Markups and Misallocation with Trade and Heterogeneous Firms

Ariel Weinberger*

October 21, 2014

aweinberger@ucdavis.edu
University of California, Davis

Abstract

With general preferences, a monopolistic competition equilibrium can be inefficient in the way inputs are allocated towards production. This paper formalizes the welfare impact of reallocation of quantities across firms (within an industry) by comparing real income growth with the hypothetical case of no misallocation in quantities – as is the case when consumer preferences are CES and producers choose a constant markup. In contrast, my monopolistic competition model is consistent with variable markups such that reallocations initiated by aggregate shocks that affect firms' demand or cost curves can impact allocative efficiency. Open economy shocks, even by raising production overall, can have opposing consequences for the misallocation distortion depending on the adjustment of the market power distribution. Using firm-level data from Chile for 1995-2007, a period with large terms of trade gains, I find that average productivity gains are not necessarily associated with gains in allocative efficiency because firms pass-through productivity gains into markups. From industry-year variation, there is also evidence that industries that import a larger share of their inputs have higher markup dispersions and become more misallocated as a result of appreciations compared to "open" sectors that compete globally in the sale of their output.

*This work has benefited from the tireless support and guidance of Robert Feenstra, Katheryn Russ and Ina Simonovska. I also appreciate comments by Deborah Swenson, Colin Carter, Paul Bergin, Lucas Herrenbrueck and other participants at UC Davis seminars and the Midwest International Trade Conference at IUPUI.

1 Introduction

An important conclusion from the last decade of trade theory is that tougher competition raises the average productivity of the set of producing firms. In fact, the big breakthrough of the Melitz (2003) heterogeneous firm model is to add reallocation of market share to the most efficient firms, i.e. selection, as a further aggregate productivity gain. However, the nature of the reallocation is simplified by using Constant Elasticity of Substitution (CES) preferences (Dixit and Stiglitz, 1977) that result in market outcomes that are identical to what would be chosen by a social planner given the aggregate domestic environment. The reallocation when using CES is therefore always welfare improving, and the market equilibrium is shown to be allocatively efficient[1]: the allocation of variety and the quantity produced of each variety maximizes welfare given an endowment of inputs.

The goal of this study is to estimate the market power distortionary effects on the allocation of production across firms over time by examining the joint movement of prices and quantities across Chilean firms in response to changes in competitive pressures and costs. Literature on growth and productivity has pinpointed that within-industry production misallocation, or allocative inefficiency, is an important reason for cross-country productivity differences (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008). Basu and Fernald (2002) expand on Solow productivity gains – akin to shifting out a country's production possibility frontier – to include welfare-improving movements along this frontier that can be measured using real income. In this paper I nest the growth rate of allocative efficiency measured in terms of real income gains in a framework with heterogeneous firms that wield market power based on their productivity. Consistent with the aggregate productivity literature, a distortion inflicts a wedge between total revenues and total output. Aggregate shocks initiate a reallocation of production that contributes to allocative efficiency changes through the effects on the distribution of market power. Even without price and quantity data at the firm-level, I produce a structural estimate of growth in allocative efficiency at the industry level and motivate it using first and second order comparative statics for firms.

I relate allocative efficiency to the CES feature of constant markups, and investigate an economy such that the market equilibrium with heterogeneous firms is not necessarily efficient. I use the separable utility function framework underlying CES preferences, but generalize it to allow for a utility function that is not homothetic. In a variable elasticity (VES) framework with monopolistic competition, variable markups imply a production

[1] Dhingra and Morrow (2012) and Feenstra and Kee (2008) have shown this in a setting with firm heterogeneity.

allocation that does not equalize relative marginal utilities with relative production costs, as market power allows highly productive firms to only partially pass-through cost advantages. Higher markups by these firms maps onto lower aggregate income relative to the allocatively efficient benchmark and thus creates an aggregate distortion. Dhingra and Morrow (2012) also incorporate this distortion into welfare and characterize qualitative properties of misallocation. I build on that paper because I produce a quantitative measure of changes in misallocation using aggregate data. Further, I incorporate shocks to the input and output markets that motivate time series variation in allocative efficiency and can be tested with data.

Switching to a VES framework thus allows for a richer welfare analysis by capturing the difference between growth rates of aggregate real income and physical production. Since these measures coincide in a constant markup environment, I link deviations in the growth of real income and physical production to changes in the market power distortion and show that this distortion is present in the welfare decomposition of a representative consumer.[2] The first goal of this work is to measure the annual change in the distortion by measuring real income growth and comparing it to the case when misallocation must be zero. Then, I allow for exogenous changes in the competitive environment – which affects firm demand elasticities – and exogenous shocks to productivity to motivate the underlying factors behind changes in misallocation. These correspond to changes in aggregate variables but the firm-level responses are heterogeneous. For example, if a shift in demand elasticities results in initially high-markup firms lowering their prices relatively more, then inputs must be reallocated to these firms as they move down their demand curve. In the empirical section I measure how these shocks correspond to allocative efficiency at the industry level. The most striking result is that Chile's large terms of trade appreciation, by lowering input costs[3], resulted in greater misallocation in industries reliant on imported inputs.

My paper contributes to the literature by quantifying a change in misallocation that is due exclusively to heterogeneity in market power but differentiates itself from Edmond et al. (2012) and Holmes et al. (2012) which both assume nested CES preferences. In contrast to these two studies, my measure of misallocation in Section 4.2.2 cannot be summarized entirely by the markup distribution, but on the total covariance of prices and quantities. I also do not assume a specific functional form for utility (beyond additive separability), and maintain the monopolistic competition environment as in Krugman

[2]I do not necessarily measure *all* of the change in allocative efficiency as a component of welfare because it does not take into account optimal entry.

[3]Lower inputs costs corresponds to productivity increases in the model.

(1979) and Melitz (2003). These theoretical works, along with empirical studies of trade liberalization based on the monopolistic competition framework, usually focus on average productivity and reallocation of market shares to more productive firms. Although that source of gain is still present in this model, variable markups imply that total production also depends on how market power allows firms to under-/over-produce relative to the optimum.

In order to connect the theory to the empirical work, I take the structural specification of changes in allocative efficiency and relate it to firm-level reallocation. These changes are based on assuming that an equilibrium exists in each year and the allocation of production changes annually. Firm-level markups themselves are governed by the production allocation, and firms make production decisions taking aggregate variables as given. Shocks to the open economy will affect misallocation through their distributional effects on firm-level markups within a sector. As mentioned above, I allow for two reduced-form effects that can result from globalization. One possibility is that there is tougher competition on domestic firms, or an equivalent shock that shifts firm-level demand elasticities. A separate possibility is to lower the marginal cost of production for domestic firms through cheaper intermediate inputs. The former forces firms to lower their markups and results in a smaller dispersion of markups because the bigger firms reduce their markups relatively more. The latter allows firms to charge higher markups due to incomplete pass-through into prices, resulting in larger markup heterogeneity because pass-through is smaller for high-markup firms.

After defining the quantity misallocation, I bring in evidence for the importance of taking account where quantity is being allocated. I use firm-level Chilean data from 1995-2007, a period that includes both trade liberalization aspects in the emergence of the open economy, and large terms of trade gains due to the increase in demand for Chile's raw commodity exports. This allows me to test the predictions of the model in response to reductions in marginal costs and tougher competition. Firms can be categorized as importers of intermediate inputs, exporters of final goods, as well as both or none. The working hypothesis is that importers become more productive and face less competition with a real exchange rate appreciation. Exporters and other domestic producers might face stricter competition in this case, as well as when output tariffs are reduced. I focus on the terms of trade gain because of the large magnitude of the appreciation.

Although the distortion I measure is at the industry level, it is the product of aggregating firm-level responses. At the firm level, I show that the terms of trade appreciation raises markups for importers especially. The response is largest for firms with initially high markups, which implies the overall market power distortion increases. In fact, the terms of trade shock is followed by a large rise in the dispersion of markups and there is evidence that importing sectors drive this result. However, I do not rely on firm-level TFP and markup estimates to measure industry misallocation. Using the structural specification of misallocation, I measure changes in allocative efficiency using aggregated data and show that the reallocation witnessed at the firm level is consistent with allocative efficiency findings at the industry level.

The industry allocative efficiency results are obtained using variation across time and two-digit ISIC sectors. The most robust result is that industries dominated by firms that rely on imported inputs become more misallocated in response to a positive terms of trade shock. For example, comparing the extreme case of an industry that imports 100% of inputs and does not export their products with an industry whose share of imports in inputs equals the share of exports to sales[4], a 1% larger growth in the terms of trade leads to a 6 percentage points larger decline in allocative efficiency in the former industry. By being able to produce at a lower marginal cost, firms, and more-so the initially high-markup firms, increase production by less than would be implied in a constant markup model. For sectors that compete in final goods (export-oriented and import-competing), the terms of trade gain leads to a modest increase in allocative efficiency. Those sectors are also affected the most by trade liberalization. A sector that does not import, but exports all of its output, has allocative efficiency growth 2 percentage points larger than an industry whose share of imports in inputs equals the share of exports to sales, in response to a 1% decrease in the growth of output tariffs.

My findings are consistent with studies on competition, variable markups, and pass-through but provide aggregate implications that have not been discussed in this context. Related studies find that tougher competition forces firms to lower prices and raises average productivity, and that pass-through of costs to prices is below one (DeLoecker et al. (2012)). Relatedly, Amiti et al. (2012) find that the most productive firms import the most and also have the lowest pass-through. This is consistent with the terms of trade shock in Chile raising total production but also increasing the degree of misallocation because productive firms raise their markups the most. On the efficiency side, theoretical models have explored variety and scale trade-offs (Chamberlin, 1933; Vives, 2001), but not necessarily misallocation of quantity among existing producers, which requires firm heterogeneity.

The rest of this paper is organized as follows. Section 3 delineates the theoretical framework and Section 4 differentiates between growth in real income in the CES and VES models. Section 5 provides predictions for aggregate movements in misallocation based

[4]In Section 7 I delineate the reasons for this definition

on two distinct ways that open economy shocks drive reallocation. Section 6 describes the data and Section 7 presents the empirical results applied to Chilean firm-level and aggregate data from 1995-2007. Section 8 concludes and discusses the composition of importers and exporters at the country level in relation to misallocation. First, however, I survey the related literature in the following section.

2 Related Literature

Theoretical trade models have explored variable markups to generalize welfare gains from trade, though the earlier literature concentrates on the decrease in the average markup in search of a "pro-competitive" effect as in Krugman (1979). When there is free entry, competition decreases average markups and increases aggregate productivity as firms increase their scale and move down their average cost curves. This is possible with symmetric firms and for this reason should be separated from the quantity-misallocation distortion present in this paper that is the result of the interaction of incomplete pass-through and firm heterogeneity. Feenstra and Weinstein (2010) use a Translog expenditure function to measure the pro-competitive plus variety effects from increased global competition. Arkolakis et al. (2012) look at a broader class of variable markup models and point out that with non-homothetic demand there is an extra welfare term due to the interaction of incomplete pass-through and firm heterogeneity.

Misallocation has recently been introduced into models with CES preferences and heterogeneous firms in oligopolistic competition based on Bernard et al. (2003) and Atkeson and Burstein (2008). de Blas and Russ (2010), Holmes et al. (2012) and Edmond et al. (2012) all focus on welfare gains of tougher competition when the distribution of markups plays a key role. de Blas and Russ (2010) and Holmes et al. (2012) generalize the assumptions on the productivity distribution in the model of Bernard et al. (2003) to find implications on firm-level markups and overall welfare. Holmes et al. (2012) derive a welfare decomposition that includes allocative efficiency but holds only for homothetic tastes. Peters (2011) applies a model with head to head Bertrand competition and one top supplier for each differentiated good to relate markups to the productivity difference between the two most productive firms. Edmond et al. (2012) use nested CES utility to make firm-level markups endogenous to competition and quantify welfare gains that result from lowering misallocation, which is due exclusively to the level and dispersion of markups. Finally, misallocation can be attributed to supply-side wedges by adding an additive cost as in Khandelwal et al. (2013). In my model with VES preferences and monopolistic competition, misallocation is due to non-homotheticity on the demand side. It allows for more flexible demand, a distortion mapping to the aggregate productivity literature, and an intuitive application to incomplete pass-through. The latter feature not only receives empirical support, but opens up the possibility of cost shocks that lower allocative efficiency in addition to tougher competition lowering the market power of high-markup firms.

The distortions present in this model will remind the reader of Hsieh and Klenow (2009) (HK), which models firms' production choices given they face output and capital distortions. The distortions mean that firms optimally choose non-equal marginal products even though they face identical factor prices. This generates different revenue productivities (TFPR) even though in an undistorted world these would be the same for all firms. Misallocation in HK is due to firms with higher production efficiency (TFPQ) being too small as they hire too few inputs due to distortions. Reallocating inputs to firms with higher TFPR would lower their price and hence lower their TFPR. In the HK framework firms are heterogeneous in productivity but markups are constant due to CES preferences. A similar measure of allocative inefficiency is derived in Basu and Fernald (2002): some firms are producing too little (much) and choose revenue productivities that are too high (low). My paper establishes a new way to observe deviations from allocative efficiency. The non-equalization of firms' marginal rates of transformation occurs endogenously through non-homothetic tastes. I depart from the previous literature by taking a demand side approach to misallocation, with a clear mechanism for a more efficient resource allocation. The distortion that leads to an inefficient allocation of resources is the market power heterogeneity that arises due to a non-constant demand elasticity. HK assume constant markups, so it is not possible for the firm's market power to affect misallocation. Basu and Fernald (2002) and Petrin and Levinsohn (2012) use a very similar measure of reallocation, but there is no demand structure. In section 4.3 I discuss how aggregate productivity and welfare in these models relates to a VES economy with monopolistic competition.

This paper fits into the Aggregate Productivity Growth (APG) literature that decomposes APG into growth in average firm productivity and reallocation. I argue that reallocation increases welfare if inputs are reallocated to where they have the highest social valuation in terms of marginal utility. This argument is made in Basu and Fernald (2002), Petrin and Levinsohn (2012) and Basu et al. (2010). In these papers, the markup is the gap between marginal revenue productivity of an input and the cost share of that input in the total input cost. If aggregate productivity is linked to the aggregate value added in the economy, then there is an aggregate productivity gain (APG) when inputs are reallocated towards firms with markups above the mean markup. This methodology fits in my trade model of monopolistic competition. When production is reallocated to high markup firms, these firms jointly lower their price and increase their production. It will be shown that this raises aggregate real income. Therefore, this APG can be directly linked to welfare gains. I separate from the existing literature by examining the contribution of the open economy to industry misallocation. The focus is on applying this theory to open economy shocks that affect firms distinctively and how this aggregates to a distortion at the industry level. Additionally, whereas the above papers tend to regard prices and markups as fixed, I follow the joint movements of prices and quantities.

Empirically, my results are closest to DeLoecker et al. (2012). Studying the trade liberalization of India, they find large productivity gains for manufacturing firms but also an incomplete pass-through of those gains into consumer prices. Although they do not incorporate these findings into a model with aggregate productivity implications, it is evident a CES model would over-state the gains from trade. This is what I find, that firm-level productivity gains are not necessarily passed through to aggregate productivity.

Pavcnik (2002) and Bartelsman et al. (2013) also attempt to measure productivity growth through reallocation in developing countries, though they focus on different sufficient statistics. Their method uses a decomposition of weighted average plant-level productivity from Olley and Pakes (1996). The decomposition of aggregate productivity is the sum of unweighted average productivity and the covariance of market share and firm productivity. This latter term is an indicator of overall misallocation since aggregate productivity increases when more productive firms make up a higher total share of the market (as in Melitz (2003)).[5] This methodology is consistent with aggregate Solow residuals, so it misses the part of reallocation that is due to misallocation (captures only the selection effects). Amiti and Konings (2007), Kasahara and Lapham (2013) and Goldberg et al. (2010) show how a significant part of the productivity gains from liberalization are a result of cheaper and more abundant intermediate inputs. I show that this was true also for Chilean firms.

3 Model: Variable Elasticity and Allocative Efficiency

In this section, I describe the Variable Elasticity of Substitution (VES) framework. I follow Dhingra and Morrow (2012) and Zhelobodko et al. (2012) to set up the environment and highlight how the markups distribution is the driving factor behind allocative inefficiency. This framework is based on the demand elasticity being non-constant with respect to quantity produced. More productive firms (producing a differentiated good with

[5] Trefler (2004) shows that Canadian firms became more productive due to the NAFTA reforms, though the study does not examine reallocation.

a lower marginal cost) charge lower prices as usual and their optimal markup depends on their price elasticity of demand. Preferences are general as I do not choose a functional form for utility, but assume that it is additively separable across products. Although this allows for any range of demand elasticities, I restrict myself to preferences where the **inverse** demand elasticity is increasing with quantity.[6] For this reason more productive firms will have more market power and charge higher markups than their less productive counterparts. Whenever possible I use the notation of Dhingra and Morrow (2012).

I will refer to each "economy" as individual sectors at the 2 digit ISIC level. In each sector there is a measure of differentiated varieties produced by a single-product firms that are imperfect substitutes. The VES form is applied for preferences within a sector, where consumers demand differentiated goods: $U(M_e, q) = M_e \int u(q(c))dG$, where M_e represents the mass of entering varieties, c is the marginal cost and index of a differentiated good, $q(c)$ represents the individual consumption of a representative consumer of a good indexed by marginal cost c drawn from a distribution G[7], and consumers integrate over a unit bundle of goods. It is possible to allow for sector heterogeneity in parameter values within this framework, but this general setup is relevant for all manufacturing sectors. Therefore, from now on I focus on a "representative sector" since all the qualitative results are identical across sectors. The empirical section will use variation at the sector-year level to examine the testable predictions.

Given the budget constraint, we can solve for $p(q(c)) = \frac{u'(q(c))}{\delta}$ where the shadow price of income is $\delta = M_e \int_0^{c_d} u'(q(c))q(c)dG$. From here we can establish the following relationship between the inverse demand elasticities and markups:

$$\mu(q) = |\frac{qu''(q)}{u'(q)}| = |dlnp(q)/dlnq| = (p(c) - c)/p(c) \tag{1}$$

I refer to this as the markup rate, though I will mostly use the price-cost ratio for markups. The first step to working with this framework is to establish a socially optimal allocation under the general VES preferences and compare this to the market equilibrium. Although the social planner maximizes overall utility given the resource constraint, individual firms will solve a profit maximization problem. Using the optimal price condition above, the market equilibrium is such that each firm maximizes revenues that depend only on the representative utility function. From Dhingra and Morrow (2012) (DM) we

[6]This is the case most often chosen in the literature, which Mrazova and Neary (2013b) call "Marshall's Second Law of Demand". It is also the pro-competitive case in Krugman (1979). I am partial to Paul Krugman's words that to get reasonable results, "I make this assumption without apology".

