NBER WORKING PAPER SERIES

FIRM PERFORMANCE IN A GLOBAL MARKET

Jan De Loecker
Pinelopi Koujianou Goldberg

Working Paper 19308
http://www.nber.org/papers/w19308

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2013

We would like to thank Tim Bresnahan for early conversations on the topic, and Channing Verbeck Jr. for excellent research assistance. This paper is not intended to be a comprehensive literature review. Instead, we discuss a selected set of papers from the perspective of our empirical framework that nests the productivity and pass-through literatures spanning Industrial Organization, International Trade and International Macro. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Firm Performance in a Global Market
Jan De Loecker and Pinelopi Koujianou Goldberg
NBER Working Paper No. 19308
August 2013
JEL No. D22,D40,F10,L1

ABSTRACT

In this article we introduce an empirical framework to analyze how firm performance is affected by increased globalization. Using this framework we discuss recent work on measuring the impact of various shocks firms face in the global marketplace, such as reductions in trade costs (through lowering tariffs and abolishing quotas). Our analytical framework nests most empirical approaches to estimating the impact of trade and industrial policies on firms active in international markets. We identify outstanding issues surrounding the identification of the underlying mechanisms and conclude with suggestions for future research.

Jan De Loecker
Department of Economics
307 Fisher Hall
Princeton University
Princeton, NJ 08544-1021
and NBER
jdeloeck@princeton.edu

Pinelopi Koujianou Goldberg
Yale University
Department of Economics
37 Hillhouse Ave.
P.O. Box 208264
New Haven, CT 06520-8264
and NBER
penny.goldberg@yale.edu

1 Introduction

Few topics have been researched as extensively as the relationship between trade openness and firm, or industry, performance. A Google Scholar search on the keywords "trade and productivity"returns 1,920,000 papers! The purpose of this article is twofold: First, we develop a general framework for discussing the large empirical literature on the topic. Second, we summarize the main insights from work to date. While much of the empirical research we draw upon rests on theoretical models, we do not review theoretical work directly, except to the extent that this theoretical work informs empirical specifications and helps identify underlying mechanisms.

Our point of departure is the voluminous empirical literature on the effects of trade on firm productivity. Economists have always postulated that one of the main benefits of opening up markets to foreign competition is to make firms more "efficient" and have proceeded to estimate the effects of trade liberalizations, or globalization more generally, on efficiency. Early empirical attempts predate theoretical work in this area and for the most part lack a sound theoretical justification for why one would expect efficiency to increase with exposure to foreign markets. With the development of models of firm heterogeneity, however, this literature has experienced a renaissance. There are two main strands. The first one investigates the performance of exporters and relies primarily on cross-sectional comparisons in order to assess whether exporters are more efficient than firms that sell to domestic markets only. The second strand focuses on trade liberalization episodes and investigates whether industry productivity increases in the post-reform period, where the productivity gains can arise either because of within-firm improvements or because of reallocation of economic activity towards more efficient firms.

While this literature has established some interesting patterns, we show that with few exceptions, it has been loose in its use of the term "productivity". What it actually delivers is a measure of "firm performance" or "profitability". The distinction between "productivity" and "profitability" is important; the latter will depend not only on physical efficiency, but also on prices, which will reflect product differentiation and markups in addition to costs. The framework we develop allows one to explicitly trace these components. This decomposition offers the advantage that one can link improvements in firm performance to specific mechanisms though which globalization affects firms. Understanding these mechanisms is important for assessing the welfare and distributional effects of trade openness; for example, a trade liberalization that improves firm performance by inducing improvements in physical efficiency has different implications from a liberalization that makes firms better off by increasing their profits.

The realization that measured firm performance captures markups as well as physical efficiency naturally leads to two other literatures that were developed in different contexts; the large Industrial Organization literature on imperfect competition and the International literature on incomplete (exchange rate) pass-through. The first explicitly investigates the measurement and determinants of markups (e.g., the role of market structure, product differentiation and demand elasticities); the second focuses on how a certain type of cost shock (i.e., exchange rate changes) is passed through to prices. Firms competing in global markets are faced with a different market (and likely also demand) structure, and may as a result change their prices and markups or their product characteristics. A tariff reduction for example intensifies the import competition domestic firms face and we would expect them to reduce their prices and markups in response. This is the usual pro-competitive effect one associates with trade liberalizations. At the same time, a reduction in tariffs allows firms to purchase imported intermediates that they use as inputs in their production at lower prices; these reductions in input prices are tantamount to a cost shock facing the firm, similar to the exchange rate changes considered in the International literature. Whether or not this cost shock will be passed through to prices will depend on many factors that reflect demand and market structure conditions.

Our review concludes that there is one robust finding that emerges from this literature: globalization improves industry performance. However, there is less of a consensus on how these improvements arise. Many studies find that improvements are generated through reallocation of market shares towards better performing firms, while others document significant within-firm improvements. A number of papers find that the effects of input tariff liberalization dominate those of output tariff liberalization in some developing countries. Perhaps more importantly, the majority of studies do not distinguish between physical efficiency and price/markup effects. Hence, it is not clear that the aforementioned improvements represent true productivity gains as opposed to increases in market power. A couple of recent papers that have emphasized this distinction show that the effects on markups and prices are significant. However, it is too early to know whether the results from these studies generalize to other contexts. This is an exciting area for future research and we hope that the current article will help guide future endeavors in this direction.

The remainder of the article is organized as follows. In section 2 we introduce a general framework that can be used to discuss existing work and clarify what economic theory suggests should be measured, and what is measured in practice. We use this framework to decompose "firm performance" into its components. In section 3, we discuss the particular mechanisms through which trade openness is expected to affect each of these components. Section 4 briefly summarizes the insights obtained from work to date and section 5 concludes.

2 A Framework for Measuring Performance

Performance at the firm level is measured in many different ways. Such ways include accounting measures of profitability, the Lerner index, sales per input and total factor productivity. While correlated, the various measures capture different aspects of firm performance, and exposure to a global market is not expected to affect these aspects in the same way. In this section, we first use a simple regression framework to summarize the large empirical literature on the topic. Next, we introduce an empirical framework, based on a production function, that allows us to reinterpret the commonly used measures of performance and decompose them into their underlying determinants: demand and cost primitives, as well as the market structure of the industry under study.

2.1 Measuring Performance

We characterize a large literature – spanning Industrial Organization, International Economics and International Macroeconomics – by considering firm performance (π_{it}) as the residual in a regression of sales (s_{it}) on input expenditures ($\boldsymbol{e}_{it}$).[1] Applied researchers typically consider a log-linear relationship between sales and expenditures:

$$s_{it} = \boldsymbol{e}'_{it}\boldsymbol{\beta} + \pi_{it}, \tag{1}$$

where $\boldsymbol{\beta}$ is the vector of coefficients. The data we have in mind track firms indexed by i for which we observe sales and input use, both expressed in monetary terms (i.e., sales and expenditures), over time t. The residual π_{it} as a measure of performance is closely related to profits since we subtract expenditures from sales.

The fundamental research question economists have been trying to answer over the last few decades is: How do changes in international competition affect the performance of firms (π_{it}), industries, and ultimately the overall welfare of countries or regions? Trade liberalization episodes, preferably exogenous to firms in an industry, are the natural place

---

[1] The vector $\boldsymbol{e}_{it}$ typically includes expenditures on labor, intermediate inputs (materials), and capital. Unless noted otherwise, small letters denote the log of the corresponding variable throughout this article.

to turn to in order to analyze this question empirically by relating performance measures to changes in trade protection, such as tariff declines, lifting of quota restrictions, or removal of antidumping duties.

Equation (1) is a natural point of departure for the literature that typically utilizes firm- or plant-level data across many different sectors of one or more economies. Such data tend to be readily available for a large set of countries and time periods. In contrast, the case study approach commonly employed in modern Industrial Organization makes use of much more detailed information on product-level prices and product- or firm-specific cost variables to compute performance measures - at the cost that the approach is only feasible for the specific industries and countries for which this information is available. Another approach towards measuring performance would be to use accounting data to measure performance as operating profits. However, this article is concerned with recovering measures of firm performance that accurately reflect *economic* costs and benefits and it is well known that accounting profits provide poor measures of economic profits.[2]

With few exceptions, the existing literature has viewed equation (1) as the empirical analogue of a production function and interpreted the residual π_{it} as a measure of total factor productivity (TFP). As we argue below, under a set of restrictive assumptions, equation (1) is indeed a production function with $\boldsymbol{\beta}$ the estimated production function coefficients. One possible set of such assumptions features firms that produce a homogeneous product and are active in an industry characterized by perfectly competitive output and input markets. In this case, the relationship between sales and expenditures is equivalent to the relationship between physical output and inputs, and consequently π_{it} is an estimate of (physical) productivity. However, in the general case, the performance residual π_{it} will capture much more than productivity.

When we try to understand how increased international competition impacts firm performance, recovering the residual of (1) is only the first step; in order to identify the underlying mechanism(s) we need to understand the underlying components of π_{it}. To this end, we need to be more precise about what π_{it} actually measures. We turn to this issue in the next section, and use a production function approach to interpret the performance residual (π_{it}).

---

[2] See Schmalensee (1989) for a detailed discussion of the problems of using accounting profits as a measure of firm performance.

2.2 Firm Performance: A View from the Production Side

Using a basic production function[3] that relates quantities produced (q_{it}) to a vector of physical input usage ($\boldsymbol{x}_{it}$) allows us to express sales in a (perhaps) more familiar fashion as:

$$\begin{aligned} s_{it} &= \boldsymbol{x}'_{it}\boldsymbol{\alpha} + \omega_{it} + p_{it} \\ &= \boldsymbol{e}'_{it}\boldsymbol{\alpha} + \omega_{it} + p_{it} - \boldsymbol{z}'_{it}\boldsymbol{\alpha}, \end{aligned} \tag{2}$$

where we relied on the definition of sales: $s_{it} = q_{it} + p_{it}$; a standard Hicks-neutral production function: $q_{it} = \boldsymbol{x}'_{it}\boldsymbol{\alpha} + \omega_{it}$, with $\boldsymbol{\alpha}$ and ω_{it} the vector of production function coefficients and productivity, respectively[4]; and the definition of input expenditures: $\boldsymbol{e}_{it} = \boldsymbol{x}_{it} + \boldsymbol{z}_{it}$, with $\boldsymbol{z}_{it}$ the vector of input prices.

