Misallocation Under Trade Liberalization

Yan Bai
University of Rochester
NBER

Keyu Jin
LSE
CEPR

Dan Lu
University of Rochester

July 11, 2019

Abstract

We analyze the impact of trade liberalization in an economy with distortions. In an open-economy model that features firm-level heterogeneity in both productivity and distortions, trade can lead to welfare losses by exacerbating the misallocation of resources. We characterize welfare under misallocation in an open economy, and derive an analogue to the well-known ACR result. Input and output shares become the additional sufficient statistics needed to infer welfare. A quantitative assessment using Chinese manufacturing data shows that trade integration has led to a TFP loss and a significantly smaller welfare gain compared to the standard numbers implied by the ACR formula (12.7%).

Keywords: Capital and labor wedges, misallocation, trade liberalization, gains from trade, industrial policy

JEL classification: E23 F12 F14 F63 L25 O47

1 Introduction

The question of how much developing countries benefit from opening up to goods trade is a time-honoured subject, both of practical import and intellectual interest. Much has been understood about the nature and type of gains to trade, thanks to the remarkable progress made in the field of international trade in recent decades. Less clear, however, is why certain developing countries have benefited from trade more than others, and why certain countries have seemingly not benefitted much at all.[1] Theoretical models suggest that developing countries have the most to gain from trade. One mechanism that has recently received much attention is that if trade liberalization can induce reallocation of resources from less to more productive firms, aggregate productivity and welfare will rise in turn.

But developing countries are different in another respect: they are also subject to prevalent policy and institutional distortions. Examples include taxes and subsidies to certain firms, implicit guarantees and bailouts, preferential access to land and capital, industrial policy and export promotion policies—common themes in developing countries. These distortions are believed to be key reasons why inefficient but politically-connected private firms and state-owned companies (SOE) in China have survived and even thrived. Implicit and explicit support for these firms combined with limited exit mechanisms for many SOEs have weakened firm selection effects, the upshot of which is a drag on aggregate productivity. Many believe that joining the WTO can potentially alleviate some of these problems, by inviting more foreign direct competition and helping the more productive firms expand.

How effective is this mechanism? Can trade necessarily improve allocations? Does trade necessarily lead to welfare gains for developing countries? These issues are far from obvious, as alluded to by Rodriguez-Clare (2018), " [a] complication that may matter for the computation of the gains from trade is the presence of domestic distortions." Thus, the goal of the paper is to provide an answer to this question using the discipline of a general equilibrium model of trade that incorporates idiosyncratic firm-level distortions. We then use Chinese manufacturing data to quantify the effect of trade liberalization, as well as to conduct empirical investigations relevant for our mechanisms.

[1] For example, Waugh (2010) shows, in large sample of countries, that poor countries do not systematically gain more from trade.

Our modelling framework incorporates firm-specific distortions into a two-country Melitz model. There are two dimensions of heterogeneity at the firm-level: productivity and distortions. These distortions are assumed to be exogenous output wedges or factor wedges, following Hsieh and Klenow (2009), henceforth HK (other works include Baily, Hulten, and Campbell (1992), Restuccia and Rogerson (2008), Bartelsman, Haltiwanger, and Scarpetta (2009), etc). They drive differences in the marginal products across firms. Different from HK, however, our model allows for firm entry and exit, and international trade. The endogenous mechanism of entry and selection is crucial in our setting and is what can bring about efficiency losses associated with trade.

The paper makes three contributions. First, it derives a general expression for welfare under trade that encompasses the case with and without distortions. We show that a special case in which there is only heterogeneity in productivity, which follows a Pareto distribution, yields the well-known result of ACR (after Arkolakis, Costinot, and Rodríguez-Clare (2012)); a special case in which there is only heterogeneity in distortion with Pareto distribution yields an analogue to ACR under misallocation. These two are special cases in which firm selection is driven either solely by productivity or solely by distortions. This explicit expression we derive for the special case yields insight on exactly why there could be a welfare loss associated with trade. A crucial sufficient statistic is the gap between the aggregate input and output share. If the required inputs used for production is greater than the output share it yields, then the reduction in allocative efficiency arising from a reallocation of resources can bring about a welfare loss. In the special case, welfare is always lower in an open compared to a closed economy.

In the more general case, productivity and distortions are competing in their impact on firm selection. The relative strength of the two depends on their joint distribution, and micro-level information still matters for welfare. The impact of allocative efficiency arising from a reallocation of resources can again be captured by the input-output gap, along with additional structural microeconomic parameters.

These results resonate with a crucial result in Baqaee and Farhi (2018), which shows that the aggregate impact of micro-shocks in an economy with distortions depends crucially on the change in allocative efficiency. Their analysis takes place in a closed-economy setting,

with an emphasis on firm-specific productivity shocks, while in our open economy model the main shock is a trade cost shock.[2]

Contrary to the mechanism underpinning the Melitz (2003) model and its extensions, i.e. that trade can induce a reallocation of resources from low productivity to high productivity firms, the presence of distortions can bring about the opposite and exacerbate misallocation. The reason is simple: distortions (for instance, tax and subsidies) act as a veil to a firm's true productivity. A firm may be producing in the market not because it is inherently productive, but because it is sufficiently subsidized. A mass of highly-subsidized but not adequately productive firms will export and expand, at the cost of other more productive firms. The high productivity/ high tax firms which were marginally able to survive in the domestic market would be driven out as the other firms gain market share and drive up costs. In other words, the selection effect which brings about efficiency gains in the Melitz-type model is no longer based solely on productivity; it is determined jointly by firm productivity and distortions. Trade may thus *lower* the average productivity of firms. Depending on the prevalence and nature of distortions, different countries may have different experiences with trade liberalization.

Second, we conduct an empirical investigation on some of the key implications of our model. Although the study can apply to any group of countries, we focus on Chinese manufacturing data for the simple reason that this country is well known for its prevalent State interventions and policies; that a body of work has shown (see references below) that idiosyncratic distortions explain the majority of the dispersion in marginal products; and that trade liberalization has been an important recent phenomenon. The exercise is also well-suited to compare with HK's findings on how distortions affect aggregate productivity for China—-with the addition of endogenous firm selection and trade.

Employing Chinese manufacturing firm-level data, we first use a 'measured wedge' to demonstrate some of its key features. These measured wedges reveal that there is substantial misallocation at the firm level. These features and patterns are also consistent with common perception on what types of firms might enjoy special privileges, and hence have

[2]This paper analyzes the macroeconomic impact of a microeconomic shock in the presence of distortions, in a closed-economy setting. They provide a general formula for aggregating microeconomic shocks in general equilibrium economies with distortions.

'lower taxes'. We then evaluate whether key implications of the model receive empirical support. Specifically, the model implies that selection will drive a positive relationship between productivity and taxes, for the reason that firms facing higher taxes must be more productive to survive. This is true in the data—there is a large dispersion of the measured wedges across firms, as well as a robust positive relationship between wedges and productivity.[3] Second, the model indicates that given productivity, firms that face lower taxes will tend to export. Data indicates that controlling for productivity, exporters indeed have lower wedges.