[7]$c \in (0, c_d]$, where c_d is the highest possible cost with positive demand.

can set up the two maximization problems, one solves for the social optimum and the other solves for the market equilibrium:

$$\text{Social: } \max_{q,M_e,c_d} M_e^{opt} \int_0^{c_d^{opt}} u(q(c))dG(c) \,\text{s.t.}\, L \geq M_e^{opt} \left[\int_0^{c_d^{opt}} (cq(c)L + f)\, dG(c) + f_e \right] \tag{2}$$

$$\text{Market: } \max_{q,M_e,c_d} M_e^{mkt} \int_0^{c_d^{mkt}} u'(q(c))q(c)dG(c) \,\text{s.t.}\, L \geq M_e^{mkt} \left[\int_0^{c_d^{mkt}} (cq(c)L + f)\, dG(c) + f_e \right] \tag{3}$$

where the full allocation consists of equilibrium values for $\{M_e, q(c), c_d\}$. c_d represents the cost cutoff at which firms find it profitable to produce, M_e is the measure of entry and the production allocation is the vector of $q(c)$ for all $c \in (0, c_d]$. Let λ be the shadow price in the social allocation problem and δ the market allocation multiplier. Both are aggregates and I assume firms take them as given.

There is a sector-specific zero profit condition such that expected profits net of sunk costs are zero: $\int \pi(c)dG = f_e$. In the language of Dixit and Stiglitz (1977), the social optimum is a "constrained optimum" since firms need to be compensated for the chance of losing the entry cost and not producing. Even when the entry, productivity cutoff and quantity allocation are such that social welfare is maximized, this will not be the perfectly competitive limit as prices will be greater than average costs. Below I show that the markup at the constrained optimum is the same for all firms.

Under the social problem, the social planner sets a quantity such that $u'(q^{opt}(c)) = \lambda c$ for all firms indexed by c, their marginal cost. λ, like the multiplier in the market economy, is a sector aggregate that represents the shadow value of resources. At these quantity allocations, if firms set prices as a constant of the marginal utility, then $p(q^{opt}(c)) = \frac{\lambda}{\delta}c$. Then, under the socially optimal allocation all firms set prices as a constant over marginal cost. The markup for all firms is the difference between λ and δ, or the difference between the maximum welfare per capita and the maximum real aggregate revenue per capita given representative consumers' utility function.

Then, under VES demand, the following is true:

Proposition 1. *A social planner guarantees a constrained optimum by setting all quantities such that the marginal utility for every good, and hence the price-cost ratio, is constant across firms. This means that the dispersion of prices and marginal utilities is always equal to the dispersion in costs within a sector. The benchmark for allocative efficiency is a degenerate dispersion of markups. A positive markup dispersion across firms is evidence of misallocated resources.*

In the market equilibrium, firms charge variable markups. Following the first order conditions of Equation 3, for all firms: $u'(q(c))+u''(q(c))q = \delta c$, or $u'(q(c))[1-\mu(q(c))] = \delta c$, where μ is the markup rate ($(p-c)/c$). Given that $p = u'(q(c))/\delta$:

$$u'(q(c)) = \frac{\delta c}{1-\mu(q(c))} \tag{4}$$

$$p(q(c))\delta = \frac{\delta c}{1-\mu(q(c))} \Leftrightarrow p(q(c)) = \frac{1}{1-\mu(q(c))}c \tag{5}$$

Under VES preferences, the price is not a constant over marginal cost because $\mu(q(c))$ is a function of firm-varying productivity (or marginal cost). In other words, market power is heterogeneous across firms within a sector.[8] As in Basu and Fernald (2002), when market power is heterogeneous firms do not equate marginal rates of transformation. As expressed in DM and related to Feenstra and Kee (2008), the social and market allocations are aligned only when utility is defined by CES preferences, where prices and marginal utilities are a function of a constant over marginal cost and the market allocation mirrors a constrained optimum. Outside of CES, the markup is heterogeneous across firms.

Having established the social optimum, we can compare it to each firm's production decision from the first order conditions. Combining the market and social first order conditions the following must hold:

$$\frac{\lambda}{\delta} = \frac{1}{1-\mu(q(c)^{mkt})}\frac{u'(q(c)^{opt}}{u'(q(c)^{mkt})} \tag{6}$$

By definition, market and social equilibrium quantities are the same when:

$$u'(q(c)^{mkt}) = u'(q(c)^{opt}) \tag{7}$$

$$\implies \frac{\lambda}{\delta} = \frac{1}{1-\mu(q(c)^{mkt})} \tag{8}$$

The LHS is the (constant) markup all firms charge at the constrained optimum, and the RHS is the market equilibrium. Equation 8 is a condition that needs to hold for every firm in order for the sector to be at a socially optimal allocation.

I define a distortion due to misallocation by the movement of the market allocation away from the socially optimal allocation, and show below how the distribution of markups will affect the distortion.

[8]In this decreasing demand elasticity case, more productive firms have more market power and higher markups.

4 Quantifying Allocative Inefficiency

Dhingra and Morrow show that the VES model leads to distortions not present in the standard CES model because the market equilibrium is socially optimal only when preferences are CES. Although they characterize qualitative properties of misallocation, this paper will focus on quantifying the welfare losses between the social and market allocations. Here I calculate a part of the difference in total welfare by using the definition of real income per capita. I compare the CES allocation with the allocation that results from VES utility, with the restriction that the demand functions are such that the demand elasticity decreases with sales.[9] I decompose real income per capita into physical production and a covariance term that includes firm markups and input expenditures. Part of the welfare loss in the market equilibrium is therefore represented by the dispersion in markups. Intuitively, a dispersion in markups distorts the quantity allocation of the producing firms and lowers aggregate real income. This setup will be useful in quantifying real income effects to domestic firms from global shocks, as explored in Section 5.

My analysis will be less general than DM in that I eliminate one of the three allocation choices. The available firm-level data is not equipped to measure consumer variety gains, and using a general framework that does not assume a functional form for $u(q(c))$ means there is no obvious way to correct welfare for variety gains.[10] Therefore M_e will be taken as given in all equilibria so that an allocation is summarized solely by the vector $q(c)$ and cutoff cost, c_d. In the analysis below, growth rates are annual, with the equilibrium in each year defined by a new quantity and cutoff allocation set.

I start with decomposing welfare with homothetic utility, which is proportional to revenue. Generalizing preferences to the VES form, I first follow the proof from Basu and Fernald (2002) and take the total derivative of the utility function to show utility increases when inputs are reallocated to firms with higher markups (as in Basu and Fernald (2002)). Then, in order to get a measure of misallocation in the data, I compare real income in the socially optimal and market allocation.[11] By following the joint movement of prices and quantities, real income is shown to decrease in the market equilibrium when cost advantages are not passed through completely to prices which distorts quantity produced relative to the CES benchmark. Notice that this framework is consistent with the results of Edmond et al. (2012) and Arkolakis et al. (2012), who both find that it is the *joint dis-*

[9]This standard assumption in the trade literature assures that more productive firms charge higher markups.

[10]In the homothetic case for example this is done using the Ideal Price Index that measures cost of living.

[11]With non-homothetic preferences revenue is no longer proportional to revenue.

tribution of markups and production that matters.[12] Finally, although real income is not equivalent to utility, I decompose utility to show what are the (separable) components that I miss in my measure.

4.1 Utility with CES

To start I explore the social allocation where aggregate real revenue is proportional to welfare because $u(q) \propto qu'(q)$, which means we can relate utility to aggregate real revenue. From the definition of preferences and the consumer budget constraint, $M_e \int_0^{c_d} u'(q(c))q(c)dG(c)$ is proportional to aggregate real revenue (more commonly written as $M_e \int_0^{c_d} p(q(c))q(c)dG(c)$) which can be expanded as:

$$U = M_e L \int u(q)dG \propto M_e L \int_0^{c_d} u'(q(c))q(c)dG(c) \tag{9}$$

$$\propto \lambda M_e L \int_0^{c_d} p(q(c))q(c)dG(c) \tag{10}$$

$$\propto \lambda M_e L \int_0^{c_d} \frac{1}{1-\mu(q(c))} cq(c)dG(c) \tag{11}$$

$$\propto \lambda \Big(\int_0^{c_d} \frac{1}{1-\mu(q(c))} dG(c) \Big) \big(L - M_e G(c_d) f - M_e f_e \big) \tag{12}$$

where the last line uses the budget constraint and that $Cov(\frac{1}{1-\mu(q(c))}, cq(c)) = 0$ when the sub-utility function is homothetic. Welfare is proportional to the average markup times the total labor used for production. As shown in Dhingra and Morrow (2012), the market allocation maximizes utility when $\frac{p(c)}{c} = \frac{1}{1-\mu(q(c))}$ is constant.

When sub-utility is not homothetic, the market allocation price/cost ratio is a function of productivity: $p(c)/c = \frac{1}{1-\mu(q^{mkt}(c))}$, where $\mu(q^{mkt}(c))$ is the inverse demand elasticity the firm faces. Utility and aggregate revenue diverge, $Cov(\frac{1}{1-\mu(q(c))}, cq(c)) \neq 0$, and the market allocation no longer maximizes utility. The rest of this section defines how to measure changes in this distortion caused by misallocation. Since revenue is proportional to welfare only when preferences are homothetic, I quantify a part of welfare loss of the market versus social allocation by measuring the loss of aggregate *real income* that results from variable markups. In Subsection 4.4 I fit this loss in real income into the total change in utility from a change in allocation of production. But first, I show that changes in utility arises with the reallocation of inputs to higher price firms as in Basu and Fernald (2002).

[12] Alternatively, the intuition is that the whole distribution of markups matters, not the unweighted mean.

4.2 VES and Market Power Distortions

4.2.1 Relationship to Basu and Fernald (2002)

To relate changes in the markup distribution to welfare, I first derive an expression straight from the utility function following Basu and Fernald (2002). I take the total derivative of the utility function with the assumption that each year the economy is at a new equilibrium of production.

$$U(q(c), M_e) = M_e L \int_c u(q(c)) dG(c) \tag{13}$$

$$dU = M_e L \int_c u'(q(c)) dq(c) dG(c) = \lambda M_e L \int_c p(q(c)) dq(c) dG(c) \tag{14}$$

The second line is the total derivative of utility function $U(q(c), \bar{M}_e)$ (fixed entry) with a shock to the production of surviving firms, and substitutes prices for marginal utility $(u'(q(c)) = \lambda p(q(c)))$.

Then rewrite Equation 14 in terms of means and variances:

$$dU = \lambda M_e L \big[\int_c p(q(c)) dG(c) \int_c dq(c) dG(c) + \lambda M_e L \text{Cov}(p(q(c)), dq(c)) \big] \tag{15}$$

To see how the markups contribute to the distortion I show the relationship between changes in utility and markups under general preferences, again assuming the economy is hit with a shock that affects the allocation of production. I work with the absolute markup, consistent with the measure from the data, which corresponds with the previous notation as: $\frac{p(q(c))}{c} = \frac{1}{1-\mu(q(c))}$.

$$dU = M_e L \int_c u'(q(c)) dq(c) dG(c) = \lambda M_e L \int_c \frac{p(q(c))}{c} c dq(c) dG(c) \tag{16}$$

$$= \lambda M_e L \int_c \frac{1}{1-\mu(q(c))} c dq(c) dG(c) \tag{17}$$

$$dU = \lambda M_e L \left[\int_c \frac{1}{1-\mu(q(c))} dG(c) \int_c c dq(c) dG(c) + \text{Cov}\left(\frac{1}{1-\mu(q(c))}, c dq(c) \right) \right] \tag{18}$$

In this case the conditions for the covariance term are obvious since it must be constant if $u(q(c))$ is CES.

Lemma 1. *The last term in Equation 18 is constant when the markup distribution is degenerate.*

The sufficient condition is that $u(q(c))$ is CES.

From Equation 18, the change in utility is proportional to not only growth in production, but also to changes in the covariance of the markup and input expenditure. In the CES case, the covariance term is zero because cost heterogeneity is optimally passed through to prices so that markups are constant. Only the average markup and average productivity will be affected. In the VES case with the assumption of decreasing demand elasticity, the covariance term is negative and the distribution of markups affects real income. Shocks to the production allocation can raise utility depending on whether more production goes towards the high markup firms. In the process, their markup decreases as they move down their demand curve and increase production.

Equation 18 provides no term that can be taken to the data and interpreted as allocative efficiency. Next I show how available aggregate data can serve this purpose.

4.2.2 Quantifying the Quantity Misallocation using Real Income

From Equation 10, let aggregate revenue, $R = M_e L \int_0^{c_d} p(q(c))q(c)dG(c)$. I will work with the conditional distribution of $g(c)$ on $(0, c_d]$, defined as follows:

$$h_d(c)dc = \begin{cases} \frac{g(c)}{G(c_d)}dc & \text{if } c \leq c_d, \\ 0 & \text{if } c > c_d \end{cases} \tag{19}$$

It will be useful to define the average price level $P \equiv \int_0^{c_d} p(q(c))h_d(c)dc$ and aggregate physical production sold $Q \equiv LM \int_0^{c_d} q(c)h_d(c)dc$.[13] $q(c)$ stands for the consumption of an individual variety by a representative consumer. I will decompose aggregate revenue in terms of mean and variances and let $\text{Cov}(p, q) = \int_0^{c_d}(p(q(c)) - \tilde{p})(q(c) - \tilde{q})h_d(c)dc$, with $\tilde{p}$ and $\tilde{q}$ the average price and average quantity respectively.[14] I use that to examine real income by dividing both by the average price:

$$R = ML \left[\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc\right] + ML\left[\text{Cov}(p, q)\right] \tag{20}$$

$$\frac{R}{P} = ML \int_0^{c_d} q(c)h_d(c)dc + ML\left[\text{Cov}(p, q)/\int_0^{c_d} p(q(c))h_d(c)dc\right] \tag{21}$$

$\frac{R}{P}$, real income, will be denoted by W. The last term, as will be shown shortly, is a residual that will capture allocative efficiency since it represents the deviation of real income from

[13] $M = M_e G(c_d)$ is the measure of firms that produce.

[14] $\tilde{p} = \int_0^{c_d} p(q(c))h_d(c)dc$; $\tilde{q} = \int_0^{c_d} q(c)h_d(c)dc$.

physical production. Equation 21 can be further expanded substituting for W and Q, and then taking logs to get growth rates:

$$\frac{W}{Q} = 1 + \frac{\text{Cov}(p,q)}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc} \tag{22}$$

$$ln\left(\frac{W}{Q}\right) = ln\left(1 + \frac{\text{Cov}(p,q)}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc}\right) \tag{23}$$

$$\Delta ln\left(\frac{W}{Q}\right) \approx \Delta\left(\frac{\text{Cov}(p,q)}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc}\right) \tag{24}$$

The last line uses the approximation that $ln(1+x) \approx x$. The way I will identify changes in allocative efficiency is by establishing that Equation 24 is zero in the case of allocative efficiency. Therefore, I can use the observable left hand side to measure whether the market equilibrium is getting closer or farther from efficiency.

To get an intuitive interpretation, I define the terms in Equation 24 with respect to the allocation variables. There are two changes to the allocation over time: inputs/production are reallocated as firms change their production quantity (vector of q(c)); and the cutoff cost (c_d) changes given the new competitive environment. The former effect will not affect the covariance term when utility is CES: the distribution of prices is equal to the distribution of marginal costs in that case, so reallocating quantity will have no affect on the distribution of prices. However if the cutoff cost changes, this impacts the distribution depending on the assumed distributional properties. Therefore I take the simplifying case where $G(c)$ is Pareto, a case where truncation does not change the shape of the distribution. This does not hold, for example, in the log-normal[15], though it is the case that a truncated log-normal distribution is very similar to the Pareto.[16] In Appendix A I take the case of CES preferences and Pareto distribution of costs and show that the terms in Equation 24 are zero, which establishes the following Lemma:

Lemma 2. *The RHS of Equation 24 is equal to zero when i) the distribution of prices stays unchanged and ii) a change in the cutoff cost does not affect the shape of the price and/or quantity distribution. The conditions that $u(q(c))$ is CES and $G(c)$ is Pareto fit the two requirements as shown in Appendix A. Therefore, $\Delta ln(\frac{W}{Q}) = 0$ in the allocative efficiency case of $u(q(c))$ CES.*

[15]From Head (2011): the truncated log-normal has an expected value: $E[x|x > x_0] = e^{(\mu+0.5\sigma^2)}\frac{1-\Phi(z_0-\sigma)}{1-\Phi(z_0)}$ with $z_0 = \frac{ln(x_0)-\mu}{\sigma}$. That last ratio in the expected value is the effects of the truncation.

[16]Eeckhout (2004) shows that log-normal is the right approximation to a distribution where the tails look Pareto but the shape parameter is sensitive to the choice of a minimum truncation point. In my data, the distribution of productivity does seem sensitive to the truncation, and I will experiment with both distributions to measure dispersion.

Given that the market power distortion exists only in the non-efficient market equilibrium, I label the change in the covariance term as ΔAE, with the interpretation that it tracks movements in real income that can only occur when misallocation is present:

$$\Delta(AE) = \Delta ln\left(\frac{W}{Q}\right) \tag{25}$$

With allocative efficiency, changes in real income are captured completely by the growth in an index of physical production. Change in real income not captured in physical production is therefore due to changes in allocative efficiency. Although I do not have firm level data on prices and quantities, the distortion term can be inferred using aggregate data on growth in real income and physical production. The growth of real income has an empirical counterpart in the data consistent with the assumptions that input prices are taken as given and prices reflect the marginal utility of a representative consumer. This is the Aggregate Productivity Growth (APG) measure used by Basu and Fernald (2002) and Petrin and Levinsohn (2012) defined by total growth in (deflated) value added within an industry, and corrected for the growth in primary inputs.[17]. For the aggregate price level, the Chilean statistical agency provides 4-digit ISIC industry deflators. Furthermore, I will use a real production index provided by the same agency that conducts the annual firm census. This survey tracks only a subset of the census of firms, but gets data on physical production (divorced from prices). This is used to produce an index of production at the 3-digit ISIC level that is available from 1995-2007 and allows me to track annual growth in physical production by incumbent firms. There is enough data therefore to infer the change in the misallocation distortion as implied by Equation 25.