This simple framework allows us to immediately connect our structural equation (2) to the profitability residual in (1) by grouping the last three terms in the equation above to $\pi_{it} = \omega_{it} + p_{it} - \boldsymbol{z}'_{it}\boldsymbol{\alpha}$. Hence, we can rewrite (2) as an equation relating sales to expenditures:

$$s_{it} = \boldsymbol{e}'_{it}\boldsymbol{\alpha} + \pi_{it}. \tag{3}$$

Equations (1) and (3) are almost identical up to the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. We distinguish between the structural coefficients of the production function, $\boldsymbol{\alpha}$, and the coefficients obtained after estimating an equation such as (1). It is only under a set of very restrictive assumptions that these two vectors will be identical.

Consider equation (1) and assume that the researcher is willing to make the economic assumption that there is neither output nor input price variation across firms.[5] Even in this case, OLS estimation of equation (1) on a panel dataset of firms is not expected to yield the vector of coefficients ($\boldsymbol{\alpha}$) due to the well known simultaneity and selection biases. Both biases arise from the likely correlation between inputs and unobserved productivity (ω_{it}). This is a well known problem when estimating production functions that has been addressed in the literature and we refer the reader to a recent overview by Ackerberg, Benkard, Berry, and Pakes (2007) for a detailed discussion. For the remainder of this article, we assume that these biases can be appropriately dealt with

---

[3] To simplify notation, we base our discussion on a Cobb-Douglas production function, but our framework generalizes to any other functional form.

[4] For the well-known Cobb-Douglas production function in three inputs (labor (l), materials (m) and capital (k), we get: $q_{it} = \alpha_l l_{it} + \alpha_m m_{it} + \beta_k k_{it} + \omega_{it}$.

[5] Time variation in both output and input prices can be accommodated by either deflating the variables appropriately using industry-wide deflators, or by including a full set of year fixed effects. In any case, the price variables would not be indexed by i.

when estimating the underlying structural parameters. Our point is that if output and input prices do not vary across firms, it is possible in principle to recover both the production function coefficients $\boldsymbol{\alpha}$ and firm productivity ω_{it} by applying the techniques suggested in the recent literature.

Consider equation (1) once more, but let us now allow for both output and input price variation. In this case, the structural error π_{it} contains two more components in addition to productivity: the output price (p_{it}) and the vector of input prices ($\boldsymbol{z}_{it}$). Relying on sales and expenditure data will clearly not deliver an estimate of productivity, nor will it deliver the vector of production function coefficients. In this case, $\boldsymbol{\beta}$ is a vector of coefficients describing the mapping from expenditures to sales.

To see why estimation of equation (1) will lead to biased coefficients in the presence of output and/or input price variation, let us explicitly introduce deflators and rewrite (2) to reflect the usual practice in empirical work in this area. Let $\overline{p}_t^I$ be the price deflator for industry I, $\bar{\boldsymbol{z}}_t^I$ a vector of industry-specific input price deflators, and $p_{it}^* = p_{it} - \overline{p}_t^I$; $\boldsymbol{z}_{it}^* = \boldsymbol{z}_{it} - \bar{\boldsymbol{z}}_t^I$; $s_{it}^* = s_{it} - \overline{p}_t^I$; and $\boldsymbol{e}_{it}^* = \boldsymbol{e}_{it} - \bar{\boldsymbol{z}}_t^I$ denote deflated output price, input prices, sales and expenditures respectively. Researchers typically take the following version of (2) to the data:

$$\begin{aligned} s_{it}^* &= \boldsymbol{e}_{it}^{*\prime}\boldsymbol{\alpha} + \omega_{it} + p_{it}^* - \boldsymbol{z}_{it}^{*\prime}\boldsymbol{\alpha} \\ &= (\boldsymbol{x}_{it} + \boldsymbol{z}_{it}^*)'\boldsymbol{\alpha} + \omega_{it} + p_{it}^* - \boldsymbol{z}_{it}^{*\prime}\boldsymbol{\alpha}, \end{aligned} \tag{4}$$

Consider first the case where there is output price, but no input price variation across firms. In this case, $\boldsymbol{z}_{it}^* = 0$, and the terms related to input price variation drop out from (4). This is the typical case considered in Industrial Organization studies which assume that input prices are equalized across firms (once regional differences are controlled for). Output price variation captured by p_{it}^* will generally be correlated with $\boldsymbol{e}_{it}^*$ (which is equal to $\boldsymbol{x}_{it}$ in this scenario): ceteris paribus, we expect firms that charge higher prices to sell lower quantities, which in turn implies lower input quantities. Hence, the correlation between p_{it}^* and $\boldsymbol{x}_{it}$ is likely to be negative, leading to a downward bias in the estimates of the coefficients $\boldsymbol{\alpha}$ as well as the returns to scale. This bias was the focus of the work of Klette and Griliches (1996), as well as De Loecker (2011). We refer to it as the *output price bias*. A closely related point was made by Katayama, Lu, and Tybout (2009) who allow for input prices to vary across firms, but presume that this input price variation can be completely controlled for when one observes firm-specific wages (so that $\boldsymbol{z}_{it}^*$ can be treated as an observable). As we argue below, this assumption is strong – even when wages are observed, the prices of other inputs (e.g., materials) are typically not observed, while a firm-specific price for capital is never available.

Second, consider the case where there is no output price variation, but input prices vary across firms. This case may seem irrelevant in practice, but as we show below, it corresponds to the case where we observe firm-specific output prices and hence can control for output price variation, while input price variation remains uncontrolled for. In this case, it is evident from (4) that input price variation will lead to a strong negative bias in the estimated coefficients; a firm that faces higher input prices will have higher input expenditures that will not lead to higher physical output. We refer to this bias as the *input price bias.* In contrast to the output price bias, the input price bias has received no attention in the literature so far; the only study we are aware of that has attempted to address it is De Loecker, Goldberg, Khandelwal, and Pavcnik (2012) (DLGKP hereafter).

Realistically, the data will be characterized by both output and input price variation, so that estimates of equation (1) will suffer from both output and input price bias. Unless these two biases interact in a way so as to offset each other, they will lead to biased estimates of the production function coefficients. Further, even conditional on the parameter vector that the estimation delivers in this case, one will be able to recover only a (biased) estimate of the composite residual π_{it}.

The message so far is that with sales and expenditure data alone, one cannot, in general, recover the underlying components of firm performance or identify productivity. As always in empirical work, there are two ways out of such a situation: either one collects more data or one makes additional assumptions that will allow identification.

2.2.1 Additional Data

One might be tempted to conclude at this point that the solution to all issues discussed in this section would be to collect more data on firms' output and input prices. Obviously, more data are always preferable (at a minimum, one can ignore them). However, the introduction of additional data creates its own challenges; while more data may help alleviate some of the problems discussed above, they are not a panacea.

Output prices tend to be more readily available than input prices in firm or plant-level surveys. However, these data are often unit values (derived by dividing revenues by quantities over a period of time), and suffer from the well documented problems associated with unit values. Further, their use in the context of production function estimation poses additional challenges associated with differences in the units in which quantities are recorded across firms and products. Finally, the attempt to exploit output prices forces the researcher to explicitly confront issues that are specific to multiproduct firms; output prices are recorded at the product level, which calls for estimation of production functions at the product level. However, input expenditures are only recorded at the firm level. Hence, even when output prices are available, estimation of production functions for multiproduct firms is not possible unless one adopts one of three approaches: (a) eliminate multiproduct firms from the sample and focus on single-product firms only; (b) aggregate product prices to the firm level and conduct the analysis at the firm level; (c) devise a mechanism for allocating firm input expenditures to individual products and conduct the analysis at the product level.

Each of these approaches has its drawbacks: Given that multiproduct firms account for a significant fraction of output in the manufacturing sector, eliminating them from the analysis is hard to defend; approach (b) requires one to specify a demand system that allows aggregation in a consistent manner and creates the need for additional assumptions; similarly, approach (c) requires spelling out the assumptions needed in order to allocate input expenditures across products.

The fact that the use of output price data requires assumptions is not necessarily a weakness, especially since the assumptions one needs in this context are either already made (albeit implicitly) in the existing literature or are weaker than the ones required under alternative approaches. Nevertheless, our point is that more data do not eliminate the need for assumptions and structure.

Assuming the challenges outlined above can be dealt with, the use of output prices should allow one to eliminate the output price bias we discussed above. Syverson (2004) and Foster, Haltiwanger, and Syverson (2008) are examples of studies that have accomplished this successfully. They rely on a selected set of plausibly homogeneous good industries (ready-mixed concrete for example) and exploit output price data in order to separate out price variation from productivity. An implicit assumption in their framework is that input prices do not vary across firms. This assumption is indeed plausible in the context of the homogeneous product industries they consider; for example, it is plausible to assume that (conditional on region), the input prices ready-mixed concrete producers face are the same. In this setting, the only bias present in (4) is the output price bias, and this bias can be successfully dealt with when one observes output prices.