The third contribution is to use our estimated model to quantify the impact of trade on welfare and aggregate productivity. We run counterfactual experiments for local changes in trade cost to compute the TFP and welfare effects of trade liberalization. We also run counterfactual experiments for domestic reforms. Our main conclusion is that the trade gains are much smaller when taking into account distortions; that there is a TFP loss of 3% as opposed to a TFP gain of 13.3% in the case without distortions.

It is important to point out that in the quantitative analyses we do not use directly the empirically-measured wedges, observed correlations, or distributions in the data to assess the impact of trade on welfare in the presence of distortions. The reason is that the *observed* statistics are not the *underlying* ones: existing firms have been subject to selection and thus their observed distributions are not the true ones. The same reasoning goes for the observed correlation between productivity and wedges. Both selection and fixed costs can drive a positive relationship between the two. For these reasons, the approach adopted in the quantitative exercises is to estimate the underlying joint distribution of wedges and productivity, costs of producing and exporting so as to match the observed patterns of firms' outputs, inputs, and exports. Importantly, the difficulty in inferring these wedges directly from the data means that a structural model in conjunction to micro data is needed. On this basis, we evaluate how the presence of distortions change the impact of trade on

[3]Different from Hsieh and Klenow (2009), this correlation matters. Typically, this correlation is taken to be exogenous. In works that assume a perfect correlation between distortions and productivity (such as Costa-Scottini (2018) and Ho (2010)), there are always TFP and welfare gains to trade, in contrast to losses that can arise from our model. We demonstrate why this correlation cannot be taken to be exogenous when there is firm selection and fixed cost. Moreover, this assumption counters key empirical facts. We discuss this at the end of Section3.

productivity and welfare, and how much trade has contributed to Chinese growth in a decomposition exercise. This contrasts with the reduced-form approach adopted in Berhou et al. (2018), which uses empirically measured revenue productivity to assess the impact of trade reforms.[4]

Thus, a key point is that micro structure is needed to assess welfare gains to trade when distortions drive misallocation across firms. This departs from the seminal contribution of Arkolakis, Costinot, and Rodríguez-Clare (2012), which shows that total gains from trade are identical in a large class of models, neatly summarized by the aggregate trade elasticity. In our model, the presence of idiosyncratic distortions can deliver large quantitative and even qualitatively different results compared to ACR estimates.

Another point stressed in this paper is that the 'misallocation of resources' goes beyond the observed misallocation among a set of operating firms. Because policy distortions also act as a barrier to entry (and exit), there is also misallocation among potential entrants and incumbents– firms that should have entered the market in an efficient economy that couldn't, and firms that should have otherwise exited but have not. We show that theoretically, this unobserved misallocation can be as significant as the observed one. This echoes a few empirical studies that have demonstrated the importance of entry and exit for China's growth.[5]

The source of misallocation of resources is not the focus of our work, though we take the particular stance that firm-specific 'distortions' due to policy and institutional features is a major cause of misallocation in the case of China. Distortions there manifest themselves in the form of substantial privileges of state owned enterprises over private firms, of connected private firms, or of firms belonging to particular locations. Specific policies that can drive these wedges include implicit subsidies such as soft budget constraints, favorable costs of capital, preferential tax treatments and implicit guarantees. Firms with

[4]Berhou et al. (2018) theoretically and empirically assesses the impact of trade reforms for 14 European countries and 20 industries over the period 1998-2011. They find that trade reforms have ambiguous effects on measured revenue productivity in the theory, and it is positive in the data. However, this measured revenue productivity does not map onto actual welfare as it measures productivity, fixed costs *and* distortions. We access the impact of trade on actual welfare and productive capacity in China.

[5] Biesebroeck and Zhang (2012) find that the entry and exit account for 72 percent of Chinese manufacturing productivity growth. The creation and selection of new firms in China's non-state sector has been particularly important.

political connections having access to special deals and receiving substantial benefits are also widely documented (see Guo, Jiang, Kim and Xu (2014) and Bai, Hsieh, Song (2019)). Wu (2015) conducts an empirical analysis and finds that policy distortions can be explained by investment promoting programs that favor such firms.

There is also substantial evidence coming from a number of papers that idiosyncratic firm-distortions due to policy and institutional features account for a large part of the observed dispersion in marginal products across firms in China. In principle, misallocation can arise from a variety of factors. They could be technological frictions, such as adjustment costs, information frictions, financial frictions, or markups. However, various works using different approaches to disentangle the various factors have come to similar conclusions that policy distortions are elemental. Wu (2015) find that they account for the majority of the observed misallocation of capital, as opposed to financial frictions. Using a different approach and modeling framework, David and Venkateswaran (2017) find also that firm-specific distortions, rather than technological or information frictions, account for the majority of the observed dispersions in marginal products. Bai, Lu and Tian (2017) disciplines financial frictions with firms' financing patterns, sales distribution and change of capital. They find that financial frictions cannot explain the observed relation between firms' measured distortions and size. Furthermore, we conduct empirical analyses to show that these wedges are systematically related to certain firm characteristics, in a way that echoes prior findings. To ensure that measured dispersion is not due to measurement error, we use three alternative approaches utilizing panel data to demonstrate the minor role it plays.[6]

Many of these distortions are presumably idiosyncratic and unrelated to trade. Examples are in the legion: Chinese firms such as the car manufacturing company Chery, have enjoyed easy access to land and capital from the Wuhu government. Foxconn, the world's largest electronics contracter manufacturer has enjoyed substantial tax breaks from many provinces including industrial land at significantly discounted prices. Tesla has recently received free land and subsidies from the local government of Shanghai. A recent study by Chen and Kung (2019) demonstrate the firms that are connected with political elites were

[6]However, Bils and Klenow 2017 cannot rule out multiplicative measurement errors.

able to obtain land at 80 to 90 percent discount over the period 2004-2016.

Thus, the focal point in this paper is on domestic policy distortions. Still, one can ask whether some of the large dispersion of marginal products reflects endogenous distortions—those that can potentially change with trade liberalization. As a robustness check we examine a model of endogenous distortions with variable markup, and ask whether trade can mitigate these distortions and the misallocation of resources. Section 5.1 takes up a variable markup model. We show that these models 1) yield some obvious counterfactual predictions on the relationship between exporters and wedges; 2) that markups alone also explains little of the dispersion in wedges. To match the observed correlation and dispersion one would still need to include exogenous distortions. Moreover, the attendant pro-competitive effects in a model with endogenous markups may be 'elusive', as pointed out by Arkolakis et al. (2017).[7]

This paper focuses specically on the firm selection channel associated with trade, and how that drastically alters when there is misallocation. As such, the framework abstracts from other types of gains to trade, such as trade-induced technological diffusion (Alvarez, Buera, and Lucas (2014) and Buera and Oberfield (2015)), adoption (Perla, Tonetti, Waugh (2015) and Sampson (2015)) and innovation (Atkeson and Burstein 2010). While these mechanisms in principle work to increase the gains to trade, the quantitative impact is not so clear. Perla et al. (2015) and Atkeson and Burstein (2015), for instance, find that trade gains are not too different from ACR gains.[8] These positive effects are important but are abstracted from the current model. If firm innovation is to be introduced into the baseline model, one would need to consider the fact that distortions not only affect production decisions, but potentially also innovation decisions. These considerations go beyond the scope of this paper. We also do not consider how trade can reduce domestic distortions, for example if concurrent domestic reforms are requisite for joining the WTO, or if quotas

[7] This paper makes the point that when more productive firms expand at the expense of less productive ones, thanks to trade, the aggregate markup tends to rise. Thus, overall, trade models with endogenous markups do not necessarily generate higher gains from trade.