Lastly, it is useful to try to understand the relationship between Equations 25 and 18. To do so, I go back to revenue and decompose the price-quantity covariance further to bring in markups. The new expression for real income (again using $P = \int_0^{c_d} p(q(c))h_d(c)dc$

[17]By the national income accounting identity, I use that the sum of value added is equal to the sum of final demand in an industry.

and $\frac{p}{c} = (\frac{1}{1-\mu(c)})$:[18]

$$R = ML\left[\int_0^{c_d} \frac{p(q(c))}{c} cq(c)h_d(c)dc\right] \tag{26}$$

$$\frac{R}{P} = LM\int_0^{c_d} cq(c)h_d(c)dc \int_0^{c_d} \frac{1}{c}h_d(c)dc + \left[\frac{LM}{P}\int_0^{c_d} cq(c)h_d(c)dc\left[\text{Cov}(p,\frac{1}{c})\right]\right] \tag{27}$$

$$+\frac{LM}{P}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq)\right] \tag{28}$$

I can separate out aggregate quantity from the first term on the right hand side. Since $Q = LM\int_0^{c_d} q(c)f(c)dc$, then $LM\int_0^{c_d} cq(c)f(c)dc\int_0^{c_d}\frac{1}{c}f(c)dc = Q - LM\text{Cov}(\frac{1}{c}, cq)$. I substitute this into the last equation and then once again come up with an equation for $\frac{W}{Q}$:

$$\frac{R}{P} = Q - LM\left[\text{Cov}(\frac{1}{c}, cq)\right] + \frac{LM}{P}\left[\int_0^{c_d} cq(c)h_d(c)dc\left[\text{Cov}(p,\frac{1}{c})\right]\right] + \frac{LM}{P}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq)\right] \tag{29}$$

$$\Delta ln\left(\frac{W}{Q}\right) = \Delta\left[\frac{LM}{PQ}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq) + \int_0^{c_d} cq(c)h_d(c)dc\left[\text{Cov}(p,\frac{1}{c})\right]\right] - \frac{LM}{Q}\left[\text{Cov}(cq,\frac{1}{c})\right]\right] \tag{30}$$

To identify misallocation from Equation 25 it is the first difference of this term that must be zero. Again, if the distribution is immune to truncation then the first difference must be zero if $u(q(c))$ is homothetic. Comparing to Equation 25, the price-quantity covariance is decomposed to separate out productivity ($\frac{1}{c}$), markups ($\frac{1}{1-\mu}$), total input cost (cq) and prices. Again it is important to notice that an increase in allocative efficiency occurs when there is a reallocation to high markup firms, in this case $\Delta\text{Cov}(\frac{1}{1-\mu(c)}, cq) > 0$ (of course using only this term would omit the simultaneous changes in the other two terms on the left hand side).

4.3 Discussion

The above decomposition identifies the role of the markup distribution in the market equilibrium with variable markups. The misallocation distortion that I measure in this paper differs from Hsieh and Klenow (2009) by breaking the linkage between physical productivity and welfare. Reallocation raises income and welfare not necessarily by reallocating production to more productive firms, but raising the production of firms with

[18] Appendix B shows the step-by-step decomposition.

a high social valuation. As in Hsieh and Klenow (2009), misallocation depends on firm-level distortions taking economy aggregates as given. In fact, the distortion I identify is not possible without firm heterogeneity.

The focus on average firm productivity in the previous literature makes sense in a CES world where the covariance term above is constant. Changes in the real income would be tracked by the change in aggregate weighted firm-level productivity as in Pavcnik (2002). This is not the case in Equations 18 and 25 as an increase in quantity is not necessarily associated with more production value. Using my firm-level data, I will track aggregate value added and decompose the growth in value added into growth in quantity produced and the change in the allocative efficiency term.

An important question regarding the quantification of distortions is where resources should be reallocated to improve social efficiency. This is an unsettled topic in the literature. Channeling the theory of Olley and Pakes (1996) and Melitz (2003), Pavcnik (2002) (and more recently Bartelsman et al. (2013)) focused on the covariance of market share and productivity to summarize allocative inefficiency (higher covariance implies better allocative efficiency/aggregate productivity). However Petrin and Levinsohn (2012) argue that this is not a correct measure of welfare gains due to reallocation as firms with higher productivity are not necessarily those that will have the highest social return from extra inputs. If firms minimize costs then the markup is the gap between the value of the marginal product and the marginal cost (or the factor price). Reallocating inputs from a firm with a lower than average markup to a firm with higher than average markup gets us closer to equalization of the marginal rates of transformation. Firms have high markups because their revenue productivity is too high, or input usage too low (as in Hsieh and Klenow (2009)). Although the distortions are different in my model (firm distortions are due to too-much/too-little market power), a similar argument can be made: reallocating inputs towards more productive firms is only optimal to the point where their marginal rates of transformation would be equalized. Heterogeneous markups indicate non-equalization of marginal rates of transformation, and reallocation works in equalizing these, as in Hsieh and Klenow (2009) and Basu and Fernald (2002).

In the VES model, "over/under-producing" is a result of market power and pass-through. In a constant markup model there is full pass-through so more productive firms pass on their lower costs to prices by increasing their quantity until the markup is the same as less productive firms. Due to incomplete pass-through, some of the cost advantages are passed through to markups instead, which is why low cost firms have higher markups. To do this, the low-cost firms reduce their production, and so produce less than under the optimal allocation. High cost firms choose low markups by producing more than is optimal. Take the case of two heterogeneous firms indexed by c and c′, with $c < c'$. Then, with incomplete pass-through: $\frac{p}{p'} > \frac{c}{c'}$, in clear contradiction to the socially optimal condition present in the CES that $\frac{p}{p'} = \frac{c}{c'}$.

4.4 Total Welfare Decomposition

Equation 25 captures the changes in misallocation due to real income effects from reallocation of production within a given measure of heterogeneous firms. However this does not capture the full implications of allocative inefficiency on welfare. It has been known since Dixit and Stiglitz (1977), summarized in Vives (2001), and expanded in Mrazova and Neary (2013a) that an inefficiency still exists with homogeneous firms due to a distortion in the number of available varieties.[19][20] I will now decompose the full welfare expression in my model to express clearly how I ignore this part of misallocation. Starting from $U(M_e, q) = M_e L \int u(q(c))dG(c)$, I will bring in the "elasticity of utility": $\epsilon(q) = \frac{\partial u(q)}{\partial q}\frac{q}{u(q)}$, the proportional increase in utility given an increase in the quantity of a variety. Then, as in Dhingra and Morrow (2012), the (utility-weighted) average elasticity of utility is $\bar{\epsilon} = \frac{\int \epsilon(q)u(q)}{\int u(q)}$. Using this definition, total utility is now $U(M_e, q) = \frac{1}{\bar{\epsilon}} M_e L \int u'(q(c))q(c)dG(c)dc$. Then, with δ as the marginal utility of income, $u'(q(c)) = \delta p(q)$. I follow the same steps as subsection 4.2.2 to decompose revenue within the welfare function:

$$U = ML\frac{\delta}{\bar{\epsilon}} \int p(q)q(c)h_d(c)dc \tag{31}$$

$$= ML\frac{\delta}{\bar{\epsilon}} \left[\frac{PQ}{ML}\right] \left[\frac{W}{Q}\right] \tag{32}$$

$$\Delta ln(U) = \Delta ln(\delta) + \Delta ln(1-\bar{\epsilon}) + \Delta ln\,(P) + \Delta ln\,(Q) + \Delta ln(AE) \tag{33}$$

Related to the result in Equation 12, the second and last terms in Equation 33 are zero with CES preferences. The last term, my measure of allocative efficiency, is zero by Lemma 2 with CES preferences if the distribution of costs is Pareto. The second term is a part of allocative efficiency that I do not capture, and is zero when the sub-utility function is CES.[21] Vives (2001) refers to $(1 - \epsilon(q))$ as "the proportion of social benefits not captured by revenues when introducing a new variety." A marginal entrant would lower the per-

[19]These studies all establish allocative inefficiency with symmetric firms, in contrast to my focus on the allocative inefficiency due solely to firm heterogeneity.

[20]Chamberlin (1933) also argued for "excess capacity" which resulted in excess entry.

[21]In which case: $\frac{1}{\bar{\epsilon}} = \frac{1}{1-\bar{\mu}} = \frac{\sigma}{\sigma-1}$, with σ the constant elasticity of substitution.

capita quantity sold by each incumbent firm, and therefore move $\epsilon(q)$ depending on the functional form of $u(q)$: for example in the linear demand case without the non-separable term, $(1-\epsilon(q))' > 0$, so the extra entry lowers the social benefits of the incumbent firms (the "business stealing" effect), while concurrently raising variety and welfare through higher revenues. Under the CES benchmark allocation these two effects cancel each other out.

The first term is the Lagrange multiplier of the budget constraint, whose inverse in ACDR represents the choke price. In fact, in the demand system of ACDR, they show that the percentage change in the choke price is equal to the percentage change in prices for homothetic preferences (though they ignore the CES case). This can be generalized for Equation 33 so that $\Delta ln(\delta)$ is equal to the negative of $\Delta ln(P)$ in the CES case. Therefore in that case we are left with $\Delta ln(U) = \Delta ln(Q)$[22], which explains why the literature has thus far concentrated on quantifying TFP gains. With VES preferences, there are extra terms that need to be accounted for to measure welfare gains. In Appendix D I show the decomposition of welfare in response to lower trade costs in the model of ACDR. In that decomposition a distortion exists in the VES model that is not present with the homothetic CES and Translog preferences. The focus of this paper is on this distortion: the last term in Equation 33. The next section details the separate competition and cost shocks that I study, in contrast to the 2 country model with variable trade costs of ACDR.

5 Global Shocks and Misallocation

The model above is informative about the firm-level distortions that cause misallocation and how to reallocate production to reduce this distortion. I assumed that there exists an equilibrium allocation at each point in time in which firms make decisions given the aggregate environment. Next I will investigate how changes in the aggregate environment affect the reallocation of production and the implications for allocative efficiency. The strategy is to fit into a reduced-form approach aggregate shocks that affect the equilibrium allocation. Changes in the domestic environment can affect firms through either i) their residual demand curve or ii) their marginal cost.

Changes in the residual demand curve can be due to tougher competition (or conversely, being more insulated from competition) through a larger market size. This changes the slope of the residual demand curve and leads firms to adjust prices (and quantities) for a given cost. In standard trade models with CES preferences, competition leads to a stan-

[22] Q is defined as $LM \int_0^{c_d} q(c)f(c)dc$. With heterogeneous firms, Q can rise due to selection.

dard selection effect that increases welfare through a higher average productivity. Aside from the competition effect, trade policy and other global shocks affect the marginal costs of firms. This type of trade gain has gotten a lot more attention in the recent literature on imported intermediate goods and technology adoption. Higher terms of trade, lower inputs tariffs, or better access to intermediate goods markets, lower the costs of production for domestic firms. When this effect has been investigated and combined with CES consumer preferences, cost decreases are fully passed on to prices.

However, with VES preferences, allocative efficiency comes into play though imperfect pass-through. Take for example an increase in market size. Demand elasticities increase for all firms by a constant, lowering prices and increasing quantity. There is the standard competition effect: the lowest productivity firms get selected out (this will raise average productivity). Additionally with VES, competition affects firms' demand curves heterogeneously because demand elasticies are a function of sales. I show in Section 5.2 that this reallocates production from less to more productive firms because more productive firms lower markup relatively more. Marginal cost shocks affect the production distribution in the opposite way. Not all of the marginal cost decrease is passed through to a pure aggregate productivity gain for the economy as some of that is eaten up by the increase in allocative inefficiency due to the pass-through into markups. More productive firms are able to increase their markup relatively more, resulting in relatively more production going to the less efficient firms.

A reduced-form approach allows for a more general framework than just integrating output tariffs into the model. Focusing exclusively on output tariffs can confound the tougher competition and lower costs, so that the two channels above can cancel each other out. Below I outline how each channel affects the markup distribution. Though the two shocks can happen simultaneously, in the empirical section I can identify the shock using the firm or industry's exposure to competition.

5.1 Misallocation and Markup Dispersion

Given Equations 18 and 30, the next step is to establish how markup dispersion drives the market power distortion. This is evident from the definition of the correlation:

$$\text{Cov}\left(\frac{1}{1-\mu(q(c))}, cq(c)\right) = \text{corr}\left[\frac{1}{1-\mu(q(c))}, cq(c)\right] \sqrt{\int_0^{c_d} \left(\frac{1}{1-\mu(q(c))}\right)^2} \sqrt{\int_0^{c_d} (cq(c))^2} \tag{34}$$

The second term on the right hand side is the standard deviation of the markup distribution. In the empirical section I show that both the correlation term and the markup dispersion are important in driving covariance.

5.2 Global Shocks and Markups

I analyze competition shocks that can change the slope of firms' residual demand curves as well as shifts in the marginal cost distribution of domestic producers that are possible through increased trade and cheaper inputs. The average productivity responses from these two shocks are well known: tougher competition lowers the cutoff cost for domestic producers, which leads to an overall higher value of production; while lower marginal costs allow firms to lower prices and increase quantity produced. However the impact on the markup distribution has not yet been explored. Allowing for average productivity to increase with either type of shock, I differentiate between shocks that also increase allocative efficiency (I call these *pro-resource allocation*) and those that dampen welfare gains by reducing allocative efficiency (*anti-resource allocation*). I adopt a method introduced by Mrazova and Neary (2013b) to compare the distributional changes at the new equilibria and determine whether a shock is pro- vs anti-resource allocation.

I assume that an equilibrium exists at each point in time, which allows for a cross-section comparison in a given equilibrium (how do firms change their markup?), and a time-series analysis by comparing across equilibria given changes in the domestic environment (after the shock how do firm-level responses affect the aggregate?). At each equilibrium I can analyze selection and competition effects by comparing across producers that are differentiated by their productivity/marginal cost. Since misallocation can be summarized using the covariance of markups and input expenditure, I will focus on the effect of two kinds of shocks on firm-level markups. First, a pure globalization shock (without taking trade costs into consideration) occurs with a shock to the number of firms (M_e) or the market size (L). Intuitively, this shifts the demand elasticities that firms face. In a two country model with trade costs, changes in demand elasticities can also be a result of lower import tariffs which allow for more foreign entry. However that case would also have to take into account lower costs of importing inputs.[23] Second, I allow for lower costs of production/efficiency improvements for domestic firms that can result from cheaper inputs. In contrast to globalization, this shock is identified by movements in a firm's supply curve. In the aggregate, the cumulative effects are reflected in a dif-

[23] DeLoecker and Goldberg (2013) actually differentiate between shocks to the residual demand curve and shocks to the marginal cost curve as responses to output and input tariffs changes respectively.

ferent allocation of production (the q(c) vector that itself summarizes markups), and an adjustment in the cutoff productivity to retain the labor (i.e. budget) constraint.

Both scenarios above will shift firm markups as each shock affects the pricing decision of the firm. In the globalization scenario, larger market size/tougher entry imply an increase in the marginal utility of income ($\frac{\partial \lambda}{\partial L} > 0$). Since $p(c) = \frac{u'(q(c))}{\lambda}$, prices shift for all firms by a constant proportion.[24] Let $p_i(\lambda', c_i)$ represent the price decision of firm i after a globalization shock.

To examine the second case, I introduce imported inputs as a source of production with a constant labor requirement. To give the firm marginal cost more structure, let the cost using only domestic inputs be expressed as a labor requirement to produce one unit: $c_i(\varphi_i) = \frac{a}{\varphi_i}$, for a constant a, and φ_i the firm's draw from a productivity distribution. With trade, firms can also import inputs at a labor requirement of $\frac{a(\tau-1)}{\varphi_i}$. The total marginal cost of production then becomes $c_i(\tau, \varphi_i) = \frac{a\tau}{\varphi_i}$, where τ is a scalar in the marginal cost curve that represents the cost of importing inputs. A shock that lowers the cost of imported inputs scales down $a\tau$. This allows for a productivity shock that lowers production cost and allows firms to increase markups with incomplete pass-through. The impetus for this mechanism can be a reduction in production costs through endogenous productivity increases,[25] terms of trade gains (of course a terms of trade loss would just imply an increase in τ), and lower input tariffs.[26]

Taking both effects into consideration, let price be represented by $p_i(\lambda, a\tau/\varphi_i)$ and Equation 5 is rewritten to express the markup as:

$$\frac{p_i(\lambda, a\tau/\varphi_i)}{a\tau/\varphi_i} = \frac{1}{1-\mu(\lambda, \tau, \varphi_i)} = m_i(\lambda, \tau, \varphi_i) \tag{35}$$

Markups are a function of one firm primitive and two aggregate variables that identify the domestic environment. There is a continuum of firms endowed with productivity φ_i, leaving changes in allocation equilibria (vector $q_i(c(\tau, \varphi_i))$ and c_d) due solely to movements in λ and $a\tau$. Although changes in either aggregate acts as a constant shifter to the demand or supply curve, the firm-level response is of course heterogeneous because markups are a function of sales.

The goal is not to measure productivity growth but to separate growth of allocative efficiency from productivity growth in total real income. Next, I show how a pro-resource

[24] Marginal utility of income and market size/entry are aggregates and taken as given at the firm-level.

[25] Melitz and Redding (2014)

[26] $\frac{\partial q}{\partial c}\frac{c}{q} < -1$, so that a reduction in marginal costs will increase the equilibrium individual consumption of each variety and increase its markup.

allocation shock is the result of an increase in λ (e.g. globalization), and an anti-resource allocation shock results from a decease in $a\tau$ (e.g. firms see a reduction in costs). Both events can of course occur simultaneously, but which effect dominates depends on industry characteristics. In the empirical tests, I compare across specific firm- and industry-characteristics in a differences-in-differences type approach, choosing characteristics that can imply dominance by one of the two particular shifters.