However, the focus on homogeneous product industries leaves us with the bulk of economic activity unaccounted for. The set of industries characterized by substantial product differentiation comprises a large share of economic activity. In such industries, controlling for output prices alone is insufficient; differentiated products require differentiated inputs, so that we would expect input prices to vary across firms, even when these firms are located in the same region and even when input markets are perfectly competitive. Input prices are typically unobserved. Controlling for output prices in this case will eliminate the output price bias, but will leave the input price bias intact and will in fact make its consequences for estimation more salient. In this case, (4) becomes:

$$q_{it} = (\boldsymbol{x}_{it} + \boldsymbol{z}^{*}_{it})'\boldsymbol{\alpha} + \omega_{it} - \boldsymbol{z}^{*\prime}_{it}\boldsymbol{\alpha}. \tag{5}$$

The problem with trying to estimate this version of (4) is immediately apparent: without controlling for the (unobserved) variation in input prices, the coefficients $\boldsymbol{\alpha}$ will be seriously biased. The seriousness of this problem is demonstrated in DLGKP: the authors estimate a physical production function for Indian manufacturing that relates physical output to (deflated) input expenditures. When input price variation is not controlled for, the coefficients $\boldsymbol{\alpha}$ often seem nonsensical and have the wrong sign. Yet, these apparently nonsensical results make a lot of sense in the presence of input price bias: consider for example two firms that produce shirts and use the same technology; however, one firm uses silk as in input to produce silk shirts, while the other firm uses less expensive cotton to produce (less expensive) cotton shirts. Suppose that both firms produce the same number of shirts in a period. If we relate the number of shirts produced to (deflated) expenditures on materials, we would find that the firm that uses higher expenditures (silk) produces as much as the firm that uses lower expenditures (cotton). Hence, the coefficient on materials will be negative. Note that if one had not corrected for the output price bias in this case and had used deflated revenues instead of quantities as the left hand side variable, the problem would have been less transparent as silk shirts would be associated with higher revenues so that higher expenditures would have led to higher revenues. Of course, the fact that the problem is in this case less transparent does not mean that the problem does not exist.

Ultimately, the source of the problem is that products and inputs are differentiated. DLGKP address this issue by introducing a control function for the (unobserved) input prices. This is based on the premise that conditional on regional variation, input price differences reflect quality differences in the inputs across firms. DLGKP go on to argue that the assumptions underlying this control function are weaker than the ones required under alternative approaches that develop full-fledged models of product differentiation (with assumptions on particular demand functions and market structure). Nevertheless, their approach once again demonstrates the need for assumptions; data alone do not eliminate the need for economic structure.

Finally, one might argue that if we were able to observe in addition to output, input prices for every single input, then we would be able to estimate the structural equation (2) with only a limited set of assumptions. While in theory this is an appealing prospect, in practice, it is unlikely that we will ever be able to control for input price variation using data. To our knowledge, the only survey that contains information on prices of materials purchased is the one for Colombian manufacturing firms; but even there, the firm-specific price of capital is not observed, while utilizing the detailed input price information for materials is a challenge itself due to a variety of measurement issues. We conclude that while the use of additional data on prices can improve on certain aspects of estimation and identification, the need to introduce assumptions remains. The question is not whether or not one needs assumptions, but which set of assumptions is less restrictive.

2.2.2 Additional Assumptions

Let us return to equation (3) and consider the typical case where we have a standard dataset containing (deflated) sales and expenditures $(s^*_{it}, \boldsymbol{e}^*_{it})$ for a panel of firms in a given industry. We now ask: under which assumptions can we recover the structural parameters and the components of firm performance when estimating equation (3)?

One case that allows us to recover productivity (ω_{it}) has already been discussed: the case where there is neither output, nor input price variation across firms. The assumptions required to produce this lack of price variation are strong. Therefore, we ask whether it is possible to achieve the same result with a set of less restrictive assumptions; specifically we consider a setting where there is both output and input price variation and ask under which conditions these two types of variation will interact in such a way that the output price bias exactly offsets the input price bias, that is $p^*_{it} - \boldsymbol{z}^{*\prime}_{it}\boldsymbol{\alpha} = 0$. We show that the following assumptions are required for the price bias to be completely eliminated:

1. The industry is characterized by monopolistic competition (MC)
2. Firms produce a horizontally differentiated product and face the same constant elasticity of substitution (CES) demand system
3. Production is characterized by constant returns to scale (CRS)
4. Input price variation (across firms and time) is input neutral, such that $z^{*h}_{it} = \lambda_{it}$ $\forall h = \{1, 2, ..., H\}$, with H the total number of inputs.

The first two assumptions imply that firms will pass through costs completely to prices, and therefore any input price variation, both in the cross section and in the time series, will be completely reflected in output price variation, so that $p^*_{it} = z^*_{it}$.

Assumption 4 is required to make sure that in any cross section of firms the input price variation is restricted to a scalar. The CRS assumption is important to guarantee that $\sum_h \alpha_h = 1$, and consequently the input price variation exactly offsets the output price variation.[6]. To see the impact of this assumption, let $\sum_h \alpha_h < 1$. This would lead to a price error of $p^*_{it} - \sum_h \alpha_h \lambda_{it}$, which, even under complete pass-through, would leave $1 - \sum_h \ \alpha_h \lambda_{it}$ in the error term biasing the estimates of interest.

While the above assumptions are somewhat weaker than the assumption of no output or input price heterogeneity, they are still restrictive and inconsistent with a substantial body of work in International and Macroeconomics that has documented incomplete pass-through of cost shocks to prices.[7] Hence, the question arises whether there are alternative assumptions that would allow for a more satisfying treatment of unobserved prices. We briefly discuss the set of assumptions underlying two recent papers that explicitly address these biases, De Loecker (2011) and DLGKP.

Output price heterogeneity. De Loecker (2011) works with standard data in which output and input prices are unobserved. He assumes away input price variation across firms[8] and focuses on addressing the output price bias. To this end, he explicitly introduces a demand system in his empirical model; in the particular application he considers, he works with CES, but his approach works also for alternative demand systems as long as one can relate log quantity to log prices and additional demand shifters; such demand systems include the nested logit and the random coefficient demand model as in Berry, Levinsohn, and Pakes (1995), where log prices enter the indirect utility function.

The basic idea behind De Loecker's approach is to use the demand structure to express the (unobserved) price variation p^*_{it} as a function of observables, in his case, an industry quantity index q_{It} and a set of product dummies $\boldsymbol{D}_i$. De Loecker (2011) shows that this gives rise to the following version of (4):

$$s^*_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \beta_I q_{It} + \boldsymbol{D}'_i \boldsymbol{\delta} + \widetilde{\omega}_{it}, \tag{6}$$

where $\beta_I = \frac{1}{|\eta_I|}$ with η_I denoting the elasticity of demand of industry $I, \boldsymbol{\beta} = \frac{\eta_I+1}{\eta_I}\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}$ denoting the production function coefficients, and $\widetilde{\omega}_{it} = \frac{\eta_I+1}{\eta_I}\omega_{it}$. Estimation of

---

[6] Consider a two-input production function with labor and capital (l, k). Under Assumption 4, the structural error term can be written as $p^*_{it} - \alpha_l w^*_{it} - \alpha_k r^*_{it} = p^*_{it} - \alpha_l \lambda_{it} - \alpha_k \lambda_{it} = p^*_{it} - (\alpha_l + \alpha_k)\lambda_{it}$, where w and r denote the (log) prices of labor and capital respectively, asterisks are used to denote deviations from industry averages and λ_{it} is a scalar that captures the input-price neutral variation of input prices relative to the input price indexes ($\lambda_{it} = w^*_{it} = r^*_{it}$). CRS guarantees that $(\alpha_l + \alpha_k) = 1$ and with complete pass-through, $p^*_{it} = \lambda_{it}$

[7] See Goldberg and Knetter (1997) for an overview of this literature.

[8] This was plausible in the context of Belgian textile producers who are geographically concentrated.

the above equation delivers the demand elasticity η_I, along with the true production coefficients $\boldsymbol{\alpha}$, and allows one to separate the physical productivity ω_{it} from the output price variation $(\beta_I q_{It} + \boldsymbol{D}_i'\boldsymbol{\delta})$.

The results are consistent with one's priors; the production function coefficients obtained after correcting for output price bias are larger in magnitude (consistent with the presence of a downward bias when output price bias is not controlled for) and suggest increasing returns to scale. The approach rests crucially on assuming a demand system, but assumptions on the demand system and cost pass-through are implicit whenever one estimates (4) without controlling for the output price bias.

Output and input price heterogeneity. DLGKP work with a different dataset from India that contains information on product-specific prices (i.e., unit values). As noted earlier, the availability of output price data is not a silver bullet, at least not when one considers a large set of differentiated product industries. While the output price data allow DLGKP to eliminate the output price bias without resorting to any assumptions, they still need to address the input price bias. To do so, they assume that the only source of input price variation across firms (apart from regional differences) is quality differentiation; this assumption rules out imperfect competition in input markets. Further, they assume an output quality production function that displays complementarities in the qualities of inputs and output: higher quality output demands higher quality inputs, and high quality inputs are complements to each other. Under these assumptions, they show that conditional on regional variation, input prices will be a function of output quality, which can be proxied through a flexible polynomial in output prices, market shares, product dummies, and interactions thereof. This polynomial represents a control function for input prices and is consistent with a large set of alternative demand and market structures and the main models used in Industrial Organization and International Trade. Given that output prices are observed, the output price variation is eliminated from both the left and right hand sides of equation (4), which becomes:

$$q_{it} = \boldsymbol{e}_{it}^{*\prime}\boldsymbol{\alpha} + \omega_{it} + z_t(p_{it}, \boldsymbol{ms}_{it}, \boldsymbol{D}_i, G_i), \tag{7}$$

where p_{it} is the output price of the firm, $\boldsymbol{ms}_{it}$ is a vector market share variables (including unconditional and conditional market shares), $\boldsymbol{D}_i$ captures product dummies and G_i denotes firm location. The function $z_t(.)$ serves here as a control for the unobserved input price variation $\boldsymbol{z}_{it}^{*\prime}\boldsymbol{\alpha}$.

Conditional on productivity and input price variation captured by $z_t(.)$, we obtain the correct structural production function parameters by considering how variation in physical input use maps into variation in physical output.

