Trade, Markup Heterogeneity and Misallocations*

Paolo Epifani†
Bocconi University and KITeS

Gino Gancia‡
CREI, UPF and CEPR

August 2010

Abstract

Markups vary widely across industries and countries, their heterogeneity has increased overtime and asymmetric exposure to international trade seems partly responsible for this phenomenon. In this paper, we study how the entire distribution of markups affects resource misallocation and welfare in a general equilibrium framework encompassing a large class of models with imperfect competition. We then identify conditions under which trade opening, by changing the distribution of markups, may reduce welfare. Our approach is novel both in its generality and in the emphasis on the second moment of the markup distribution. Two broad policy recommendations stand out from the analysis. First, whenever there is heterogeneity in markups, be it due to trade or other distortions, there is also an intersectoral misallocation, so that the equilibrium can be improved upon with an appropriate intervention. This suggests that trade liberalization and domestic industrial policy are complementary. Second, ensuring free entry is a crucial precondition to prevent adverse effects from asymmetric trade opening.

JEL Numbers: F12, F15.

Keywords: Markups, Dispersion of Market Power, Procompetitive Effect, Trade and Welfare.

*We would like to thank Giancarlo Corsetti, Jonathan Eaton, Nezih Guner, Winfried Koeniger, Omar Licandro, Enrico Sette, Guido Tabellini, Jaume Ventura, Fabrizio Zilibotti and an anonymous referee for helpful comments and discussion. We also thank participants at the EEA Annual Congress (Barcelona, 2009), ITSG Meeting (Cagliari, 2009), RIEF Meeting (Aix-an-Provence, 2009), ESSIM (Tarragona, 2008), XXXII Symposium of Economic Analysis (Granada, 2007), Universidad Carlos III de Madrid, Banco de España, the UPF Faculty Lunch, the Macro Workshop at IEW, Zurich University and the IADB-ELSNIT (Paris, 2006). All errors are our own. Gino Gancia acknowledges financial support from the Barcelona GSE, the Government of Catalonia (2009SGR1157) and the ERC Grant GOPG 240989. Part of this research was done while Gino Gancia was visiting as Research Associate the Chair of Political Economy and Macroeconomics of the IEW-University of Zurich. We thank the IEW for the hospitality.

†Department of Economics and KITeS, Università Bocconi, via Sarfatti 25, 20136, Milano (Italy). E-mail: paolo.epifani@unibocconi.it

‡CREI, Universitat Pompeu Fabra and CEPR, Ramon Trias Fargas 25-27, 08005, Barcelona (Spain). E-mail: ggancia@crei.cat

"What is relevant for the general analysis is not the *sum* of individual degrees of monopoly but their *deviations*"

A.P. Lerner

1 Introduction

Monopoly power varies widely across industries and firms. Data on 4-digit US manufacturing industries show that Price-Cost Margins (PCMs), a common measure of markups, range from 1% in the first percentile of the distribution, to 60% in the 99th percentile. Cross-country evidence suggests these asymmetries to be even larger in less developed economies. Moreover, monopoly power varies systematically with exposure to international competition. For instance, the average PCM is a slim 13% in US manufacturing, producing goods that are typically traded, versus a fat 33% in nontraded business sector services. The conventional wisdom is that, as the process of globalization continues, competition among firms participating in international markets will intensify, thereby alleviating the distortions associated with monopoly power. The latter presumption is not however granted, because the exact mapping between the economy-wide distribution of markups and the extent of misallocations is not an obvious one.

The textbook partial-equilibrium analysis of the deadweight loss from monopoly seems to imply that market power, by raising prices above marginal costs, is always distortionary. Yet, this reasoning neglects the fact that, as pointed out by Lerner (1934) and Samuelson (1947), in general equilibrium misallocations depend on *relative* rather than *absolute* prices. If all prices incorporate the same markup, Lerner noted that relative prices would signal relative costs correctly and, absent other sources of inefficiencies, would lead to the optimal allocation. This suggests that distortions only come from the *dispersion* of market power across firms and industries. The implications of this principle are far-reaching. For instance, it implies the seemingly paradoxical result that an *increase* in competition in industries with below-average markups, such as those producing tradeable goods, is deemed to *amplify* monopoly distortions. It also raises warnings that the increased heterogeneity in observed measures of market power across industries may indicate growing misallocations. Yet, when market power is coupled with free entry, so that markups affect the equilibrium number of firms, new welfare effects arise. For instance, when firms produce differentiated products and consumers like variety, it is desirable that markups be high enough to induce the socially optimal level of entry.

The aim of this paper is to study how the economy-wide distribution of markups distorts

the allocation of resources in a general framework that encompasses a large class of models of imperfect competition. In doing so, we revisit and qualify the Lerner condition that markups should be uniform across industries and illustrate when and how the level and dispersion of markups matter. Our second goal is to relate the degree of monopoly power in an industry to the presence of foreign competition and to study how trade can affect welfare by changing the dispersion of market power. In particular, we are interested in finding under what circumstances asymmetric trade liberalization may turn out to be welfare reducing.

To this end, we build a model with a continuum of industries that are heterogeneous in both costs and demand conditions. Firms may produce homogenous or differentiated goods and entry may or may not be restricted. By comparing the market equilibrium with the one chosen by a benevolent social planner, we identify the misallocations due to the entire markup distribution. The results crucially depend on the assumptions about entry. If entry is restricted, we confirm Lerner's principle that markup symmetry across industries and firms is a necessary and sufficient condition for efficiency. Whenever this condition is violated, there is an *intersectoral misallocation* whereby relatively less competitive industries underproduce and relatively more competitive ones overproduce relative to the socially optimal quantity. Perhaps surprisingly, we find the extent of this misallocation to be stronger when the elasticity of substitution between industries is high and we show that its welfare cost may be quantitatively significant. We also show that trade liberalization affecting only some industries may have adverse welfare effects when it raises markup heterogeneity. In other words, contrary to the conventional wisdom, procompetitive losses from trade are possible.

With free entry, the level of markups matters too and the Lerner condition about symmetry turns out to be necessary, but not sufficient, for efficiency. Moreover, contrary to the previous case, we show that in general there exists no markup distribution capable of replicating the first best allocation. This means that policy interventions aimed at controlling prices only are not enough to correct all the distortions and other instruments, such as subsidies, are needed. Moreover, we find that some heterogeneity in markups, despite the misallocation it induces, may well be welfare improving. Finally, we show that free entry makes procompetitive losses from trade unlikely, even when trade increases markup dispersion.

Two general policy recommendations stand out from our analysis. First, whenever there is heterogeneity in markups, be it due to trade, regulations or differential ability to collude across sectors, there is an intersectoral misallocation. Industries with above-average markups always underproduce (either in terms of output per firm or of product variety), so that the equilibrium can be improved upon with an appropriate intervention. This also suggests that

trade liberalization and domestic industrial policy are complementary. Second, free entry is an important condition to prevent asymmetric trade liberalization from having possibly adverse welfare effects. A novel implication of our findings is that the relative competitiveness of the industries affected by trade liberalization matters. This observation should be taken into account when designing trade policy. Interestingly, our analysis may also help rationalize the often heard concern that trade may be detrimental in countries (especially the less developed ones) where domestic markets are not competitive enough. The reason is not that domestic firms are unable to survive foreign competition (as emphasized by the infant-industry theory), but rather that international competition may inefficiently increase asymmetries across industries in the economy.

This paper makes contact with three strands of literature. The first studies monopoly distortions in general equilibrium and includes classics such as Lerner (1934), Samuleson (1947), Dixit and Stiglitz (1977), but also more recent works by Neary (2003), Koeniger and Licandro (2006), Bilbiie, Ghironi and Melitz (2006) and others. Although all these papers made important contributions, they all present a collection of special cases. Our approach is more general in modelling preferences, imperfect competition (with and without free entry) and sectorial asymmetries. We believe that such a unified framework is key to understanding how monopoly distortions interact with more specific modelling assumptions.

Second, this paper is related to the literature studying the welfare effects of trade in models with imperfect competition, including the works of Brander and Krugman (1983), Helpman and Krugman (1985) and more recently Eckel (2008). The observation that, in the presence of distortions, trade might have adverse welfare effects is an application of the second-best theory and goes back to Bhagwati (1971) and Johnson (1965).[1] Yet, what we find more interesting is the more specific insight, so far neglected, that trade can affect welfare by changing the cross-sectoral dispersion of market power.

Third, this paper relates to a recent literature on the macroeconomic effects of misallocations. Noteworthy contributions by Banerjee and Duflo (2005), Restuccia and Rogerson (2006), Hsieh and Klenow (2009), Song, Storesletten and Zilibotti (2009) and Jones (2009) provide striking evidence that wedges distorting the allocation of resources between firms and industries within a country are quantitatively very important in explaining low aggregate productivity, particularly in less developed economies. Yet, they leave to future research the task

[1] Bhagwati (1971) and Johnson (1965) where the first to argue that trade can lower welfare if it exacerbates an existing distortion. Later studies have examined sufficient conditions for positive gains from trade in the presence of various distortions. See, for example, Eaton and Panagariya (1979) and references therein.

of identifying the origin of such wedges. Our paper contributes to this line of investigation by studying how asymmetries in market power, that appear to be especially large in poor countries, may be one source of misallocations.[2]

The paper is organized as follows. Section 2 documents a number of little known stylized facts that motivate our analysis. Section 3 builds a general theoretical framework that encompasses the most popular models of imperfect competition. Section 4 studies the welfare effects of markup dispersion and trade when entry is restricted. Section 5 extends the analysis to the case of free entry. Section 6 concludes.

2 Motivating Evidence

In this section, we document a number of little known stylized facts that motivate our theoretical investigations: (1) markups vary widely across sectors and their dispersion has increased overtime; (2) asymmetric exposure to trade seems to be a likely explanation for the rise in markup heterogeneity; (3) markup asymmetries are systematically related to the level of economic development, with less asymmetries in wealthier countries.

Following a vast empirical literature (see, e.g., Roberts and Tybout, 1996; Tybout, 2003; Aghion et al., 2005), we use price-cost margins (PCMs) as a proxy for market power.[3] To compute them, we draw production data from the OECD STAN database[4], the CEPII 'Trade, Production and Bilateral Protection' database[5] and the NBER Productivity database by Bartelsman and Gray. The latter is the most comprehensive and highest quality database on industry-level inputs and outputs, covering roughly 450 US manufacturing (4-digit SIC) industries for the period 1958-1996. Price-cost margins are computed as the value of shipments (adjusted for inventory change) less the cost of labor, capital, materials and energy, divided by the value of shipments.[6] Capital expenditures are computed as $(r_t + \delta)K_{it-1}$, where K_{it-1} is

[2] Basu and Fernald (2002) stress the distinction between aggregate productivity and aggregate technology and recognize that technology improvements could reduce welfare if they lead sectors with smaller-than-average markups to increase input use. Yet, they do not pursue this idea, nor they relate it to trade liberalization. Gancia and Zilibotti (2009) shows how markup asymmetries may also distort the development of technology in dynamic models with innovation.