[8] In Perla et al. (2015), there is trade-induced within-firm productivity improvements. However, their aggregate growth effects come with costs—losses in variety and reallocation of resources away from goods production. Thus, the aggregate effect on welfare is similar to ACR gains. Atkeson and Burstein (2010) show that general equilibrium effects limits the first-order effects on aggregate productivity even when there is firm-level innovation.

are removed (see Khandelwal, Schott, and Wei 2013). As a robustness check, however, we allow for firms to face a different distribution of distortions when they start to export and examine welfare and efficiency gains therein.

A key message of this paper is that in order for developing countries to reap the full gains of trade, simultaneous or antecedent domestic reforms aimed at reducing policy distortions may be crucial. Based on our structural model, trade contributed only 8% to growth over the period of 1998 to 2005 in China. This echoes with the findings of Tombe and Zhu (2018) that find similar numbers despite employing a very different framework with migration flows. The policy implication drawn from this framework is consistent with works indicating that policies aimed to neutralize domestic distortions may be complementary to trade liberalization (Chang, Kaltani and Loayza (2009) and Harrison and Rodríguez-Clare (2010)). It counters other claims that trade liberalization should take precedence owing to positive firm selection (Asturias, Hur, Kehoe, and Ruhl 2016).[9]

In sum, this paper shows that experiences of trade liberalization in developing countries should not be considered to be independent of micro-level distortions to which they are subject. Our paper demonstrates that the presence of policy distortions have a first-order quantitative effect on the gains to trade. The organization of the paper is as follows: Section 2 derives a theoretical framework of trade gains under misallocation. Section 3 investigates patterns of measured wedges among Chinese manufacturing firms. Section 4 provides a quantitative assessment on the impact of trade liberalization under misallocation, with various extensions of the benchmark framework. Section 6 concludes.

[9]They show that the best sequence of reforms is to first decrease trade costs, then to improve contract enforcement, and, finally, to decrease the cost of firm creation.The reason is that an increase in competition leads to an expansion of productive firms and crowding out of less efficient ones. By liberalizing international trade first so as to impose firm selection early, inefficient firms are prevented from entering later when contract enforcement and firm entry costs are reformed. In contrast, we show that the selection mechanism is substantially weakened in the presence of distortions.

2 Theoretical Framework

2.1 Baseline Model

The world consists of two large open economies. In each country i, there is a measure L^i of identical consumers. The two economies can differ in population, L, which is immobile across countries and inelastic in supply.

Consumers. A representative consumer in the Home country chooses the amount of final goods C in order to maximize utility $u(C)$, subject to

$$PC = wL + \Pi + T,$$

where wL is labor income, Π is dividend income, and T is the amount of lump-sum transfers received from the government.

Final Goods Producers. Final goods producers are perfectly competitive, and combine intermediate goods using a CES production function

$$Q = \left[\int_{\omega\in\Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}},$$

where σ is the elasticity of substitution across intermediate goods, and Ω is the endogenous set of goods. The corresponding final goods price index is thus

$$P = \left[\int_{\omega\in\Omega} p(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{1-\sigma}},$$

where $p(\omega)$ is the price of good ω in the market. The individual demand for this good is thus given by

$$q(\omega) = \frac{p(\omega)^{-\sigma}}{P^{-\sigma}} Q.$$

Intermediate Goods Producers. There is a competitive fringe of potential entrants that can enter by paying a sunk entry cost of f_e units of labor. Potential entrants face uncertainty about their productivity in the industry. They also face a stochastic revenue wedge τ, which can be seen as a tax (>1) or subsidy (<1) on every pq earned. Once the sunk entry cost is

paid, a firm draws its productivity φ and τ from a fixed joint distribution, $g(\varphi, \tau)$ over $\varphi \in (0, \infty), \tau \in (0, \infty)$. Productivity and the revenue wedge remain the same after entry, but firms face a constant exogenous probability of death δ, which induces steady-state entry and exit of firms in the model.

Firms are monopolistically competitive. Production of each intermediate good entails fixed production cost of f units of labor and a constant variable cost that depends on firm productivity. The total labor required to produce $q(\varphi)$ units of a variety is therefore: [10]

$$\ell = f + \frac{q}{\varphi}.$$

Productivity φ is idiosyncratic and independent across firms. The existence of a fixed production cost means that only a subset of firms produces—those that draw a sufficiently low productivity cannot generate enough variable profits to cover the fixed production cost. If firms decide to export, they face a fixed exporting cost of f_x units of labor and iceberg variable costs of trade τ_x, which is greater than 1. Firms with the same productivity and distortion behave identically, and thus we can index firms by their (φ, τ) combination.

An intermediate goods firm thus solves the following problem

$$\max_{p,q} \quad \frac{pq}{\tau} - \frac{w}{\varphi} q - wf$$

subject to the demand function

$$q = \frac{p^{-\sigma}}{P^{-\sigma}} Q, \tag{1}$$

henceforward suppressing ω for convenience. Firms are infinitesimally small, and thus take the aggregate price index as given. Equating the after-tax marginal revenue with marginal costs yields the standard result that equilibrium prices are a mark-up over marginal costs:

$$p = \frac{\sigma}{\sigma - 1} \frac{w\tau}{\varphi}. \tag{2}$$

[10] We can easily extend the production including capital, $k^{\alpha}\ell^{1-\alpha}$. The unit cost for producing q or fixed cost is $\alpha^{-\alpha}(1-\alpha)^{\alpha-1} w^{1-\alpha} r_k^{\alpha}$ where r_k is the rental cost of capital. In our model, we introduce one heterogeneous distortions at the firm level, and our τ includes distortions that increase the marginal products of capital and labor by the same proportion as an output distortion. In the data, there are distortions that affect both capital and labor and distortions that change the marginal product of one of the factors relative to the other. In our quantitative exercises, we include both capital and labor, and the distortions on both factors.