In sum, while the bulk of empirical work has focused on estimating the sales generating function (1), recent papers have focused on estimating the structural equation (2) by correcting for the output and/or input price biases. In all cases, the correction involves not only additional data, but also explicit statement of the assumptions under which the correction is valid.

2.2.3 A Classification of Existing Work

In Table 1, we classify existing papers on the broad subject of globalization and firm performance based on how they deal with unobserved output and input prices. The columns and rows indicate which kind of price variation is controlled for.

Table 1: Output and input prices: a classification

		Output price (p_{it})	
		$\bar{p}_t^I$	p_{it}^*
Input price (z_{it})	$\bar{z}_t^I$	Case A	Case B
	z_{it}^*	Case C	Case D

Note: We remind the reader of our notation: $\overline{p}_t^I$ is the price deflator for industry I, $\bar{\boldsymbol{z}}_t^I$ a vector of industry-specific input price deflators, and $p_{it}^* = p_{it} - \overline{p}_t^I$ and $\boldsymbol{z}_{it}^* = \boldsymbol{z}_{it} - \bar{\boldsymbol{z}}_t^I$ denote deflated output price and input prices respectively.

Case A: Standard Framework: Most existing work use output and input price deflators common across firms that capture industry-wide movements in output and input prices. The only price variation occurs in the time series, and any variation away from these industry-wide deflators will introduce the output and input price biases discussed above.

Case B: De Loecker (2011): In the setting considered in De Loecker (2011), price variation across firms is controlled for by explicitly introducing a demand system. On the input side, it is assumed that firms face common input prices.

Case C: Observing Output Prices. In this case, input prices vary across firms, but output prices do not. As we noted above, this case corresponds to a specification estimated in DLGKP for expositional purposes, where physical quantities are regressed against deflated input expenditures, as in equation (5).

Case D: Pass-through literature and DLGKP: Perhaps the only literature that has allowed for both output and input price variation to characterize firms active in international markets is the pass-through literature[9]. This literature recognizes both that prices vary across firms and that firms respond incompletely and differentially to cost shocks; this incomplete response generates not only price dispersion across firms, but also for a particular firm, price dispersion across destinations. The cost shocks considered in the pass-through literature are exchange rate changes, but the insights of this literature are equally applicable to other cost shocks, including input price shocks or changes in tariff and other trade policies.

In contrast to the literature described above that takes a production function approach, the pass-through literature is based on a demand-based approach where assumptions on the demand side, market structure and firm behavior are combined in order to derive measures of firm performance. While this literature offers the advantage of a much richer treatment of heterogeneity and product differentiation, it comes at the cost of an extensive set of assumptions. These assumptions seem defendable in the context of case studies of particular industries, where knowledge of the institutional details can guide the choice of the appropriate structure, but are more controversial when applied to a large cross-section of industries. We therefore do not devote any space to discussing this literature here, but focus on the production function-based approach instead. From a production function perspective, to our knowledge, only DLGKP develop a framework that accounts for both output and input price variation and estimate equation (7). Interestingly, the insights obtained using their approach turn out to be consistent with the main insights of the pass-through literature; not only do the findings indicate substantial heterogeneity across firms, the authors find evidence consistent with incomplete pass-through of cost shocks to prices.

2.2.4 An Example

We conclude this subsection with an example. We have a sample of 318 single-product Indian textile producers over the period 1989-2003 for which we observe (deflated) sales and expenditures on labor, intermediate inputs and capital; in addition we observe

---

[9] Representative papers include Goldberg and Verboven (2001); Nakamura and Zerom (2010); Goldberg and Hellerstein (2013); Berman, Martin, and Mayer (2012); Amiti, Itshkhoki, and Konings (2013).

product-level prices.[10]

We consider a standard Cobb-Douglas production function and highlight the output and input price biases using four distinct specifications: (1) an OLS regression of sales against expenditures; (2) an OLS regression of quantity against expenditures; (3) an OLS regression of quantity against expenditures and a control function for input prices that includes only the output price (this control function is a special case of the control function used in DLGKP; it rests on a vertical differentiation model of consumer demand where output price is a sufficient statistic for quality and hence input price variation); and (4) a special case of DLGKP that assumes an AR(1) process for productivity and, as in specification (3), a control function for input prices that depends only on output price. Table 2 lists the estimated coefficients.

Table 2: Estimated coefficients: evaluating the price biases

	Revenue	Physical Output		
Coefficients	(1)	(2)	(3)	(4)
labor	0.162	-0.029	0.171	0.227
materials	0.812	0.576	0.834	0.634
capital	0.035	-0.514	0.010	0.140

Note: The data were constructed by DLGKP. We observe 318 producers of textiles (industry code PNIC 17). We omit the standard errors, but all coefficients are significant at the 1% level, as is common in the estimation of production functions. Specifications (1), (2) and (3) are OLS regressions of the relevant dependent variables (revenue in (1) and physical output in (2) and (3)) on deflated expenditures, plus a polynomial in output price in (3). Specification (4) estimates the coefficients using GMM, where we rely on lagged inputs as instruments exploiting the variation in adjustment costs in labor and capital, while allowing current productivity shocks to affect current material choices. See DLGKP for further discussion.

Specification (1) generates perhaps the most familiar numbers for the various coefficients. This OLS regression is a useful way to describe the underlying data and check that the estimates are within the range of existing studies; there is a long list of papers that estimate production functions using different datasets, and therefore the literature in this area has settled on what are reasonable looking estimates.

In the second specification, we consider the case discussed in equation (5): physical quantity is projected onto deflated expenditures, leaving input price variation uncontrolled for. We get results that are hard to interpret at first, such as negative labor and

---

[10] We only consider single-product producers to demonstrate the price bias in isolation. Using multi-product firms requires a treatment of the unobserved input allocations, which is not the focus of this article. See the discussion under section 2.1.1, option (a).

capital coefficients. But our framework actually predicts, or at least is consistent with, negative coefficients. Just as in our shirt production example, we find that firms that spend more on labor, produce less output, but generate more sales.

In specification (3) we stick to a simple OLS regression, but add a third order polynomial in the firm's output price. We recover coefficients similar to specification (1).[11] This OLS regression is a very useful *diagnostic check* for the problem at hand: by merely including output prices, while ignoring all other well known identification problems, we generate "'plausible" production function coefficients[12]. Of course, the proper benchmark should by no means be specification (1). However, the fact that we get positive output elasticities for all three inputs, and that the returns to scale now look sensible ($\sum_h \hat{\alpha}_h = 1.015$), is reassuring.[13]

Finally, in specification (4) we allow for both unobserved productivity shocks to affect input choices and for heterogeneous input prices. This specification can be thought of as a special case of DLGKP: we control for serially correlated productivity using an AR(1) process for productivity, and for unobserved input price variation by including a control function in output prices as in specification (3). The coefficients are similar to those in specification (3), but there are a few differences as expected. By controlling for unobserved productivity differences, we undo the negative correlation between capital and productivity and recover a substantially higher capital coefficient; further, we find a lower coefficient on materials due to the positive correlation between input use and productivity. Table 2 does not list the remaining structural parameters describing the relationship between input prices and output price – i.e., $z_t(p_{it})$. But it is interesting to note that we find that $\frac{\partial z_t(.)}{\partial p_{it}} > 0$ for all firms and time periods, confirming that output prices are positively correlated with input prices[14].

---

[11] We obtain almost identical coefficients if we include output price in linear fashion.

[12] We have found output price to have a first order effect when we correct for input price variation. The use of output price as the sole control can be justified on the basis of a vertical differentiation model, where price is a proxy for output quality, and hence also input quality and input prices. Additional variables, such as market shares and product dummies, make the control function more general and consistent with a larger set of demand models, but lead to very similar results. See DLGKP for more discussion on the estimation and identification of this control function $z_t(.)$.

[13] The OLS regressions do not control for the standard simultaneity bias, so even if we observed q_{it} and $\boldsymbol{x}_{it}$ we would not recover the production function coefficients $\boldsymbol{\alpha}$.

[14] This positive correlation is also documented in Kugler and Verhoogen (2012) who have the advantage of directly observing some input prices (e.g., wages and the prices of materials) as well as output prices in their Colombian data.

2.3 Interpreting the Performance Residual (π_{it})

Equation (2) will be useful for the remainder of this paper for two reasons. First, it allows us to highlight the potential channels through which globalization can affect firm performance: firms competing in international markets are likely to adjust their scale of production ($\boldsymbol{x}_{it}$ and hence $\boldsymbol{e}_{it}$), productive efficiency (ω_{it}), prices and associated markups (p_{it}), as well as product and input quality (reflected in both p_{it} and $\boldsymbol{z}_{it}$). Second, we can classify almost all studies on the subject based on which component(s) of profitability they have focused on in each instance.

Existing work has, in one form or another, studied how episodes of trade reform affect productivity, prices and quality across a wide range of countries, industries and time periods. Below we argue that a large part of the literature has only recovered the impact on profits π_{it}, without decomposing it into the separate effects on the underlying factors. But in most instances, we care about the exact mechanisms through which trade affects firm performance, as these have different distributional and potentially also aggregate welfare implications.

Our production function framework suggests ways for identifying the components of the structural error π_{it}. After estimating the structural production function and demand parameters, De Loecker (2011) recovers separate estimates of productivity and markups, and consequently estimates of (average firm-level) prices. This is accomplished by committing to a particular demand system and market structure – i.e., CES demand paired with monopolistic competition. However, even when we are not willing to make assumptions on the underlying demand system and market structure, we may still be able to recover, if not the levels, at least the changes in markups using the production function framework. If researchers observe prices, they can further obtain estimates of marginal costs.