[3] An important advantage of PCMs is that they can vary both across industries and overtime. An alternative approach would be to estimate markups from a structural regression *a là* Hall (1988). One problem with this approach is that, to estimate markups across industries or over time, either the time or industry dimension is to be sacrificed, implying that markups have to be assumed constant overtime or across industries.

[4] This database allows to compute PCMs for broad aggregates of traded and nontraded industries for a sample of OECD countries.

[5] This dataset is based on information from the World Bank, the OECD and the UNIDO. It allows to compute PCMs across broad manufacturing aggregates for a sample of developed and developing countries.

[6] Due to data availability, we do not net out capital expenditures and inventory change when using the OECD and CEPII datasets.

the capital stock, r_t is the real interest rate and δ is the depreciation rate.[7] As a proxy for trade exposure at the industry level, we use the openness ratio, defined as the value imports plus exports, taken from the NBER Trade database by Feenstra, divided by the value of shipments.

2.1 Markup heterogeneity across sectors and overtime

We start by showing asymmetries in markups across broad sectorial aggregates. Using economy-wide data for the US in 2003 (from the OECD dataset), we find that the average PCM equals 33% in the business sector services (mostly nontraded industries), 28% in agriculture (a heavily protected industry) and 13% in manufacturing. As for services, the average PCM equals 24% in the transport and storage industry, 28% in post and telecommunication, 38% in finance and insurance, 48% in electricity, and reaches a peak of 66% in real estate activities. Interestingly, in the renting of machinery and equipment industry, selling nontraded services, the average PCM equals 41.5%, whereas in the machinery and equipment industry, producing traded manufacturing goods, the average PCM is 9.5%. These huge asymmetries in price-cost margins between traded and nontraded industries immediately suggest that markups may crucially depend on the degree of tradeabililty of an industry's output, and hence that asymmetric exposure to international competition may be an important determinant of markup heterogeneity across industries.

Perhaps surprisingly, exposure to international competition varies dramatically also among manufacturing industries. Figure 1 reports the time evolution of the openness ratio for selected 2-digit SIC industries within US manufacturing. Note that, at one extreme, the leather industry has increased its trade share from 4% in the late 50s to 230% in the mid 90s. Other industries, such as miscellaneous products or apparel, show a similar upward trend in the trade share. At the other extreme, however, there are industries, such as printing, fabricated metal products or food, whose openness ratio has increased by only a few percentage points over the past 40 years. More generally, when considering the entire distribution of the trade share across 4-digit manufacturing industries, we find that the openness ratio increased by only 6 percentage points in the first quintile of the distribution (from 1.2% in 1958 to 7.2% in 1994), and by more than 47 percentage points in the fourth quintile of the distribution (from 10.2% to 57.7%). These figures suggest that, by affecting some industries more than others, trade opening may have

[7] The US real interest rate, drawn from the World Bank-*World Development Indicators*, has a mean value of 3.75 percent (with a standard deviation of 2.5 percent) over the period of analysis. As for the depreciation rate, the values for δ used in the empirical studies generally vary from 5% for buildings to 10% for machinery. We choose a value of 7%, implying that capital expenditures equal, on average, roughly 10% of the capital stock.

increased asymmetries in market power.

INSERT FIGURE 1 HERE

Figures 2 and 3 provide suggestive evidence consistent with this conjecture. The former reports the evolution of the standard deviation of PCMs across 450 US manufacturing industries (broken line) and the average trade openness of the same industries (solid line) over the periods 1959-96 and 1958-94, respectively. It is immediate to see that, starting in the mid 70s, the dispersion of PCMs shows a relentless increase. Moreover, the standard deviation of PCMs and the average openness chase each other closely. The simple correlation between the two series equals 0.90 (0.40 after removing a linear trend). In Figure 3, we replace the first moment of the openness ratio with its second moment, again across 4-digit industries. Note that the standard deviation of trade openness closely follows the standard deviation of PCMs; the simple correlation between the two series is again very high, as it equals 0.88 (0.45 for the detrended series). Thus, a first look at the data suggests that markup heterogeneity has increased overtime and that growing asymmetries in trade exposure may be partly responsible for it.

INSERT FIGURES 2 AND 3 HERE

2.2 Trade and markup heterogeneity

We now look for more systematic evidence on the link between trade openness and the dispersion of market power. To this purpose, we exploit information contained in the cross-sectional and temporal variation in the NBER datasets[8] to construct the following time-varying, industry-level proxies for the dispersion of markups and trade openness: for each 3-digit SIC industry, we compute the standard deviation of PCMs and the standard deviation of the openness ratio among the 4-digit industries belonging to it. Next, we run Fixed-Effects regressions of the former on the latter to test whether markup heterogeneity increases systematically in those 3-digit industries where trade exposure becomes more asymmetric. The

[8] We focus on US manufacturing because of the high level of disaggregation of industry data on sales and costs provided by the NBER dataset. This should provide a lower bound for the effects we aim to illustrate, as international trade may raise the dispersion of markups also by increasing asymmetries between manufacturing and service industries.

main results are reported in Table 1. In column 1, we run a univariate regression of the standard deviation of PCMs on the standard deviation of the openness ratio and find that the two variables are strongly positively correlated, with a t-statistic around 8. In column 2, we add a full set of time dummies to control for spurious correlation due to time effects, e.g., the deregulation of the US economy initiated by the Carter Administration in the mid '70s. The coefficient of the standard deviation of openness is somewhat reduced but is still very precisely estimated, with a t-statistic of 5.

INSERT TABLE 1 HERE

The cross-industry dispersion of price-cost margins may also be affected by technological characteristics. Although Fixed-Effects estimates implicitly account for time-invariant technological heterogeneity, our results may still be driven by asymmetric technical change. Hence, to control for variation in industry technology, in column 3 we add the standard deviation (again within 3-digit SIC industries) of total factor productivity (TFP), skill-intensity (H/L) and capital-intensity (K/Y).[9] While these controls are generally significant, they leave the sign and significance of our coefficient of interest unaffected. In column 4, we run a most severe test by adding a full set of sector-specific linear trends. Notwithstanding the loss of identifying variance, the results are unchanged.

In column 5, we add the average value of all RHS variables to check whether the correlation between the second moments of PCMs and openness is driven by variation in the first moment of our covariates. It is not. Interestingly, the coefficient of average openness is small and insignificant, which suggests that the strong positive correlation between average openness and the standard deviation of PCMs illustrated in Figure 2 was mediated by the induced increase in the standard deviation of openness. In column 6, we therefore instrument the standard deviation of the openness ratio with its mean value to see how the rise in the second moment of openness attributable to a rise in its first moment affects the dispersion of markups. Estimation is by Two-Stage Least Squares. In the first stage regression for the standard deviation of openness, not reported to save space, average openness is found to be a strong instrument for its standard deviation, with a coefficient of 1.65 and a t-statistic of 11.5. In the second stage regression, the coefficient of the standard deviation of openness is instead

[9] Our measure of TFP is TFP5, from the Bartelsman and Gray's database. Skill-intensity is proxied by the ratio of non-production to production workers, and capital-intensity by plant and equipment per unit of output.

the same as in the OLS regression, with a slightly larger standard error. These results are consistent with the idea that globalization increases asymmetries in trade exposure, which in turn increase markup heterogeneity across industries.

Finally, in column 7 we also control for the average PCM. Although this variable is obviously endogenous, including it ensures that the second moment of the distribution of PCMs is not mechanically driven by variation in the first moment. The size and significance of the coefficient of interest are unaffected. Moreover, the coefficient of the average PCM is negative and significant, consistent with the idea that procompetitive forces may have induced a simultaneous fall in average markups and an increase in their dispersion across industries.

2.3 Economic development and markup heterogeneity

Finally, we show how the dispersion of markups is correlated with the level of economic development. To this purpose, we have computed the standard deviation of PCMs across three-digit ISIC manufacturing industries for a sample of 49 countries in the year 2001 (from the CEPII dataset). In Figure 4, we plot the log standard deviation of PCMs against the log of real per capita GDP. Note that higher-income countries are characterized by a significantly lower dispersion of PCMs. This stylized fact is even stronger when considering asymmetries in the PCMs between manufacturing and services. In Figure 5, we plot the log difference between the average PCM in services and manufacturing for a sample of 22 OECD countries in the year 2002. Note, again, that more developed countries are characterized by much lower asymmetries in the PCMs. We thus conclude that misallocations due to asymmetries in market power seem potentially relevant for understanding economic performance.

INSERT FIGURES 4 AND 5 HERE

3 A General Model of Imperfect Competition

A preliminary step for a comprehensive analysis of the distortions caused by an entire markup distribution is to build a tractable multi-sector model of imperfect competition that is general enough. This is the goal of the present section. We start by presenting a convenient representation of preferences, technology and market structure that encompasses as special cases a large class of models used in the literature. Next, we will use this model as a workhorse to study

three issues: (1) the misallocation arising in a market equilibrium, (2) when and how regulations affecting markups can replicate the first best equilibrium and, finally, (3) the welfare effects of asymmetric exposure to international trade. In modeling trade, we will focus on a symmetric-country case that will allow us to discuss the procompetitive effect of liberalization. In the interest of clarity, however, we begin our investigation with the closed economy.

3.1 Preferences and Technology

We focus on economies that admit a representative agent whose utility function can be used for normative purposes. We assume that there is a unit measure of agents (implying that averages coincide with aggregates), each supplying one unit of labor inelastically.[10] Preferences are given by the following CES utility function:

$$W = \left[\int_0^1 {C_i}^{\alpha} di\right]^{1/\alpha}, \quad \alpha \in (-\infty, 1], \tag{1}$$

where C_i is the sub-utility derived from consumption of possibly differentiated varieties produced in industry $i \in [0,1]$, and α governs the elasticity of substitution between industries, $\sigma = 1/(1-\alpha)$. Maximization of (1) subject to a budget constraint yields relative demand:

$$\frac{P_i}{P_j} = \left(\frac{C_j}{C_i}\right)^{1-\alpha}, \tag{2}$$

where P_i and P_j denote the cost of one unit of consumption baskets C_i and C_j, respectively.

To preserve tractability, we assume that varieties within a given industry are symmetric, so that in equilibrium each will be consumed in the same amount. This assumption is in line with our focus on between-industry rather than within-industry heterogeneity and is not essential.[11] It is particularly useful in that it allows us to use a simple and general reduced-form representation for the sub-utility derived from consumption in a given industry. Specifically, C_i is given by:

$$C_i = (N_i)^{\nu_i+1} c_i, \tag{3}$$

where c_i is consumption of the typical variety in industry i and N_i is the number of available

[10] We take labor supply as inelastic for simplicity. The effects of competition and trade when labor supply is elastic are extensively discussed in Bilbie, Ghironi and Melitz (2006) and Corsetti, Martin and Pesenti (2007). In these models, imperfect competition also distorts the trade-off between work and leisure.