To relate the pro- and anti-resource allocation effects to misallocation I start with the second case from above. Letting $m_i(\lambda,\tau,\varphi_i) = \frac{1}{1-\mu_i(\lambda,\tau,\varphi_i)}$, firm-level responses to a shock are given by $\frac{\partial m_i(\lambda,\tau,\varphi_i)}{\partial \tau}$, and the reallocation effects can be interpreted as $\frac{\partial m_i^2(\lambda,\tau,\varphi_i)}{\partial \tau \partial \varphi_i}$. The first comparative static is trivial: the direction of the markup for each firm after the shock. The interpretation for the latter is the firm-specific sensitivity of the markup in response to the shock based on the firms' primitive marginal cost. If $\frac{\partial m_i^2(\lambda,\tau,\varphi_i)}{\partial \tau \partial \varphi_i} > 0$, markup differences across low versus high productivity firms get smaller at higher τ. This follows the method of Mrazova and Neary (2013b), who use the second derivative to establish super/sub-modularity. In general, the function $m_i(\lambda,\tau,\varphi_i)$ is supermodular in τ and φ_i (for a given λ) if:

$$\Delta_{\varphi_i} m_i(\lambda,\tau_1,\varphi_i) \leq \Delta_{\varphi_i} m_i(\lambda,\tau_2,\varphi_i) \text{ when } \tau_1 \geq \tau_2 \tag{36}$$

$$\text{where } \Delta_{\varphi_i} m_i(\lambda,\tau,\varphi_i) = m_1(\lambda,\tau,\varphi_1) - m_2(\lambda,\tau,\varphi_2) \text{ for } \varphi_1 \geq \varphi_2 \tag{37}$$

Notice $\Delta_{\varphi_i} m_i(\lambda,\tau,\varphi_i)$ is always positive. Super-modularity holds when $\frac{\partial m_i^2(\lambda,\tau,\varphi_i)}{\partial \tau \partial \varphi_i} > 0$. Therefore the markup difference between two firms differentiated by their productivity/marginal cost gets smaller or larger depending on the change in τ.

Going back to Equation 35, $\frac{\partial m_i(\lambda,\tau,\varphi_i)}{\partial \tau} < 0$, or markups decrease with τ. By differentiating this with respect to firm-specific productivity, we get the differential effects of a change in τ for firms with different marginal costs. With CES preferences, it can be shown that $\frac{\partial m_i^2(\lambda,\tau,\varphi_i)}{\partial \tau \partial \varphi_i} = 0$.[27] With my assumption about preferences in Section 3, it can be shown that $\frac{\partial m_i^2(\lambda,\tau,\varphi_i)}{\partial \tau \partial \varphi_i} > 0$.[28] Therefore at lower τ's, there is a *bigger* markup difference between a low cost and a high cost firm. This means that lowering τ (firms being able to charge higher markups) is more beneficial for *low-cost* firms as they can increase their markup relatively more by passing through less of their cost decreases to prices. In reallocation terms, inputs are reallocated relatively to initially low markup firms for this result to hold.

[27] See Appendix C for the derivation. The second step of showing super/sub-modularity (differentiating with respect to φ_i) has to be worked through a bit.

[28] In Mrazova and Neary's terminology, this is equivalent to markups being super-modular with respect to trade costs when demand is "less convex" than CES.

This setup makes the simplifying assumption that importing requires a constant labor requirement.[29] A conceivable setup is that of Gopinath and Neiman (2012) with a fixed import cost, which allows for non-homothetic import demand. Their paper focuses on productivity changes in response to shocks in the ability to import, and not market power or allocative efficiency. I allow for a productivity increase through a larger Q, but focus on the concurring change in misallocation due to market power. Additionally, in Amiti et al. (2012) larger firms import more and are the most likely to take advantage of a reduction in τ. In the empirical section I use information about the share of imports in a firm's material cost and study the distributional effects of market power for a given share of imports. I should point out also that the anti-resource allocation result should only be exacerbated if it is the most productive firms that get the cost decreases. Finally, more surviving firms due to the reduction in the cutoff cost can be a factor through the utility of variety but is unlikely to affect aggregate revenue since in the model these are very small firms.

In a Krugman (1979) globalization episode where the market size (indexed by L) expands, $\frac{\partial q_i}{\partial L}\frac{L}{q_i} < -1$ (equilibrium consumption of each variety decreases), and demand elasticities increase. The same result occurs when entry intensifies (M_e increases) and the probability of surviving to produce decreases. With markups a positive function of individual consumption, this is equivalent to a decrease in the firm-level markup for all varieties. Using the same super/sub-modularity method as above, tougher competition not only lowers the average markup but also leads the lower cost firms to decrease their markup more than high cost firms.[30]. This can be shown using $p_i(\lambda', c_i(\tau, \varphi_i)) = \frac{\lambda'\eta(c_i)}{\lambda'\eta(c_i)-1}c_i(\tau, \varphi_i)$ where $\eta(c_i)$ is the original demand elasticity faced by a firm with marginal cost c_i (before the globalization shock). λ' is an aggregate that shifts demand elasticities for all firms with $\lambda' > 1$ when L or M_e increase, and $c_i = \frac{a\tau}{\varphi_i}$ is constant for each firm since there is no shock in τ. Then $\frac{\partial p_i(\lambda, c_i(\tau,\varphi_i))}{\partial \lambda'} < 0$: an upward shift in demand elasticities lower prices for all firms. Furthermore as shown in Appendix C, $\frac{\partial p_i^2(\lambda, c_i(\tau,\varphi_i))}{\partial \lambda' \partial c_i} < 0$, meaning that prices are sub-modular in λ' and c_i. Again, this relies on the assumption that demand elasticities decrease with sales (in the CES case $\frac{\partial p_i^2(\lambda, c_i(\tau,\varphi_i))}{\partial \lambda' \partial c_i} = 0$). The price difference between a low and high-cost firm gets *smaller* with bigger upward shifts in demand elasticities – markups differences tighten – and the reallocation implication is that higher markup firms increase production relatively more as they move down their demand curve.

A globalization episode will lower the cost cutoff, eliminating firms with the lowest

[29] An example of a setup where τ is the same for all firms is Acemoglu et al. (2012). In that case the foreign country plays the role of a general-purpose technology.

[30] A result similar to Melitz and Ottaviano (2008)

markups. Aside from the quantity gains from a Melitz-type selection, tougher competition can lead to welfare gains by reallocating the inputs originally used by the "selected" firms to a production of larger valuation.

5.3 Testable Predictions

Incorporating the global shocks allows for testable predictions. The main question of interest is how the distinct global shocks, either through costs or competition, affect aggregate misallocation at the industry level.

In the empirical section I will establish that the observed firm level reallocation is consistent with growth rate in allocative efficiency. Subsection 5.2 allows for a theoretical foundation in the firm-level responses to shocks in costs or competition. With the assumptions on demand, reallocation of production can be inferred from the observed markup response and this is allows for the channel that links the shocks to aggregate misallocation.

Hypothesis 1. *A "favorable" cost shock is anti - resource allocation as it reallocates inputs to initially low markup firms. Globalization, defined as increased competition, is pro - resource allocation as quantity production is shifted relatively to highly-valued products.*

This hypothesis is tested by the consistency of the observed firm-level responses with the growth in aggregate allocative efficiency as defined in subsection 4.2.2. In the following empirical analysis I use Chilean data to measure growth in allocative efficiency at the 2-digit industry level and markups using production function estimation. In Section 7, I establish that the shocks are consistent at the micro level with the predicted changes in markups and at the macro level with the implied changes in allocative efficiency. In the next section, I describe the data and important open economy measures for Chile.

6 Data and Background Information

6.1 Data Description

I combine a Chilean firm level panel data from 1995-2007 with aggregate statistics from this same period. The firm level data is provided by Encuesta Nacional Industrial Anual (ENIA, National Industrial Survey) and collected by the National Institute of Statistics (INE). It covers a census of manufacturing firms, ISIC (rev. 3) classification 15-37, with more than 10 workers. There are approximately 5,000 firm level observations per year and firms are tracked across time with a unique identification number. Each firm provides detailed economic data such as total sales, number of workers, value of fixed capital, expenditures on intermediate inputs, etc. Importantly, firms also report the value of inputs that are imported from abroad and what value of their total sales is exported. The percentage of firms that export and the fraction that import are both consistently around 20% throughout the data span.

From Section 4, there are three aggregate measures that I use in Equation 25: revenue, prices, and quantity. The growth in real income (revenue over prices) can be computed by aggregating deflated value added of all firms within an industry. Value added is at the firm level, and the ENIA provides sales and input deflators at the 4-digit ISIC level. Therefore the most disaggregated measure of real income growth available is at the 4-digit ISIC. For physical production, I use an index provided by the INE at the 3-digit level. This index follows a subset of firms with bases in 1989 (for the 1995-2002 data) and 2002 (used for the 2003-2007 data).

Other macro and open economy data is taken from a variety of sources. The Central Bank of Chile provides manufacturing GDP, nominal exchange rate and aggregate export and import data. Detailed export and import data at the 4-digit level is provided in the world trade flows database of Feenstra et al. (2005). I compute a real effective exchange rate as a geometric average of relative prices using trade weights and output prices provided by the Penn World Tables.[31] Terms of trade plus alternative import and export data can be obtained from World Development Indicators (WDI) at the World Bank. The World Integrated Trade Solutions (WITS) database has detailed tariff data that I aggregate to the 4-digit level. It provides data from both the World Trade Organization (WTO) and Comtrade. In the main specification I use applied rates reported by Comtrade. To measure input tariffs, which I define below, I use a 3-digit[32] input output matrix provided by the Chilean Central Bank in its National Accounts publications of 1996 and 2003.

6.2 Open Economy Summary Statistics

The time period examined in this paper is subsequent to the big trade reform in Chile that occurred in the late 1970's (and studied in Pavcnik (2002)). Although Chile has been a WTO member since 1995, in the period under analysis it underwent several important trade liberalization episodes. The decrease in average tariffs and signings of various trade agreement were concurrent with an increase in the share of exports to manufacturing

[31] There is also a real effective exchange rate provided by the IFS database.

[32] I concord industry descriptions by hand to match my ISIC revision 3 data.

GDP. Part of this was demand driven as Chile gained from the inflation in commodity prices that was likely due to the increased demand from emerging countries. For Chile this was especially important in the copper industry, which constitutes almost half of its export value. The result was a large terms of trade gain starting in 2003, which was later followed by a large increase in imports, driven especially by intermediate inputs.[33]

Figure 1 shows the average applied tariff rate from the Comtrade database. In the time span of the data, average applied tariffs in the manufacturing sector decrease from 11% to below 2%.[34] This drop is mostly homogeneous across industries.

[Figure 1 about here.]

Aside from the average tariffs above, the many trade agreements signed by Chile are anecdotal evidence of its trade liberalization. Appendix F lists these agreements.

Apart from trade reform, Chile also experienced a large shock to its real exchange rate during this period. Figure 2 describes a terms of trade index taken from the WDI (right axis), and the annual log differences in a real effective exchange rate that I construct (PWT) compared to an IMF index (left axis). The terms of trade increases starting in 2003, which coincides with the incline in import and export values at about the same time (next paragraph). I expect the exchange rate to play a role in my analysis as Chile goes through a sustained depreciations and appreciations during my data span. [35] Chile experienced an appreciation in 1997, a sustained depreciation from 1999-2003, and a sustained appreciation 2004-2006 led by the terms of trade gain.

[Figure 2 about here.]

[Figure 3 about here.]

Since my data spans firms the manufacturing sector, I investigate how manufacturing specifically is affected by liberalization and subsequent terms of trade gains. Although the Mining industry is not included in my data, the Basic Metal industry is significantly affected by the price of copper, so I drop it from my analysis below.[36] Figure 3 plots manufacturing exports and imports as a ratio of total manufacturing value added. Exports

[33] Desormeaux et al. (2010) establishes that firms and households import a significant amount of their intermediary inputs. In current work with Felipe Lucero, I use customs data to examine firm level imports in Chile.

[34] Using the Most Favored Nation (MFN) tariffs instead of the applied rates, rates only decrease to 6%.

[35] I use Penn World Tables to calculate Chile's production price index relative to its top trade partners and take a geometric average using trade shares as weights. I compare it to the IMF real effective exchange rate index.

[36] Berthelon (2011) documents that Chilean export performance from $1990-2007$, even taking out copper industries, shows growth in the extensive margin and diversification of products as well as partners.

and imports are gross flows from Feenstra et al. (2005) (so they can be greater than total manufacturing value added). I sum flows only for manufacturing industries (ISIC industries 15-37 after I concord with ISIC rev.3). I also report manufacturing exports/imports excluding the Basic Metal industry. Manufacturing exports as a ratio of manufacturing GDP rises sharply starting in 1999, though imports do not rise until 2004. Exports climb before the terms of trade gain, evidence of a push towards exports and help from the depreciation of the Peso, while imports seem to react to the terms of trade gain through the higher purchasing power. This pattern still holds after eliminating the Basic Metal industry, though there is a big drop-off in exports/GDP. For the manufacturing firms that I consider, importing is as, or even more important than the export side.[37] Although there is evidence of both export and import growth, it seems that export earnings are the initial impetus, with the demand for intermediate inputs driving imports.

7 Empirical Analysis

In this section I test the model predictions about reallocation and the aggregate misallocation consequences to connect firm level behavior with aggregate data. I split the section into specification and results, with the former defining measures that I summarize in results.

The misallocation distortion that I estimate in this paper occurs because the allocation of production is not based only on firm level productivity but also firm level market power. Market power is positively related to productivity, allowing more productive firms to under-produce in order to increase profits through a high markup. A way to gauge the evolution of this distortion in the data is to measure the markup estimates directly. In the next subsections I summarize the method to calculate firm level markups and then show suggestive evidence by concentrating on the aggregate average and dispersion of markups. This can be compared to the time series of aggregate allocative efficiency as described in Section 4.2.2. Later, I turn to a regression analysis with differential treatment groups to investigate how pro- and anti- resource allocation effects determine productivity growth and allocative efficiency.

[37] Imports are not affected by excluding Basic Metal.

7.1 Empirical Specification

7.1.1 Production Function Estimation and Markups

I have defined allocative efficiency as a degenerate markup distribution. I now investigate to what extent we see markup dispersion in Chile, and how this compares across industries and years. I use the method from DeLoecker and Warzynski (2012) to first calculate production function coefficients ala Ackerberg et al. (2006) (ACF), in itself an extension of the seminal contributions of Olley and Pakes (1996) and Levinsohn and Petrin (2003) (OP and LP). I then use the coefficients to estimate firm-level markups. The production function must follow the following functional form:

$$Q_{it} = F(X_{it}, K_{it}; \beta)exp(\omega_{it})$$

β is the vector of output coefficients, ω_{it} is a firm's (i) productivity at time t, ϵ_{it} the measurement error, and X_{it} the set of variable inputs (e.g. labor and materials). Given data constraints, Q_{it} is deflated total sales.[38] I take logs and use a Gross Output, Translog production function:

$$y_{it} = \beta_l l_{it} + \beta_{ll} l_{it}^2 + \beta_k k_{it} + \beta_{kk} k_{it}^2 + \beta_m m_{it} + \beta_{mm} m_{it}^2 + \beta_{lk} l_{it} k_{it} + \beta_{lm} l_{it} m_{it} + \beta_{km} k_{it} m_{it} + \beta_{lkm} l_{it} k_{it} m_{it} + \omega_{it} + \epsilon_{it}$$

l, k, m refer to labor, capital and intermediate inputs respectively. I estimate each 2-digit industry separately, using 4-digit industry input and output deflators provided by the Chilean Statistics Institution (INE). Notice that this Translog production specification allows for heterogeneous firm level output coefficients.[39] Importantly, I incorporate exporter and importer dummies into the ACF methodology as state variables to the firms' production decisions. This allows exporters and importers to follow a different production technology, following the strategy of Kasahara and Rodrigue (2008) (they add an importer dummy as a state variable), and DeLoecker and Warzynski (2012) (they use export status similarly). Specifically, in the first step of the ACF procedure for the production function estimation, I add imports and exports into the intermediate input demand func-

[38] Labor is the number of total workers. I combine skilled and unskilled although they can be split up using a subjective classification of labor categories. Capital and materials are both expressed as total deflated value of the input.

[39] Given the production function above, the output elasticity of materials for example is: $\theta_{it}^m = \beta_m + 2\beta_{mm} m_{it} + \beta_{lm} l_{it} + \beta_{km} k_{it} + \beta_{lkm} l_{it} k_{it}$. βs are constant by sector for all years, however notice that θ_{it}^m depends on firm and year specific input values. Output elasticities are therefore firm and year specific.

tion of the firm.[40][41] Furthermore, these dummy variables are used in the estimation of survival probabilities (using a Probit function) that control for non random exit of firms as a determinant of next-period productivity.[42]

Table 1 shows the production function coefficients and the median markup in all industries. I estimate firm level markups from the gap (or "wedge") between the output elasticity of materials (θ_{it}^m) and the cost share of materials (α_{it}^m) in total costs. The only assumption necessary is that firms minimize costs, so that the output elasticity is then set equal to its cost share. Markups could also be estimated using the same gap in the labor input, though labor requires more adjustment costs than materials and is less variable. This would make it a worse measure of markups, but I do compare some results to using the labor "wedge" as well. Specifically, my markup measure, at the firm-time level, is represented by:

$$\mu_{it} = \frac{P_{it}}{MC_{it}} = \frac{\theta_{it}^m}{\alpha_{it}^m} \tag{38}$$

From Table 1, markups and returns to scale are highest in ISIC industries 31-34 as well as 24-27, which encompass the machinery and chemical plus metal industries respectively. The median markup is consistent with past estimates, at 26%.

[Table 1 about here.]

7.1.2 Regression Specification

In the regression analysis of Subsection 7.2.4, I start at the firm-level with a framework similar to Pavcnik (2002) and Amiti and Konings (2007), that study, respectively, how output and input tariffs affect firm revenue TFP. I am interested in testing the results in Section 5 with respect to the distributional effects of aggregate shocks. I then aggregate to the industry level to use my measures of ΔW, ΔQ, and ΔAE as introduced in Section 4. I show that the distributional effects and aggregate outcomes are consistent with the model.

My data has information on whether a firm is an importer/exporter plus the respective value. I interact this information with macroeconomic shocks that include trade liberalization variables and the terms of trade. Information on imports and exports is important because competition and cost shocks affect firms depending on their exposure.

[40] For a full account of the 2-step procedure see Olley and Pakes (1996), Levinsohn and Petrin (2003), or Ackerberg et al. (2006).

[41] Or in the Olley and Pakes (1996) framework, the investment demand function. This gets inverted to get a non-parametric function for the unobserved productivity shock.

[42] See Olley and Pakes (1996) for a full discussion about the necessity to account for exit/survival.

The general framework is:

$$y_{ijt} = \alpha_i + \alpha_t + \alpha_j + \beta\tau_{j,t} + \gamma\mathbb{1}_{ijt} + \psi\tau_{j,t} * \mathbb{1}_{ijt} + \zeta X_{ijt} + u_{ijt} \tag{39}$$

α_t and α_j represent time (t) and industry (j) fixed effects respectively (and in some cases I add firm (α_i) fixed effects). There is a trade liberalization variable (τ_{jt}) that can be output tariffs, input tariffs, or terms of trade, plus a firm- or industry-level indicator, $\mathbb{1}_{ijt}$. This indicator can take the form of an exporter/importer dummy or a share of exports in total sales/share of imports in inputs. Following Ekholm et al. (2012), I create a "Net Exposure" variable that I describe below. The main variable of interest is the interaction of the trade variable with the firm/industry indicator. Therefore the framework is a difference-in-difference approach with the import/export dummy or net exposure variable as the treatment group. Lastly, X_{ijt} includes other firm/industry characteristics.[43] The outcome variable, y_{ijt}, is the log markup ($ln(m_{ijt}(\lambda_t, \tau_t, q_{ijt}(c_{ijt})))$) at the firm-level. At the industry-level, the outcome variable becomes y_{jt} and I discuss the measures below.

