology and this is mainly due to the fact that the Cobb-Douglas specification fits the data better and delivers more precise and reliable estimates of the production function parameters and, therefore, of the markups as well. Furthermore, I conduct a test to verify whether the average and median markups are statistically bigger than one. In fact, since the markup in this context is defined as price over marginal cost, a meaningful markup should be equal to or above one. For almost all the industries and under any specification I cannot reject the null hypothesis, meaning that the markups are statistically bigger than one at a 5 percent significance level. This result, though, needs to be considered cautiously because the confidence intervals obtained by block-bootstrap used for inference are not tight.

Figure 4: Distribution of plant-level markup estimates



(a) Beverages (b) Clothing and Apparel

(c) Wood and Furniture (d) Iron and Steel

In conclusion, the striking differences between the industry-level and plant-level markup estimates highlight the following important point. Relaxing the constant markup assumption across firms and allowing for time varying and heterogeneous (among firms) productivity shocks leads to more precise and substantially higher markups.

7.3 Markups and Trade Liberalization

In this section I rely on the models described in Section 4.2 to analyze the impact of trade liberalization on the price-cost margins at the industry and plant level.

7.3.1 Industry-Level Analysis

To perform the industry-level analysis I use the results of the markups estimation presented in Section 7.1 obtained with the simple dual approach and the results obtained with the structural approach estimating a Cobb-Douglas technology and using investment as a proxy for productivity. In fact, only these two sets of results are directly comparable since one of the requirements to estimate (3.25) is for the production function to be linearly homogeneous and the test on the returns to scale of the Cobb-Douglas production function

Table 16: Plant-level markup estimates, translog technology

	Olley&Pakes (full sample)			Levinson&Petrin (full sample)			Levinson&Petrin (no *maquiladoras*)		
Industry	Mean	Med.	St.D.	Mean	Med.	St.D.	Mean	Med.	St.D.
1	2.98	2.07	4.04	3.89	2.99	7.12	3.51**	2.51	4.84
2	2.37**	1.63	2.79	2.61	1.87	3.03	1.62	1.22	2.72
4	1.27**	1.11**	0.80	4.89**	4.42**	3.60	3.16**	2.71**	2.50
5	2.17**	2.20**	1.20	1.76**	1.61**	1.25	2.01**	1.91**	1.22
6	2.35**	1.88**	3.15	1.84**	1.39**	2.25	1.79**	1.33**	2.50
7	2.17**	2.16**	1.22	1.83**	1.43**	2.42	1.87**	1.54**	2.50
8	4.90	1.35**	94.62	0.89	1.52	17.36	0.96	1.43	8.32
9	1.06**	0.51**	3.51	1.61	1.03	3.36	1.53	0.91	3.34
10	1.45**	1.42**	0.71	1.44	1.34	0.66	1.46	1.33	0.96
11	2.74	1.82	3.78	0.63	0.55	1.74	1.17	0.65	2.01
12	3.10	2.89	3.45	1.26	1.57	5.18	3.31	3.28	1.68
13	2.22**	2.37	1.17	1.87**	1.65**	1.60	1.90**	1.71**	1.63
14	2.12	1.60	2.72	3.49**	2.97**	3.22	3.11**	2.57**	3.68
16	1.24	1.57**	1.47	1.45	1.62	1.80	1.52	1.47	8.82
17	0.69	0.60	0.45	0.44	0.22	0.89	0.48	0.27	1.13
18	1.85	1.22	3.33	2.53	1.45	8.21	2.43	1.27	7.53
19	1.20	0.75	2.27	1.13	0.97**	1.90	1.66**	1.12**	2.15
20	2.61**	2.69	2.23	1.48**	1.22**	1.81	1.50	1.27	1.95

Note: The stars indicate significance level ($^{**}p < 0.05$), i.e. zero is not contained in the 95 percent confidence interval obtained by block-bootstrapping.

in (6.1) confirmed that there is statistically significant evidence of constant returns to scale most of the industries. Recall that the markups obtained with the simple dual approach are directly estimated at the industry-level. On the other hand, the price-cost margins obtained with the structural approach are estimated at the plant level, thus in this part of the analysis I collapse these results to the annual average markup in each sector. The other explanatory variables included in (4.1), i.e. Herfindahl index, capital-output ratio and measures of trade exposure, are constructed by aggregating and averaging across individual firms in each sector in each year. Models 1 and 2, which include industry-specific dummy variables, should explain the temporal variation within each industry while models 3 and 4, with only year dummy variables, are supposed to capture the variation between sectors. Note also that the measures of trade exposure reflect the extent of trade liberalization, i.e. a decrease in the quota coverage or in the tariff rate implies an increase in trade openness and foreign competition. Therefore, a positive coefficient associated with these trade indicators describes a negative effect of the trade reforms on the markups and is expected in the presence of import discipline. The regression results are reported separately for each type of trade liberalization instrument in Tables 17-20.