Optimal profits are then

$$\pi = \sigma^{-\sigma}(\sigma-1)^{\sigma-1}P^{\sigma}Q\tau^{-\sigma}w^{1-\sigma}\varphi^{\sigma-1} - wf. \tag{3}$$

It immediately follows that given the fixed cost of production, there is a zero-profit cutoff productivity below which firms would choose not to produce, and exit the market. Thus, a firm would choose to produce only if $\varphi \geq \varphi^*(\tau)$. This cutoff productivity level satisfies

$$\varphi^*(\tau) = \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1}\left[\frac{wf}{P^{\sigma}Q}\right]^{\frac{1}{\sigma-1}} w\tau^{\frac{\sigma}{\sigma-1}}. \tag{4}$$

The cutoff productivity is now a function of the firm-specific distortion, and differs across firms facing different levels of distortions. Firms with a higher tax τ will have a higher cutoff for productivity. This means that low productivity firms that would have been otherwise excluded from the market can now enter the market and survive if sufficiently subsidized.

The government's budget is balanced so that

$$T = \int_{\omega\in\Sigma}\left(1-\frac{1}{\tau}\right)\ p(\omega)q(\omega)d\omega,$$

where Σ is the endogenous set of home products.

With trade, firms now have the option of exporting abroad. If a Home firm exports to the Foreign economy, it solves the following problem

$$\max\ \frac{p_x q_f}{\tau} - \frac{w}{\varphi}\tau_x q_f - wf_x$$

subject to the foreign demand function

$$q = \frac{p_x^{-\sigma}}{P_f^{-\sigma}}Q_f,$$

where P_f and Q_f denote the aggregate price index and demand in Foreign. Given the same

constant elasticity of demand in the domestic and export markets, equilibrium prices in the export market are a constant multiple of those in the domestic market:

$$p_x(\varphi, \tau) = \frac{\sigma}{\sigma - 1} \frac{w\tau_x \tau}{\varphi},$$

The optimal profit from servicing the foreign market,

$$\pi_x = \sigma^{-\sigma}(\sigma - 1)^{\sigma-1} P_f^{\sigma} Q_f \tau^{-\sigma} (w\tau_x)^{1-\sigma} \varphi^{\sigma-1} - w f_x, \tag{5}$$

yields an optimal cutoff for exporting:

$$\varphi_x^*(\tau) = \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma - 1} \left[\frac{w f_x \tau_x^{\sigma-1}}{P_f^{\sigma} Q_f} \right]^{\frac{1}{\sigma-1}} w\tau^{\frac{\sigma}{\sigma-1}}. \tag{6}$$

Consumer love of variety, a fixed production cost and additional fixed cost of exporting, mean that firms would never export without also selling in the domestic market. There are, hence, two cutoff productivities relevant for the domestic economy: one for entering the domestic market as given by (4) and one for entering the foreign market, as given by (6). To the extent that taxes τ are constant across firms, the ratio $\varphi_x^*(\tau)/\varphi^*(\tau)$ is a constant and is greater than 1 so long as $\frac{\tau_x^{\sigma-1} f_x}{f} \frac{P^{\sigma} Q}{P_f^{\sigma} Q_f} > 1$. Analogously, firms in the Foreign country, which draw their productivity from a distribution $g_f(\varphi, \tau)$, are subject to two cutoff productivities, one for servicing their domestic market, and one for exporting to the Home economy

$$\varphi_f^*(\tau) = \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma - 1} \left[\frac{w_f f}{P_f^{\sigma} Q_f} \right]^{\frac{1}{\sigma-1}} w_f \tau^{\frac{\sigma}{\sigma-1}}, \tag{7}$$

$$\varphi_{xf}^*(\tau) = \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma - 1} \left[\frac{w_f f_x \tau_x^{\sigma-1}}{P^{\sigma} Q} \right]^{\frac{1}{\sigma-1}} w_f \tau^{\frac{\sigma}{\sigma-1}}. \tag{8}$$

where w_f denotes the foreign wages, and the fixed cost of producing and exporting are assumed to be identical in the two economies.

The steady-state industry equilibrium features a constant mass of firms entering and producing, along with a stationary ex-post distributions of productivity and taxes among

operational firms. With a constant level of productivity fixed upon entry, and a constant independent probability of firm death δ, the stationary ex-post distribution for productivity and taxes is a truncation of the ex-ante productivity-tax distribution, $g(\varphi, \tau)$, at the zero-profit cutoff productivity given by Eq.4:

$$\mu(\varphi, \tau) = \frac{g(\varphi, \tau)}{\int_0^\infty \int_{\varphi^*(\tau)}^\infty g(\varphi, \tau) \, d\varphi d\tau} \tag{9}$$

if $\varphi \geq \varphi^*(\tau)$; and 0 otherwise. The denominator is the probability of successful entry, denoted as

$$\omega_e = \int_0^\infty \int_{\varphi^*(\tau)}^\infty g(\varphi, \tau) \, d\varphi d\tau. \tag{10}$$

The probability of exporting conditional on entry is given by

$$\omega_x = \int_0^\infty \int_{\varphi_x^*(\tau)}^\infty \mu(\varphi, \tau) \, d\varphi d\tau = \frac{\int_0^\infty \int_{\varphi_x^*(\tau)}^\infty g(\varphi, \tau) \, d\varphi d\tau}{\int_0^\infty \int_{\varphi^*(\tau)}^\infty g(\varphi, \tau) \, d\varphi d\tau}.$$

In an equilibrium with positive entry, the free entry condition requires that

$$\int \int_{\varphi^*(\tau)} \pi(\varphi, \tau) g(\varphi, \tau) d\varphi d\tau + \int \int_{\varphi_x^*(\tau)} \pi_x(\varphi, \tau) g(\varphi, \tau) d\varphi d\tau = \delta w f_e. \tag{11}$$

The first term is the expected profits from domestic sales conditional on entry, multiplied by the probability of entry. The second term is the expect profits from export sales conditional on exporting, multiplied by the probability of exporting. The free entry condition requires that their sum be equal to the entry costs (in terms of labor).

The free entry condition (11), combined with optimal profit functions (3) and (5) gives an expression for the price index P:

$$P^{1-\sigma} = \left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} \left[M \int_0^\infty \int_{\varphi^*(\tau)}^\infty \left(\frac{\varphi}{w\tau}\right)^{\sigma-1} \mu(\varphi, \tau) \, d\varphi d\tau + M_f (\tau_x)^{1-\sigma} \int_{\varphi_{xf}^*(\tau)}^\infty \left(\frac{\varphi}{w_f \tau}\right)^{\sigma-1} \mu_f(\varphi, \tau) \, d\varphi d\tau \right] \tag{12}$$

where M and M_f denote the measure of operating firms in Home and Foreign.