[11] Extending our results to models featuring firm heterogeneity, such as Melitz (2003) and Melitz and Ottaviano (2008), would be an interesting exercise that we leave for future work. Unfortunately, combining between and within sector heterogeneity complicates substantially the analysis, making it convenient to study them separately.

varieties, equal to the number of firms in industry i. The parameter ν_i in (3) captures the preference for variety and is allowed to vary across industries. From (3), a greater variety N_i is associated with higher utility whenever $\nu_i > 0$.

Given the price of a single variety p_i, the industry price index P_i, equal to the minimum cost of one unit of C_i, can be found setting expenditure in industry i equal to the value of demand, $P_i C_i = N_i p_i c_i$. Substituting (3) yields:

$$P_i = N_i^{-\nu_i} p_i. \tag{4}$$

Each variety is produced by a single firm. Firms are owned by the totality of consumers so that any positive profits or losses are rebated, but the exact form of redistribution is irrelevant in our representative-agent economy. Production requires a fixed cost f_i and a marginal cost $1/\varphi_i$ in units of labor. Firms charge a price equal to a markup over the marginal cost:

$$p_i = \frac{\mu_i(\cdot)\, w}{\varphi_i}, \tag{5}$$

where w is the wage rate and $\mu_i(\cdot) \geq 1$ denotes the markup function. In general, the equilibrium markup is a function of the price elasticity of demand perceived by each firm, ϵ_i:

$$\mu_i = \left(1 - \frac{1}{\epsilon_i}\right)^{-1} \quad \text{with} \quad \epsilon_i \equiv -\frac{\partial \ln y_i}{\partial \ln p_i},$$

where y_i is production by a given firm. The perceived elasticity may in turn depend on the number of firms in an industry and/or the elasticity of substitution in consumption across goods. We impose the restriction $\epsilon_i > \sigma$, implying that goods are more substitutable within industries than between industries.

The markup may also depend on regulations that affect market *contestability*. For example, there might be a competitive fringe of firms that can copy and produce any variety without incurring the fixed cost, but at a higher marginal cost. The higher marginal cost may capture the lower expertise of outsiders, but can also depend on entry regulations that make production more costly for external competitors. In this case, firms may be forced to charge a limit price below (5) and equal to the marginal cost of the external competitors, in order to keep them out of the market.

In what follows, we do not impose any restriction on the markup function so as to preserve generality. Moreover, to ease notation, we will denote the markup simply as μ_i, with the understanding that it represents a function rather than a parameter. Rather than providing

a list of examples of markup functions, we just recall some of the most common reasons for markup heterogeneity. In models with differentiated products ($\nu_i > 0$), these include: (1) cross-industry differences in the within-industry elasticity of substitution among varieties, as in monopolistic competition *a lá* Dixit-Stiglitz with a continuum of firms; (2) cross-industry differences in the (low) integer number of firms and a common elasticity of substitution, as in Dixit-Stiglitz with a discrete number of firms or in Atkeson and Burstein (2008). With homogeneous products ($\nu_i = 0$), variable markups may instead result from: (3) Cournot or Bertrand competition, as in Epifani and Gancia (2006) and Bernard et al. (2003). In cases (2) and (3), more firms leads to higher competitive pressure and lower markups, i.e., $\partial\mu_i/\partial N_i < 0$. In sum, our framework can encompass the most common models of imperfect competition, describing environments where firms produce homogeneous or differentiated goods and compete in quantity or price.[12]

4 Restricted Entry

We consider now the case in which the number of firms per industry N_i is exogenously given, so that entry and exit are not allowed. Although free entry might be a reasonable assumption in many industries, entry restrictions are fairly common too, particularly in less developed countries. For instance, the number of active firms may depend on the presence of government regulations, such as licences. Restricted entry may also provide an adequate description of a short-run equilibrium in which entry has not taken place yet and fixed costs are sunk, making exit never optimal. Be as it may, this case is useful to understand the effects of trade and monopoly power in a situation when firms make pure profits.

Note that when the number of firms is not a choice variable, the fixed cost in production has no bearings on the efficiency property of the equilibrium. Therefore, without loss of generality, we simplify the exposition by setting $f_i = 0$. Recall also that profits are rebated to consumers.

4.1 Market Equilibrium

We start by characterizing the *laissez-faire* equilibrium. Denote L_i as the number of workers employed in industry i. By virtue of symmetry and the absence of fixed costs ($f_i = 0$),

[12] See the working paper version, Epifani and Gancia (2009), for specific examples. Other models of imperfect competition that can be represented within our approach include monopolistic competition with translog demand in Feenstra (2003), the generalization of Dixit-Stiglitz preferences by Benassy (1998), Melitz and Ottaviano (2008), and models of price competition with differentiated products that use the "ideal variety" approach, such as Salop (1979), Lancaster (1979) and Epifani and Gancia (2006).

production by a given firm is $y_i = \varphi_i L_i / N_i$. Then, imposing market clearing ($y_i = c_i$) into (3) we obtain:

$$C_i = \varphi_i L_i \left(N_i\right)^{\nu_i}. \tag{6}$$

The allocation of labor across sectors can be found using (2), (4), (5) and (6):

$$\frac{L_i}{L_j} = \left(\frac{\mu_j}{\mu_i}\right)^{\frac{1}{1-\alpha}} \left(\frac{\Phi_i}{\Phi_j}\right)^{\frac{\alpha}{1-\alpha}}, \tag{7}$$

where $\Phi_i \equiv \varphi_i N_i^{\nu_i}$ is a measure of aggregate "productivity" at the industry level, taking into account that consumption delivers a higher utility in industries where there are many firms and a strong preference for variety. As expected, whenever goods are gross-substitutes ($\alpha > 0$), more productive industries hire more workers. Further, for any finite value of α, more competitive industries (low μ_i) also employ more workers. Integrating (7) and imposing labor market clearing yields:

$$L_i = \frac{\left(\mu_i\right)^{\frac{1}{\alpha-1}} \left(\Phi_i\right)^{\frac{\alpha}{1-\alpha}}}{\int_0^1 \left(\mu_j\right)^{\frac{1}{\alpha-1}} \left(\Phi_j\right)^{\frac{\alpha}{1-\alpha}} dj}. \tag{8}$$

Finally, substituting (8) into (6) and then into preferences (1), we obtain the utility of the representative agent:

$$W = \left[\int_0^1 \left(L_i \Phi_i\right)^{\alpha} di\right]^{\frac{1}{\alpha}} = \frac{\left[\int_0^1 \left(\mu_i^{-1} \Phi_i\right)^{\frac{\alpha}{1-\alpha}} di\right]^{\frac{1}{\alpha}}}{\int_0^1 \left(\mu_i^{-1} \Phi_i^{\alpha}\right)^{\frac{1}{1-\alpha}} di}. \tag{9}$$

This is our welfare measure. Inspection of (9) immediately reveals that utility is homogeneous of degree zero in markups: multiplying all μ_i by any positive constant leaves welfare unaffected. In other words, as originally argued by Lerner (1934), in this economy welfare is independent of the average markup.[13]

[13] An important assumption behind this result is that labor supply is inelastic. The reason is that markups lower wages below the marginal product of labor (MPL) and thus distort the work-leisure decision. For example, in the case $\Phi_i = 1$, $\forall i \in [0,1]$ we can show that:

$$w = \left[\int_0^1 \mu\left(i\right)^{-\alpha/(1-\alpha)} di\right]^{(1-\alpha)/\alpha} < MPL = 1,$$

where the latter equality follows from the fact that, with $\Phi_i = 1$, labor productivity is equal to one. The strength of this distortion depends upon the elasticity of labor supply. However, as already noted by Lerner, even in this setting a markup on leisure (or leisure goods) would restore his principle that only dispersion matters.

4.2 Social Planner Solution

To illustrate the distortions that may arise in the market equilibrium, we now solve for the allocation that maximizes the utility of the representative agent, subject to the resource constraint of the economy. This is equivalent to solving the following planning problem:

$$\max_{L_i} \; W = \left[\int_0^1 (L_i \Phi_i)^{\alpha} \, di \right]^{\frac{1}{\alpha}}, \tag{10}$$

subject to the resource constraint:

$$\int_0^1 L_i di = 1.$$

Taking the ratio of any two first order conditions yields the optimal labor allocation:

$$\frac{L_i}{L_j} = \left(\frac{\Phi_i}{\Phi_j} \right)^{\frac{\alpha}{1-\alpha}}. \tag{11}$$

Comparing (11) with (7) we see immediately that, for any finite α, the decentralized equilibrium is Pareto-efficient if and only if $\mu_i = \mu_j$, $\forall i, j \in [0, 1]$.

We summarize these results in the following proposition:

Proposition 1 *When the number of firms is exogenous, welfare is homogeneous of degree zero in markups. A necessary and sufficient condition to replicate the first best allocation is that markups be identical across all industries.*

4.3 The Cost of Heterogeneity: Intersectoral Misallocations

If a uniform markup is sufficient to replicate the optimal allocation, what is then the cost of asymmetric market power? From (8), it can be shown that the market equilibrium entails underproduction in industries where the markup is above the following productivity-weighted average:

$$\mu^* \equiv \left[\frac{\int_0^1 (\mu_j)^{\frac{1}{\alpha-1}} (\Phi_j)^{\frac{\alpha}{1-\alpha}} \, dj}{\int_0^1 (\Phi_j)^{\frac{\alpha}{1-\alpha}} \, dj} \right]^{\alpha-1},$$

and overproduction in industries where $\mu_i < \mu^*$.[14] Thus, the problem is one of an *intersectoral misallocation* whereby less competitive industries attract a sub-optimally low number of

[14] Interestingly, this means that monopoly power is associated to overproduction in industries where $1 < \mu_i < \mu^*$. Thus, the conventional wisdom that a monopolist always produces less than the socially optimal quantity turns out to be wrong.

workers. This happens because high markups compress the wage bill.

The welfare cost of the misallocation due to markup heterogeneity depends in an interesting way on the curvature of the utility function (1). To see this, suppose that the μ_i can be approximated by a log-normal distribution and, to isolate the effect of markup heterogeneity, consider the case $\Phi_i = \Phi$. Then, (9) can be rewritten as:[15]

$$\ln W = \ln \Phi - \frac{\text{var}\left(\ln \mu\right)}{2\left(1-\alpha\right)}, \tag{12}$$

showing that markup dispersion is more costly when goods are highly substitutable (high α).

The effect of substitutability across goods on the monopoly distortion induced by asymmetric markups is not an obvious one. On the one hand, a high substitutability means that the cost of overproduction in some industries is small: indeed, this cost goes to zero as goods become perfect substitutes. On the other hand, equation (7) shows that, for a given asymmetry in markups μ_i/μ_j, a high substitutability magnifies the misallocation of labor towards the more competitive industries. It turns out that the latter effect dominates, so that perhaps counter-intuitively a flatter curvature of the utility function leads to a higher cost of markup dispersion. On the contrary, (11), (7) and (12) show that, as we approach the Leontief case ($\alpha \to -\infty$), the intersectoral misallocation disappears.