Tables 17 and 18 report the regression results obtained estimating (4.1) by OLS and using the industry-level markups recovered with the simple dual approach as a dependent variable. It is easy to see that, using either the quota coverage or the average tariff rate as trade indicators, very few coefficients in these regressions are significant and this result is exacerbated in models 3 and 4 which include only year dummy variables. The lack of significance is further confirmed by the adjusted R^2 which is quite low, although all the models are globally significant as verified by the F-statistic.

In both cases, with quota and tariff, one of the few significant coefficients is the one associated with the capital-output ratio in models 1 and 2. The sign of this coefficient is unexpectedly negative however, since only the temporal variation is picked up in the model with industry dummy variables, this result may be reflect underutilization of installed capacity during the recession, which was prevalent for most of the sample period.

As for the measures of trade exposure, the coefficient associated with the tariff rate (Table 17, model 3) is positive and highly significant indicating that the markups tend to be lower the more the openness to trade. The coefficient on the interaction term between the trade indicator and the capital-output ratio in model 2 is also consistently significant with both quota and tariff. Its positive sign is again evidence of trade discipline and suggests that industries with a higher capital-output ratio are more likely to experience a reduction in margins as a consequence of trade liberalization. Nonetheless, because of the overall very low explanatory power, the regression results reported in Tables 16 and 17 cannot be viewed as strong evidence of an impact of trade on the markups. In addition, these results confirm that the simple dual approach used to obtain the markups at the industry level is inadequate since the markup estimates are imprecise, and in many cases insignificant, and this compromises any further analysis conducted with those estimates.

I now turn to the regressions reported in Tables 19 and 20 whose results are also obtained estimating (4.1) by OLS with the annual average markup in each sector, recovered from the structural plant-level estimation, as dependent variable. First note that in these regressions the level of significance is substantially higher, especially when the industry dummy variables are included (models 1 and 2). Thus, even if aggregated at the industry level, the markups coming from the structural estimation appear to perform much better. However, a substantial part of the explanatory power comes from industry effects as demonstrated by the R^2 which greatly increases from models 3 and 4 to models 1 and 2.

This outcome possibly reflects sector-specific industrial characteristics, policies, entry barriers or technological differences that are not captured by the other explanatory variables. The year effects are always negative, with both quota and tariff, in model 1 and 2 and are also negative in models 3 and 4 for the majority of the years considered. This result may capture the fact that during the period of analysis the Mexican economy faced difficult challenges that negatively impacted the profitability of the firms. The industry dummy variables, when included, are significant in many industries and their sign are consistent across all the specifications.
When quota is used as a trade indicator, the coefficient on the capital-output ratio has the expected positive sign in every model and in model 3 and 4 it is also highly significant, implying that industries with a higher capital share of output have higher price-cost margins. On the other hand, when tariff is used as a measure of trade exposure, this coefficient is still positive and significant when industry dummies are left out, but turns negative, although insignificant, when industry effects are controlled for. The coefficient associated with the Herfindahl index, when significant (model 3 with both quota and tariff) is positive confirming a higher rate of profit in more concentrated industries.
The coefficients on quota coverage and tariff rate are both positive and significant in model 3 indicating that price-cost margins decrease as trade protections are removed. In model 1, however, the same coefficients are not significant, suggesting that differences in the level of protection across sectors seem to be more relevant than variation over time. Adding interaction terms reveals a more complex picture. The net impact of quota coverage and tariff rate as well as their interaction with the Herfindahl index and the capital-output ratio are not significant in explaining the temporal variation (model 2), but the interaction between quota and Herfindahl index (Table 19, model 4) and the interaction between tariff and capital-output ratio (Table 20, model 4) are significant. Specifically, the interaction term for quota coverage and Herfindahl index is positive and significant implying that the profitability of the most concentrated industries is likely to decrease when trade is liberalized. Conversely, the interaction term for tariff rate and capital-output ratio in negative and highly significant suggesting that the trade reforms have a negative impact on the margins of the industry with the lowest capacity.

In summary, the industry-level analysis provides some evidence of import discipline, i.e. lower protection generated lower profitability in the Mexican manufacturing industries. This pattern is clearly established across sectors but not equally clearly over time. Moreover, the importance of using reliable estimates (in this case markups) to correctly evaluate economic policies is emphasized by the much better performance of the markups estimated within the structural framework with respect to the ones estimated relying on the simpler approach.

7.3.2 Plant-Level Analysis

To examine the intra-sectoral variation in price-cost margins I estimate (4.2) by OLS using as the dependent variable the plant-level markups. Recall that these markups were recovered from the structural estimation of my preferred specification, i.e. Cobb-Douglas technology with investment as a proxy for productivity. The explanatory variables are also calculated at the plant level with the exception of the trade indicators, quota coverage in model 1 and average tariff rate in model 2, which are only available at the industry level. The regression results are reported in Table 21.