7.2 Results

This section starts with suggestive evidence in summarizing the distribution of firm level markups. The last subsection details the results of the regression framework outlined in Section 7.1.2.

7.2.1 Markup Moments

The majority of the literature on variable markups has focused on average markups due to a "pro-competitive" effect (Feenstra and Weinstein, 2010). Here I show the evolution of both the average and dispersion of markups.[44] Figure 4 plots the mean markup computed by sector and reports a sectoral value added-weighted average. The figure shows that there is a reduction in the mean markup at the beginning of the period but that it rises starting in 1997.

[Figure 4 about here.]

[43]These include: Industry Herfindahl index, index of "import competition", a dummy for whether a foreign entity owns more than 10% of the firm, and the Rauch classification of differentiation in the industry.

[44]I drop the top and bottom 1% of firms (sorted by markups) in each year-sector and also the Basic Metal industry which would drive the results if it were included. It does not seem to matter how much I eliminate in terms of outliers. I have also dropped up to the top and bottom 3% of firms without a change in qualitative results.

Next I turn to the markup distribution, which is the new contribution to the reallocation analysis relative to the rest of the gains from trade literature in the monopolistic competition framework. I use the standard deviation of log markups as my measure, though the results would be qualitatively similar using the Pareto shape parameter.[45] In addition to using the material input markup wedge, I also use the labor coefficient-cost share wedge as a separate measure. Figure 5 takes all firms in a given sector, with the markup dispersion calculated annually excluding the Basic Metal industry, and averaged, where the weights are defined by sector value added. This is therefore an economy-wide measure of markup dispersion using only the dispersion within sectors. The dispersion has a similar pattern as average markups,[46] though the increase in dispersion is more consistent with the story in Section 5. The dispersion gets smaller through 2002, and then spikes up in 2003. The FTAs signed by Chile starting in the mid-1990s and the rise in trade are compatible with a globalization episode that reduces markup dispersion. The spike in markup dispersion coincides with the terms of trade shock and large increase in the value of imports.[47] If firms are responding heterogeneously to increases in productivity in a way that increases dispersion in market power, misallocation could be a source of *dampening* the positive welfare effects that arise from an increase in the value of Chile's production.

[Figure 5 about here.]

7.2.2 Markup Dispersion versus Productivity Dispersion

A useful comparison is to look at the markup dispersion versus productivity dispersion. In Hsieh and Klenow (2009), markups are constant and misallocation is given by the dispersion in revenue productivity. I find that the results above would not hold if revenue productivity dispersion were used as an indicator for changes in allocative efficiency. For example, I compare the markup distribution versus the TFP distribution in 1995 and 2005, since this is where the biggest difference should be seen. Figure 6 shows the distribution of log markups on the left in 1995 and 2005. The right panel is the distribution of TFP across all firms in 1995 and 2005. In the markup distribution we can see that the fatness of the distribution is larger in 2005 as expected given the markup dispersion results above.

[45] I calculate the Pareto parameter using the procedure outlined in Head et al. (2014). These results are available upon request.

[46] This is consistent with the theory in Section 5, where anti-resource allocation shocks that raise markups for all firms also lead to higher dispersion due to the shape of the markup function.

[47] The pattern is similar whether I use labor or materials to calculate the markup.

However this is not evident in the TFP distribution: it has definitely shifted to the right but with no noticeable change in the dispersion.[48]

[Figure 6 about here.]

7.2.3 Aggregate Allocative Efficiency

In this section I use the measures of misallocation from Equations 25 and 30 applied at the industry level, aggregating to the 2-digit level.[49]

Figure 7 shows real income growth (labeled W) and physical production growth at the aggregate manufacturing level.[50] Appendix E discusses the calculation of real income growth. As with the economy as a whole, manufacturing production slows down between 1998-2002 and really picks up starting in 2004. Real income growth is mostly higher than production up until 2003 and then is lower in 2004, 2006 and 2007. The aggregate data implies that reallocation pre-2003 induced better allocative efficiency. Without sustained growth in quantity produced, the value of production grew in almost every year. This trend was reversed after the terms of trade shock. Now the evidence points towards a reallocation that is lowering allocative efficiency. The regression results will illuminate the mechanisms underlying these aggregate measures.

In the next section I will go beyond the contemporaneous correlation evidence to regression results. I expect industries that rely on imported intermediates to benefit more in terms of productivity and cost-advantages. An increase in average productivity would present itself through more physical production, but not necessarily the income component. Industries that export a greater portion of their output, or face import competition on output sold domestically, should instead face tougher competition and this would induce pro-resource allocation behavior.

[Figure 7 about here.]

7.2.4 Regression Results

Tables 2- 5 show firm-level responses to changes in open economy variables. The reported independent variables can be interpreted as follows. $Importer * Exp = 0$ is an indicator

[48]This result still holds if I use only firms that are active throughout the whole time span and ignore new firms after 1995.

[49]Quantity data is available at 3-digit, but due to some inconsistencies in appending pre- and post-2002 data, I aggregate up to two digits.

[50]Each is calculated at the 3-digit group level and I aggregate to the manufacturing level using value added shares by group. I eliminate groups in industry 27 (basic metals) since this industry represents more than half of total value added after 2004.

for firms that import a positive amount of inputs and do not export any output (similar interpretation for $Exporter * Imp = 0$). The share of imports in total material inputs is "Imported Share" and exports over total sales is "Exported Share". The "Net Exposure" variable is the difference between export share and import share for a firm. Making this distinction is important because open economy shocks will affect importers and exporters differently. Section 5 details how the characteristic of a firm can determine its response to globalization and productivity shocks that happen simultaneously. A negative net exposure identifies a firm that imports a larger share of its imports than its export share of sales. This type of firm is most likely insulated from globalization as it is likely to reduce costs without necessarily competing in the global market. Similarly, a firm has positive exposure if exporting is more important than its' importing. In the following tables, I interact all these different types of indicator variables with the terms of trade (TOT) and output/input tariffs.

Table 2 uses only output tariffs and terms of trade as explanatory variables for firm-level markups. The coefficient of output tariffs in Column (1) implies that a 10 percent decrease in tariffs raises markups by .24% on average.[51] Importing firms are affected the most by TOT changes and exporters are affected the most by lower tariffs. For example, importers who are not exporters have higher markups at larger values of TOT (appreciations). Column (3) shows that a 10 percent increase in the terms of trade increases markups by .39 percent more for importers who do not export relative to the rest of firms. This is not the case for exporters. For exporters, the most significant outcome is that they gain the most in productivity terms in response to lower tariffs (TFP results available upon request). Finally, notice that I use the terms of trade in a place where the real effective exchange rate (REER) should have a similar interpretation. I ran the same regressions using the REER instead and the results are very similar.

Table 3 combines regressions from the previous two tables but uses import/export shares, as well as the net exposure (Columns (1) and (3)), instead of dummies. Section 5 motivated the need to separately identify tougher competition and cost shifters. A positive exposure in this case exposes firms to more global competition. The case of cheaper inputs is consistent with a cost shifter, and in this case I expect firms with negative exposure to be the ones affected. The negative coefficient on the interaction between terms of trade and net exposure in the first column means that a higher terms of trade (TOT) increases markups for firms that have negative exposure (input importers). This fits with the earlier results, and columns (2) and (4) show that a higher import share raises markups

[51] This positive effect on markups reflects the positive effect on revenue productivity which cannot be separately identified from markups.

and TFP when TOT increases. Exporters suffer large decreases in TFP due to TOT gains. Results are more noisy for tariffs, but it is the case that a lower tariff raises TFP for firms that are exposed to the export market. This is consistent with tariff reductions affecting those firms competing in the global market for final goods. Again, the same analysis can be done using REER instead of terms of trade and the results are qualitatively similar. This is consistent with the theory, where importers gains from a REER appreciation while exporters suffer due to tougher competition.

[Table 2 about here.]

[Table 3 about here.]

I also construct input tariffs using output tariffs and a three-digit input-output (IO) matrix provided by the Chilean Central Bank. The availability of IO matrices in this period is limited to 1996 and 2003, so I assume the intermediate input shares of each industry are constant throughout 1995-2001 and 2002-2007. Input tariffs are constructed as in Amiti and Konings (2007): a weighted average of output tariffs, with the weights based on the cost shares of each input used in the industry at the 3-digit level.[52] Tables 4 and 5 differentiate between input and output tariffs to look at the effects of firm-level TFP and markups. The former table uses only the indicator dummies, with the only result being that lower input tariffs raise markups. This is consistent with the theory because lower input tariffs are beneficial for importers just like the TOT. The large amount of zeros is probably due to the very little variation in tariff reductions in Chile. Most tariff reductions are manufacturing-wide, so that input tariffs are almost identical to output tariffs (the correlation between the two is .99). Table 5 only differentiates from the prior table in using shares instead of dummies. From the first column, there is evidence that lower input tariffs lead to higher markups for negatively exposed firms. In Column (2), by decomposing the net exposure, there is evidence that lower input tariffs raise markups for importers and lower output tariffs reduce markups for exporters. Overall importing firms gain from lower input tariffs as predicted by the model and exporting firms become more productive and lower their markup in response to lower output tariffs. However because of the insignificant results, I will concentrate on just terms of trade and output tariffs at the industry level.

[Table 4 about here.]

[52]There are 74 products in the matrix, which I concord to the 3-digit ISIC level manually using product descriptions. This is slightly more disaggregate than the 2-digit IO table available from the STAN Database.

[Table 5 about here.]

The next step is to show that the distributional effects follow the predictions in Section 5. I have found evidence thus far for the first moment predictions, especially for cost reductions raising markups. I now examine the extent to which these findings are spread across the distribution of firms. The results of two separate strategies are expressed in Table 6. The first two columns interact the "TOT*Net Exposure" and "OutputTariff*Net Exposure" interactions with an indicator of whether a firm is in the top 30% of the markup distribution in a base year. I use 1995 and 2002 as base years. Column (1) shows that for a given exposure to competition, terms of trade appreciations have a bigger effect for firms that have larger markups initially (the Top 30% dummy is one). The coefficient in the second row is significant at the 10% level and is clearly larger for top firms compared to the first row. In the second pair of columns, I multiply the interaction of interest with the firm markup in the base year. Again, having a higher base markup leads to clearly larger effects in response to terms of trade shocks.[53] The coefficients are very small with respect to output tariffs, which suggests that lower tariffs either did not exert a strong competition effect or the drop in prices is being absorbed by the concurrent drop in marginal costs.

[Table 6 about here.]

The firm level regressions mostly affirm the predicted reallocation effects of Section 5. In Section 4 I described the aggregate measures that are a result of reallocation across existing producers. I turn now to the industry-level analysis, of which the main measure of interest is the growth rate of allocative efficiency. The method is similar to the firm-level analysis in that I compare sectors (at 2-digit ISIC aggregate) who import the highest percentage of their inputs with sectors that are more open (export more and compete with imports). I also replace exporters with a measure of "Openness", the sum of exports and imports of final goods into an industry divided by total industry sales. Lower output tariffs affect the industries that import final goods and therefore compete with domestic firms, so I expect these industries to face fiercer competition.

The main outcome variable of interest is the implied growth rate in misallocation, ΔAE, from Equation 25 (the residual from $\Delta ln(W) - \Delta ln(Q)$). I add $\Delta ln(Q)$ as an outcome, as well as $\Delta Cov(markup, inputs)$, which in Equation 30 is one of the components

[53] In a quantile regression, it is firms in the 60th-75th percentile that are the largest winners in terms of markup increases.

of the allocative efficiency variable. I also investigate $\Delta ln(TFPR)$ as an outcome, which is the growth rate of revenue productivity and the aggregate productivity variable used in the past literature (i.e. Pavcnik (2002)). The interaction terms include the same sector characteristics as before, interacted with the growth rate in terms of trade, output tariffs, and input tariffs. One concern is that the ΔAE variable is created using a physical quantity measure that does not cover the census of manufacturing firms. The Chilean statistical agency uses a fixed subset of firms for this measure, which means that firms who do not produce for at least 6 years in a row are most likely not present in the measurement.[54] To account for this, I eliminate firms that do not produce for 6 consecutive years to measure the outcome variables.

From columns (1)-(2) in Table 7, the main result is that when the TOT increases, industries with a larger fraction of importers (that are not exporters) suffer in terms of allocative efficiency. In Column (1) there is evidence also that acceleration in the growth of TOT increases allocative efficiency in "open" industries. Lower output tariffs have no allocative efficiency effects in open industries, as reflected in the third row of Column (2). Unsurprisingly, both importers and exporters have higher physical production ($\Delta ln(Q)$) at higher terms of trade (Column(4)). The outcome variable in Column (3) is revenue productivity ($\Delta ln(TFPR)$)[55] and it tends to move in accordance with ΔAE. This could mean that revenue productivity and deflated value added are a similar measure empirically. The penultimate column uses the markup-input expenditure covariance as a measure of misallocation in place of ΔAE[56], and the results with respect to the terms of trade are similar to column (2). This is reassuring that the growth rate in allocating efficiency is properly estimated. [57]

As with the firm regressions, I repeat this analysis using export and import shares in Table 8. The shares are now the average firm share at the sectoral level. The first column illustrates that globally exposed industries have positive growth rates in allocative efficiency due to an increase in the growth rate of the TOT. One way to interpret this coefficient is to compare industries with different extreme values of net exposure. For example, an industry with firms that import all of their inputs but do not export will have a net exposure of -1. Net exposure of 0 means the ratio of exports to sales is equal

[54] The physical quantity index is computed 1989-2002 with a base in 1989 and 2002-2007 with a base in 2002.

[55] This is industry productivity, equivalent to the measure in Pavcnik (2002) and Olley and Pakes (1996).

[56] A higher covariance increases AE according to Equation 30.

[57] The preceding results can be re-done by replacing the Terms of Trade with the Real Effective Exchange Rate (REER). These two variables contain very similar information. The regression results are very similar, and the interpretations the same, when replacing one for the other.

to the ratio of imports to total inputs (or it could signify no import or exports). Therefore the coefficient in the first row of Column (1) is interpreted as the net exposure of 0 industry having allocative efficiency growth than is 6 percentage points larger than the industry with net exposure of -1 in response to a 1% increase in the growth of the terms of trade. As expected, the importing industries become more misallocated with terms of trade gains. Positive exposure industries also become more efficient when output tariffs decrease, an indicator of tougher competition. An interpretation of the coefficient in the second row is that an industry with net exposure of 1 (all sales are exported without importing inputs) has a growth rate of allocative efficiency that grows 1.96 percentage points more than the reference industry with net exposure of 0 in response to a 1% decrease in the growth rate of output tariffs. Column (2) decomposes exposure into the export and import shares. It is the larger import share that drives reductions in misallocation in response to TOT shocks, though a larger exported share does not seem to increase allocative efficiency with lower output tariffs.

Another way to interpret the magnitude of these results is to create a binary variable for "exposure." Given the sector averages, I define an industry as negatively exposed ($NegativeExposure = 1$) if the average net exposure is less than -0.1. The results are shown in Column (3). An industry labeled as negatively exposed to globalization has a growth in allocative efficiency 0.63 percentage points lower in response to a 1 percentage point increase in the growth rate of the TOT. This is statistically significant, and given that the average allocative efficiency growth is 1.6%, is important economically as well.

Staying on Table 8, Columns (4) and (5) examine the effect on revenue productivity growth and physical production respectively. It is the negatively exposed industries that increase their production after increases in the TOT, though this is not significant. Revenue productivity again has the same sign as allocative efficiency. I find that exposed industries raise quantity with reduction in output tariffs, though their revenue productivity is lower. Again, Columns (6) and (7) confirm the allocative efficiency results with the covariance of markups and input expenditure.

[Table 7 about here.]

[Table 8 about here.]

Finally, the following tables test more directly for the firm characteristics that are associated with allocative efficiency growth, though not necessarily causes. From Section 4, ΔAE should be driven by inputs between reallocated towards firms with higher markups. This is in contrast to the reallocation term in the CES literature which is production between reallocated to more productive firms.[58] Table 9 regresses ΔAE on the covariance term in Equation 18 plus other covariance terms. Cov(tfp,mkt share) is the reallocation term in Pavcnik (2002) and Bartelsman et al. (2013) where part of the growth in measured average revenue productivity (revenue TFP or TFPR) is market share going towards more productive firms. This statistic summarizes welfare-improving reallocation due to selection under CES preferences. I calculate growth in TFPR and use it as an outcome variable in Column (1). As expected, Cov(tfp,mkt share) is important in explaining increases in TFPR. The markup-input expenditure covariance actually decreases TFPR. Columns (2)-(4) use the markup covariance plus TFP-market share as predictors of physical production growth and ΔAE respectively. I use both total wages and cost of materials to calculate the former covariance (though I show only results using material inputs since using wages gives the same conclusion). Cov(tfp,mkt share) does not seem to contribute to growth in physical production, while a lower market power distortion (higher markup and input covariance) also takes away from production. Results in Column (3) confirm that the markup-input expenditure covariance is an important part of the implied allocative efficiency growth variable. In Column (4) I add the TFP-market share covariance as well as a TFP-input expenditure covariance to the prior regression. It is still the case that a higher markup-inputs covariance is associated with higher allocative efficiency.

[Table 9 about here.]

To delve more deeply into the ΔAE term, I also check that it behaves as expected with respect to the distribution of markups. I expect the market power distortion to get worse as the markup distribution increases since the low cost firms are able to absorb more of their productivity advantages as markups. Table 10 shows that misallocation increases (ΔAE is negative) as markup dispersion increases. This is of course not surprising since we have already seen that the markup covariance term decreases with markup dispersion as predicted by the model. TFPR dispersion is the measure in Hsieh and Klenow (2009) that measures misallocation. It comes in with the right sign in Column (2), though it is only the markup dispersion that is significant and only in Column (1).