Goods market clearing. The assumption of a balanced trade results in

$$MP_f^{\sigma}Q_f(\tau_x w)^{1-\sigma}\int_0^{\infty}\int_{\varphi_x^*(\tau)}^{\infty}(\frac{\varphi}{\tau})^{\sigma-1}\mu\left(\varphi,\tau\right)d\varphi d\tau = M_f P^{\sigma}Q(\tau_x w_f)^{1-\sigma}\int_0^{\infty}\int_{\varphi_{xf}^*(\tau)}^{\infty}(\frac{\varphi}{\tau})^{\sigma-1}\mu_f\left(\varphi,\tau\right)d\varphi d\tau \tag{13}$$

Labor market clearing. In a stationary equilibrium in which the mass of firms M are constant in both economies, the labor market condition yields

$$M = \frac{L}{\sigma\left(\delta\frac{f_e}{\omega_e} + f + \omega_x f_x\right)} \tag{14}$$

where $\omega_e = \int_0^{\infty}\int_{\varphi^*(\tau)}^{\infty} g\left(\varphi,\tau\right)d\varphi d\tau$ is the probability of entry, and ω_x is the probability of exporting given by

$$\omega_x = \int_0^{\infty}\int_{\varphi_x^*(\tau)}^{\infty}\mu\left(\varphi,\tau\right)d\varphi d\tau = \frac{\int_0^{\infty}\int_{\varphi_x^*(\tau)}^{\infty} g\left(\varphi,\tau\right)d\varphi d\tau}{\int_0^{\infty}\int_{\varphi^*(\tau)}^{\infty} g\left(\varphi,\tau\right)d\varphi d\tau}.$$

A detailed derivation of the model is provided in Appendix 7.1.

Normalizing the Home country wage rate to 1, there are eleven equations, the zero cutoff productivities for domestic production and exporting (4), (6), and its foreign counterparts, the free entry conditions (11) along with its foreign counterpart, the definition of the Home and Foreign price indices(12), and a goods market clearing/balanced trade equation(13), along with the measure of firms (14) and its foreign counterpart. These equations yield the equilibrium consisting of eleven unknowns $\{\varphi^*(\tau)$, $\varphi_x^*(\tau)$, $\varphi_f^*(\tau)$, $\varphi_{fx}^*(\tau)$, P, P_f, Q, Q_f, w_f, M, $M_f\}$.

Proposition 1. *The allocations, entrants, and cutoff functions* $\{Q, Q_f, M, M_f, \varphi^*(\tau), \varphi_x^*(\tau)\}$ *are independent of mean wedge* $\bar{\tau}$. *Prices* $\{P, P_f, w_f\}$ *change proportionally with* $\bar{\tau}$, *i.e.* $P(\bar{\tau}_1)/P(\bar{\tau}_2) = \bar{\tau}_1/\bar{\tau}_2$, *similarly for* P_f *and* w_f.

2.2 Welfare and TFP under Symmetric Equilibrium

We proceed to show that under distortions, trade can induce a TFP loss: if more productive firms also face higher distortions, then resources can flow from high to low productivity

firms. Trade can also lead to the exiting of more productive firms, for the same reason. The size of the loss therefore depends on the degree of correlation between productivity and the wedges, ρ, as well as the correlation with the dispersion of the wedges σ_τ.

To illustrate the impact of domestic distortions on welfare and efficiency gains to trade, we start out by considering a symmetric equilibrium in which the two economies are identical—- facing the same level of domestic and trade distortions. We also make the assumption that φ and τ is jointly log-normally distributed, with means $\bar{\phi}$ and $\bar{\tau}$, standard deviations σ_φ and σ_τ, and correlation ρ.

In this symmetric equilibrium, aggregate TFP is synonymous with welfare in each economy. TFP is given by

$$TFP = \frac{\sigma-1}{\sigma}\left[M\int\int_{\varphi^*(\tau)}\left(\varphi\frac{MPL}{MPL_i}\right)^{\sigma-1}\mu\left(\varphi,\tau\right)d\varphi d\tau + M\int\int_{\varphi_x^*(\tau)}\left(\frac{\varphi}{\tau_x}\frac{MPL}{MPL_i}\right)^{\sigma-1}\mu\left(\varphi,\tau\right)d\varphi d\tau\right]^{\frac{1}{\sigma-1}}. \tag{15}$$

Equation (15) shows that the source of TFP loss in the presence of firm-level distortions can arise from a misallocation of resources, captured by dispersions in MPL/MPL_i, and a misallocation caused by selection and entry mechanisms as captured by M, φ^*, φ_x^* being different from their respective efficient levels.

We next derive a full expression for changes in welfare associated with changes in trade costs. In the absence of distortions, as shown in Arkolakis, Costinot, and Rodríguez-Clare (2012), welfare changes can be inferred using two variables across a wide class of trade models: (i) changes in the share of expenditure on domestic goods; and (ii) the elasticity of bilateral imports with respect to variable trade costs (the trade elasticity). Different trade models can have different micro-level predictions, sources of welfare gains, and different structural interpretations of the trade elasticity. But conditional on observed trade flows and an estimated trade elasticity, the welfare predictions are the same. Of course, the generality of this formulation relies on a certain set of macro-level restrictions. Melitz and Redding (2015) (henceforth MR) show that under more general distribution functions for productivity, the trade elasticity is no longer invariant to trade costs and across markets, and therefore no longer a sufficient statistic for welfare. Micro-level information is still important.

The following proposition gives a general expression for welfare, derived in Appendix 7.1:

Proposition 2. *With firm level distortions, the change of welfare associated with an iceberg cost shock is*

$$
\begin{aligned}
d\ln Q = \frac{1}{\gamma_{sd}+\sigma-1}\Big[& -d\ln\lambda \qquad (ACR) \\
& + d\ln M_e \qquad (MR) \\
& + \frac{\sigma}{\sigma-1}(\gamma_{sd}-\gamma_d)d\ln M_e \\
& - \left(\sigma-1+\frac{\sigma\gamma_{sd}}{\sigma-1}\right)d\ln\lambda + \left(\sigma-1+\frac{\sigma\gamma_d}{\sigma-1}\right)d\ln S_d\Big]
\end{aligned} \tag{16}
$$

where λ is the share of domestic output; S_d is the share of labor used in producing domestic goods; γ_d is the elasticity of the cumulative domestic market share for firms above the cutoff, with respect to the cutoff; and γ_{sd} is the same elasticity for S_d, with respect to the cutoff. [11]

1. Without distortions, $S_d = \lambda$ and $\gamma_{sd} = \gamma_d$. If productivity follows a Pareto distribution with parameter θ, $\gamma_d = \theta - \sigma + 1$, and $dlnM_e = 0$. Hence, $dlnW = \frac{1}{\theta}[-dln\lambda]$ as in ACR.

2. Under a general distribution function and without distortions, $S_d = \lambda$, $\gamma_{sd} = \gamma_d$, but they are not constant, and $dlnM_e \neq 0$. Hence, $dlnW = \frac{1}{\gamma_d(\varphi^*)+\sigma-1}[-dln\lambda + dlnM_e]$. Micro structure matters for $\gamma_d(\varphi^*)$ and welfare as in MR.