First note that the plant-level models are globally significant as indicated by the F-statistic, but explain only a small fraction of the plant-level variation in price-cost margins as confirmed by the relatively low value of the adjusted R^2 (approximately 1.13). This is nonetheless a common outcome of regressions performed on large micro-level dataset as the one used here. The year dummy variables are always negative but insignificant while the industry dummy variables are significant in many cases with both positive and negative signs.

The coefficient associated with market share is positive and highly significant suggesting that a rise in its market share increases the price-cost margin of a plant but at a decreasing rate since the coefficient on the squared share is, on the other hand, negative and significant. The coefficient on capital-output ratio is positive and highly significant in both models implying that, as expected, an increase in capacity has a positive effect on the profitability of a plant, however this effect becomes marginal when the capacity is large as demonstrated by the very small magnitude of the coefficient on the squared capital-output ratio.

As for the trade indicators, the coefficients on quota coverage and tariff rate are both insignificant, implying that there is no evidence that the trade reforms affected the price-cost margins of the Mexican manufacturing plants. This is not a particularly surprising result since the high number of firms populating the manufacturing sectors should have imposed some degree of internal competitive pressure prior to the trade liberalization.

I also estimate the same regressions with the plant-level markups estimated under different specifications, i.e. translog technology with both investment and intermediate inputs demand as proxies for productivity, as dependent variable. The results, not reported here, are in line with those presented in Table 21. However, the overall significance of the models is lower presumably because, as already mentioned, the translog production function specification delivers unreliable markup estimates.

7.4 Markups for exporters

The previous section presented the analysis conducted on markups considering measures of imports liberalization. In this section, on the other hand, I focus on characterizing the relation between markups and export status. Moreover, since the structural framework allows for estimating both markups and productivity at the plant level, I further explore the role of productivity in the profitability of the Mexican manufacturing plants.

I first estimate (4.3) by OLS using the logarithms of the plant-level markups estimated using my preferred specification, i.e. Cobb-Douglas technology with investment as a proxy for productivity, as the dependent variable. The explanatory variables are capital and labor use, a full interaction of year and industry dummy variables, and, of course, a dummy variable indicating export status. After obtaining an estimate for the coefficient associated with the exporter dummy, ψ_1, and the constant term ψ_0, I perform a test on the significance of the nonlinear combination of the parameters $\psi_1 + \exp(\psi_0)$ which captures the level markup difference for exporters. Finally, I re-estimate (4.3) adding the estimated productivity ω_{jt} in order to directly control for differences in productivity and verify whether there is still evidence of a markup premium for exporter. Specifically, the second regression is given by: $\ln(\mu_{jt}) = \psi_0 + \psi_1 E_{jt} + \psi_2 \omega_{jt} + z_{jt}\rho + \varepsilon_{jt}$.

In both cases, with and without the additional productivity control, I obtain that the the level markup difference $\mu_E = \psi_1 + \exp(\psi_0)$ is positive, 0.019 not controlling for productivity and 0.016 controlling for productivity respectively, but insignificant. However, since the number of exporting firms in the Mexican manufacturing industries is not very high in the years considered and, most importantly, the extent of exporting is quite limited for the majority of the exporters, I try to verify whether the markup premium exists for intensive exporters, i.e. firms that export a high percentage of the value of their output. To do so I calculate for each exporter the ratio of exports value over output value and substitute the export dummy in (4.3) with another dummy, $E_{H_{jt}}$, that indicates export intensity. Specifically, $E_{H_{jt}}$ is equal to one if firm j is in the 75th or above percentile of the export-output ratio distribution, i.e. if firm j is an intensive exporter. Table 22 shows the results obtained estimating by OLS the modified version of (4.3) with the export intensity dummy, not controlling (model 1) and controlling for productivity (model 2).

First the coefficient associated with the export intensity dummy variable is always positive and significant, with and without including productivity in the regression suggesting that exporting has a positive impact on price-cost margins. Also the coefficient associated with productivity in model 2 is positive and highly significant, meaning that productivity contributes to firms' profitability. In addition, the level markup difference for intensive exporters μ_{E_H} is positive significant in both model. More precisely, I obtain a significantly estimated μ_{E_H} of 0.0588 in model 1 which implies that intensive exporters have a level markup premium of approximately 6 percent. In model 2 the estimated μ_{E_H} is significant and equal to 0.0539 meaning that, even controlling for productivity, the intensive exporter have a level markup premium of approximately 5.4 percent. Note that controlling for productivity in this context means to control for differences in marginal costs, if ψ_2 (the coefficient on productivity) picks up cost heterogeneity fully, so that the coefficient on the intensive exporter dummy picks up the variation in average prices between intensive exporter and the other firms (low exporters and non exporters). However, because the productivity used in this regression was estimated as the residual of a value added production function, it may not contain only differences in costs but also unobserved quality differences in both inputs and output, as well as others market power effects. Nonetheless it is important to emphasize that an intensive exporter effect is still present even once differences in productivity are accounted for. This result is therefore consistent with the recent international trade literature predicting a positive relation between markups and exports status, especially when intensive exporters are considered.