De Loecker and Warzynski (2012) show how to recover markups from production data. The essential insight is that for any variable input free of adjustment costs the markup drives a wedge between the input's output elasticity and the input's revenue share. The latter is directly observed in the data; the former is not, but can be estimated. In the context of our Cobb-Douglas production function, the markup μ_{it} for firm i at time t is given by:

$$\mu_{it} = \alpha_v \frac{S_{it}}{E_{it}^{v}}, \tag{8}$$

The subscript v stands for *variable.* Depending on the application, variable inputs can include labor, electricity, or any other intermediate input.

To illustrate this approach, let us consider the same data on Indian textile producers.

Let materials be a variable input in production. Using (8) we compute markup as: $0.634 \times \frac{S_{it}}{E^m_{it}}$, and obtain an average markup of 1.30 with a standard deviation of 0.65. The markup distribution suggests considerable variation, with the 25th percentile firm breaking even with a markup of about 1, while the 75th percentile firm makes substantial profits (excluding fixed costs) with a markup of about 1.42.

The markup calculation relies on estimates of the production function, which in turn deliver estimates of productivity. The specifics will depend on the data at hand, and we refer the reader to De Loecker and Warzynski (2012) and DLGKP for detailed discussions of the issues that arise in the context of different data sets.

In sum, the production function framework has the potential to generate separate measures of productivity and markups without commitment to specific demand and market structure assumptions as is common in the demand-oriented IO literature on imperfect competition. Once the components of profitability are identified, one can examine how globalization affects each of these components. This is the subject of the next section.

3 Mechanisms

We use the framework introduced above to discuss the main mechanisms through which participation in international markets, and in particular trade reforms, affect performance. We denote trade reforms by T_{it} and allow them to affect firms differently over time. This specification allows for firms in an industry to produce different products for instance, so that they end up facing different rates of protection. Whether or not there is variation across firms is crucial for any identification strategy that tries to recover the causal effect of T_{it} on firm performance and its components. The mechanisms we discuss below can be broadly classified into two categories: mechanisms that induce changes within firms and hence affect firm-level components of profitability; and mechanisms that induce reallocation of economic activity across firms in an industry. In the latter case, firm-level profitability may be unaffected by trade, but trade-induced reallocation of resources from less to more profitable firms can still lead to better performance at the industry level.

3.1 Within-Firm Changes

Firms participate in international markets both as producers/sellers of goods and as buyers of intermediate inputs used in the production of these goods. Trade policies may affect both aspects of firm activity. Specifically, changes in the protection of final products (e.g., reductions in output tariffs) will affect the competition domestic producers face, and changes in the protection of intermediate inputs (e.g., input tariffs) on the other hand will affect the costs of production. The channels through which trade reforms affects firms will accordingly depend on the specific nature of the trade policy changes, and in particular on whether these affect output versus input markets. Therefore, in the course of our discussion we will often find it necessary to make a distinction between "output" and "input"oriented trade policies (for expositional purposes, we will base our discussion on *tariffs* given that these are easily measured, but in principle the arguments apply to any other trade policy).

We use equation (2) as the basis of our discussion to differentiate between mechanisms that affect: (a) the firm-level productivity ω_{it}; (b) the expenditures e_{it} and their components (input quantities x_{it} and prices z_{it}); (c) output prices p_{it} and markups; and 4) none of the above, but induce within-firm reallocation in multi-product firms.

3.1.1 Reduction of X-inefficiencies and Management Practices

The perhaps most advocated argument for opening up a country to foreign markets is that exposure to international competition increases the efficiency of the previously protected domestic producers. In terms of our framework, this channel would lead to an increase in the physical efficiency ω_{it}. But why would the efficiency of these producers increase? A popular argument is that intensified competition will reduce *X-inefficiencies* at the firm level. Although intuitive, this argument has little theoretical appeal in its simplest form; why were firms willing to leave money on the table prior to the trade reforms? A potential answer is that in practice, reduction of X-inefficiencies is costly, and therefore it takes an increase in competition for firms to find it profitable to undertake the actions necessary to become more efficient. For example, in the face of intensified competition, firms might find it necessary to replace old, inefficient managers by more competent ones, or adopt better management practices. While these considerations feature prominently in casual discussions of trade and productivity, we are aware of no theoretical work that formalizes these arguments.

From an empirical point of view, this mechanism suggests a reduced form way of introducing trade policy by making physical efficiency a function of trade policy:

$$\omega_{it} = \omega(T_{it}). \tag{9}$$

Note that the above mechanism suggests that the relevant policy is one that affects the output side of the firm, since the motivation for reducing X-inefficiencies arises from the exposure to intensified competition; in the case of tariffs for example, the relevant measure of trade policy would be output tariffs.[15]

3.1.2 Feedback Effects

The reason we expect firms to increase their productivity in response to a trade shock is that we expect them to undertake *actions* in order to become more efficient. Some of these actions may be unobservable to the researcher in which case they will be subsumed in the residual ω_{it} (this is for example the case if firms replace current managers by better ones or adopt better management practices and these actions are not reflected in changes in expenditures). But in many instances, improvements in productivity will be associated with actions that are observable, for example, investment in new technologies, R&D, entry in export markets, etc. In these cases, the law of motion of productivity should be modified so as to explicitly allow for these actions to affect productivity:

$$\omega_{it} = g(\omega_{it-1}, A_{it-1}) + \xi_{it}. \tag{10}$$

The term A_{it-1}, denoting any action undertaken by the firm in order to increase its productivity, is lagged given that it likely takes time for actions to take effect. Of course, the fact that these actions (investment; R&D; exporting, etc.) *are allowed* to affect productivity does not mean that they will in fact do so. The above law of motion is entirely consistent with a finding that the action undertaken by the firm did not have an effect on productivity ultimately; hence, it does not assume the result. Nevertheless, if one believes that a certain action is likely to affect productivity, it is imperative to include it in the law of motion.

Productivity enhancing actions will typically be correlated with the inputs in the production function, so that their omission from the law of motion will generate a omitted variable bias. This is the main point of De Loecker (2013) and it can be made clear using the example of a productivity enhancing investment. The investment will affect not only productivity, but also the capital stock. Suppose that we do not allow investment (the action) to affect productivity though the law of motion. Then the estimation of the production function will suffer from an omitted variable bias that will generate a biased capital coefficient; specifically, given that higher investment will likely be associated with higher capital, we would expect an upward bias in the capital coefficient estimate. Moreover, a second stage regression relating productivity to investment would tend to

---

[15] Below we discuss the importance of including both contemporaneous and lagged trade policy variables to accommodate the role of expectations and dynamics.

understate the role of investment for the same reason; given that investment was not included in the first stage that estimated the production function, the improvement in firm performance will be attributed to the higher capital and not to the productivity improvement.

In conclusion, recognizing that productivity evolves endogenously in response to firms' actions calls for a modified law of motion for productivity at a minimum. Ideally, one would like to supplement this law of motion with an explicit model of how the actions A are determined. This is done for example in recent papers by Bustos (2011) and Aw, Roberts, and Xu (2011). The relevant actions are "exporting" and "adoption of new technology" in the first paper, and "exporting" and "R&D" in the second, and in both cases the authors use structural models to show how these actions are determined and how they respond to trade liberalization.

3.1.3 Input Side

Input Expenditures In equation (2) we explicitly distinguished between the two components of expenditures (e_{it}): physical input use (x_{it}) and input prices ($\boldsymbol{z}_{it}$). Both components will typically vary across firms and will be affected by globalization; for example, a reduction of input tariffs will have a direct effect on the prices of imported inputs, and hence $\boldsymbol{z}_{it}$. There are three distinct reasons for firm-specific input prices ($\boldsymbol{z}_{it}$), even for firms active in narrowly defined industries: (a) pure geographical variation in input prices (for example, local labor markets and constrained labor mobility imply regional differences in wages); (b) variation in input quality leading to differences in input prices; and (c) firm specific input prices due to monopsony power in input markets. The literature on production function estimation has typically ignored or assumed away heterogeneity in input prices due to quality differences across producers or to imperfect competition in input markets.

To highlight the forces generating variation in expenditures across firms let us write: $e_{it}^h = x_{it}^h + z_{it}^h(\nu_{it}^h, G_{it})$ for a given input h. The term ν_{it}^h refers to the "quality" of input h.[16] We collect all other firm-specific factors determining input prices, including the firm's geographic location, in G_{it}, which will capture among other things the (re)location of economic activity (plant closings, offshoring, etc.) induced by a firm's exposure to global

---

[16] We use the term "quality" to capture differences in observed and unobserved attributes of a given input – for example, skill differences across workers. Whether one can measure input quality will depend on the data at hand. Standard firm-level data usually provide us with the total use of intermediate inputs in dollars, and sometimes even with physical units, but will typically not record input characteristics or direct measures of quality.

markets or other shocks.

Four distinct and largely disconnected literatures have focused on how globalization affects the various components of input expenditures. First, the *trade and productivity literature* has set out to measure how the transformation of physical inputs (x) to output (q) changes with increased foreign competition; indirectly, this literature also deals with the scale effects that operate through inputs (x). Second, the *trade and quality literature* has focused on whether producers upgrade (or downgrade) the quality of their products and inputs in response to increased exposure to international trade[17]. Third, the *trade and labor literature* has focused on how globalization affects workers of different skills. Finally, the *literature on multinationals and offshoring* has investigated how globalization affects firms' locational choices (part of G_{it}). In all these cases, exposure to global markets affects e_{it} directly.

Imports of New Intermediate Inputs A different mechanism through which globalization can affect firm performance is highlighted in Halpern, Koren, and Szeidl (2011). So far, we have abstracted from the fact that the materials used in the production, m, are a composite of many different domestic and imported intermediate inputs. Acknowledging the aggregation underlying m suggests an additional channel for performance improvements: if the reduction in trade costs leads to the import of new intermediate inputs, then we would expect an increase in production beyond the one predicted by the increase in expenditures. This increase will be more pronounced if the new inputs are of higher quality compared to the ones previously used, but the argument does not rest on quality improvements: as long as the production technology exhibits a "taste for variety", a larger number of imported inputs will imply higher output.