3. With homogenous productivity and if $1/\tau$ follows a Pareto distribution with parameter θ, $\gamma_d = \frac{\sigma-1}{\sigma}(\theta-\sigma+1)$ and $\gamma_{sd} = \frac{\sigma-1}{\sigma}(\theta-\sigma)$, hence $dlnW = \frac{\sigma}{\sigma-1}[dlnS_d - dln\lambda]$. Changes in the share of expenditure on domestic goods and the share of labor used in producing domestic goods are sufficient statistics for welfare change. In addition, the open economy has an unambiguously lower

11

$$\gamma_d = \frac{\int\left(\frac{\varphi^*(\tau)}{\tau}\right)^{\sigma-1} g(\varphi^*,\tau)\varphi^* d\tau}{\int\int_{\varphi^*(\tau)}(\frac{\varphi}{\tau})^{\sigma-1} g(\varphi,\tau)d\varphi d\tau}$$

is also known as the hazard function for the distribution of log firm size within the domestic market.

$$\gamma_{sd} = \frac{\int\left(\frac{\varphi^*(\tau)}{\tau}\right)^{\sigma-1}/\tau g(\varphi^*,\tau)\varphi^* d\tau}{\int\int_{\varphi^*(\tau)}(\frac{\varphi}{\tau})^{\sigma-1}/\tau g(\varphi,\tau)d\varphi d\tau}$$

is the hazard function for the distribution of log after-tax firm size (which is also proportional to the distribution of firm input) within the domestic market. See Appendix A for proof and details.

welfare than in the closed economy.

4. In general, for a local change in trade cost, information on the change of domestic (sales and employment) shares and the measure of entrants, the joint distribution of firms sales and inputs, and the marginal firms (hence we know γ_d and γ_{sd}) are sufficient for computing the associated welfare change.

The above proposition encapsulates welfare results for four different cases. The special case in which there is only heterogeneity in productivity, Pareto-distributed, gives rise to the ACR formula (point 1). The case under a more general productivity distribution gives rise to MR (point 2). With misallocation, the special case in which there is only heterogeneity in distortions, Pareto distributed, gives an analogue formula to ACR, where sufficient statistics can capture welfare (point 4).

The third and last terms in the welfare equation are the additional terms brought about by misallocation. These two terms capture the change in allocative efficiency due to resource reallocation. The resource reallocation is both one across existing firms (last term), and along the entry/exit dimension (third term). Without distortions, a firm's share of input is equal to its share of output, so that in aggregate, $S_d = \lambda$. In the presence of distortions, the two are no longer equal. The gap between input and output shares is informative about changes in allocative efficiency. If the change in required inputs exceeds the change in output it produces, i.e. $dlnS_d < dln\lambda$, resources reallocation has induced an efficiency loss.[12] This reallocation comes from the flows that take place among a given set of producers/exporters, the intensive margin, captured by $(\sigma - 1)(dlnS_d - d\ln\lambda)$; it also comes from a reallocation induced by the fact that cutoffs (due to distortions) have changed, the extensive margin, and the effect of which is summarized by the terms with the hazard rates (the elasticity of the cumulative domestic market share for firms above the cutoff, with respect to the cutoff) γ_{sd} and γ_d.

In the special case, welfare is unequivocally lower in the open economy compared to the closed economy (proof in Appendix). The intuition is as follows: under homogenous productivity, the efficient allocation for the closed economy dictates that firms have identical market shares— either all firms export or none of them export when the economy

[12]Note that $d\ln S_d \leq 0$ and $d\ln\lambda \leq 0$ in response to a trade cost shock reduction.

opens up to trade. Hence, the ex-post efficient allocation should also be equal shares for all firms. However, with distortions, the relatively subsidized firms produce more than in the efficient case, with the dispersion of sales (employment) reflecting the distortions. When opening up to trade, it is the relatively subsidized firms that export, in turn making the firm's distribution even more skewed—misallocation is exacerbated. The share of labor required in producing domestic goods ends up being less than the domestic output share. When firm selection is purely driven by distortions, allocative efficiency deteriorates when moving from autarky to an open economy. Welfare for local changes in trade costs, as stated in the proposition, displays a U-shape pattern: there is first a welfare loss for small reductions in trade cost, and then a welfare gain when trade costs fall even further. When trade costs are small, firm selection in general is less important (as at zero trade cost all firms export), and welfare starts to rise. However, it is still lower than in autarky.

These implications carry a similar flavor to findings in Baqaee and Farhi (2018), to which we draw an analogy in order to further fix ideas in our open economy. Their work studies the macroeconomic impact of microeconomic shocks to productivity and wedges, and show that the effect on output can be decomposed into a "pure technology effect" and a "resource allocation effect". The former is the change in output holding fixed the share of resources going to each user; the latter is the change in output resulting from the reallocation of shares of resources across users. To invoke a closer comparison to theirs, one can derive a similar result in our closed-economy setting. The output impact of a micro shock can also be decomposed into these two effects. [13]

This decomposition squarely demonstrates why an increase in a firm's productivity may not raise aggregate productivity: if a firm is relatively subsidized, then firm i's labor share is larger than its sales share, and its expansion means that the resource allocation component is negative. The same reasoning applies for the trade cost shock. Associated with a lowering of trade costs is a positive "pure technology" effect (the first two terms), as well

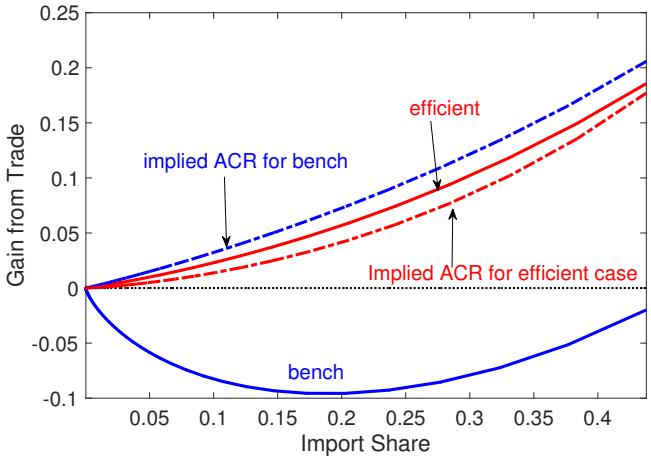
[13]The aggregate impact of an exogenous firm-specific productivity shock is given by $\frac{dlnY}{dln\varphi_i} = \frac{q_i^{\sigma-1}}{\int q_j^{\sigma-1}dj} + (\sigma-1)[\frac{q_i^{\sigma-1}}{\int q_j^{\sigma-1}dj} - \frac{q_i^{\sigma-1}\frac{1}{\tau_i}}{\int q_j^{\sigma-1}\frac{1}{\tau_j}dj}] = \frac{p_iq_i}{PQ} + (\sigma-1)[\frac{p_iq_i}{PQ} - \frac{\ell_i}{L}]$, which shows that it depends on firm i's sales share and the gap between its sales share and labor share. A firm is relatively subsidized if $1/\tau_i$ is larger than the average level of distortion.

as a reallocation effect (the last two terms of welfare in Proposition 2), where γ_d and γ_{sd} and the change of the share of labor summarize the reallocation of resources towards or away from more distorted firms. When a relatively subsidized firm expands due to trade, allocative efficiency deteriorates, and it is welfare-reducing.