Table 17: Regression estimates at the industry level with industry-level markup as the dependent variable and quota coverage as the trade liberalization indicator

Variable	Model 1	Model 2	Model 3	Model 4
Independent				
Intercept	1.394 (1.952)	2.791 (2.016)	0.267 (0.626)	0.324 (0.806)
H	-11.228 (23.00)	-21.935 (23.53)	4.422 (7.586)	5.710 (10.19)
QUOTA	-0.460 (1.757)	-3.081 (3.106)	1.385 (1.197)	1.331 (2.366)
KQ	-0.531 (0.314)*	-1.608 (0.549)***	0.177 (0.231)	0.016 (0.402)
H*QUOTA		8.967 (41.13)		-7.001 (37.36)
KQ*QUOTA		2.847 (1.211)**		0.531 (1.062)
Year dummy				
1985	2.497 (1.143)**	2.132 (1.168)*	1.637 (0.841)**	1.581 (0.855)*
1986	-0.606 (0.705)	-0.674 (0.691)	-0.796 (0.598)	-0.773 (0.607)
1987	0.225 (0.610)	0.345 (0.600)	0.009 (0.563)	0.059 (0.578)
1988	0.250 (0.572)	0.217 (0.560)	0.315 (0.547)	0.333 (0.554)
1989	0.068 (0.561)	-0.044 (0.552)	0.213 (0.545)	0.219 (0.552)
Industry dummy				
1	-0.171 (1.643)	-0.416 (1.675)		
2	2.294 (1.277)*	2.461 (1.277)*		
4	-0.061 (1.347)	-0.251 (1.321)		
5	0.171 (0.958)	0.365 (0.942)		
6	0.742 (0.954)	0.872 (0.937)		
7	0.868 (1.075)	1.168 (1.062)		
8	0.479 (1.191)	0.555 (1.166)		
9	-0.916 (1.566)	-1.179 (1.540)		
10	0.598 (1.024)	0.376 (1.009)		
11	2.244 (0.935)**	2.707 (0.936)***		
12	2.875 (1.143)***	3.911 (1.202)***		
13	0.098 (1.179)	0.219 (1.155)		
14	0.603 (1.064)	1.404 (1.095)		
16	-0.139 (1.039)	-0.164 (1.017)		
17	0.464 (1.028)	0.933 (1.026)		
18	-0.183 (1.037)	-0.154 (1.020)		
19	-0.166 (0.977)	0.098 (0.962)		
N. of Observations	108	108	108	108
Root MSE	1.573	1.539	1.623	1.638
Adjusted R^2	0.256	0.288	0.208	0.194
F-statistic	2.480	2.600	4.510	3.570
Prob> F	0.001	0.000	0.000	0.000

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). Model 1 includes year and industry dummy variables. Model 1 includes year and industry dummy variables as well as the interactions between quota and Herfindahl index and quota and capital-output ratio. Model 3 includes only year dummy variables. Model 4 includes year dummy variables and the interactions between quota and Herfindahl index and quota and capital-output ratio.

Table 18: Regression estimates at the industry level with industry-level markup as the dependent variable and average tariff rate as the trade liberalization indicator

Variable	Model 1	Model 2	Model 3	Model 4
Independent				
Intercept	-1.516 (2.041)	1.962 (2.314)	-0.884 (0.727)	-0.941 (1.280)
H	-3.121 (21.78)	-30.882 (27.64)	1.989 (7.288)	0.193 (18.63)
TARIFF	14.372 (4.391)***	1.822 (7.155)	8.699 (2.863)***	8.908 (6.504)
KQ	-0.464 (0.294)	-2.850 (0.859)***	0.342 (0.230)	0.494 (0.623)
H*TARIFF		99.348 (106.5)		9.179 (98.80)
KQ*TARIFF		9.085 (3.128)***		-0.665 (2.552)
Year dummy				
1985	-0.329 (0.989)	-0.557 (0.955)	0.682 (0.771)	0.680 (0.792)
1986	-2.857 (0.870)***	-3.023 (0.836)***	-1.916 (0.701)***	-1.932 (0.715)***
1987	-0.727 (0.613)	-0.734 (0.589)	-0.515 (0.573)	-0.530 (0.581)
1988	0.495 (0.543)	0.470 (0.521)	0.409 (0.527)	0.401 (0.534)
1989	0.133 (0.528)	-0.052 (0.510)	0.209 (0.525)	0.208 (0.531)
Industry dummy				
1	1.070 (1.553)	0.437 (1.515)		
2	1.809 (1.084)*	2.251 (1.069)**		
4	0.249 (1.269)	0.130 (1.247)		
5	-0.596 (0.880)	-0.454 (0.851)		
6	0.005 (0.880)	0.164 (0.846)		
7	0.615 (1.010)	0.634 (0.991)		
8	1.099 (1.136)	0.988 (1.108)		
9	0.373 (1.525)	-0.076 (1.480)		
10	0.910 (0.968)	0.732 (0.948)		
11	1.915 (0.885)***	1.994 (0.849)**		
12	4.119 (1.135)***	6.055 (1.265)***		
13	0.436 (1.113)	0.433 (1.085)		
14	1.876 (1.069)*	2.994 (1.118)***		
16	0.548 (0.997)	0.472 (0.974)		
17	1.289 (0.999)	1.772 (0.986)*		
18	-0.011 (0.977)	0.091 (0.968)		
19	0.489 (0.926)	0.773 (0.899)		
N. of Observations	108	108	108	108
Root MSE	1.480	1.420	1.563	1.580
Adjusted R^2	0.394	0.208	0.265	0.251
F-statistic	3.220	3.580	5.830	4.580
Prob> F	0.000	0.000	0.000	0.000