It is also interesting to observe that the distortion induced by markups differs fundamentally from distortions that manifest themselves as higher production costs. To see this, consider the case of no markup dispersion. When $\mu_i = \mu$, the welfare function becomes:

$$W = \left[\int_0^1 \left(\Phi_i\right)^{\frac{\alpha}{1-\alpha}} di\right]^{\frac{1-\alpha}{\alpha}}. \tag{13}$$

Assuming Φ_i to be log-normal and using the properties of log-normal distributions (twice), welfare becomes:

$$\ln W = \ln \mathbb{E}\left(\Phi\right) + \left(\frac{2\alpha-1}{1-\alpha}\right)\frac{\text{var}\left(\ln \Phi\right)}{2}. \tag{14}$$

Intuitively, welfare is an increasing function of average productivity, $\ln \mathbb{E}\left(\Phi\right)$. Perhaps more surprisingly, (14) shows that, despite the symmetry in preferences, the variance of productivity Φ_i becomes welfare increasing when $\alpha > 0.5$, i.e., when the elasticity of substitution between

[15] To derive (12), recall that, if $x \sim \log Normal$, then:

$$\ln \mathbb{E}\left(x^n\right) = n\mathbb{E}\left(\ln x\right) + \frac{n^2 \text{var}\left(\ln x\right)}{2}.$$

goods is greater than two. This happens because, when α is sufficiently high, agents can easily substitute consumption from unproductive industries to high Φ_i ones.[16]

In the general case, welfare is a complex function of the entire distributions of both μ_i and Φ_i. Although it is difficult to make more precise statements regarding the impact of a particular change in those distribution and their correlation, from (9) we can derive a simple formula that can be used to measure welfare given data on μ_i and Φ_i:

$$W^{\alpha} = \frac{\mathbb{E}\left(\overline{\mu}^{\alpha}\right)\mathbb{E}\left(\overline{\Phi}\right) + \operatorname{cov}\left(\overline{\mu}^{\alpha}, \overline{\Phi}\right)}{\left[\mathbb{E}\left(\overline{\mu}\right)\mathbb{E}\left(\overline{\Phi}\right) + \operatorname{cov}(\overline{\mu}, \overline{\Phi})\right]^{\alpha}}, \tag{15}$$

where $\overline{\Phi} = (\Phi_i)^{\alpha/(1-\alpha)}$ and $\overline{\mu} = \mu\left(i\right)^{1/(\alpha-1)}$.

We summarize the main findings of this section in the following proposition:

Proposition 2 *Markup heterogeneity introduces an intersectoral misallocation, whereby industries with below-average markups overproduce, and industries with above-average markups underproduce. The extent and the cost of this misallocation are proportional to the elasticity of substitution between industries.*

4.4 Procompetitive Losses From Trade

We now open the model to international trade to show that, when entry is restricted, trade integration tightening competition in some industries may amplify monopoly distortions. Moreover, *the effect can be so strong that an equilibrium with trade may be Pareto-inferior to autarky for all the countries.* Although rather extreme, this example is illustrative of the neglected principle that trade can affect welfare by changing the cross-sectoral dispersion of market power. A noteworthy corollary is that the characteristics of industries affected by trade liberalization and particularly their competitiveness relative to the rest of the economy are important factors to correctly foresee the effects of globalization. In turn, the result that welfare may fall with trade liberalization is an application of second-best theory. As pointed out by Bhagwati (1971) and Johnson (1965), if trade induces a contraction of industries that were already underproducing compared to the optimum, it exacerbates an existing distortion and may thus lower welfare.

To isolate the point we want to make, we adopt the following simplifying assumptions. First, we consider a world populated by $M > 1$ identical countries so as to abstract from

[16] Interestingly, this also suggests that price dispersion may be beneficial when it originates from technology and the elasticity of substitution is high enough. See also Jones (2009) on the role of complementarity in amplifying industry-level distortions.

specialization effects. Second, to remove any unnecessary heterogeneity, we normalize the number of firms in each country to one ($N_i = 1$), and set $\varphi_i = 1$ for all i. Third, we assume that in some industries goods can be freely traded, while in others trade costs are prohibitive. Accordingly, the unit measure of sectors is partitioned into two subsets of traded and nontraded industries, ordered such that industries with an index $i \leq \tau \in [0, 1]$ are subject to negligible trade costs while the others, with an index $i > \tau$, face prohibitive trade costs. This simple description of imperfect trade integration accords well with the evidence that trade volumes are high in some industries and very low in others. We consider two complementary aspects of international integration: (1) an increase in the range τ of traded industries and (2) an increase in the number M of trading partners. Finally, we assume the markup to be a negative function of the number of competing firms in a given industry. This immediately delivers the procompetitive effect of trade, as the number of firms is one in nontraded industries and $M > 1$ in the others.[17]

For convenience, we denote the relative markup in nontraded industries as $x \equiv \mu(1)/\mu(M)$. Under our assumptions, x is greater than one and increasing in M. After some straightforward substitutions into equation (9), we obtain:

$$W = \frac{\left[1 - \tau + \tau \left(xM^{\nu}\right)^{\alpha/(1-\alpha)}\right]^{1/\alpha}}{1 - \tau + \tau \left(xM^{\nu\alpha}\right)^{1/(1-\alpha)}}. \tag{16}$$

This expression shows that welfare is a function of the measure of traded industries, τ, the number of trading countries, M, and the markup asymmetry x between open and closed industries.

INSERT FIGURE 6 HERE

Figure 6 plots welfare as a function of τ for the case $\nu = 0$ (solid line) and the case $\nu > 0$ (broken line). In the first case (corresponding for example to Cournot competition), the economy attains the same level of welfare in autarky ($\tau = 0$) and when trade is free in all industries ($\tau = 1$). For any intermediate case, an equilibrium with trade is Pareto inferior to autarky. The intuition for this result should be by now clear. When $\nu = 0$, there is no gain from consuming foreign varieties and the only effect of trade is to lower markups in industries

[17] The procompetitive effect of trade, whereby exposure to international competition reduces markups features prominently in Krugman (1979) and Melitz and Ottaviano (2008), among others. See Chen, Imbs and Scott (2009) for recent evidence.

exposed to foreign competition. In both extreme cases, $\tau = 0$ and $\tau = 1$, markups are uniform across industries and there is no distortion. As τ moves from zero to one, trade breaks this symmetry: it increases markup dispersion as long as open industries are a minority and lowers it afterwards. Moreover, when $\nu = 0$ and $\tau \in (0, 1)$ it is easy to see that welfare declines with an increase in the number of trading countries, M. The reason, again, is that a larger number of international competitors increases the markup asymmetry between traded and nontraded industries.

When $\nu > 0$, consumers derive a higher utility from the possibility to buy foreign varieties. In our model, we can think of this variety effect as capturing any source of gains from trade that is independent of the procompetitive effect. As the figure shows, in this case an equilibrium with some trade *might* still be Pareto inferior to autarky when τ is low, for the gains from small volumes of trade might be too low to dominate the price distortion (this happening for low enough ν). However, when τ is large enough, the gains from variety will eventually dominate the (falling) cost of misallocations. With gains from trade of any sort, the equilibrium with full integration ($\tau = 1$) must necessarily dominate autarky.

Even when a high $\nu > 0$ assures positive gains from trade, when liberalization increases markup dispersion, it generates or exacerbate the intersectoral misallocation discussed above. What can then be done to counteract this negative effect of market integration? We have seen that the first-best solution is attained with a uniform markup. Thus, if trade lowers markups in some sectors, competition policy might be used to match the change in market power in nontraded sectors too. If competition policy cannot be used, the first best solution can still be achieved by giving an appropriate subsidy to industries producing nontraded goods.

Note also that the likelihood that trade is harmful increases with x and that positive gains from trade will surely materialize if an economy is perfectly competitive ($x = 1$). In other words, the potential for welfare losses is higher when domestic markets are not competitive enough and trade brings large asymmetries between industries selling in world markets and the rest of the economy. These considerations may be particularly relevant for less developed countries, suggesting that in some cases *promoting competition may be a prerequisite to ensuring positive gains from trade.*

In sum:

Proposition 3 *With an exogenous number of firms, procompetitive welfare losses from trade are possible when trade increases markup dispersion.*

4.5 A Simple Quantitative Exercise

The example discussed in the previous section was admittedly provocative. We now show, however, that a simple quantitative exercise suggests the welfare cost of markup heterogeneity to be potentially large when entry is restricted. To this end, we use our model, together with the evidence and the data discussed in Section 2, to compute the cost of markup dispersion across US manufacturing industries, relative to a first-best allocation in which markups are instead uniform. Using equation (9), and denoting by W^{FB} welfare in the first-best allocation, we obtain:

$$\frac{W}{W^{FB}} = \frac{\left[\int_0^1 \left(\mu_i^{-1}\Phi_i\right)^{\frac{\alpha}{1-\alpha}} di\right]^{1/\alpha}}{\int_0^1 \left(\mu_i^{-1}\Phi_i^{\alpha}\right)^{\frac{1}{1-\alpha}} di \cdot \left(\int_0^1 \Phi_i^{\frac{\alpha}{1-\alpha}} di\right)^{\frac{1-\alpha}{\alpha}}}. \tag{17}$$

Computing (17) requires an empirical measure of the productivity index $\Phi_i = \varphi_i N_i^{v_i}$ for each industry i. To build such measure, we proceed as follows. Equations (4) and (5) imply:

$$\Phi_i = \frac{\mu_i w}{P_i}. \tag{18}$$

From utility maximization, we obtain:

$$C_i = \left(\frac{P_i}{P}\right)^{\frac{1}{\alpha-1}} W = P_i^{\frac{1}{\alpha-1}} E, \tag{19}$$

where E is total expenditure, $P = \left[\int_0^1 P_i^{\frac{\alpha}{\alpha-1}} di\right]^{\frac{\alpha-1}{\alpha}}$ is the ideal price index associated to (1), and the latter equality follows from choosing P as the numeraire. Equation (19) allows us to express the unobserved industry price index P_i as a function of the observed expenditure share on an industry's products, $\theta_i \equiv P_i C_i / E$:

$$P_i = \theta_i^{\frac{\alpha-1}{\alpha}}. \tag{20}$$

Finally, using (20) into (18) gives:

$$\Phi_i = w \mu_i \theta_i^{\frac{1-\alpha}{\alpha}}. \tag{21}$$

Note that W/W^{FB} is homogeneous of degree zero with respect to all the Φ_i. Hence, without any loss of generality, we can disregard the factor w, which is constant across sectors, and calibrate Φ_i using data on markups, μ_i, and expenditure shares, θ_i, for the US 4-digit SIC industries over the period 1959-1994. In particular, to compute μ_i, we use the definition of

price-cost margins (PCMs) in Section 2 and set $\mu_i = (1 - PCM_i)^{-1}$.[18] θ_i, is instead computed as the value of an industry's shipments plus net imports, divided by the total expenditure on manufacturing goods, using again the NBER datasets. Calibrated in this way, Φ_i accounts for factors, such as parameters of technology but also preferences, that affect expenditure shares other than relative markups.