2.3 Numerical Example

To unpack the theoretical results and to provide more intuition for the mechanisms that underpin these results, we next turn to a numerical example of the benchmark model. The joint distribution between productivity and distortions is taken to be joint log-normal. First, we show that using ACR to infer welfare changes when distortions are present gives rise to markedly different results. Figure 1.1 compares welfare under the efficient case, the benchmark case with distortions, and the welfare gains in each case using the ACR formula. The figure shows that ACR invariably predicts gains to trade, rather than losses. In the absence of distortions, ACR is a good approximation for welfare in the efficient case. But using ACR under distortions leads to a large departure: our benchmark results in this example predicts welfare losses rather than gains. Using aggregate observables to infer welfare gains as in ACR can thus be very misleading in the presence of distortions.

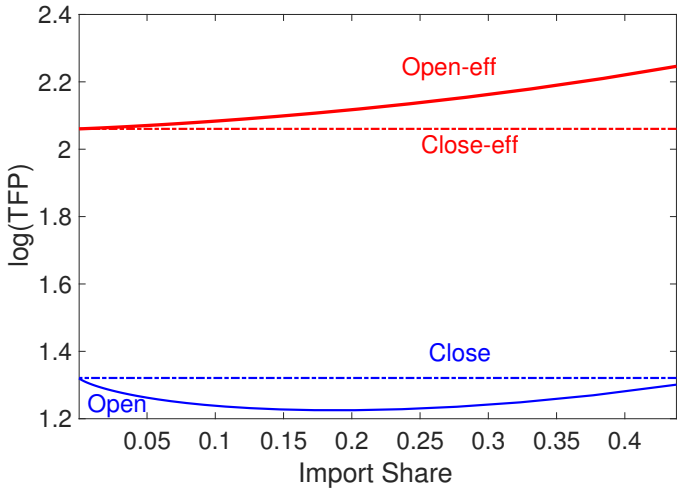
Figure 1.1: Welfare Changes from Trade



Second, the example illustrates that TFP (Eq. 15) can fall when the economy opens up to trade. Figure 1.2 plots the level of TFP against import shares under the alternative scenarios: the efficient case without distortions, the case with distortions, and when the

economy is closed or open. Three observations immediately follow: 1) that there is a TFP loss in the case with distortions compared to the case without; 2) opening up to trade leads to productivity gains in the efficient case; however, 3) opening up engenders a productivity *loss* in the presence of distortions. In order to understand the mechanisms behind these results, it is useful to first examine the closed-economy case.

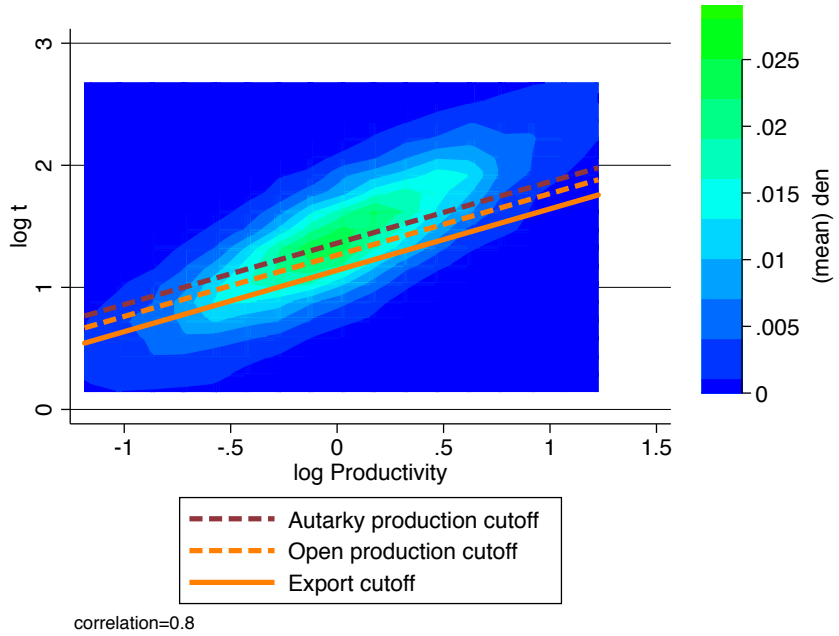
Figure 1.2: TFP Changes from Trade



Third, we examine the precise mechanisms that explain how trade can reduce aggregate efficiency. In the same numerical example, Figure 1.3 illustrates how distortions affect firm selection. The density of firms is shown by a heat map of firms that lie along a positively sloped tax-productivity line under a case with correlation of φ and τ of $\rho = 0.8$. It is clear that the productivity cutoff for production and exports is no longer determined solely by productivity, but also by domestic distortion. Only firms below the cutoff line can operate. In this figure, a large mass of highly-productive firms are excluded from servicing the market altogether. As the economy opens up, the cutoff line is shifted further downward. Even if firms have the same level of productivity, some with higher taxes may be displaced while those with lower ones will survive. This downward shift of the cutoffs allows for some low productivity and high subsidy firms to survive and gain market share.

Another way to show the impact on selection is to examine firms' market share. The two panels in Figure 1.4 plot the market share of firms, both in the closed and open economy.

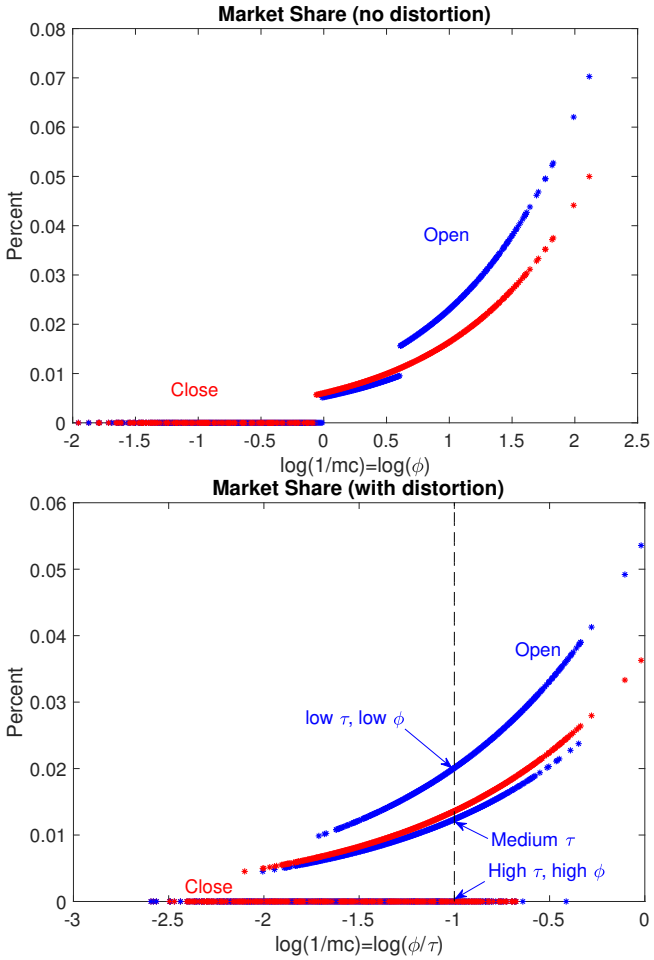
Figure 1.3: Selection Effects



The upper panel is the case without distortions. Firms with the same productivity level have the same marginal cost; their market share, above a cutoff productivity, rises with their productivity. Comparing the blue and red lines show that above the export cutoff, more productive firms have higher market shares in the open economy than in the closed economy, demonstrating that these firms expand under trade liberalization. This happens at the cost of displacing other less productive firms' market share, or driving them out of the market entirely. Here, the example clearly demonstrates that resources move from less productive to more productive firms as an economy opens up to trade.