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). Model 1 includes year and industry dummy variables. Model 1 includes year and industry dummy variables as well as the interactions between tariff and Herfindahl index and tariff and capital-output ratio. Model 3 includes only year dummy variables. Model 4 includes year dummy variables and the interactions between tariff and Herfindahl index and tariff and capital-output ratio.

Table 19: Regression estimates at the industry level with industry-average markup as the dependent variable and quota coverage as the trade liberalization indicator

Variable	Model 1	Model 2	Model 3	Model 4
Independent				
Intercept	1.283 (0.163)***	1.278 (0.175)***	0.593 (0.180)***	0.700 (0.225)***
H	-2.051 (1.923)	-2.044 (2.055)	7.254 (2.173)***	3.667 (2.868)
QUOTA	-0.019 (0.147)	-0.026 (0.266)	1.076 (0.346)***	0.443 (0.663)
KQ	0.025 (0.026)	0.031 (0.048)	0.164 (0.067)**	0.295 (0.114)**
H*QUOTA		0.288 (3.459)		18.843 (10.24)*
KQ*QUOTA		-0.015 (0.106)		-0.438 (0.298)
Year dummy				
1985	-0.119 (0.095)	-0.115 (0.100)	-0.593 (0.243)**	-0.534 (0.241)**
1986	-0.121 (0.058)**	-0.120 (0.059)**	-0.231 (0.173)	-0.269 (0.171)
1987	-0.056 (0.050)	-0.057 (0.051)	-0.113 (0.163)	-0.186 (0.163)
1988	-0.036 (0.048)	-0.036 (0.048)	0.022 (0.158)	-0.013 (0.156)
1989	-0.057 (0.047)	-0.057 (0.047)	0.011 (0.158)	-0.013 (0.155)
Industry dummy				
1	-0.800 (0.138)***	-0.795 (0.144)***		
2	0.447 (0.106)***	0.448 (0.109)***		
4	-0.143 (0.112)	-0.142 (0.114)		
5	0.599 (0.080)***	0.598 (0.082)***		
6	1.101 (0.080)***	1.102 (0.081)***		
7	0.070 (0.090)	0.069 (0.092)		
8	-0.183 (0.099)*	-0.183 (0.100)*		
9	-0.381 (0.131)***	-0.379 (0.134)***		
10	-0.012 (0.085)	-0.010 (0.087)		
11	-0.803 (0.078)***	-0.805 (0.081)***		
12	1.037 (0.095)***	1.032 (0.103)***		
13	0.081 (0.099)	0.080 (0.100)		
14	0.074 (0.089)	0.070 (0.096)		
16	-0.202 (0.087)**	-0.202 (0.088)***		
17	-0.155 (0.086)*	-0.157 (0.088)*		
18	-0.312 (0.087)***	-0.311 (0.088)***		
19	0.247 (0.081)***	0.246 (0.083)***		
N. of Observations	108	108	108	108
Root MSE	0.131	0.133	0.470	0.462
Adjusted R^2	0.933	0.931	0.147	0.175
F-statistic	60.70	54.85	3.300	3.280
Prob> F	0.000	0.000	0.000	0.000

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). Model 1 includes year and industry dummy variables as well as the interactions between quota and Herfindahl index and quota and capital-output ratio. Model 3 includes only year dummy variables. Model 4 includes year dummy variables and the interactions between quota and Herfindahl index and quota and capital-output ratio.