I then check how sectoral markup and TFPR dispersion are affected by imported inputs and openness in Table 11. I find clear evidence that sectors with a higher share of

[58] I should emphasize though that my allocative efficiency growth does *not* have the same interpretation as reallocation in that literature. Misallocation is not present with CES preferences; reallocation gains measure the gains to average productivity due to the new competitive environment. This reallocation is still present in my model but the way it affects aggregate productivity depends how the reallocation is to high markup firms instead of high productivity firms.

imported inputs have a higher markup dispersion. Openness is associated with lower markup dispersion. I replace markup dispersion with TFPR dispersion as the outcome in Column (3) and the coefficients are the same sign as Column (1). This is more evidence that using revenue productivity will conflate actual TFP growth with increases in markups.

[Table 10 about here.]

[Table 11 about here.]

8 Conclusion

This study examines how misallocation fits into demand systems with preferences that are "less convex" than CES. The distortion that keeps the market economy away from productive efficiency is the heterogeneity in market power, and I show this effect can be important using the case of Chile. By having a benchmark of allocative efficiency, I can back out growth in misallocation that is consistent with the co-movement of prices and quantities. I then turn to open economy shocks as potential factors for changes in this market power distortion. I use a reduced form approach that allows trade liberalization and terms of trade shocks to have separate and simultaneous effects on firm markups even if they both lead to average productivity gains. This can be summarized by industry aggregates that act as shifters of firm-level pricing decisions.

Chile experiences an increase in openness and a large demand shock for its commodities that increases estimated productivity. Markup dispersion decreases until 2003, but increases significantly after the terms of trade gain for Chile. I find evidence that the increase in markup dispersion is due to firms acting heterogeneously in response to productivity increases, and this means that allocative efficiency can be a significant factor in terms of overall welfare gains/losses. In this context, the mechanism I find most compelling is incomplete pass-through of productivity gains that are heterogeneous across the firm distribution within an industry. Chile experiences real income and productivity growth throughout this period with an obvious contribution coming from expanding its exports and imports. However changes in misallocation suggest that the real income growth relative to productivity growth can be smaller/larger than what is implied by CES models. In Chile's case the mechanism for growth in income is not necessarily selection. Instead, firms are able to produce with higher revenue productivities, which allows them to raise markups. This makes the allocation of production more distorted relative to the CES benchmark.

Chile can be characterized as an exporter of natural resources, especially copper, and importer of intermediate goods. It is therefore not surprising that there is a significant benefit for Chilean firms in terms of cheaper imported inputs. On the other hand, it is not clear how much its domestic producers are affected by an increase in global competition. However, other countries, especially rich countries, could have a very different import composition. They might import mostly final goods and export goods higher up in the vertical specialization ladder. This would mean that trade liberalization can have a more dramatic effect in terms of increasing competition in the manufacturing sector, as is convincingly shown in Feenstra and Weinstein (2010). Future research should consider the importance in the composition of imports and exports to how domestic firms respond to global shocks.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., Salehi, A. T., 2012. The network origins of aggregate fluctuations. Econometrica 80 (5), 1977–2016.

Ackerberg, D., Caves, K., Frazer, G., Dec. 2006. Structural identification of production functions. MPRA Paper 38349, University Library of Munich, Germany.
URL `http://ideas.repec.org/p/pra/mprapa/38349.html`

Amiti, M., Itskhoki, O., Konings, J., 2012. Importers, exporters, and exchange rate disconnect. Working paper research, National Bank of Belgium.
URL `http://ideas.repec.org/p/nbb/reswpp/201212-238.html`

Amiti, M., Konings, J., 2007. Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia. American Economic Review 97 (5), 1611–1638.

Arkolakis, C., Costinot, A., Donaldson, D., Rodriguez-Clare, A., June 2012. The elusive pro-competitive effects of trade. Working paper, MIT.

Atkeson, A., Burstein, A., 2008. Pricing-to-market, trade costs and international relative prices. American Economic Review 98 (5), 1998–2031.

Bartelsman, E., Haltiwanger, J., Scarpetta, S., 2013. Cross-country differences in productivity: the role of allocation and selection. American Economic Review 103 (1), 305–334.

Basu, S., Fernald, J., 2002. Aggregate productivity and aggregate technology. European Economic Review 46, 963–991.

Basu, S., Pascali, L., Schintarelli, F., Serven, L., 2010. Productivity, welfare and reallocation: Theory and firm-level evidence. NBER working paper 15579.

Bernard, A., Eaton, J., Jensen, B., Kortum, S., 2003. Plants and productivity in international trade. American Economic Review 93, 1268–1290.

Berthelon, M., 2011. Chilean export performance: The role of intensive and extensive margins. Working Paper Series 615, Banco Central de Chile.

Chamberlin, E., 1933. The Theory of Monopolistic Competition. Cambridge: Harvard University Press.

de Blas, B., Russ, K., Dec. 2010. Understanding markups in the open economy under bertrand competition (16587).
URL http://ideas.repec.org/p/nbr/nberwo/16587.html

DeLoecker, J., Goldberg, P. K., Aug. 2013. Firm performance in a global market. NBER Working Papers 19308, National Bureau of Economic Research, Inc.
URL http://ideas.repec.org/p/nbr/nberwo/19308.html

DeLoecker, J., Goldberg, P. K., Khandelwal, A. K., Pavcnik, N., Mar. 2012. Prices, markups and trade reform. NBER Working Papers 17925, National Bureau of Economic Research, Inc.
URL http://ideas.repec.org/p/nbr/nberwo/17925.html

DeLoecker, J., Warzynski, F., 2012. Markups and firm-level export status. American Economic Review 102 (6), 2437–2471.

Desormeaux, J., Garca, P., Soto, C., Autumn 2010. Terms of trade, commodity prices and inflation dynamics in chile. In: for International Settlements, B. (Ed.), Monetary policy and the measurement of inflation: prices, wages and expectations. Vol. 49 of BIS Papers chapters. Bank for International Settlements, pp. 115–128.

Dhingra, S., Morrow, J., 2012. Monopolistic competition and optimum product diversity under firm heterogeneity. Working Paper.

Dixit, A. K., Stiglitz, J. E., June 1977. Monopolistic competition and optimum product diversity. American Economic Review 67 (3), 297–308.

Edmond, C., Midrigan, V., Xu, D., 2012. Competition, markups and the gains from international trade. Working Paper.

Eeckhout, J., December 2004. Gibrat's law for (all) cities. American Economic Review 94 (5), 1429–1451.

Ekholm, K., Moxnes, A., Ulltveit-Moe, K., 2012. Manufacturing Restructuring and the Role of Real Exchange Rate Shocks. Journal of International Economics 86, 101–117.

Feenstra, R., Kee, H. L., 2008. Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity. Journal of International Economics 74 (2), 500–518.

Feenstra, R. C., 2014. Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity. NBER Working Papers 19833, National Bureau of Economic Research, Inc.
URL http://ideas.repec.org/p/nbr/nberwo/19833.html

Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., Mo, H., January 2005. World trade flows: 1962-2000. Working Paper 11040, National Bureau of Economic Research.
URL http://www.nber.org/papers/w11040

Feenstra, R. C., Weinstein, D. E., Feb. 2010. Globalization, markups, and the u.s. price level. NBER Working Papers 15749, National Bureau of Economic Research, Inc.
URL http://ideas.repec.org/p/nbr/nberwo/15749.html

Goldberg, P. K., Khandelwal, A. K., Pavcnik, N., Topalova, P., November 2010. Imported intermediate inputs and domestic product growth: Evidence from india. The Quarterly Journal of Economics 125 (4), 1727–1767.

Gopinath, G., Neiman, B., 2012. Trade adjustment and productivity in large crises. Working Paper.

Head, K., 2011. Skewed and extreme: Useful distributions for economic heterogeneity. Working paper series, University of British Columbia.

Head, K., Mayer, T., Thoenig, M., Feb. 2014. Welfare and Trade Without Pareto. Sciences po economics discussion papers, Sciences Po Departement of Economics.
URL http://ideas.repec.org/p/spo/wpecon/infohdl2441-f6h8764enu2lskk9p2m9j4i07.html

Holmes, T. J., Hsu, W.-T., Lee, S., March 2012. Allocative efficiency, markups, and the welfare gains from trade. Working paper, University of Minnesota.
URL http://www.econ.umn.edu/~holmes/papers/alloc_eff_trade.pdf

Hsieh, C.-T., Klenow, P., 2009. Misallocation and manufacturing tfp in china and india. Quarterly Journal of Economics 124 (4), 1403–1448.

Kasahara, H., Lapham, B., 2013. Productivity and the decision to import and export: Theory and evidence. Journal of International Economics 89 (2), 297–316.

Kasahara, H., Rodrigue, J., 2008. Does the use of imported intermediates increase productivity? plant-level evidence. Journal of Development Economics 87 (1), 106–118.

Khandelwal, A. K., Schott, P. K., Wei, S.-J., October 2013. Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters. American Economic Review 103 (6), 2169–95.

Krugman, P. R., 1979. Increasing returns, monopolistic competition, and international trade. Journal of International Economics 9 (4), 469–479.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economics Studies 70 (2), 317–342.

Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.

Melitz, M., Ottaviano, G., 2008. Market size, trade and productivity. Review of Economic Studies 75, 295–316.

Melitz, M., Redding, S. J., 2014. Missing gains from trade? American Economic Review Papers and Proceedings.

Mrazova, M., Neary, P. J., Dec. 2013a. Not So Demanding: Preference Structure, Firm Behavior, and Welfare. Economics Series Working Papers 691, University of Oxford, Department of Economics.

Mrazova, M., Neary, P. J., January 2013b. Selection effects with heterogeneous firms. Working paper.

Olley, G., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–1298.

Pavcnik, N., 2002. Trade liberalization, exit and productivity improvements: Evidence from chilean plants. Review of Economic Studies 69, 245–276.

Peters, M., 2011. Heterogeneous markups and endogenous misallocation. Working Paper.

Petrin, A., Levinsohn, J., 2012. Measure aggregate productivity growth using plant-level evidence. The RAND Journal of Economics 43 (4), 705–725.

Restuccia, D., Rogerson, R., 2008. Policy Distortions and Aggregate Productivity with Heterogeneous Plants. Review of Economic Dynamics 11 (4), 707–720.

Trefler, D., 2004. The long and short of the canada-u.s. free trade agreement. American Economic Review 94, 870–895.

Vives, X., 2001. Oligopoly Pricing: Old Ideas and New Tools. Vol. 1 of MIT Press Books. The MIT Press.

Zhelobodko, E., Kokovin, S., Parenti, M., Thisse, J., 2012. Monopolistic Competition: Beyond the Constant Elasticity of Substitution. Econometrica 80 (6), 2765–2784.
URL http://ideas.repec.org/a/ecm/emetrp/v80y2012i6p2765-2784.html

Appendices

A Price-Quantity Covariance

This appendix establishes the result of Lemma 2 that Equation 24 is zero in the case when the sub utility function is CES and the added assumption of Pareto distribution of marginal costs. I use the definition of the covariance: $\text{Cov}(p, q) = \int_0^{c_d} (p(q(c)) - \tilde{p})(q(c) -$

$\tilde{q})h_d(c)dc$,[59] and the RHS of Equation 24, $\Delta\left(\frac{\mathrm{Cov}(p,q)}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc}\right)$. Using the definition of the covariance above, this reduces to

$$\Delta\left(\frac{\int_0^{c_d} p(q(c))q(c)h_d(c)dc}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc} - 1\right) \tag{40}$$

When preferences are CES, $p(c) = \frac{1}{1-\mu}c$ with μ constant, and $q(c) = c^{-\sigma}\left(\frac{1}{1-\mu}\right)^{-\sigma}\left(\frac{R}{\tilde{P}}\right)$ with $\tilde{P}$ the aggregate "ideal" price index and R the aggregate revenue. Additionally, $h_d(c)dc = \frac{g(c)}{G(c_d)} = \theta c^{\theta-1}c_d^{-\theta}$. Thus I can input all this information into Equation 40 and reduce the numerator and denominator separately:

$$\begin{aligned}
\int_0^{c_d} p(q(c))q(c)h_d(c)dc &= \left(\frac{R}{\tilde{P}}\right)\frac{1}{1-\mu}\left(\frac{1}{1-\mu}\right)^{-\sigma}\int_0^{c_d} cc^{-\sigma}\theta c^{\theta-1}c_d^{-\theta}dc && (41)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{1-\sigma}\theta c_d^{-\theta}\int_0^{c_d} c^{\theta-\sigma}dc && (42)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{1-\sigma}\theta c_d^{-\theta}\left(\frac{1}{\theta-\sigma+1}\right)c^{\theta-\sigma+1}\Big|_0^{c_d} && (43)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{1-\sigma}\left(\frac{\theta}{\theta-\sigma+1}\right)c_d^{-\theta}c_d^{\theta-\sigma+1} && (44)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{1-\sigma}\left(\frac{\theta}{\theta-\sigma+1}\right)c_d^{1-\sigma} && (45)\\
\int_0^{c_d} p(q(c))h_d(c)dc &= \frac{1}{1-\mu}\int_0^{c_d} c\theta c^{\theta-1}c_d^{-\theta}dc && (46)\\
&= \frac{1}{1-\mu}\theta c_d^{-\theta}\left(\frac{1}{\theta+1}\right)c^{\theta+1}\Big|_0^{c_d} && (47)\\
&= \frac{1}{1-\mu}\frac{\theta}{\theta+1}c_d && (48)\\
\int_0^{c_d} q(c)h_d(c)dc &= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{-\sigma}\int_0^{c_d} c^{-\sigma}\theta c^{\theta-1}c_d^{-\theta} && (49)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{-\sigma}\theta c_d^{-\theta}\int_0^{c_d} c^{\theta-\sigma-1}dc && (50)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{-\sigma}\theta c_d^{-\theta}\left(\frac{1}{\theta-\sigma}\right)c^{\theta-\sigma}\Big|_0^{c_d} && (51)\\
&= \left(\frac{R}{\tilde{P}}\right)\left(\frac{1}{1-\mu}\right)^{-\sigma}\left(\frac{\theta}{\theta-\sigma}\right)c_d^{-\sigma} && (52)
\end{aligned}$$

[59] Notice this also relies on productivity being unbounded above. This matters: see Feenstra (2014).

Next, combining the three above terms into Equation 40:

$$\Delta\left(\frac{\int_0^{c_d} p(q(c))q(c)h_d(c)dc}{\int_0^{c_d} p(q(c))h_d(c)dc \int_0^{c_d} q(c)h_d(c)dc} - 1\right) = \Delta\left(\frac{(\theta+1)(\theta-\sigma)}{\theta(\theta-\sigma+1)}\right) \tag{53}$$

where the term inside the parenthesis on the RHS is constant. Therefore, under the case of CES sub utility and Pareto $G(c)$, the terms in Equation 24 are zero.

B Growth in Real Income, Quantities, Productivities and Markups

Equation 25 uses the aggregate price-quantity covariance because this is what will be picked up by the difference between real income and physical production growth. However the decomposition can be expanded further. To do so, I go back to revenue and decompose prices further to bring in markups, and then get the expression for real income (again using $P = \int_0^{c_d} p(q(c))f(c)dc$ and $\frac{p}{c} = (\frac{1}{1-\mu(c)})$:

$$R = LM\left[\int_0^{c_d} \frac{p(q(c))}{c} cq(c)f(c)dc\right] \tag{54}$$

$$= LM\left[\int_0^{c_d} \frac{p(q(c))}{c} f(c)dc \int_0^{c_d} cq(c)f(c)dc\right] + LM\left[\text{Cov}(\frac{p}{c}, cq)\right] \tag{55}$$

$$= LM\int_0^{c_d} cq(c)f(c)dc\left[\int_0^{c_d} p(q(c))f(c)dc \int_0^{c_d} \frac{1}{c}f(c)dc + \text{Cov}(p, \frac{1}{c})\right] + LM\left[\text{Cov}(\frac{p}{c}, cq)\right] \tag{56}$$

$$\frac{R}{P} = LM\int_0^{c_d} cq(c)f(c)dc \int_0^{c_d} \frac{1}{c}f(c)dc + \frac{LM}{P}\int_0^{c_d} cq(c)f(c)dc\left[\text{Cov}(p, \frac{1}{c})\right] + \frac{LM}{P}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq)\right] \tag{57}$$

I can separate out aggregate quantity from the first term on the right hand side. Since $Q = LM\int_0^{c_d} q(c)f(c)dc$, then $LM\int_0^{c_d} cq(c)f(c)dc \int_0^{c_d} \frac{1}{c}f(c)dc = Q - LM\text{Cov}(\frac{1}{c}, cq)$. I substitute this into the last equation and then once again come up with an equation for $\frac{W}{Q}$:

$$\frac{R}{P} = Q - LM\left[\text{Cov}(\frac{1}{c}, cq)\right] + \frac{LM}{P}\int_0^{c_d} cq(c)f(c)dc\left[\text{Cov}(p, \frac{1}{c})\right] + \frac{LM}{P}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq)\right] \tag{58}$$

$$\frac{W}{Q} = 1 + \frac{LM}{PQ}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq) + \int_0^{c_d} cq(c)f(c)dc\left[\text{Cov}(p, \frac{1}{c})\right]\right] - \frac{LM}{Q}\left[\text{Cov}(cq, \frac{1}{c})\right] \tag{59}$$

$$\Delta ln\left(\frac{W}{Q}\right) = \Delta\left[\frac{LM}{PQ}\left[\text{Cov}(\frac{1}{1-\mu(c)}, cq) + \int_0^{c_d} cq(c)f(c)dc\left[\text{Cov}(p, \frac{1}{c})\right]\right] - \frac{LM}{Q}\left[\text{Cov}(cq, \frac{1}{c})\right]\right] \tag{60}$$

The last line is reported in the main text.

C Super/Sub Modularity

Start with $m(\lambda, \tau, \varphi) = \frac{p(a\tau/\varphi)}{a\tau/\varphi}$. From the notation in the main text, the actual production cost for a firm is $c = a\tau/\varphi$ though I keep $a = 1$ here. then: $\frac{\partial m(\lambda,\tau,\varphi)}{\partial \tau} = -\frac{p\varphi}{\tau^2} < 0$.