Computing Φ_i and W/W^{FB} also requires choosing a value for the elasticity of substitution in consumption among manufacturing goods, $\sigma = 1/(1-\alpha)$. Although available estimates of σ vary widely across studies, most of them are in the range $[2, 10]$ and a value around $\sigma = 5$ is most frequently used in quantitative exercises. We therefore set $\sigma = 2$, 5 and 10 (implying $\alpha = 0.5$, 0.8 and 0.9) as benchmark cases. We then perform two distinct exercises.

First, we assume the Φ_i to be constant overtime and compute their cross-section from (21) by taking, for each 4-digit industry, the mean value of μ_i and θ_i overtime. Then, we compute W/W^{FB} over the period 1959-94 using (17). This exercise allows us to isolate the welfare loss induced by the change in the distribution of markups, holding constant preferences and other technological factors. The results are reported in the left panel of Table 2. For an intermediate value of the elasticity of substitution ($\sigma = 5$), utility falls by 3.6 percent in the period of analysis. For $\sigma = 2$, the welfare cost of the observed rise in markup dispersion is lower (1.3 percent), whereas it is much larger (8.1 percent) for $\sigma = 10$.

Second, we allow the Φ_i to be time-varying and compute them year by year. We use these time-varying coefficients again into (17) to obtain the value of W/W^{FB}. This exercise provides information on the overall welfare cost of changes in the markup distribution taking into account that other parameters (such as tastes and technology) have also affected expenditure shares simultaneously. The results are in the right panel of Table 2, showing that in this case the welfare cost of markup dispersion is even larger. This indicates that the evolution of the exogenous parameters contained in the Φ_i has contributed to amplify monopoly distortions. In particular, over the period of analysis, relative utility falls by 2.5 percent for $\sigma = 2$, by 9.8 percent for $\sigma = 5$, and by as much as 24.3 percent for $\sigma = 10$.

INSERT TABLE 2 HERE

[18] Price-cost margins are defined as the value of shipments minus costs divided by the value of shipments. Thus:

$$PCM_i = (p_i y_i - \frac{w}{\varphi_i} y_i)/p_i y_i = 1 - 1/\mu_i,$$

where $p_i y_i$ is the value of a firm's shipments in industry i, and $(w/\varphi_i)\, y_i$ is its variable cost. Although in our simple model labor is the only variable cost, we also net out materials and capital expenditures in our empirical definition of price-cost margins. This avoids spurious variation in the PCMs due to variation in materials intensity and capital intensity across industries.

In sum, these numerical exercises suggest that, abstracting from any effect that changes in the average PCMs may have had, the observed increase in markup heterogeneity can entail substantial welfare costs. How much of these costs can then be attributed to trade integration? Given that the impact of trade on markup dispersion can be ambiguous, the answer to this question is ultimately an empirical one. Although identifying and quantifying how trade has affected the markup distribution goes beyond the scope of the current paper, the evidence discussed in Section 2 does suggest that trade may be a major determinant of observed asymmetries in market power.[19] We therefore conclude that markup heterogeneity matters for misallocations and that procompetitive losses from trade may be more than a theoretical *curiosum* in this class of models.

5 Free entry

So far, firms are making positive profits and barriers to entry prevent potential competitors from challenging incumbent firms and sharing the rents. Without those barriers, entry will take place until pure profits are driven to zero. We now allow for this possibility with the aim of extending our results to a widely-used class of models where entry is free. This exercise will lead to remarkably different conclusions, which qualify some of Lerner's original statements.

We start by presenting the market equilibrium and show that, contrary to the previous case, welfare is a function of the average markup too. Next, we compare it with the social planner solution and discuss the inefficiencies that arise in the decentralized equilibrium. Finally, we study the effect of trade between identical countries and argue that, while procompetitive losses are now unlikely, asymmetric trade liberalization may still exacerbate misallocations, thereby providing a rationale for policy intervention.

5.1 Market Equilibrium

We now reintroduce the fixed cost of production, f_i, defined in units of labor, and let the number of firms vary so as to guarantee that each breaks even. In this way, in equilibrium all operating profits are used to cover the fixed cost:

$$p_i y_i - \frac{y_i w}{\varphi_i} = f_i w_i.$$

[19] In particular, if we rerun the regressions in Table 1 computing the standard deviation of markups and other variables across all the 4-digit SIC industries, we obtain that the increase in trade openness (or its standard deviation) over the period 1959-94 can explain the entire observed increase in markup dispersion across US manufacturing industries. The results are available upon request.

Substituting p_i from (5) and rearranging gives:

$$y_i = \frac{\varphi_i f_i}{\mu_i - 1}. \tag{22}$$

As is well-known, the free-entry condition pins down uniquely firm size.

Given firm size, the number of active firms must be proportional to the amount of labor employed in each industry. More precisely, the demand for labor in industry i is:

$$L_i = N_i \left[\frac{y_i}{\varphi_i} + f_i \right].$$

Substituting (22) we obtain:

$$N_i = \frac{\mu_i - 1}{f_i \mu_i} L_i. \tag{23}$$

Finally, to solve for L_i, we manipulate the demand equation (2) to yield:

$$\frac{C_i P_i}{C_j P_j} = \left(\frac{C_j}{C_i} \right)^{-\alpha} = \frac{L_i}{L_j},$$

where the latter equality follows from the fact that, with a fixed cost in units of labor and without extra-profits, industry revenue equals the wage bill. Using (3), the market clearing condition $c_i = y_i$, (22) and (23), we obtain:

$$\frac{L_i}{L_j} = \left(\frac{\varphi_i N_i^{\nu_i} \mu_j}{\varphi_j N_j^{\nu_j} \mu_i} \right)^{\frac{\alpha}{1-\alpha}}. \tag{24}$$

Note that (24) differs from the analogous condition in the model without free entry (7). The key reason is that, in the model with free entry, the entire industry revenue is used up to pay workers (both for the fixed and variable production costs), while in the other case a fraction of revenue is captured by profits.[20] Combining (22), (23) and (24) we obtain an equation linking the number of firms in any two industries to relative firm size and other exogenous parameters:

$$\frac{N_i^{1-\alpha-\alpha\nu_i}}{N_j^{1-\alpha-\alpha\nu_j}} = \left(\frac{y_j}{y_i} \right)^{1-\alpha} \frac{\varphi_i}{\varphi_j} \frac{\mu_j}{\mu_i}. \tag{25}$$

This condition will turn out to be useful below.

[20] A second, less important, difference is that, in the model without entry we set the fixed cost to zero. Of course, the equilibrium allocation of labor would coincide in the two models if we had the same fixed cost and if the exogenous number of firms without entry happened to be equal to the equilibrium number of firms with entry.

To understand the role of markups in the model with free entry, consider for the moment the simpler Cobb-Douglas case corresponding to $\alpha = 0$. Then, using (22), (23) and (24), we can write welfare as:

$$\ln W = \int_0^1 \ln\left(N_i^{\nu_i+1} y_i^{\alpha}\right) di = \overbrace{\int_0^1 \nu_i \ln\left(\frac{1-\mu_i^{-1}}{f_i}\right) di}^{variety} + \overbrace{\int_0^1 \ln \varphi_i \mu_i^{-1} di}^{f-saving}. \tag{26}$$

This equation shows that the level of markups has now both positive and negative direct effects on welfare. The first term in (26) is increasing in μ_i and captures the fact that a high profit margin stimulates entry, thereby raising the number of firms and welfare so long as variety has value ($\nu_i > 0$). The second term in (26) is instead decreasing in μ_i and captures the fact that entry implies that more productive resources are taken by the fixed costs. Thus, contrary to the restricted-entry case, when $\nu_i > 0$ markups now pose a trade-off between diversity and fixed costs at the industry level.

5.2 Social Planner

What are the distortions in the market equilibrium? Is markup asymmetry desirable or does it impose any welfare costs? To answer these questions, we now compute the allocation that maximizes the utility of the representative agent. A benevolent social planner chooses N_i and $y_i = c_i$ so as to solve:

$$\max_{N_i, y_i} W = \left[\int_0^1 \left(N_i^{\nu_i+1} y_i\right)^{\alpha} di\right]^{1/\alpha},$$

subject to the resource constraint:

$$\int_0^1 N_i \left(\frac{y_i}{\varphi_i} + f_i\right) di = 1.$$

The Lagrangian for the above program is:

$$\mathbb{L} = \left[\int_0^1 \left(N_i^{\nu_i+1} y_i\right)^{\alpha} di\right]^{1/\alpha} - \lambda \left[\int_0^1 N_i \left(\frac{y_i}{\varphi_i} + f_i\right) di - 1\right],$$

and the first order conditions for an optimum are:

$$\begin{aligned} \frac{\partial \mathbb{L}}{\partial N_i} &= 0 \rightarrow W^{1-\alpha} (\nu_i + 1) y_i^{\alpha} N_i^{(\nu_i+1)\alpha - 1} = \lambda \left(\frac{y_i}{\varphi_i} + f_i\right), \\ \frac{\partial \mathbb{L}}{\partial y_i} &= 0 \rightarrow W^{1-\alpha} N_i^{(\nu_i+1)\alpha} y_i^{\alpha-1} = \frac{\lambda N_i}{\varphi_i}. \end{aligned}$$

Substituting the second first order condition into the first yields:

$$y_i = \frac{\varphi_i f_i}{\nu_i}. \tag{27}$$

Taking the ratio of the second first order condition in industries i and j delivers:

$$\frac{N_i^{1-\alpha-\alpha\nu_i}}{N_j^{1-\alpha-\alpha\nu_j}} = \left(\frac{y_j}{y_i}\right)^{1-\alpha} \frac{\varphi_i}{\varphi_j}. \tag{28}$$

Comparing the optimal firm scale (27) to the market outcome (22), we see that the two coincide when:

$$\mu_i - 1 = \nu_i, \tag{29}$$

that is, when the markup is equal to the preference for variety, as in the Dixit-Stiglitz case. However, comparing (28) with (25), we also see that the optimal allocation of resources across industries requires $\mu_i = \mu_j$, that is, a uniform markup. When the preference for variety is unequal across industries (the most realistic case) these two requirements are incompatible and we thus have the following impossibility result:[21]

Proposition 4 *When entry is free and the preference for variety is heterogeneous across industries ($\nu_i \neq \nu_j$ for at least two $i, j \in [0, 1]$), there exists no markup distribution such that the market equilibrium replicates the first-best allocation.*

These results show that, in general, a uniform markup across industries and firms is still a necessary condition for efficiency but, contrary to Lerner's original claim, it is not sufficient anymore. The reason is that profits now have a dual role: they affect the allocation of resources across industries and the equilibrium number of firms per industry. As we know from the previous section, avoiding intersectoral misallocations requires $\mu_i = \mu_j$. However, markups should also correctly signal the social value of entry and this requires higher profit margins in industries where variety is more valuable.