Table 20: Regression estimates at the industry level with industry-average markup as the dependent variable and average tariff rate as the trade liberalization indicator

Variable	Model 1	Model 2	Model 3	Model 4
Independent				
Intercept	1.335 (0.179)***	1.421 (0.217)***	0.448 (0.223)**	-0.144 (0.360)
H	-2.194 (1.928)	-3.245 (2.595)	6.367 (2.238)***	3.801 (5.109)
TARIFF	-0.259 (0.387)	-0.624 (0.661)	1.615 (0.883)*	4.399 (1.817)**
KQ	0.025 (0.026)	-0.019 (0.082)	0.188 (0.071)***	0.914 (0.178)***
H*TARIFF		4.860 (9.502)		9.583 (26.67)
KQ*TARIFF		0.160 (0.299)		-3.202 (0.733)***
Year dummy				
1985	-0.083 (0.087)	-0.087 (0.089)	-0.335 (0.237)	-0.376 (0.223)*
1986	-0.086 (0.077)	-0.090 (0.078)	-0.275 (0.216)	-0.375 (0.201)*
1987	-0.042 (0.054)	-0.043 (0.054)	-0.133 (0.177)	-0.200 (0.164)
1988	-0.041 (0.048)	-0.041 (0.048)	0.050 (0.163)	0.015 (0.151)
1989	-0.058 (0.047)	-0.062 (0.047)	0.014 (0.162)	0.013 (0.150)
Industry dummy				
1	-0.828 (0.137)***	-0.838 (0.141)***		
2	0.447 (0.095)***	0.460 (0.099)***		
4	-0.147 (0.112)	-0.144 (0.115)		
5	0.609 (0.078)***	0.607 (0.080)***		
6	1.110 (0.078)***	1.113 (0.079)***		
7	0.074 (0.090)	0.079 (0.092)		
8	-0.194 (0.100)*	-0.191 (0.102)*		
9	-0.405 (0.135)***	-0.412 (0.138)***		
10	-0.018 (0.085)	-0.017 (0.088)		
11	-0.798 (0.078)***	-0.797 (0.079)***		
12	1.012 (0.100)***	1.052 (0.118)***		
13	0.075 (0.099)	0.080 (0.101)		
14	0.052 (0.095)	0.083 (0.105)		
16	-0.213 (0.088)**	-0.211 (0.090)**		
17	-0.170 (0.088)*	-0.156 (0.091)*		
18	-0.315 (0.087)***	-0.308 (0.090)***		
19	0.232 (0.082)***	0.241 (0.083)***		
N. of Observations	108	108	108	108
Root MSE	0.131	0.132	0.484	0.447
Adjusted R^2	0.933	0.932	0.094	0.228
F-statistic	61.04	55.51	2.380	4.160
Prob> F	0.000	0.000	0.000	0.000

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). Model 1 includes year and industry dummy variables. Model 1 includes year and industry dummy variables as well as the interactions between tariff and Herfindahl index and tariff and capital-output ratio. Model 3 includes only year dummy variables. Model 4 includes year dummy variables and the interactions between tariff and Herfindahl index and tariff and capital-output ratio.

Table 21: Regression estimates at the plant level with plant-level markup as the dependent variable

Variable	Model 1	Model 2
Independent		
Intercept	1.0684 (0.1175)***	1.1154 (0.1485)***
SHARE	1.6139 (0.2531)***	1.3719 (0.3773)***
SHARE^2	-0.2433 (0.1261)*	-0.2219 (0.1275)*
KQ	0.1659 (0.0057)***	0.1658 (0.0057)***
KQ^2	-1.8e^{-4} (8.3e^{-6})***	-1.8e^{-4} (8.3e^{-6})***
QUOTA	-0.0048 (0.2349)	
TARIF		-0.2925 (0.5565)
SHARE*QUOTA	0.6111 (0.5275)	
SHARE*TARIFF		1.9790 (1.6234)
Year dummy		
1985	-0.1134 (0.1370)	-0.0654 (0.1126)
1986	-0.1257 (0.0771)*	-0.0908 (0.1012)
1987	-0.1004 (0.0649)	-0.0848 (0.0707)
1988	-0.0603 (0.0617)	-0.0614 (0.0619)
1989	-0.0602 (0.0619)	-0.0596 (0.0618)
Industry dummy		
1	-0.7805 (0.1311)***	-0.7909 (0.1254)***
2	0.2169 (0.1489)	0.2269 (0.1331)*
4	-0.1105 (0.1321)	-0.1115 (0.1318)
5	0.5857 (0.1432)***	0.5990 (0.1396)***
6	1.1255 (0.1833)***	1.1399 (0.1806)***
7	0.0333 (0.1454)	0.0415 (0.1460)
8	-0.2394 (0.1295)**	-0.2496 (0.1304)**
9	-0.3969 (0.1189)***	-0.4125 (0.1236)***
10	-0.0451 (0.1244)	-0.0510 (0.1249)
11	-1.1191 (0.1938)***	-1.1326 (0.1943)***
12	0.6382 (0.1824)***	0.6229 (0.1889)***
13	0.0650 (0.1337)	0.0627 (0.1334)
14	-0.2687 (0.1421)*	-0.2924 (0.1533)*
16	-0.2190 (0.1316)*	-0.2299 (0.1327)*
17	-0.2457 (0.1296)*	-0.2601 (0.1329)**
18	-0.3643 (0.1309)***	-0.3699 (0.1308)***
19	-0.0406 (0.1330)	-0.0515 (0.1335)
N. of Observations	11205	11205
Root MSE	1.814	1.814
Adjusted R^2	0.127	0.127
F-statistic	59.48	59.49
Prob> F	0.000	0.000