The simplest way to make this point is to consider (as in Halpern, Koren, and Szeidl (2011) or Goldberg, Khandelwal, Pavcnik, and Topalova (2010b)) a standard Cobb-Douglas production function in capital, labor, a set of intermediate inputs M, and productivity Ω. To keep notation manageable, we abstract from quality differences between imported and domestic products, and suppress the firm and time subscripts. Each intermediate input M_j is assembled from a combination of a domestic and imported

---

[17] Schott (2004) and Khandelwal (2010) are classic references on the relationship between trade and product quality, while Verhoogen (2008) is a classic reference on the effects of globalization on both product and input quality. The main message of these papers is that producers often need to change/upgrade the quality of their products in order to enter foreign markets. These changes in product quality induce changes in input quality. Through the link between output and input quality, shocks in output markets affect factor markets.

variety:

$$\begin{aligned} Q &= \Omega L^{\alpha_L} K^{\alpha_K} \Pi_{j=1}^{J} M_j^{\alpha_j} && (11) \\ M_j &= [M_{jF}^{\frac{\theta-1}{\theta}} + M_{jD}^{\frac{\theta-1}{\theta}}]^{\frac{\theta}{\theta-1}}, && (12) \end{aligned}$$

where M_{jF} and M_{jD} denote the quantities of the foreign and domestic inputs respectively, and θ is the elasticity of substitution. Let us abstract from the output and input price biases. One can show that in this setting, equation (2) takes the form:

$$s_{it}^{*} = q_{it} = \boldsymbol{x}_{it}'\boldsymbol{\alpha} + F_i(n_{it}) + \omega_{it} = \boldsymbol{e}_{it}^{*\prime}\boldsymbol{\alpha} + F_i(n_{it}) + \omega_{it}. \quad (13)$$

The term $F_i(n_{it})$ is a function increasing in the number of imported inputs n_{it}. Hence, if a trade liberalization increases the number of imported intermediates, we will expect to see a rise in *measured* productivity, $F_i(n_{it}) + \omega_{it}$. Why would a reduction in trade costs lead to the import of new intermediates? When deciding whether or not to import intermediates, firms balance the marginal cost savings associated with the new inputs against the fixed costs of importing. A reduction in the tariffs on inputs increases the cost savings associated with importing, hence leading to a larger number of imported intermediates. Note this mechanism does not suggest any improvements in the *physical* productivity ω_{it}. The issue here is analogous to the well known gains from the introduction of new products; because the standard deflators used to obtain deflated expenditures $\boldsymbol{e}_{it}^{*}$ do not account for new imported inputs, these new inputs will ultimately show as an increase in measured productivity. This mechanism likely underlies the large within-firm productivity gains found in studies that examine the effects of input tariff liberalization, such as Amiti and Konings (2007) for Indonesia and Khandelwal and Topalova (2011) for India. In fact, Goldberg, Khandelwal, Pavcnik, and Topalova (2009) and Goldberg, Khandelwal, Pavcnik, and Topalova (2010b) explicitly show that the input tariff liberalization in India led to a large increase in the number of imported inputs.

3.1.4 Price and Markup Changes

It is natural to expect that firms will adjust prices and markups when faced with a trade shock. Many trade models assume CES preferences with monopolistic competition. Under these assumptions, markups are constant; hence, trade shocks do not affect markups. Recent theoretical work has moved away from these restrictive assumptions and considered alternative demand systems [18](while maintaining for the most part the

---

[18] See for example, Melitz and Ottaviano (2008), Mayer, Melitz, and Ottaviano (2011), Feenstra and Weinstein (2010), Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012).

assumption of monopolistic competition) in order to investigate how markups and prices respond to trade liberalization.

If we observe prices and have a plausibly exogenous source of variation for the trade policy T_{it}, it is conceptually straightforward to evaluate the price effects of trade reforms. We would start with the reduced form:

$$p_{it} = p(T_{it}) + \epsilon_{it}, \tag{14}$$

where ϵ_{it} is a standard i.i.d. error term. A tougher task is to identify the specific channels leading to price changes, i.e., the cost and markup responses to changes in international competition. To highlight these channels, let us write price as the sum of marginal cost mc and markup μ:

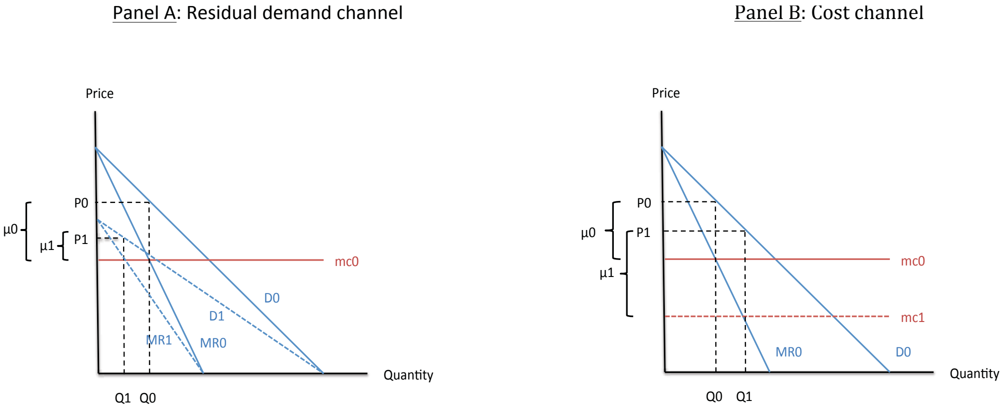
$$p_{it} = mc_{it}(q_{it}, \boldsymbol{z}_{it}, \omega_{it}) + \mu_{it}(\mathcal{D}, \mathcal{M}). \tag{15}$$

The marginal cost mc is a function of quantity produced q, the input prices the firm faces $\boldsymbol{z}$, and firm productivity ω. The markup μ will be a function of the demand structure $\mathcal{D}$ and the market structure and firm behavior which we summarize in $\mathcal{M}$. Trade shocks are expected to affect each of these components, with the specific effects depending on the particular nature of the trade shock. As noted earlier, CES preferences and monopolistic competition imply constant markups (in log terms); unless trade liberalization affects the costs mc, it will have no effect on prices. In more general setups, markups will be variable and respond to the trade reform.

To illustrate the mechanisms at work, let us consider a simple example that features linear demand and monopolistic competition. When assessing the effects of trade reforms, it is important to distinguish conceptually between reforms that affect the input markets (e.g., input tariff liberalization) and reforms that affect the output markets (e.g., output tariff liberalization). Both types of reforms are likely to affect both marginal costs and markups, albeit through different channels.

Consider a unilateral *output tariff* liberalization first. The decrease in tariffs exposes domestic producers to increased import competition. In the context of equation (15), this translates to a change in market structure and consequently, the residual demand curve facing the firm. In Panel A of Figure 1, we plot the initial equilibrium in this market, which occurs at the point at which the original marginal revenue curve intersects marginal cost. For ease of exposition, we assume that marginal cost is constant and that is not affected by trade liberalization. Trade liberalization implies intensified competition, so the residual demand curve will shift inward and become flatter. The new equilibrium occurs at a point at which the price and markup are both lower. This case corresponds to the standard intuition that trade liberalization, by intensifying competition, leads to lower prices and lower markups. Figure 1 allows one to trace the particular forces that shift price from its pre-reform level p_0 to its post-reform level p_1.

Figure 1: Price, quantity and markup response to trade liberalization


Panel A: Residual demand channel

Panel B: Cost channel

Note: Solid lines represent the initial demand and cost conditions, while dashed lines indicate the new demand and cost conditions – i.e., post trade liberalization.

Next, consider a unilateral *input tariff* liberalization. Input tariff declines will have direct effects on the marginal cost; they will reduce the input prices z_{it} and may further lead to improvements in productivity ω_{it} through the import of new intermediates discussed above. In the context of Panel B of Figure 1, this implies a downward shift in the marginal cost curve. The decline in input tariffs does not affect competition, hence we would not expect any effects on markups arising from changes in the residual demand facing the firm. However, as long as the underlying structure does not imply constant markups (as is the case with the CES), markups *will change* as a result of the incomplete pass-through of the marginal cost change to price. This is shown explicitly in Panel B of Figure 1. The marginal revenue curve is not affected by the trade reform, while the marginal cost curve shifts downward. The post-reform equilibrium is associated with a *higher* markup, though the price is lower than before. The reason is not that the environment has become less competitive. The higher markup arises as a result of the incomplete response of the price to the marginal cost change.

This (to Trade and Industrial Organization economists) apparently counterintuitive effect is completely intuitive to International Macro economists who have studied the incomplete response of prices to exchange rates. Just like *input tariff* shocks, exchange rate changes represent cost shocks to firms. It is well documented that prices respond incompletely to exchange rate shocks, a phenomenon known as *incomplete exchange rate pass-through.* Tariff and exchange rates have a symmetric effect on firms' profits; applying the insights of the exchange rate pass-through literature to input tariffs immediately yields the result that with variable markups, input tariff reductions will lead to markup increases. To our knowledge, this is an insight that the bulk of the trade literature has missed; only DLGKP address this issue explicitly and show that incomplete pass-through of input prices led to markup increases in the case of the Indian trade liberalization. Prices did decline as a result of the trade reform, but the price reductions were only a small proportion of the cost declines; the bulk of the cost reductions benefited firms in the form of higher markups.

In reality, most trade reforms, especially those that have been implemented in developing countries, combine input with output tariff liberalization (put differently, the real world trade liberalizations are characterized by a combination of Panels A and B of Figure 1). In such instances, it is particularly important to assess not only the reform's impact on prices, but also its effects on the price determinants. The reason is that the effects on input prices as well as markups can have important distributional implications.