Finally, recall that the intersectoral misallocation tends to disappear as preferences approach the Leontief case. This is true in general and the model with free entry makes no exception. In fact, taking the limit of (28) and (25) for $\alpha \to -\infty$ reveals that the two equa-

[21] Epifani and Gancia (2008) review some evidence suggesting that external economies due to love for variety differ across industries.

tions converge to the same condition:

$$\frac{y_j}{y_i} = \frac{N_i^{1+\nu_i}}{N_j^{1+\nu_j}},$$

which is of course equivalent to $C_i = C_j$. In this case, the market equilibrium converges to the first-best allocation when $\mu_i - 1 = \nu_i$.

5.3 Optimal Competition Policy, Markup Heterogeneity and Welfare

We now ask what is the markup distribution that maximizes the utility of the representative agent. In other words, we are interested in finding the constrained efficient allocation that a planner can achieve by controlling markups and without using lump-sum transfers. To find it, we use (22) and (23) to rewrite the welfare function as:

$$W = \left\{ \int_0^1 \left[\left(\frac{L_i}{\mu_i} \right)^{\nu_i+1} \left(\frac{\mu_i - 1}{f_i} \right)^{\nu_i} \varphi_i \right]^{\alpha} di \right\}^{1/\alpha} = \left\{ \int_0^1 \left[(L_i)^{\nu_i+1} \Phi_i \right]^{\alpha} di \right\}^{1/\alpha}, \tag{30}$$

where now $\Phi_i \equiv \left(\frac{1}{\mu_i}\right)^{\nu_i+1} \left(\frac{\mu_i - 1}{f_i}\right)^{\nu_i} \varphi_i$. Note that W is increasing in Φ_i. Thus, maximizing (30) is equivalent to maximizing Φ_i industry by industry. Then, the first order condition is:

$$\frac{\partial \Phi_i}{\partial \mu_i} = 0 \rightarrow \frac{\nu_i}{\mu_i - 1} = \frac{\nu_i + 1}{\mu_i} \rightarrow \mu_i - 1 = \nu_i, \tag{31}$$

which is identical to (29).[22] Thus, it is optimal to let the markup reflect the social value of entry, irrespective of the intersectoral inefficiency. This means that, so long as $\nu_i \neq \nu_j$, an increase in markup dispersion may be welfare improving if the resulting markup distribution gets closer to the one implied by condition (29). It also implies that computing the welfare cost of a given markup distribution becomes much harder, as it now requires data on ν_i.

Finally, it is instructive to consider the welfare costs of markup heterogeneity in the special case of $\nu_i = 0$, that is, when entry has no social value *per se*. This case is of particular interest because it serves as a metaphor for all models where profits are dissipated in equilibrium through rent-seeking activities that are socially wasteful. When $\nu_i = 0$, it is of course optimal to have $\mu_i = 1$; yet, this may not be feasible. What is less obvious, instead, is the cost of

[22] It follows immediately that there is no room for welfare-improving intervention on markups in the Dixit-Stiglitz case. Yet, this is admittedly not the most interesting case to study procompetitive effects.

markup dispersion. From (24) and $\int_0^1 L_i di = 1$, we obtain:

$$L_i = \frac{\left(\varphi_i \mu_i^{-1}\right)^{\frac{\alpha}{1-\alpha}}}{\int_0^1 \left(\varphi_j \mu_j^{-1}\right)^{\frac{\alpha}{1-\alpha}} dj}.$$

Substituting this into the welfare function (30) yields:

$$W = \left(\int_0^1 \left(\varphi_i \mu_i^{-1}\right)^{\frac{\alpha}{1-\alpha}} di\right)^{\frac{1-\alpha}{\alpha}},$$

which takes the same form as (13). Remarkably, this means that markup heterogeneity is welfare improving when $\alpha > 0.5$, precisely as we found for productivity heterogeneity. This result can be understood by noting that, in an equilibrium with free entry, the markup is nothing but the per unit equivalent of the fixed cost f_i. Thus, when entry is free but has no social value, the markup acts a pure cost for the economy and it affects welfare just as the marginal cost $(1/\varphi)$ in the model of Section 4 did.

We summarize the main findings in the following proposition:

Proposition 5 *With free entry, markup symmetry is a necessary, but not a sufficient condition for efficiency. Markup heterogeneity always leads to an intersectoral misallocation, but does not necessarily lower welfare: it may be welfare improving when the preference for variety is heterogeneous across industries, or when variety has no value ($\nu_i = 0$, $i \in [0,1]$) and the elasticity of substitution is high ($\alpha > 0.5$).*

5.4 The Procompetitive Effect of Trade, Welfare and Misallocations

Are procompetitive losses from trade possible when entry is free? Does asymmetric trade liberalization introduce distortions that may be corrected by policy makers? To briefly address these questions we now open the model to trade, as in section 4.4. For simplicity, we focus on the Cobb-Douglas case, i.e., $\alpha = 0$. We consider a world of M symmetric countries and we denote by N_i the number of firms per country in industry i. Consumption of a given traded variety in a given country becomes $c_i = y_i/M_i$, so that the industry consumption basket can be written as:

$$C_i = (N_i)^{\nu_i+1} (M_i)^{\nu_i} y_i.$$

Substituting this into (1) and using (22) and (23), we obtain our welfare measure:

$$\ln W = \int_0^1 \ln C_i di = \int_0^1 \ln \left[(N_i \cdot M_i)^{\nu_i} \varphi_i \mu_i^{-1} \right] di. \tag{32}$$

As before, we model imperfect market integration by allowing some industries to be closed to trade. In other words, we set $M_i = 1$ in the subset of nontraded industries.

When the markup function is such that $\mu_i = f(N_i \cdot M_i)$ with $f'(\cdot) < 0$, as in Krugman (1979), then the procompetitive effect of trade is always welfare improving. To see this, note that if trade has to lower the markup in an industry, it should also increase the equilibrium number of firms. Both the fall in μ_i and the rise in $N_i \cdot M_i$ increase W. Interestingly, this is true even when variety has no value, $\nu_i = 0$! Procompetitive losses from trade are still possible, but only when the fall in markups due to foreign competition is so strong as to reduce the equilibrium number of firms in an industry below the optimal level.[23] While this is impossible when $\mu_i = f(N_i \cdot M_i)$ with $f'(\cdot) < 0$, this outcome is conceivable if trade liberalization affects the markup not through the number of firms.

In conclusion, when there is free entry, procompetitive losses from trade seem much more unlikely. Thus, an important benefit of entry is that it prevents some of the possibly large costs identified in Section 4. Yet, asymmetric trade liberalization that increases markup heterogeneity exacerbates the intersectoral misallocation of resources and opens the way to Pareto improving intervention. In particular, there might be excessive product diversity or firm output in traded industries that can be corrected with an appropriate subsidy to nontraded industries.

We therefore conclude with the following proposition:

Proposition 6 *When entry is free and markups are a negative function of the number of firms in an industry ($\mu_i = f(N_i \cdot M_i)$, $f'(\cdot) < 0$, $i \in [0,1]$), the procompetitive effect of trade is welfare improving, even when trade amplifies misallocations due to markup dispersion and variety has no value.*

6 Conclusions

Competition is imperfect in most sectors of economic activity. By exposing firms to foreign competition, trade is widely believed to help alleviate the distortions stemming from monopolistic pricing. While this argument is often well-grounded, it neglects that in general

[23] Eckel (2008) provides conditions for this outcome. In other models, such as Melitz (2003) and Corsetti, Martin and Pesenti (2007), trade liberalization increases welfare even when the number of varieties falls.

equilibrium pricing distortions depend on both *absolute* and *relative* market power, and that a trade-induced fall in markups brings about misallocations when it raises their variance. The latter event is more than a theoretical curiosity, for market globalization affects predominantly industries that are already relatively more competitive. At the same time, misallocations across firms and industries have recently been identified as a strikingly important factor behind cross-country differences in economic performance. Studying how such misallocations may be rooted in the economy-wide markup distribution and how this may interact with trade liberalization is therefore crucial for the design of optimal trade and competition policies and for a better understanding of the welfare effects of trade opening in the presence of market power. This was the aim of our paper.

We now summarize what we view as the main results. When firm entry is restricted, we find that markup heterogeneity entails significant costs and that asymmetric trade liberalization may reduce welfare. With free entry of firms, instead, markup heterogeneity is not necessarily welfare reducing, although it generates an intersectoral misallocation that policy makers can correct. In this case, we also find that a trade-induced increase in competition is typically welfare increasing. Yet, if trade integration raises markup dispersion, the allocation of resources can be improved upon by subsidizing production in industries that remain relatively more protected. In this sense, trade liberalization and domestic industrial policy complement each other. More in general, our analysis has emphasized the neglected principle that, in order to correctly foresee the effects of trade and competition policy, the evolution of the economy-wide markup distribution has to be taken into account, and that whether entry is restricted or not makes an important difference.

By focusing on special cases, the existing literature on the topic offers a partial view only. One goal of this paper was precisely to clarify the misconceptions that may arise when restricting the analysis to special cases. A major benefit of our general framework is that it illustrates the exact role of alternative assumptions in shaping the relationship between competition, misallocations and welfare. We hope that such a unified framework may prove useful in studying other issues, such as the effects of competition on growth, and in guiding future empirical and quantitative work. In particular, while we have emphasized markup heterogeneity across industries, we think that extending the analysis to heterogeneity across firms may be at least as much important. Yet, accounting for it poses a new difficulty, in that it requires disentangling firm-level estimates of productivity from markups. We view this as a key challenge for future research.

References

[1] Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt (2005). "Competition and Innovation: An Inverted U Relationship," *Quarterly Journal of Economics* 120, 701-728.

[2] Atkeson, A. and A. Burstein (2008). "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review* 98, 1998-2031.

[3] Banerjee, A. and E. Duflo (2005). "Growth Theory Through the Lens of Development Economics," in P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth* (Amsterdam: North-Holland).

[4] Basu, S. and J.G. Fernald (2002). "Aggregate productivity and aggregate technology," *European Economic Review*, 46, 963-991.

[5] Benassy, J. P. (1998). "Is There Always Too Little Research in Endogenous Growth with Expanding Product Variety?," *European Economic Review* 42, 61-69.

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

[7] Bhagwati, J. (1971). "The generalized theory of distortions and welfare," in J. Bhagwati et al. (eds.), *Trade balance of payments and growth* (Amsterdam: North-Holland).

[8] Bilbiie, F., F. Ghironi and M. Melitz (2006). "Monopoly Power and Endogenous Variety in Dynamic Stochastic General Equilibrium: Distortions and Remedies," mimeo.

[9] Brander, J. and P. Krugman (1983). "A Reciprocal Dumping Model of International Trade," *Journal of International Economics* 15, 313-321.