Note: The stars indicate significance levels (**$p < 0.05$, ***$p < 0.01$). Model 1 includes quota as a trade liberalization indicator. Model 2 includes tariff as a trade liberalization indicator.

Table 22: Markups and export status

Variable	Model 1	Model 2
$E_{H_{jt}}$	0.0216 (0.0082)***	0.0198 (0.0076)***
ω_{jt}		0.3977 (0.0118)***
Linear restriction		
$\psi_1 + \exp(\psi_0)$	0.0588 (0.0222)***	0.0539 (0.0208)***
N. of Observations	7929	7929
Root MSE	0.287	0.268
Adjusted R^2	0.871	0.887
F-statistic	590.8	669.4
Prob$>F$	0.000	0.000

Note: The stars indicate significance levels ($^{***}p < 0.01$). Model 1 includes only the export intensity dummy variable and the set of z_{jt} controls. Model 2 includes the export intensity dummy variable, the set of z_{jt} controls as well as productivity.

8 Conclusions

In this contribution I focus on the importance of correctly estimating production function parameters and price-cost margins in order to assess differences in technology, productivity, and market power among eighteen Mexican manufacturing firms and evaluate the impact of trade liberalizing policies on their profitability.

Relying on a structural framework that corrects the simultaneity bias using investment or intermediate inputs as a proxy for unobserved productivity I estimate production function parameters. My results confirm the well establish empirical evidence that production function coefficients obtained with OLS are biased and support the argument that controlling for firm-specific productivity shocks successfully corrects this bias. In fact, compared to OLS, the structural estimation delivers a much lower labor parameter and a higher capital parameter. I also find evidence of constant returns to scale in the majority of the industries analyzed.

The second step in my empirical investigation is to use the production function estimates to recover firm-level markups adopting a structural approach in which markups are derived from cost minimization first order conditions and can be interpreted as the wedge between the cost share of production factors and their revenue share. I test the validity of this approach by comparing the firm-level markup estimates with industry-level markups obtained through a less sophisticated dual estimation approach. The price cost-margins estimated at the plant level are more reasonable in terms of magnitude and significantly higher than their industry-level counterparts. This result demonstrates that explicitly taking into account differences in productivity is crucial in assessing the extent of market power.

Finally, I exploit the fact that the sample spans over a period of dramatic reforms in the Mexican economy to quantify the impact of trade exposure on the markups. I conduct an industry-level as well as a plant-level analysis relating price-cost margins and measures of import liberalization. The industry-level evidence confirms the hypothesis of import discipline, i.e. the removal of trade protections negatively affected the profitability of domestic firms, but this evidence is not confirmed in the plant-level analysis. Nonetheless, this is not a very surprising result since the Mexican manufacturing sector was presumably quite competitive even prior to the trade policy reforms because of the large number of firms operating in this sector. In addition, I test the prediction of several recent international trade models that larger firms are likely to be more productive, thus can charge higher markups and afford to pay a sunk cost to become exporters. In the case of Mexican exporters I find a statistically significant markup premium only for "intensive" exporters, i.e. firms that export a high percentage of their output, and for these firms the premium prevails even after netting out the effect of productivity. Furthermore, as expected, productivity proves to have a positive and highly significant effect on the markups confirming that the most productive firms have, on average, higher markups.

References

[1] Ackerberg D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes", in *Handbook of Econometrics*, Vol. 6A, Chapter 64.

[2] Ackerberg D., K. Caves, and G. Frazer (2006): "Structural Identification of Production Functions", mimeo UCLA.

[3] Aw B., M. Roberts, and D. Xu (2011): "R&D Investment, Exporting and Productivity Dynamics", *American Economic Review*, 101(4), 1312-1344.

[4] Bernard A., J. Eaton, J. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade", *American Economic Review*, 93(4), 1268-1290.

[5] Blundell R., and S. Bond (2000): "GMM Estimation with Persistent Panel Data: an Application to Production Functions", *Econometric Reviews*, 19(3), 321-340.

[6] Buettner T. (2005): "R&D and the Dynamics of Productivity", mimeo LSE.

[7] Chamberlain G. (1982): "Multivariate Regression Models for Panel Data", *Journal of Econometrics*, 18(1), 5-46.