3.1.5 Within-Firm Reallocation

A different mechanism that can lead to within firm performance improvements is highlighted in Bernard, Redding, and Schott (2010) and is specific to multi-product firms: firms can improve revenue productivity that is, $\omega_{it} + p^*_{ijt}$ in the context of equation (2))by reallocating within-firm resources from the production of less profitable to the production of more profitable products, where j denotes products. This mechanism is similar in spirit to the one we will discuss in section 3.2 on the role of reallocation in increasing aggregate industry performance. The difference is that here this mechanism generates performance improvements at the firm-level. It is important to note that this mechanism does not hinge on any improvements on the physical firm productivity, ω_{it}, which remains unaffected. The distinction between physical and revenue productivity is therefore important here: it is only revenue productivity that increases and this increase is brought about entirely through reshuffling of resources across products with different profitability.

3.2 Reallocation: Aggregate Effects

So far we have focused on the potential effects of trade liberalization on individual producers. However, at the end of the day what we care about is how an industry, country or group of countries, is affected by trade. Reallocation of economic resources from less towards more profitable producers is one way in which industry- (or country) performance can increase even in the absence of any effects on individual firms. There is by now a large theoretical and empirical literature highlighting the aggregate productivity gains arising from such reallocation[19].

Collard-Wexler and De Loecker (2013) discuss some of the recent findings and point out that although we know by now that this reallocation process plays a substantial role in the data, it has been hard to identify specific forces that induce reallocation. Empirical work in international trade is perhaps the one big exception. The main advantage of studying large and arguably exogenous (at least from the perspective of an individual producer or industry) trade reforms is that they present us with exogenous shocks to the residual demand curves (in the case of output tariffs) and/or costs (in the case of input tariffs) facing domestic producers. We can then trace out how the allocation of economic activity, usually measured by the market share in a particular market/industry, changes with the change in trade policy. The reshuffling of market shares towards the more productive/profitable firms has the potential to raise aggregate performance beyond the potential individual firms' improvements discussed in the previous section. The extent to which the reallocation process is important depends, of course, on how dispersed profitability was initially, prior to the reforms.

Recently, Hsieh and Klenow (2009) put forward a simple theoretical framework to highlight this mechanism by focusing on wedges in marginal revenue products and pointing out distortions that can give rise to such wedges. Changes in both output and input tariffs fall nicely into this framework: output tariffs present standard distortions in output markets as they constrain competition, while input tariffs generate distortions in capital and intermediate input markets. The reduction of these distortions through trade reforms should lead to a more efficient allocation of resources across firms.

This reallocation mechanism is in principle simple to measure in the data: calculate the covariance of productivity and market share for each time period and see how it reacts to the trade liberalization episode. However, the above discussion should have made clear that the interpretation of the results will depend greatly on whether we rely on

---

[19] For example, this reallocation mechanism is central to the Melitz (2003) model and many other follow-up papers that feature firm heterogeneity.

actual productivity (ω_{it}), or a performance measure (such as π_{it}) that contains both cost, demand and market structure components. For example, if we relied on π_{it} and found that the market share of more profitable firms increased post trade liberalization, this increase might not be desirable from an aggregate welfare point of view if it represented a shift towards firms with more market power rather than higher efficiency.

3.3 Static Versus Dynamic Effects

So far we have not made a distinction between trade shocks that affect producers instantaneously and shocks that affect producers with a lag. Many theoretical models are static in nature or focus on steady state predictions which blur the distinction. However, for empirical models that wish to separately identify the impact of trade liberalization on the various components of performance the difference between static and dynamic effects can be important.

Let us focus on the productivity channel ($\omega(T_{it})$). It is reasonable to ask whether producers can immediately adjust their productive efficiency when faced with a change in protection. This question boils down to whether the trade reforms were expected or not, and hence whether producers had *time to adjust* prior to the actual change.

The standard working assumption is to assume that firm-level productivity moves over time according to a first-order Markov process: $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$. This specification is based on the observation that firm-level productivity (independently of how exactly it is measured) is highly persistent over time. However, the question remains whether we can think of trade liberalization as a shock that immediately affects productive efficiency. Take the example of a trade-induced reorganization of production. It arguably takes time for a firm to change the organization of production. Therefore, we would expect productivity to not react immediately or even possibly drop temporarily during the reorganization.

Therefore, it is important to let the function $\omega(.)$ be sufficiently flexible in how trade policy shocks enter, by including both contemporaneous trade policy changes as well as lags. For example, in the context of analyzing the impact of trade reforms on firm performance in Indian manufacturing, DLGKP consider a law of motion that, in addition to lagged productivity, also includes lagged output and input tariffs and the firm's lagged export status.

4 Evidence

4.1 Profitability and Feedback Effects

If there is one robust finding that the literature has delivered to date, this is that industry profitability increases with exposure to foreign competition. This relationship is documented both in the time series, in studies that exploit trade liberalization episodes to identify the effects of trade openness on firm performance, and in the cross-section, in comparisons of the performances of exporters and non-exporters. Most papers that exploit trade liberalizations focus on changes in output tariffs (Pavcnik (2002) is a classic reference), but recently, starting with the work of Amiti and Konings (2007), the focus has shifted towards input tariff liberalizations. Indeed, several studies on developing countries find the effects of input tariffs (representing direct cost shocks to firms) to be larger than those of output tariffs (which operate through changes in the competition facing firms). Further, the literature finds evidence of both within-firm performance improvements and re-allocation, with the relative importance of each channel depending on the particular setting. These findings are well documented in the literature, and we refer the reader to recent surveys by Melitz and Trefler (2012) and Melitz and Redding (Forthcoming) for a more extensive discussion.

At the cost of repeating ourselves, we emphasize once again that these findings of improved performance, though robust across countries and time, refer to profitability only, and are therefore not particularly illuminating regarding the mechanisms at work. With this caveat in mind, it is worth pointing out that studies that have allowed for endogenous productivity evolution along the lines suggested in section 3.1.2 find significant evidence of feedback effects. These feedback effects have two implications. First, in studies of trade liberalizations, they suggest that performance improvements are heterogeneous across firms, as firms with different characteristics (e.g., initial profitability levels; R&D expenditures; capital intensity, etc.) optimally choose different actions in response to trade shocks, which in turn affect their profitability. Bustos (2011), Aw, Roberts, and Xu (2011) and Lileeva and Trefler (2010) all find evidence of such heterogeneity. Second, in studies that compare exporters to non-exporters, feedback effects can lead to different conclusions regarding the relative importance of "selection" versus "learning by exporting" effects. De Loecker (2013) for example, demonstrates that in the case of Slovenia, learning by exporting seems to play an important role once one explicitly controls for the fact that entering export markets is associated with higher investment. These findings once again point to the importance of employing empirical specifications that are consistent with the underlying mechanisms one has in mind when analyzing the data.

4.2 Mechanisms Underlying Profitability

It is only recently that research has focused on unpacking the mechanisms that generate the aforementioned performance improvements. While it is still too early to draw general conclusions based on the findings of the few studies that explicitly distinguish between physical productivity and price effects, the evidence to date suggests that demand side and price effects are important and may be the primary factors generating the documented profitability increases.

To demonstrate the significance of these effects, consider the De Loecker (2011) study of the Belgian textile market. As explained earlier, De Loecker does not have price data, but introduces a demand system in order to separate productivity from price effects. Hence, one can distinguish in his framework between aggregate *profitability* (Π_t) and aggregate *productivity* (Ω_t). Table 3 performs the standard decomposition of aggregate (i.e., industry-level) productivity changes. Columns 2-4 refer to revenue productivity, while columns 5-7 refer to physical productivity. Columns 2 and 5 show the aggregate changes at the industry level; columns 3 and 6 show the within-firm component, while columns 4 and 7 capture the reallocation[20].

There are two interesting features of this table. First, the aggregate *physical* productivity change between 1994-2002 (a period that spans removal of major trade restrictions, i.e., quotas, in textiles) displayed in column 5 appears significantly lower than the change in *revenue* productivity in column 2. Hence, it seems that the usual "productivity" improvement shown in column 2 reflects primarily changes in prices. Second, column 4 suggests significant reallocation effects, from less profitable towards more profitable firms. The literature has often interpreted these effects as reallocation from less efficient towards more efficient firms. However, column 7 suggests that this interpretation is misguided: the reallocation effects computed using physical productivity are substantially smaller – almost non-existent. Hence, it appears that the reallocation documented using revenue productivity as a measure of firm performance was reallocation towards higher price and higher markup firms, not towards firms with higher efficiency.

---

[20] Levinsohn and Petrin (2012) have criticized this market-share-based measure of reallocation as being uninformative in welfare calculations. While we are sympathetic to their criticism, we simply want to compare our physical productivity based results to those one would obtain using the standard approach in the literature, and to this end, we adopt the standard practices of this literature throughout the calculations.

Table 3: Reallocation and Trade Liberalization

Year	Profitability			Productivity		
	Π_t	$\overline{\pi}_t$	cov	Ω_t	$\overline{\omega}_t$	cov
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1994	1.00	0.90	0.10	1.00	0.95	0.05
1995	0.98	0.87	0.11	0.94	0.89	0.05
1996	1.02	0.93	0.10	0.99	0.95	0.03
1997	1.09	0.97	0.13	1.02	0.97	0.05
1998	1.06	0.97	0.10	0.97	0.94	0.03
1999	1.06	0.99	0.07	0.99	0.99	0.00
2000	1.02	0.99	0.03	0.95	0.96	-0.01
2001	1.03	0.96	0.06	0.95	0.95	0.00
2002	1.05	0.96	0.09	0.97	0.95	0.02

Note: This table is based on the estimates obtained from the analysis in De Loecker (2011). Both Π and Ω are normalized to one in 1994. We follow the decomposition in Olley and Pakes (1996) whereby aggregate productivity $\Omega_t = \sum_i ms_{it}\omega_{it} = \bar{\omega}_t + \sum_i (\omega_{it} - \bar{\omega}_t)(ms_{it} - \bar{m}s_t) = \bar{\omega}_t + cov_t(ms_{it}, \omega_{it})$, with ms_{it} the market share. We apply the same decompostion to the profitability index Π_t.