[10] Chen, N., J. Imbs and A. Scott (2009). "The Dynamics of Trade and Competition," *Journal of International Economics* 77, 50-62.

[11] Corsetti, G., P. Martin and P. Pesenti (2007). "Productivity, Terms of Trade and the Home Market Effect," *Journal of International Economics* 73, 99-127.

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

[13] Eaton, J. and A. Panagariya (1979). "Gains from Trade under Variables Returns to Scale, Commodity Taxation, Tariffs and Factor Market Distortions," *Journal of International Economics* 9, 481-501.

[14] Eckel, C. (2008). "Globalization and Specialization," *Journal of International Economics* 75, 219-228.

[15] Epifani, P. and G. Gancia (2006). "Increasing Returns, Imperfect Competition and Factor Prices," *Review of Economics and Statistics* 88, 583-598.

[16] Epifani, P. and G. Gancia (2008). "The Skill Bias of World Trade," *Economic Journal* 118, 927-960.

[17] Epifani, P. and G. Gancia (2009). "Trade, Markup Heterogeneity and Misallocations," CEPR DP 7217.

[18] Feenstra, R. (2003). "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity," *Economics Letters* 78, 79-86.

[19] Gancia, G. and F. Zilibotti (2009). "Technological Change and the Wealth of Nations," *Annual Review of Economics* 1, 93-120.

[20] Hall, R. E. (1988). "The Relations Between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy* 96, 921-947.

[21] Helpman, E. and P. Krugman (1985). *Market Structure and Foreign Trade* (Cambridge, MA: MIT Press).

[22] Hsieh, C. T. and P. Klenow (2009) "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics* 124, 1403-1448.

[23] Koeniger, W. and O. Licandro (2006). "On the Use of Substitutability as a Measure of Competition," *Topics in Macroeconomics* 6-1.

[24] Kreps, D. M. and J. A. Scheinkman (1983). "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics* 14, 326-337.

[25] Krugman, P. (1979). "Increasing Returns, Monopolistic Competition and International Trade," *Journal of International Economics* 9, 469-479.

[26] Johnson, H. G. (1965). "Optimal Trade Intervention in the Presence of Domestic Distortions," in R. E. Baldwin et al. (eds.), *Trade, Growth and the Balance of Payments* (Chicago: Rand-McNally).

[27] Jones, C. (2009). "Intermediate Goods and Weak Links: A Theory of Economic Development," mimeo Stanford.

[28] Lancaster, K. (1979). *Variety, Equity and Efficiency* (New York: Columbia University).

[29] Lerner, A. P. (1934). "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies* 1, 157-175.

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

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

[32] Neary, J. P. (2003). "Globalization and Market Structure," *Journal of the European Economic Association* 1, 245-271.

[33] Restuccia, D. and R. Rogerson (2008). "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics* 11, 707-720.

[34] Roberts, M. J. and J. R. Tybout (1996). *Industrial Evolution in Developing Countries* (Oxford: Oxford University Press).

[35] Salop, S. C. (1979). "Monopolistic Competition with Outside Goods," *Bell Journal of Economics* 10, 141-156.

[36] Samuleson, P. (1947). *Foundations of Economic Analysis* (Cambridge, MA: Harvard University Press).

[37] Song, Z., K. Storesletten and F. Zilibotti (2009). "Growing Like China," *American Economic Review*, forthcoming.

[38] Tybout, J. R. (2003). "Plant-and Firm-Level Evidence on 'New' Trade Theories," in E. K. Choi and J. Harrigan (eds.), *Handbook of International Economics* (Oxford: Basil-Blackwell).