[8] Cobb C., and P. Douglas (1928): "A Theory of Production", *American Economic Review*, 18(1), 139-165.

[9] De Loecker J., and F. Warzynski (2012): "Markups and Firm-Level Export Status", *American Economic Review*, 102(6), 2437-2471.

[10] Demsetz H. (1973): "Industry Structure, Market Rivalry, and Public Policy", *Journal of Law and Economics*, 16(1), 1-9.

[11] Domowitz I., J. Hubbard, and B. Peterson (1986): "Business Cycles and the Relation between Concentration and Price-Cost Margins", *Rand Journal of Economics*, 17(1), 1-17.

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

[13] Grether J-M. (1996): "Mexico, 1985-1990: Trade Liberalization, Market Structure and Manufacturing Performance", in *Industrial Evolution in Developing Countries*, ed. by M. Roberts and J. Tybout, Oxford University Press.

[14] Griliches Z., and J. Mairesse (1995): "Production Functions: the Search for Identification", Working Paper 5067, NBER.

[15] Hall R. (1988): "The Relation between Price and Marginal Cost in U.S. Industry", *Journal of Political Economy*, 96(5), 921-947.

[16] Hall R., O. Blanchard, and G. Hubbard (1986): "Market Structure and Macroeconomic Fluctuations", Brookings Papers on Economic Activity, 2, 285-338.

[17] Hallak J. C., and J. Sivadasan (2009): "Firms' Exporting Behavior under Quality Constraints", Working Paper 14928, NBER.

[18] Hoch I. (1955): Estimation of Production Function parameters and testing for Efficiency", *Econometrica*, 23(3), 325-326.

[19] Hoch I. (1962): "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data", *Econometrica*, 30(1), 34-53.

[20] Horowitz J. (2001): "The Bootstrap", Handbook of Econometrics, Vol. 5, Chapter 52.

[21] Kehoe T. (1995): "A Review of Mexico's Trade Policy from 1982 to 1994", *The World Economy*, 18, 135-151.

[22] Klette T., and Z. Griliches (1996): "The Inconsistency of Commomn Scale Estimators when Output Prices are Unobserved and Endogenous", *Journal of Applied Econometrics*, 11(4), 343-361.

[23] Konings J., and H. Vandenbussche (2005): "Antidumping protection and markups of domestic firms", *Journal of International Economics*, 65, 151-165.

[24] Krishna P., and D. Mitra (1998): "Trade Liberalizaton, Market Discipline and Productivity Growth: New Evidence from India", *Journal of Development Economics*, 56(2), 447-462.

Kugler M., and E. Verhoogen (2012): "Prices, Plant Size, and Product Quality", *Review of Economic Studies*, 79(1), 307-339.

[25] Levinshon J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables", *Review of Economics Studies*, 70(2), 317-340.

[26] Marshack J., and W. Andrews (1944): "Random Simultaneous Equations and the Theory of Production", *Econometrica*, 12(3-4), 143-205.

Melitz M., and G. Ottaviano (2008): "Market Size, Trade, and Productivity", *Review of Economic Studies*, 75(1), 295-316.

[27] Mendershausen H. (1938): "On the Significance of Professor Douglas' Production Function", *Econometrica*, 6(2), 143-205.

[28] Mundlak Y. (1961): "Empirical Production Function Free of Management Bias", *Journal of Farm Economics*, 43(1), 44-56.

[29] Nerlove M. (1963): "Returns to Scale in Electricity Supply", in *Measurement in Economics*, ed. by C. Christ, Stanford University Press.

[30] Olley S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Industry" *Econometrica*, 64(6), 1263-1295.

[31] Pavcnik N. (2002): "Trade Liberalization, Exit and Productivity Improvements: Evidence from Chilean Plants", *Review of Economic Studies*, 69(1), 245-276.

[32] Roberts M. (1996), "Colombia, 1977-85: Producer Turnover, Margins, and Trade Exposure", in *Industrial Evolution in Developing Countries*, ed. by M. Roberts and J. Tybout, Oxford University Press.

[33] Roberts M., and J. Tybout (1996): "Industrial Evolution in Developing Countries", Oxford University Press for the World Bank.

[34] Roeger W. (1995): "Can Imperfect Competition Explain the Difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing", *Journal of Political Economy*, 103(2), 316-330.

[35] Shmalensee R. (1985): "Do Markets Differ Much?", *American Economic Review*, 75(3), 341-351.

[36] Tybout J. (1996): "Chile, 1979-1986: Trade Liberalization and its Aftermath", in *Industrial Evolution in Developing Countries*, ed by M. Roberts and J. Tybout, Oxford University Press.

[37] Tybout J., and D.Westbrook (1991): "Trade Liberalization and the Dimension of Efficiency Change in Mexican Manufacturing Industries" *Journal of International Economics*, 39(1-2), 55-78.