Because De Loecker (2011) works with a CES demand system - albeit one that allows for different demand elasticities across products within the textile industry - markups and prices at the *product/firm level* are not affected by the trade liberalization; the price and markup effects in his framework are the result of reallocation (across firms, or across products within a firm) towards firms/products with higher markups. However, we would expect trade liberalizations to also affect prices and markups at the firm/product level through both the residual demand (i.e., intensified competition) and the cost (i.e., lower prices of imported intermediates) channels. To this end, one needs to consider more general demand systems that allow for variable markups and incomplete pass-through of cost changes to prices. Such a framework is considered in DLGKP. As noted earlier, DLGKP do not commit to a particular demand or market structure, but adopt an empirical specification that nests the main models used in Trade and Industrial Organization, including those that generate variable markups. They use the Indian trade liberalization to separately identify the effects of the residual demand (reductions in output tariffs) versus cost (reductions in input tariffs) channels. The authors find that output tariff declines have the expected pro-competitive effects; they lead to lower prices and lower markups. But the striking result is that in the end, the net effect of the Indian liberalization was to *increase* markups. This increase, which at first seems at odds with the standard intuition that trade has pro-competitive effects and hence reduces markups, does not come about because of firm collusion, or any other attenuation of competition. It is the result of the incomplete pass-through of input tariff declines to prices.

These results that may seem surprising in the context of trade, are consistent with the findings of the exchange rate literature as well as the substantial Macro literature on price rigidities. Overall, studies that have attempted to explicitly address the price and markup effects associated with trade openness suggest that the demand side of the market is as important as the cost side; trade liberalizations do not lead only to (physical) productivity improvements, but also to changes in prices and markups that need to be modelled explicitly.[21]

5 Conclusions and Future Work

We conclude with some final thoughts on the state of the literature and future work. There are several strands within Trade that deal broadly with firm performance and globalization, each employing different assumptions and approaches. Unfortunately, there has been minimal cross-fertilization of ideas across these literatures up to now.

The empirical productivity literature has focused on estimating the effects of trade on firm performance, without distinguishing between physical efficiency and price/markup effects. Mainstream theoretical models in Trade often employ assumptions that imply constant markups (for example, CES preferences with monopolistic competition), and hence abstract from the potential of trade to affect markups. Models that do allow for markup effects have typically focused on the pro-competitive effects of trade, paying little attention to the markup effects that arise as a result of incomplete pass-through of trade-induced cost reductions to prices. But this type of incomplete pass-through has been precisely the focus of the large literature on exchange rate pass-through, which tries to understand how prices and markups respond to exchange rate shocks. Its insights have never been applied to trade liberalizations, despite the fact that (input) tariff reductions and exchange rates have similar effects on firm profits. Finally, there have been case studies of the effects of trade liberalization on firm performance in particular industries (e.g., automobiles[22]) that rest on estimation of structural industry models in the IO tradition, but the results of these studies do not readily generalize to the economy at

---

[21] A similar point emphasizing the role of the demand side has been made in several papers by Foster, Haltiwanger, and Syverson (2008, 2012), but in the context of the domestic market.

[22] See Goldberg (1995) and Berry, Levinsohn, and Pakes (1999).

large.

We believe that the time is ripe for the methods and insights of these separate literatures to be combined in theoretical and empirical work in this area, and hope that this article represents a small step in this direction. Indeed, there are encouraging signs. Recent empirical work on the effects of trade liberalizations has tried to distinguish between efficiency and markup effects yielding novel insights. Current work on the effects of trade on markups makes an explicit distinction between the competition and the pass-through channels. But much more research, on different countries and different time periods, is needed before we will be able to draw general conclusions. Further, the last few years have seen the emergence of an exciting new literature on assessing the aggregate gains from trade under alternative modeling assumptions. This literature that is primarily, though not exclusively, theoretical has emphasized the role that functional form assumptions, especially ones with implications for markup adjustment, play in evaluating the gains from trade. Careful empirical work that is motivated by and consistent with theoretical models, but does not depend heavily on restrictive functional form assumptions, can play an important role in informing the assumptions of the models used to assess the welfare gains from trade.

References

Ackerberg, D., L. C. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, chap. 63. Elsevier, 1 edn.

Amiti, M., O. Itshkhoki, and J. Konings (2013): "Importers, Exporters and Exchange Rate Disconnect," mimeo.

Amiti, M., and J. Konings (2007): "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 97, 1611–1638.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2012): "The Elusive Pro-Competitive Effects of Trade," mimeo.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011): "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 101, 1312–44.

Berman, N., P. Martin, and T. Mayer (2012): "How do Different Exporters React to Exchange Rate Changes?," *The Quarterly Journal of Economics*, 127, 437–492.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-Product Firms and Product Switching," *American Economic Review*, 100, 70–97.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–90.

——— (1999): “Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy,” *American Economic Review*, 89, 400–430.

Bustos, P. (2011): “Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms,” *American Economic Review*, 101, 304–40.

Collard-Wexler, A., and J. De Loecker (2013): “Reallocation and Technology: Evidence from the U.S. Steel Industry,” *NBER Working Papers*, (18739).

De Loecker, J. (2011): “Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity,” *Econometrica*, 79, 1407–1451.

——— (2013): “Detecting Learning by Exporting,” *American Economic Journal: Microeconomics*, 5, 1–21.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2012): “Prices, Markups and Trade Reform,” *NBER Working Papers*, (17925).

De Loecker, J., and F. Warzynski (2012): “Markups and Firm-Level Export Status,” *American Economic Review*, 102, 2437–71.

Feenstra, R. C., and D. E. Weinstein (2010): “Globalization, Markups, and the U.S. Price Level,” *NBER Working Papers*, (15749).

Foster, L., J. Haltiwanger, and C. Syverson (2008): “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?,” *American Economic Review*, 98, 394–425.

Foster, L., J. C. Haltiwanger, and C. Syverson (2012): “The Slow Growth of New Plants: Learning about Demand?,” *NBER Working Papers*, (17853).

Goldberg, P. K. (1995): “Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry,” *Econometrica*, 63, 891–951.

Goldberg, P. K., and R. Hellerstein (2013): “A Structural Approach to Identifying the Sources of Local Currency Price Stability,” *Review of Economic Studies*, 80, 175–210.

Goldberg, P. K., A. Khandelwal, N. Pavcnik, and P. Topalova (2009): “Trade Liberalization and New Imported Inputs,” *American Economic Review*, 99, 494–500.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010b): “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India,” *Quarterly Journal of Economics*, 125, 1727–1767.

Goldberg, P. K., and M. M. Knetter (1997): “Goods Prices and Exchange Rates: What Have We Learned?,” *Journal of Economic Literature*, 35, 1243–1272.

Goldberg, P. K., and F. Verboven (2001): “The Evolution of Price Dispersion in the European Car Market,” *Review of Economic Studies*, 68, 811–48.

Halpern, L., M. Koren, and A. Szeidl (2011): “Imported Inputs and Productivity,” *CeFiG Working Papers*, (8).

Hsieh, C.-T., and P. J. Klenow (2009): “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124, 1403–1448.

Katayama, H., S. Lu, and J. R. Tybout (2009): “Firm-level productivity studies: Illusions and a solution,” *International Journal of Industrial Organization*, 27, 403–413.

Khandelwal, A. (2010): “The Long and Short (of) Quality Ladders,” *Review of Economic Studies*, 77, 1450–1476.

Khandelwal, A., and P. Topalova (2011): “Trade Liberalization and Firm Productivity: The Case of India,” *Review of Economics and Statistics*, 93, 995–1009.

Klette, T. J., and Z. Griliches (1996): “The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous,” *Journal of Applied Econometrics*, 11, 343–61.

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

Levinsohn, J., and A. Petrin (2012): “Measuring Aggregate Productivity Growth Using Plant-Level Data,” *RAND Journal of Economics*, 43, 705–725.

Lileeva, A., and D. Trefler (2010): “Improved Access to Foreign Markets Raises Plant-Level Productivity... for Some Plants,” *Quarterly Journal of Economics*, 125, 1051–1099.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2011): “Market Size, Competition, and the Product Mix of Exporters,” *NBER Working Papers*, (16959).

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

Melitz, M. J., and G. I. Ottaviano (2008): “Market Size, Trade, and Productivity,” *The Review of Economic Studies*, 75, 295–316.

Melitz, M. J., and S. J. Redding (Forthcoming): “Heterogenous Firms and Trade,” in *Handbook of International Economics*. Elsevier, 4th edn.

Melitz, M. J., and D. Trefler (2012): “Gains from Trade When Firms Matter,” *Journal of Economic Perspectives*, 26, 91–118.

Nakamura, E., and D. Zerom (2010): “Accounting for Incomplete Pass-Through,” *Review of Economic Studies*, 77, 1192–1230.

Olley, S., and A. Pakes (1996): “The Dynamics of Productivity in the Telecommunications Industry,” *Econometrica*, 64, 1263–98.

Pavcnik, N. (2002): “Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants,” *Review of Economic Studies*, 69, 245–76.

Schmalensee, R. (1989): “Inter-industry studies of structure and performance,” in *Handbook of Industrial Organization*, ed. by R. Schmalensee, and R. Willig, vol. 2, chap. 16. Elsevier, 1 edn.

Schott, P. K. (2004): “Across-product Versus Within-product Specialization in International Trade,” *Quarterly Journal of Economics*, 119, 646–677.

Syverson, C. (2004): “Market Structure and Productivity: A Concrete Example,” *Journal of Political Economy*, 112, 1181–1222.

Verhoogen, E. A. (2008): “Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector,” *Quarterly Journal of Economics*, 123, 489–530.