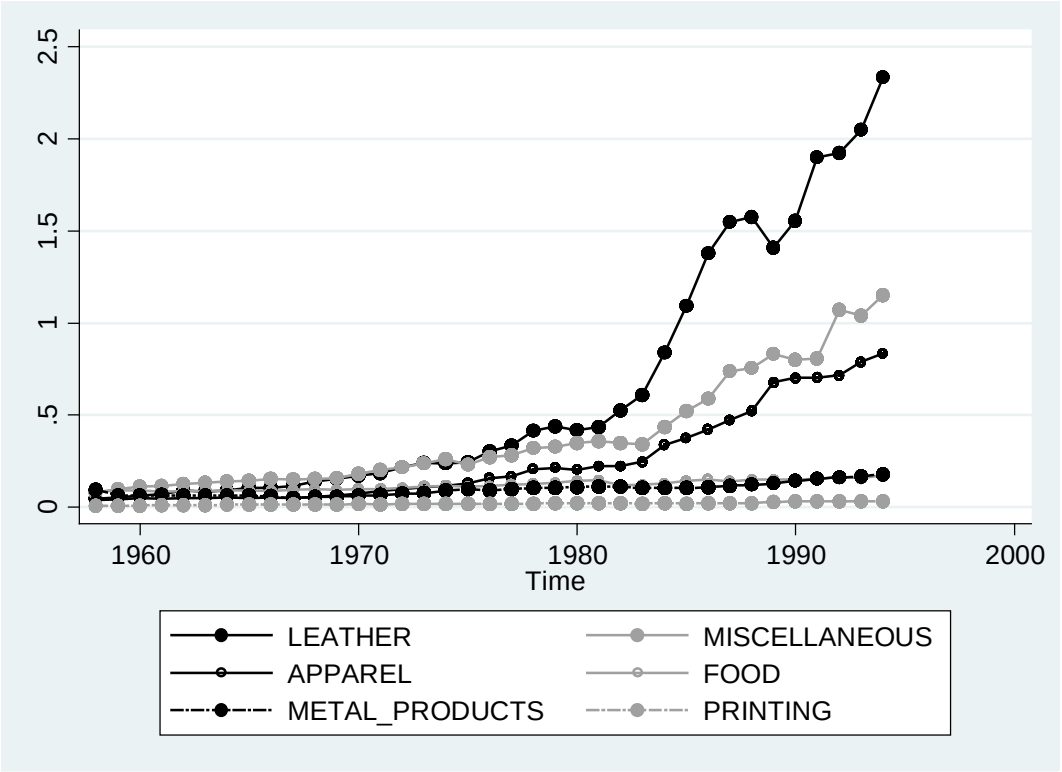

Figure 1

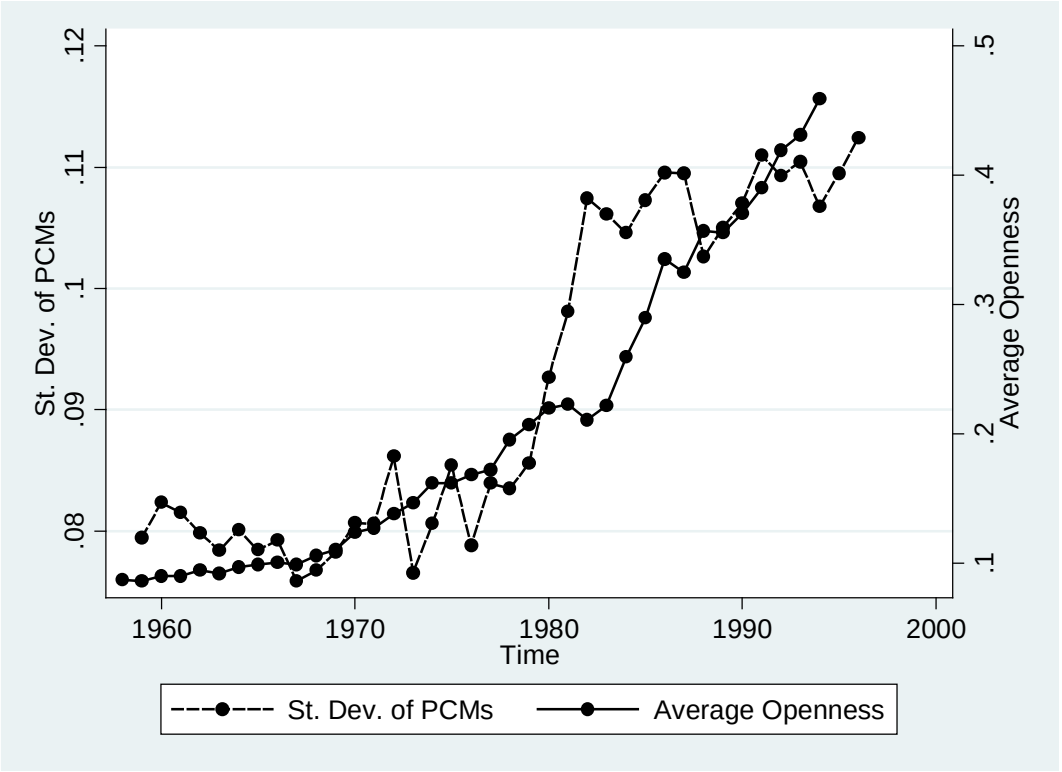

Figure 2

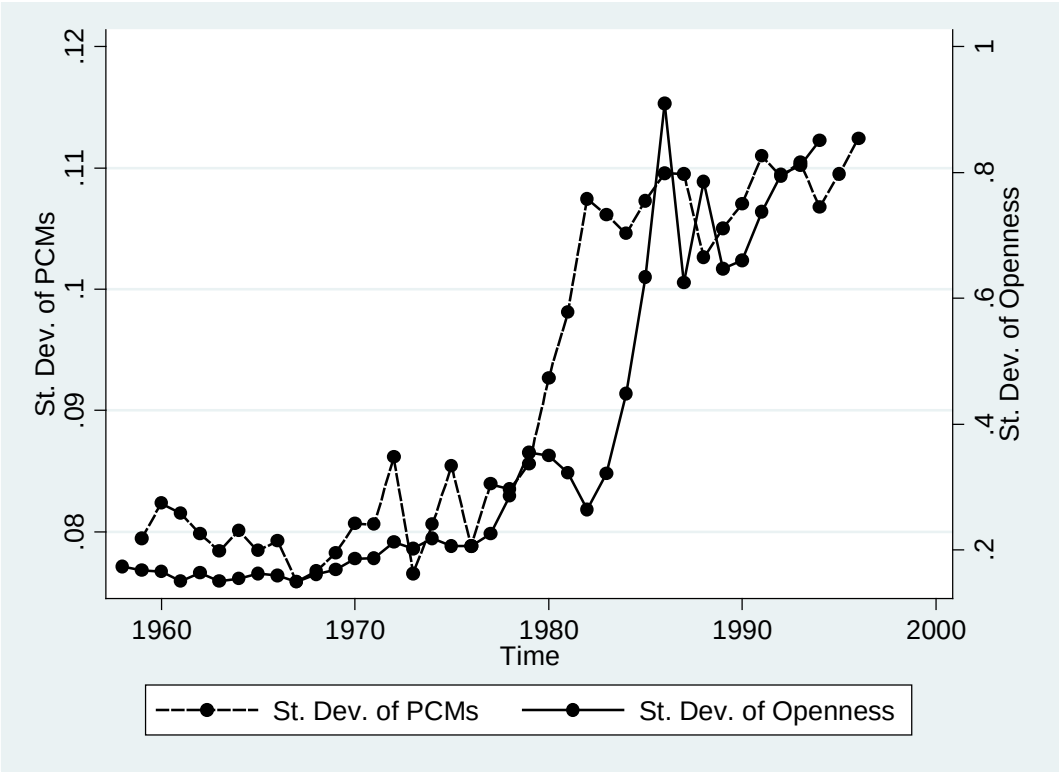

Figure 3

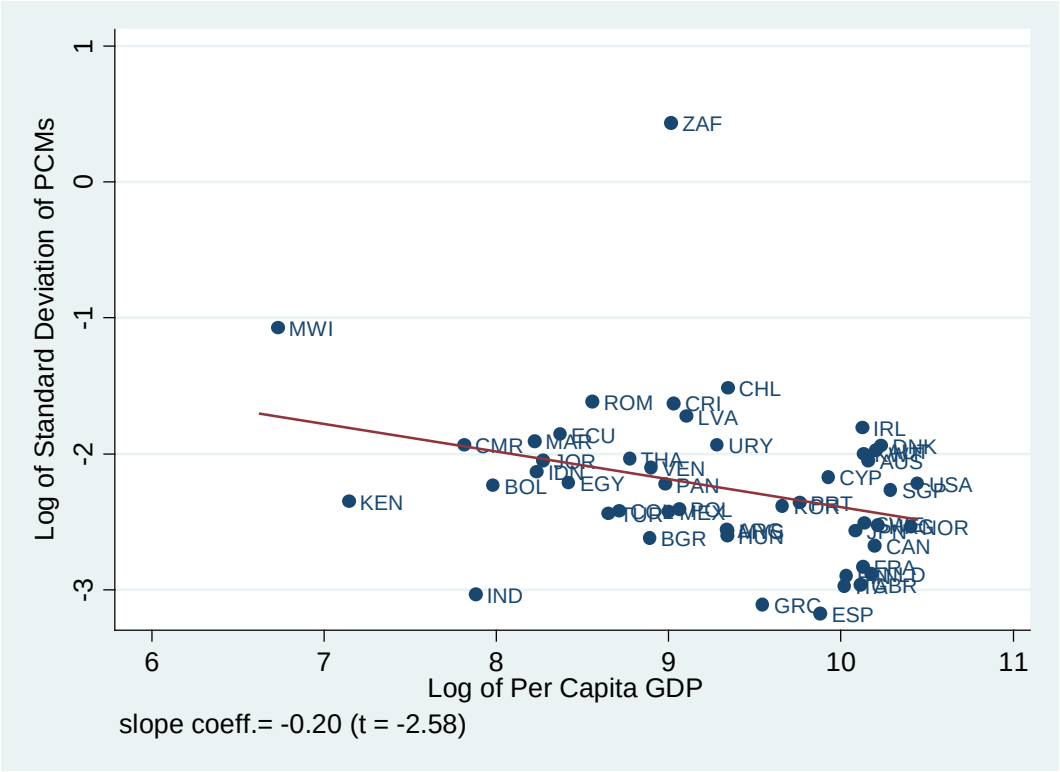

Figure 4

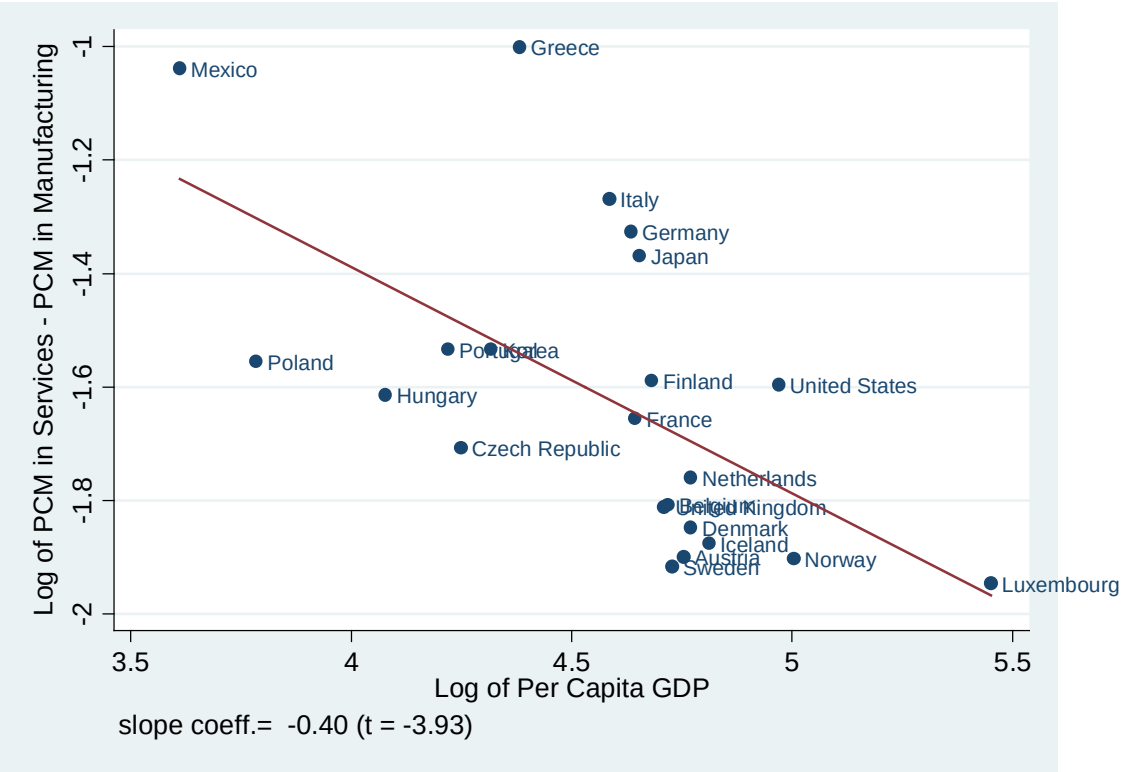

Figure 5

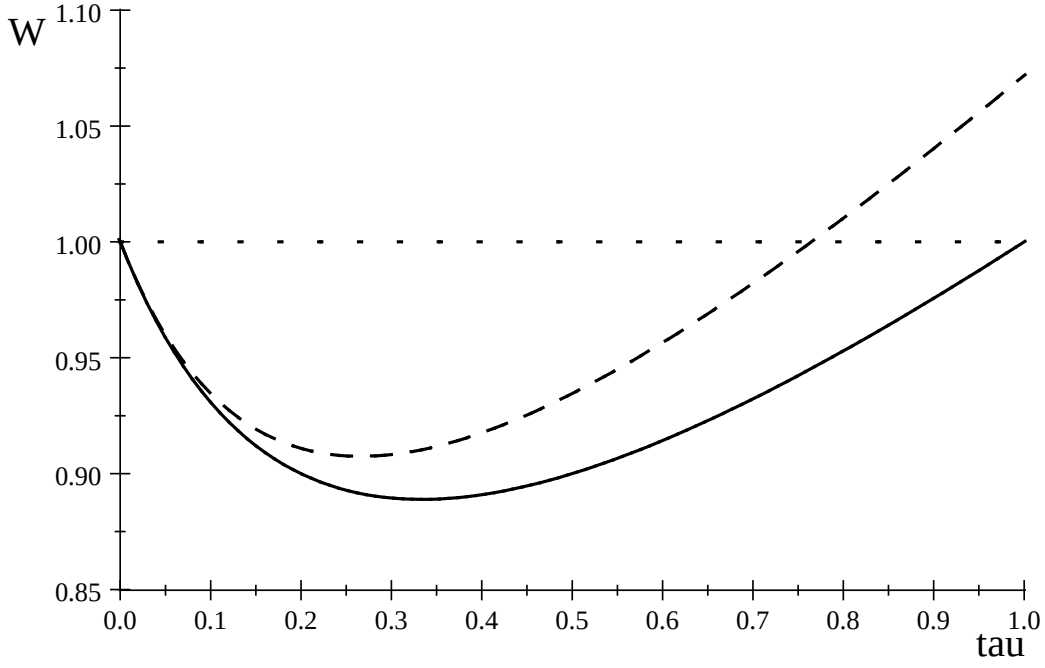

Figure 6. Trade and Welfare: solid line ν=0, broken line ν>0

Table 1. Fixed-Effects Regressions for the Standard Deviation of PCMs

Dependent Variable: Standard Deviation of PCMs within 3-digit SIC Industries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
St.Dev. Openness	0.016***	0.010***	0.010***	0.012***	0.012***	0.012**	0.011***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.005]	[0.003]
St.Dev. TFP			0.041***	0.063***	0.072***	0.072***	0.072***
			[0.008]	[0.009]	[0.010]	[0.009]	[0.010]
St.Dev. H/L			0.008	0.031***	0.032***	0.032***	0.035***
			[0.007]	[0.008]	[0.010]	[0.010]	[0.010]
St.Dev. K/Y			0.001***	-0.000*	-0.000**	-0.000**	-0.000***
			[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Average Openness					0.000		-0.002
					[0.011]		[0.011]
Average TFP					-0.033***	-0.033***	-0.005
					[0.010]	[0.010]	[0.011]
Average H/L					-0.000	-0.000	-0.006
					[0.013]	[0.013]	[0.013]
Average K/Y					0.000**	0.000**	0.000
					[0.000]	[0.000]	[0.000]
Average PCM							-0.136***
							[0.024]
Time Dummies		YES	YES	YES	YES	YES	YES
Industry-Specific Trends				YES	YES	YES	YES
Observations	3456	3456	3456	3456	3456	3456	3456
# 3-digit SIC industries	96	96	96	96	96	96	96
R-squared (within)	0.03	0.11	0.14	0.39	0.39	0.39	0.41

Notes: Fixed-Effects (within) estimates with robust standard errors in parentheses. ***,**,* = significant at the 1, 5 and 10-percent levels, respectively. The mean and standard deviation of all variables is computed within 3-digit SIC industries. In column 6, estimation is by Instrumental Variables, with the second moment of openness instrumented with its first moment. Data sources: NBER Productivity Database (by Bartelsman and Gray) and NBER Trade Database (by Feenstra).

Table 2. Welfare Cost of Markup Dispersion

	Constant Φ_i			Time-varying Φ_i		
	W/W^{FB} (1959)	W/W^{FB} (1994)	$\Delta W/W^{FB}$ (94-59)	W/W^{FB} (1959)	W/W^{FB} (1994)	$\Delta W/W^{FB}$ (94-59)
$\sigma = 2$ ($\alpha = 0.5$)	0.983	0.97	-1.3%	0.986	0.961	-2.5%
$\sigma = 5$ ($\alpha = 0.8$)	0.947	0.913	-3.6%	0.957	0.863	-9.8%
$\sigma = 10$ ($\alpha = 0.9$)	0.854	0.785	-8.1%	0.88	0.666	-24.3%