I now turn to the reallocation effects. I borrow some notation from Mrazova and Neary (2013b). Notably, I go back to indexing firms by their marginal cost, c ($1/\varphi$), in order to write $m(\lambda, \tau, \varphi) = \frac{p}{\tau c}$. This is to be consistent with the way they differentiate across firms by the marginal cost, and use the following notation and results:

1. $p(q(c)) = p$
2. $\epsilon = -\frac{p}{q(c)p'}$ (elasticity of demand)
3. $\rho = -\frac{(q(c))p''}{p'}$ (convexity of demand)
4. $\frac{dq}{dc}c = \frac{\tau c}{2p' + q(c)p''}$ (from the curvature of the marginal revenue curve)
5. $2p' + q(c)p'' = -\frac{p}{q(c)\epsilon^2}\left(\epsilon - 1 - q(c)\epsilon_q\right) > 0$. This is positive only when demand function is "log-concave," because in this case ϵ_q ($\frac{\partial \epsilon(q(c))}{\partial q(c)}$) is negative (elasticity of demand decreases with sales).
6. $\frac{d\epsilon}{dc} = \epsilon_q \frac{dq}{dc}$ (this is $\eta'(c)$ when $\eta(c)$ is the demand elasticity as in the text).
7. $\tau c = \frac{\epsilon - 1}{\epsilon} p$

I measure reallocation by looking at how the changes in τ leads to differential effects depending on firm marginal cost (which determines sales).

$$\frac{\partial m(\lambda,\tau,\varphi)^2}{\partial\tau\partial c} = -\frac{1}{\tau^2 c^2}\left(p'\frac{dx}{dc}c - p\right) \tag{61}$$

$$= \frac{1}{\tau^2 c^2}\left(p - p'\left(\frac{\tau c}{2p' + q(c)p''}\right)\right) = \frac{1}{\tau^2 c^2}\left(p - p'\left(\frac{\frac{\epsilon-1}{\epsilon}p}{-\frac{p}{q(c)\epsilon^2}(\epsilon - 1 - q(c)\epsilon_q)}\right)\right) \tag{62}$$

$$= \frac{1}{\tau^2 c^2}\left(p - p'\left(\frac{-\epsilon(\epsilon-1)q(c)}{\epsilon - 1 - q(c)\epsilon_q}\right)\right) = \frac{1}{\tau^2 c^2}\left(p + \frac{p'q(c)\epsilon(\epsilon-1)}{(\epsilon - 1 - q(c)\epsilon_q)}\right) \tag{63}$$

$$= \frac{1}{\tau^2 c^2}\frac{p(\epsilon - 1 - q(c)\epsilon_q) + p'q(c)\epsilon(\epsilon-1)}{\epsilon - 1 - q(c)\epsilon_q} = \frac{1}{\tau^2 c^2}\frac{p(\epsilon - 1 - q(c)\epsilon_q) - p(\epsilon-1)}{\epsilon - 1 - q(c)\epsilon_q} \tag{64}$$

$$= \frac{1}{\tau^2 c^2}\frac{-pq(c)\epsilon_q}{\epsilon - 1 - q(c)\epsilon_q} > 0 \tag{65}$$

In the last line, it is ϵ_q ($\epsilon'(q(c))$) that is dependent on the curvature of demand. If $u(q(c))$ is CES, then $\epsilon'(q(c)) = 0$. With elasticity of demand decreasing with sales, $\epsilon'(q(c)) < 0$ and we get the final result that $\frac{\partial m(\lambda,\tau,\varphi)^2}{\partial\tau\partial c} > 0$.

In the globalization case, there is a shift in demand elasticities that is due to $\frac{\partial\lambda}{\partial L} > 0$. This affects the price decision since we know that prices can be written with the firm demand elasticity governing the markup: $p(\lambda', c(\tau,\varphi)) = \frac{\lambda'\eta(c)}{\lambda'\eta(c)-1}c(\tau,\varphi)$ where $\eta(c)$ is the demand elasticity faced by a firm with marginal cost c, λ' is an aggregate that shifts demand elasticities for all firms with $\lambda' > 1$ when L or M_e increase, and $c = \frac{a\tau}{\varphi}$ is constant for each firm since there is no shock in τ. I follow the same steps as for τ above but change notation to denote demand elasticities ($\eta(c)$ in the text) again with ϵ as in Mrazova and Neary (2013b):

$$\frac{\partial p(\lambda', c(\tau,\varphi))}{\partial\lambda'} = \frac{-\epsilon}{(\lambda'\epsilon - 1)^2} < 0 \tag{66}$$

Then for the reallocation effects I measure how this shock to the demand elasticity affects firms with different marginal costs:

$$\frac{\partial p(\lambda, c(\tau,\varphi))^2}{\partial \lambda' \partial c} = \frac{-\frac{d\epsilon}{dc}(\lambda'\epsilon - 1)^2 + 2\epsilon(\lambda'\epsilon - 1)\lambda'\frac{d\epsilon}{dc}}{(\lambda'\epsilon - 1)^4} \tag{67}$$

$$= \frac{-\epsilon_q\big(-\frac{p}{q(c)\epsilon^2}(\epsilon - 1 - q(c)\epsilon_q)(\lambda'\epsilon - 1)\big) + 2\epsilon\lambda'\epsilon_q\big(-\frac{p}{q(c)\epsilon^2}(\epsilon - 1 - q(c)\epsilon_q)\big)}{(\lambda'\epsilon - 1)^3} \tag{68}$$

$$= \frac{1}{(\lambda'\epsilon - 1)^3}\left[\frac{\epsilon_q p}{q(c)\epsilon^2}(\epsilon - 1 - q(c)\epsilon_q)(3\lambda'\epsilon - 1)\right] < 0 \tag{69}$$

when, again, $\epsilon_q < 0$. Therefore prices are sub-modular in λ' and c by the definition in the text, which leads to the conclusion that a shock with $\lambda' > 1$ leads to all firms lowering prices and lower cost firms decreasing them relatively more than high-cost firms. Therefore inputs are reallocated to the more productive firms as they must produce relatively more to move down their demand curve.

D Comparison to ACDR

The allocative efficiency distortion that I measure in this paper can be linked to the distortion that separates non-homothetic demand with the homothetic Translog case in ACDR. They measure the change in total expenditure necessary to keep a constant utility level in response to a shock in variable trade costs. The main welfare decomposition in that paper can be decomposed as follows, with welfare growth interpreted as the inverse of the following growth in expenditure ($d\ln e_j$):

$$d\ln e_j = \underbrace{(1-\rho)\sum_i \lambda_{ij} d\ln(w_i\tau_{ij}) +}_{\text{Total selection effect}} \quad \underbrace{\rho \sum_i \lambda_{ij} d\ln(w_i\tau_{ij})}_{\text{(Price+Variety effect of competition)}} \quad \underbrace{+\rho\frac{\beta - 1}{1-\beta+\theta}\sum_i \lambda_{ij} d\ln(w_i\tau_{ij})}_{\text{(Distortion)}} \tag{70}$$

ρ stands for the weighted average markup elasticity with respect to marginal cost (or one minus the price-cost pass through elasticity), β represents the difference between the total price elasticity and cross-price elasticity (equal to 0 in the VES because firms do not take into account their effect on aggregate prices), and θ is the parameter that governs the dispersion of productivity assuming the distribution is Pareto.

The first term is a selection effect that represents the higher quantity that can be produced because the selected firms have a higher average productivity. As is known from tax incidence in monopolistic competition, the welfare gains from higher quantity are scaled by the price pass-through. The second term is the "pro-competitive" price and variety effects measured by Feenstra and Weinstein (2010). They show that competition leads to an overall increase in variety and lower average costs when competition with free entry increases scale.

The last term comes into effect only when $\beta < 1$, or when preferences are non-homothetic. In this case there is a reallocation of demand shares towards low-markup firms in response to lower trade costs, increasing expenditure. In the VES case, $\beta = 0$ and the magnitude of the distortion effects in response to a trade cost shock depends on the weighted average of markup elasticities and degree of firm heterogeneity. This is the distortion present in my model which features variable markups and firm heterogeneity with non-homothetic demand. I can identify this distortion with aggregate data because it is the only part of aggregate real income that does not track aggregate quantity. I do not measure the local effect around a small reduction in trade costs, but the overall result of shocks to competition and cost. In response to lower production costs, the distortion I find is similar to ACDR in that there is a relative reallocation of production towards initially low-markup firms. However I also show that a competition shock actually lowers this distortion. In that case the reallocation that happens due to selection is reinforced by the fact that high-markup firms increase their production more than low-markup firms.

E Data and Variable Definitions

Here I describe my measure of the left hand side of Equation 21, which I label W. It is equivalent to the Aggregate Productivity Growth (APG) that is used in Petrin and Levinsohn (2012) and Basu and Fernald (2002), which tracks welfare without taking into account variety. In words, W is the sum of deflated value added, subtracting out the growth in inputs. $\Delta ln(W_t) = \Delta ln(Y_t) - \Delta ln(L_t)$, where Y_t (sum of deflated value added) is real revenue if all production income goes towards final demand. $\Delta ln(L_t)$ corrects for changes in expenditure on labor.

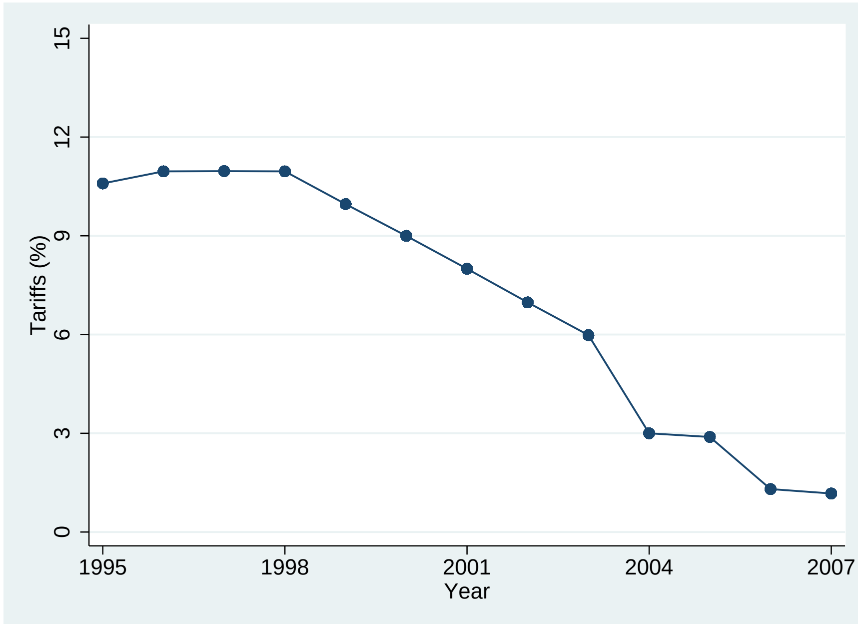
I next explain the measurement of Y_t, which measures "Final Demand." At the firm (i) level, $Y_i = Q_i - \sum_j M_{ji}$, where M_{ji} are inputs sourced from some firm, j. By the National Accounting Identity, aggregate final demand is equal to aggregate value added: $\sum_i P_i Y_i = \sum_i VA_i = \sum_i P_i Q_i - \sum_i \sum_j P_{ij} M_{ji}$.

F Trade Agreements

Below is a list of all the trade agreements signed by Chile:

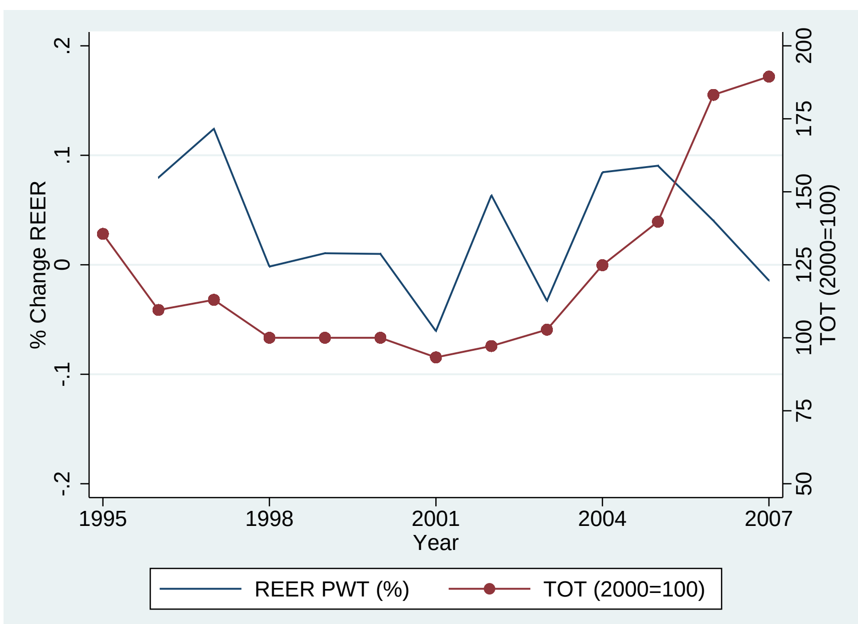
- $1990's$: Trade agreements with Canada (1996), Mexico(1998), and Central America.
- 1996: Association agreement with the Mercosur countries
- 2002: Agreements with the European Union and South Korea
- Free Trade Agreement (FTA) with the United States starting 2004. Completely free bilateral trade does not begin until 2016, but tariffs decreased immediately.
- In 2003 Chile unilaterally lowered its across-the-board import tariff to 6% for all countries with which it does not have a trade agreement.
- FTA with China signed in late 2005.

Figure 1: Average Applied Tariffs 1995-2007



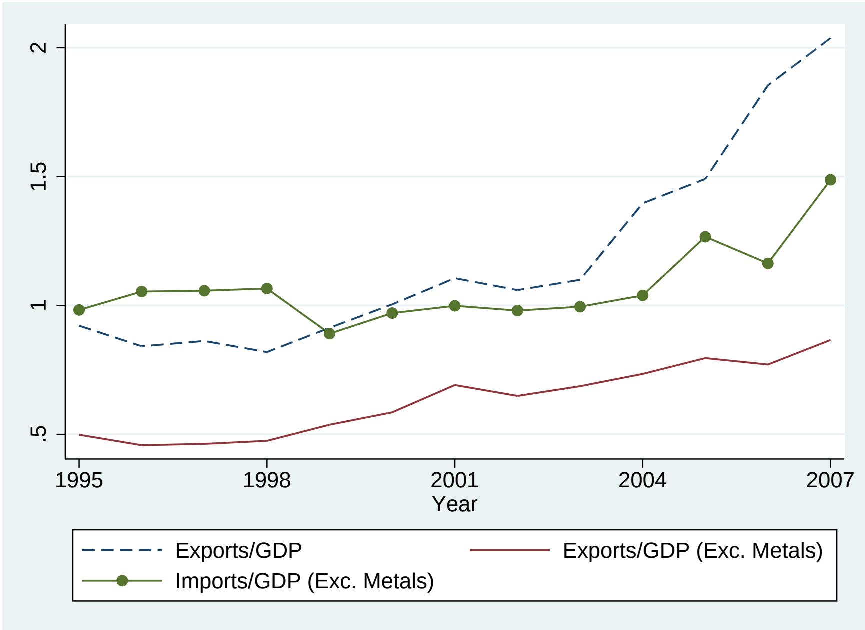
Source: Comtrade Database, downloaded from World Integrated Trade Solution (WITS). Bilateral tariffs are aggregated to 4-digit level using an unweighted average of 6-digit tariff lines, and then weighted by trade shares to get an average applied tariff rate across all trade partners.

Figure 2: Terms of Trade (2000=100) and Real Effective Exchange Rate (% change), 1995-2007



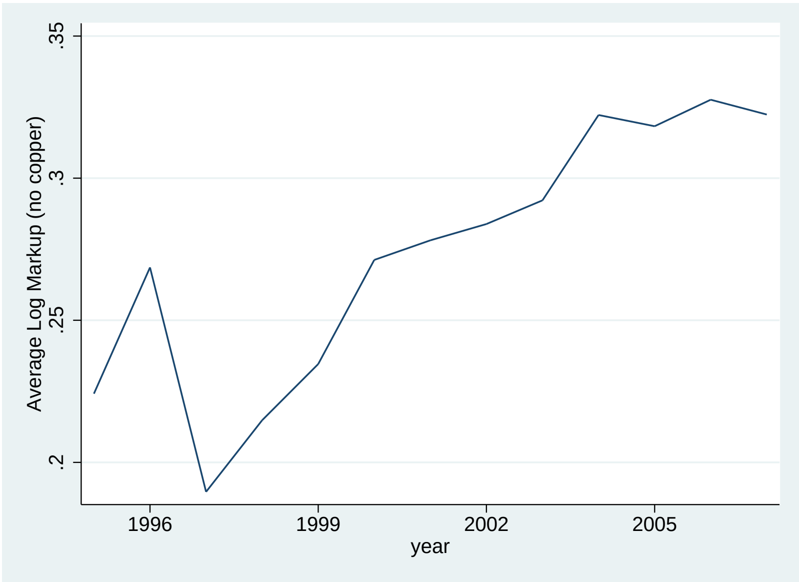
Sources: WDI Indicators, PWT 8.0, IFS Database. TOT is an index from WDI. REER IFS is an index from the IMF. I calculate REER PWT using Penn World Tables to calculate Chile's production price index relative to its top trade partners and take a geometric average using trade shares as weights.

Figure 3: Exports and Imports as a share of GDP, 1995-2007



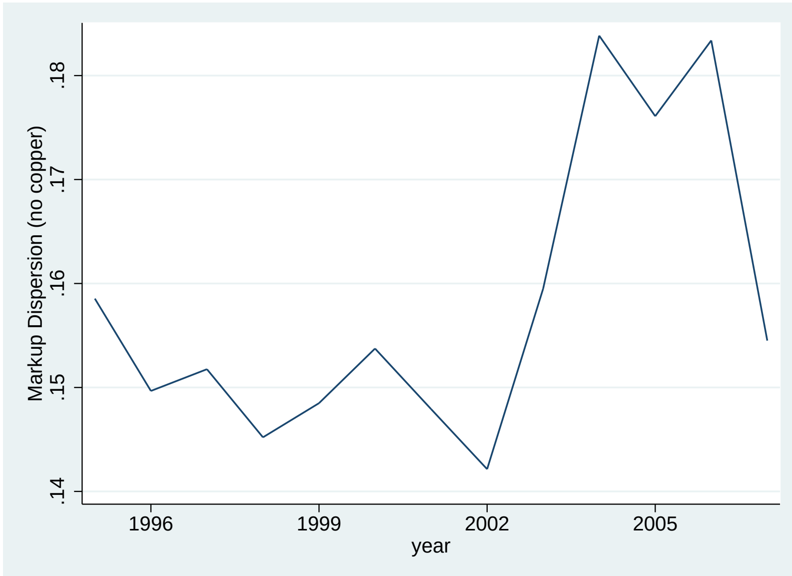
Sources: Trade data from Feenstra et al. (2005), and manufacturing GDP from Banco Central de Chile. Manufacturing GDP and manufacturing exports/imports are both in thousands of current US dollars

Figure 4: Average Market Markup



Mean calculated for each sector assuming a log normal distribution. Economy-wide average taken by weighting each sector by its value added share. I eliminate firms in the bottom and top 1% of the markup distribution.

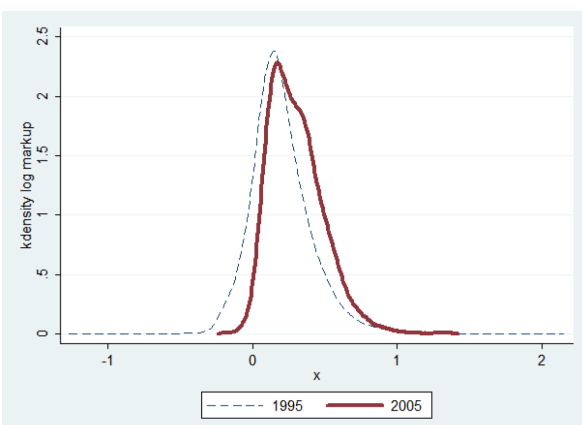
Figure 5: Markup Dispersion: Average across sectors



Markup dispersion calculated for each sector by estimating the shape parameter of a log-normal distribution using maximum likelihood. I take the economy-wide average by weighting each sector by its value added share. I eliminate firms in the bottom and top 1% of the markup distribution.

Figure 6: Markup Distribution versus TFP Distribution

(a) Markup Distribution: 1995 versus 2005



(b) TFP Distribution: 1995 versus 2005

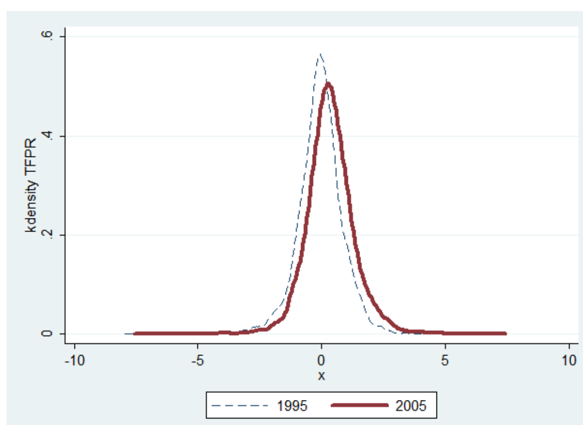
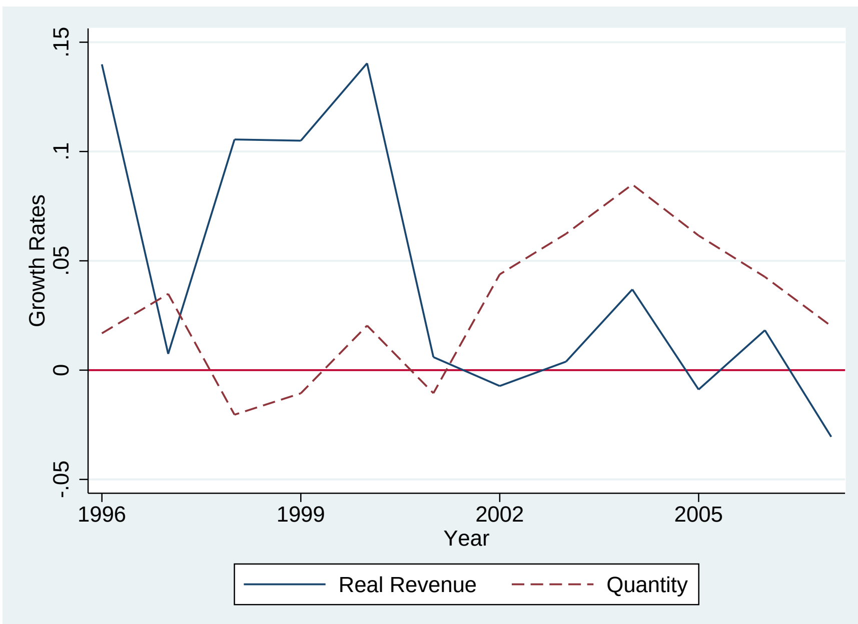


Figure 7: Real Income Growth versus Physical Production



W is the growth in the sum of deflated value added (minus primary input growth) at the 3-digit ISIC level. Economy-wide average taken by weighting each 3-digit group by its value added share. Quantity growth is taken from the physical manufacturing index provided by the ENIA at the 3-digit ISIC level with same weighting scheme.

Table 1: Factor Coefficients and Markups by 2-digit ISIC Sectors

	(1)					
	Obs	θ_L	θ_K	θ_M	Ret Scale	Median Markup
15	19475	0.218	0.073	0.757	1.048	1.192
17	3462	0.336	0.083	0.666	1.085	1.206
18	3846	0.349	0.047	0.665	1.062	1.219
19	2095	0.433	0.054	0.657	1.145	1.034
20	4382	0.240	0.051	0.773	1.064	1.264
21	1803	0.187	0.089	0.745	1.020	1.358
22	3017	0.285	0.111	0.633	1.029	1.323
24	3740	0.283	0.105	0.667	1.055	1.360
25	4085	0.221	0.072	0.734	1.027	1.352
26	2837	0.191	0.064	0.802	1.057	1.540
27	1503	0.128	0.139	0.747	1.015	1.412
28	4760	0.243	0.059	0.675	0.977	1.189
29	2923	0.508	0.098	0.489	1.095	0.993
31	1199	0.246	0.074	0.682	1.002	1.260
33	365	0.178	0.046	0.778	1.002	1.774
34	752	0.490	0.091	0.656	1.237	1.529
35	595	0.338	0.074	0.603	1.016	1.119
36	3229	0.180	0.033	0.812	1.025	1.544

Production function coefficients and median markups calculated using the methods of Ackerberg et al. (2006) and DeLoecker and Warzynski (2012) as described in the text.

Table 2: Firm Level: Differential Effect on Markups by Importer/Exporter

	Mark-up			
	(1)	(2)	(3)	(4)
TOT	-0.013			
	(0.020)			
Output Tariff	-0.024***		-0.010	-0.039
	(0.004)		(0.021)	(0.027)
TOT*IMP*Exp=0		0.023	0.039*	0.036*
		(0.026)	(0.020)	(0.019)
TOT*EXP*Imp=0			0.006	0.007
			(0.020)	(0.020)
OutputTariff*IMP*EXP=0			0.002	0.002
			(0.005)	(0.004)
OutputTariff*EXP*IMP=0			-0.000	0.000
			(0.005)	(0.005)
TFP				0.029***
				(0.005)
Fixed Effects	Sector,Firm	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm
R^2	0.173	0.176	0.197	0.218
N	49191	61490	49191	47634

Dependent variable is log markup measured using the procedure outlined in DeLoecker and Warzynski (2012). Terms of trade and output tariffs also in logs. Imp*Exp=0 signifies importers who do not export (and vice-versa for Exp*Imp=0). The following controls are used but omitted in the table output: Herfindahl Index at 4-digit industry level, a dummy if the firm is a multinational, plus fixed effects. Standard errors are clustered at the 4-digit industry level. I drop the basic metal industry (ISIC 27).

Table 3: Firm Level: Differential Effect on Markup by Degree of Exposure to Competition

	Mark-up		TFP	
	(1)	(2)	(3)	(4)
TOT*Net Exposure	-0.084**		-0.137	
	(0.033)		(0.202)	
OutputTariff*Net Exposure	-0.024		-0.021	
	(0.034)		(0.147)	
TOT*Imported Share		0.093		0.123
		(0.065)		(0.238)
TOT*Exported Share		-0.025		-0.911***
		(0.053)		(0.253)
OutputTariff*Imported Share		0.028		0.021
		(0.028)		(0.111)
OutputTariff*Exported Share		0.014		-0.315***
		(0.036)		(0.089)
Fixed Effects	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm
R^2	0.199	0.204	0.046	0.048
N	49181	49182	48224	48225

Dependent variables are log revenue TFP and log markup. Terms of trade and output tariffs also in logs. Net Exposure is defined as (Export Sales/Total Sales)-(Imported Inputs/Total material input costs). The prior two components are "Exported Share" and "Imported Share." The following controls are used but omitted in the table output: Herfindahl Index at 4-digit industry level, fraction of industry final demand that is imported, a dummy if the firm is a multinational, plus fixed effects. Standard errors are clustered at the 4-digit industry level. I drop the basic metal industry (ISIC 27).

Table 4: Firm Level: Input/Output Tariffs on Importer/Exporter

	Mark-up			TFP
	(1)	(2)	(3)	(4)
Input Tariff	-0.212*	-0.211*	-0.212*	-0.928**
	(0.110)	(0.113)	(0.113)	(0.394)
Output Tariff	0.101	0.099	0.099	0.135
	(0.097)	(0.096)	(0.104)	(0.398)
InputTariff*IMP*EXP=0		-0.000	-0.020	-0.177
		(0.017)	(0.070)	(0.323)
OutputTariff*EXP*IMP=0			-0.016	-0.033
			(0.072)	(0.284)
Fixed Effects	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm
R^2	0.164	0.162	0.164	0.040
N	47458	47458	47458	46733

Dependent variables are log revenue TFP and log markup. Input tariffs are constructed as in Amiti and Konings (2007): a weighted average of output tariffs, with the weights based on the cost shares of each input used in the industry at the 3-digit level. Input and output tariffs in logs. Imp*Exp=0 signifies importers who do not export (and vice-versa for Exp*Imp=0). The following controls are used but omitted in the table output: Herfindahl Index at 4-digit industry level, fraction of industry final demand that is imported, a dummy if the firm is a multinational, plus fixed effects. Standard errors are clustered at the 4-digit industry level. I drop the basic metal industry (ISIC 27).

Table 5: Firm Level: Input/Output Tariffs on Degree of Exposure to Competition

	Mark-up		TFP	
	(1)	(2)	(3)	(4)
InputTariff*Net Exposure	0.125		-0.059	
	(0.090)		(0.364)	
OutputTariff*Net Exposure	-0.121		0.049	
	(0.091)		(0.393)	
InputTariff*Imported Share		-0.166		-0.245
		(0.128)		(0.406)
OutputTariff*Imported Share		0.202		0.279
		(0.163)		(0.509)
InputTariff*Exported Share		0.221		-0.181
		(0.234)		(0.498)
OutputTariff*Exported Share		-0.172		0.168
		(0.211)		(0.489)
Fixed Effects	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm	Year,Sector,Firm
R^2	0.165	0.172	0.042	0.114
N	47458	47458	46546	44871

Dependent variables are log revenue TFP and log markup. Input tariffs are constructed as in Amiti and Konings (2007): a weighted average of output tariffs, with the weights based on the cost shares of each input used in the industry at the 2-digit level. Input and output tariffs in logs. Net Exposure is defined as (Export Sales/Total Sales)-(Imported Inputs/Total material input costs). The prior two components are "Exported Share" and "Imported Share." The following controls are used but omitted in the table output: Herfindahl Index at 4-digit industry level, fraction of industry final demand that is imported, a dummy if the firm is a multinational, plus fixed effects. Standard errors are clustered at the 4-digit industry level. I drop the basic metal industry (ISIC 27).

Table 6: Firm-level Markups: Distributional Effects using Initial Markup

	Top 30%		Base Year Markup	
	(Markup)	(Markup)	(Markup)	(Markup)
TOT*Net Exposure	-0.015		0.030	
	(0.015)		(0.041)	
TOT*Exposure*Top 30%	-0.065*			
	(0.038)			
OutputTariff*Net Exposure		0.009		0.002
		(0.008)		(0.009)
OutputTariff*Exposure*Top 30%		0.001		
		(0.004)		
TOT*Exposure*Base Markup			-0.172	
			(0.123)	
OutputTariff*Exposure*Base Markup				0.002
				(0.007)
Fixed Effects	Year,Firm	Year,Firm	Year,Firm	Year,Firm
R^2	0.153	0.170	0.207	0.231
N	55865	45782	41883	36142

Dependent variable is log markup. Net Exposure is defined as (Export Sales/Total Sales)-(Imported Inputs/Total material input costs). The Top 30% dummy is constructed using firms who are in the data in 1995 and/or 2002. I rank firms in these two years and set the dummy equal to 1 if a firm is in the top 30% of markups in the base year. Standard errors are clustered at the 4-digit industry level. I drop the basic metal industry (ISIC 27).

Table 7: Industry Level: Change in Aggregate Outcomes by Share of Importing Firms (using only incumbent firms)

	Δ AE		Δ Q	Δ Cov(markup,inputs)
	(1)	(2)	(3)	(4)
D.ln(TOT)	-0.109	-0.707**		
	(0.194)	(0.289)		
D.Output Tariff	-0.008	-0.048***	-0.038	-0.019
	(0.015)	(0.016)	(0.039)	(0.044)
Δ TOT*Importer (Industry Share)	**-6.672****		0.976	**-4.173***
	(2.865)		(1.014)	(2.027)
Δ TOT*Openness	**0.209**		0.080	-0.068
	(0.162)		(0.047)	(0.173)
Δ OutputTariff*Openness	0.007		0.024	-0.005
	(0.013)		(0.022)	(0.013)
Δ TOT*Openness=1		**0.015**		
		(0.339)		
Δ OutputTariff*Openness=1		-0.019		
		(0.021)		
Avg Outcome	0.016	0.016	0.029	0.006
Fixed Effects	Year,Sector	Sector	Year,Sector	Year,Sector
R^2	0.237	0.191	0.364	0.083
N	192	192	192	204

Dependent variables are ΔAE, $\Delta TFPR$, ΔQ, and Δ Cov(markup,inputs). These are all at the 2-digit ISIC level. The first three are one year growth rates with their definitions in the text. For the covariance I use first differences. $\Delta TFPR$ is the growth rate of aggregate revenue productivity as defined by Olley and Pakes (1996). Input tariffs are as above. ΔTOT, $\Delta OutputTariff$ and $\Delta InputTariff$ are all one year growth rates. I use the fraction of firms in an industry where (Imp*Exp=0)=1 as "Importer (Industry Share)". "Openness" is the sum of exports and imports of final goods into an industry divided by total industry sales. Standard errors are clustered at the 2-digit industry level. I drop the basic metal industry (ISIC 27).

Table 8: Industry Level: Change in Aggregate Outcomes by Average Industry Exposure (using only incumbent firms)

	Δ AE		Δ Q	Δ Cov(markup,inputs)	
	(1)	(2)	(3)	(4)	(5)
D.ln(TOT)	-0.402*				
	(0.212)				
D.Output Tariff	-0.051***	0.001	-0.005	0.003	-0.061
	(0.015)	(0.087)	(0.027)	(0.041)	(0.095)
Δ TOT*Net Exposure	**6.038***		-1.111	**2.496**	
	(2.295)		(0.806)	(1.586)	
Δ OutputTariff*Net Exposure	**-1.959***		-0.281***	-0.075	
	(0.922)		(0.090)	(0.134)	
Δ TOT*Imported Share		**-6.648****			-4.921*
		(3.042)			(2.754)
Δ TOT*Exported Share		1.430			-0.372
		(6.432)			(6.254)
Δ OutputTariff*Exported Share		0.375			-0.004
		(0.463)			(1.198)
Avg Outcome	0.016	0.016	0.029	0.006	0.006
Fixed Effects	Year,Sector	Year,Sector	Year,Sector	Year,Sector	Year,Sector
R^2	0.287	0.339	0.381	0.083	0.104
N	192	192	192	204	204

Dependent variables are ΔAE, $\Delta TFPR$, ΔQ, and Δ Cov(markup,inputs). These are all at the 2-digit ISIC level. The first three are one year growth rates with their definitions in the text. For the covariance I use first differences. $\Delta TFPR$ is the growth rate of aggregate revenue productivity as defined by Olley and Pakes (1996). Input tariffs are as above. ΔTOT, $\Delta OutputTariff$ and $\Delta InputTariff$ are all one year growth rates. Net Exposure is defined as (Export Sales/Total Sales)-(Imported Inputs/Total material input costs). The prior two components are "Exported Share" and "Imported Share." Standard errors are clustered at the 2-digit industry level. I drop the basic metal industry (ISIC 27).

Table 9: Industry Level: Allocative Efficiency and Covariances

	Δ TFPR	Δ Q	Δ AE	
	(1)	(2)	(3)	(4)
D.Cov(tfp, mkt share)	48.778***			
	(11.866)			
D.Cov(markup,total inputs)	-0.004			
	(0.019)			
D.Cov(tfp,mkt share)		-0.176		10.626
		(0.443)		(6.996)
D.Cov(markup,total inputs)		-0.019***	0.036*	0.034***
		(0.005)	(0.018)	(0.012)
D.Cov(tfp,total inputs)				0.160*
				(0.089)
Fixed Effects	Year,Sector	Year,Sector	Year,Sector	Year,Sector
R^2	0.540	0.315	0.100	0.169
N	564	396	417	393

Dependent variables are ΔAE, $\Delta TFPR$, and ΔQ (all one year growth rates). These are all at the 2-digit ISIC level. $\Delta TFPR$ is the growth rate of aggregate revenue productivity as defined by Olley and Pakes (1996). The independent variables are first differences. Standard errors are clustered at the 2-digit industry level. I drop the basic metal industry (ISIC 27).

Table 10: Industry Level: Allocative Efficiency and Markup Distribution

	Δ AE		Cov(markup,Δ inputs)
	(1)	(2)	(3)
D.Markup dispersion	-0.632**	-0.060	-0.078*
	(0.280)	(0.219)	(0.043)
D.Markup Average	1.179*	-0.342	-0.058
	(0.678)	(0.579)	(0.082)
D.TFPR dispersion		-0.125	
		(0.232)	
D.TFPR Average		0.792***	
		(0.216)	
Fixed Effects	Year,Sector	Year,Sector	Year,Sector
R^2	0.141	0.318	0.136
N	417	417	543

The independent variables are first differences. Standard errors are clustered at the 2-digit industry level. I drop the basic metal industry (ISIC 27).

Table 11: Industry Level: Markup Distribution for Importers and Exporters

	Markup dispersion		TFP Dispersion	Mean Markup
	(1)	(2)	(3)	(4)
Imported Share	0.321*		0.196	0.399*
	(0.187)		(0.734)	(0.228)
Openness	-0.026*	-0.027	-0.102**	-0.028
	(0.015)	(0.017)	(0.042)	(0.029)
HHI	0.196	0.249	0.092	0.011
	(0.138)	(0.162)	(0.466)	(0.173)
Importer Dummy		0.128		
		(0.161)		
Fixed Effects	Year	Year	Year	Year
R^2	0.285	0.222	0.134	0.239
N	586	586	580	586

Dependent variables are markup dispersion, revenue TFP dispersion and average markup. HHI refers to the Herfindahl Index. Standard errors are clustered at the 2-digit industry level. I drop the basic metal industry (ISIC 27).