The bottom panel shows the firm's market share in the case with distortions. Firms may share the same marginal cost and face the same potential revenues from sales even with different levels of productivity. However, their after-tax profits may differ, and thus their market share can also differ. Consider the point at which the (log) effective productivity level (φ/τ) is at -1. At this point, a firm with high, medium and low level of productivity face the same marginal costs. However, the high productivity firm is also subject to high taxes and thus low after-tax profit, and does not make the cut for production. The medium tax- medium productivity firm has positive market share but loses out to the low tax -low productivity firm when the economy opens up. Resources are reallocated from the more

Figure 1.4: Selection Effects



productive to the less productive firms. Also, there is no longer a neat line up of market shares according to productivity: there a wide range of productivities for which production is excluded.

Decomposition of TFP. The loss in TFP with misallocation can be decomposed into two components: one induced by a misallocation of resources among an observed, fixed set of operating firms, and one arising from an 'unobserved' misallocation of resources among operating and non-operating firms–that is, between potential entrants, operating firms, and displaced firms. To see this, we can decompose the deviation of TFP from its efficient level into an explicit misallocation effect—as emphasized by HK, and an implicit misallocation effect—generated by entry and selection:

$$\log TFP - \log TFP_{eff} = \underbrace{\log TFP - \log TFP_{HK}}_{\text{misallocation loss}} + \underbrace{\log TFP_{HK} - \log TFP_{eff}}_{\text{entry and selection loss}},$$

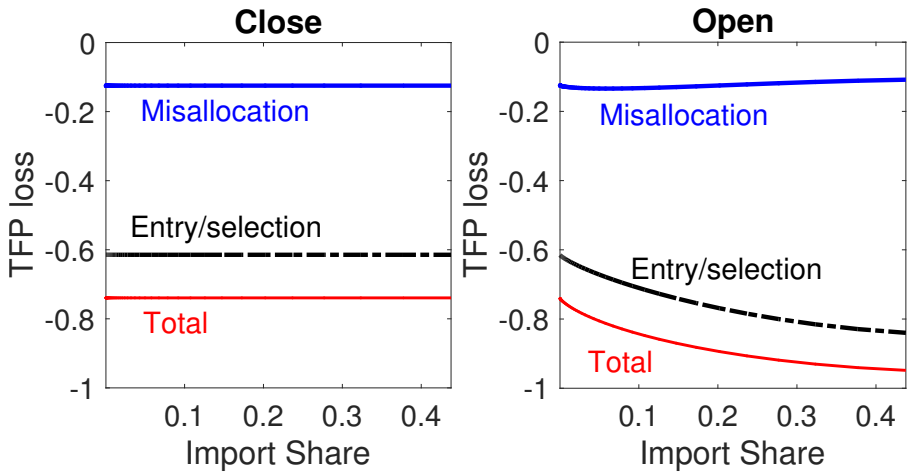
where TFP_{eff} pertains to the case without wedges: the marginal product of labor is the same across all firms and equal to the aggregate MPL, and entry is endogenously determined in this case. The HK measure of TFP corresponds to the level in the case without distortions but M, φ^*, and φ_x^* are fixed at the level with distortions:

$$\begin{aligned} TFP_{eff} &= \frac{\sigma-1}{\sigma}\left[M^{eff}\int_{\varphi^{eff*}}^{\infty}\varphi^{\sigma-1}\mu^{eff}(\varphi)\,d\varphi\right]^{\frac{1}{\sigma-1}}, \\ TFP_{HK} &= \frac{\sigma-1}{\sigma}\left[M\int_{\varphi^*}^{\infty}\varphi^{\sigma-1}\mu(\varphi)\,d\varphi\right]^{\frac{1}{\sigma-1}}. \end{aligned}$$

To illustrate the respective significance of these two components, Figure 1.5 plots the closed economy case (left panel) and the open economy case (right panel). Both cases show that the loss arising from the implicit misallocation among operating and non-operating firms is at least as significant as the explicit misallocation conventionally emphasized. Moreover, the right panel shows that these losses increase with trade, driven mostly by the loss associated with entry and selection.

Distribution of Distortions. The distribution of distortions is an important determinant

Figure 1.5: TFP loss Decomposition

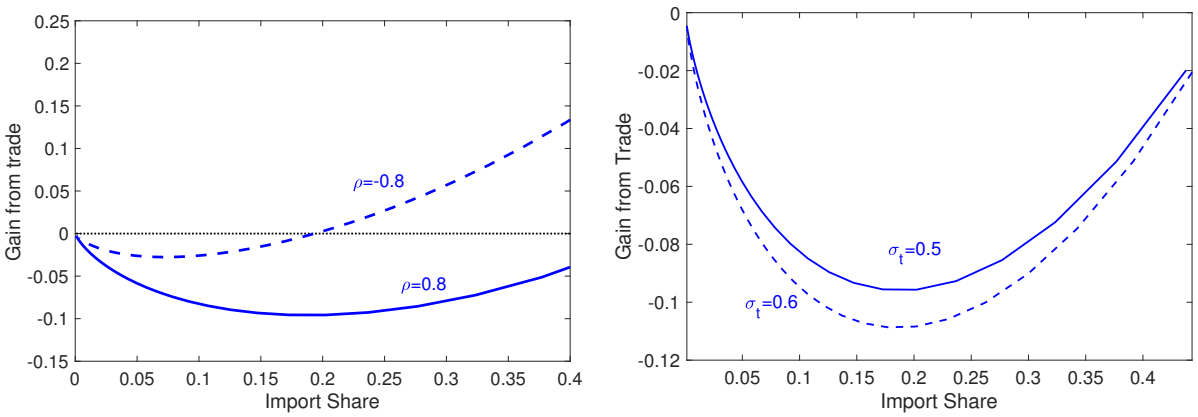


to the gains to trade and TFP changes. There are two key parameters: ρ, the correlation of τ and φ, and σ_τ, the dispersion of τ. The correlation of distortion and productivity is important insofar as a higher correlation means that more productive firms are more likely to be excluded from the market. But reductions in welfare is possible even when the correlation is negative. The reason is that there are always some productive firms that will be excluded, leading to a possible welfare loss. Figure 1.6 (a) illustrates this. It compares the gain from trade for $\rho = 0.8$, under our benchmark numerical example, and for $\rho = -0.8$, where productivity and distortion are highly negatively correlated and other parameters are the same as in the benchmark example. Under $\rho = -0.8$, the welfare gain (loss) from trade is always larger (smaller) than that in the case of $\rho = 0.8$. But when the import share is below 20%, there are still losses from trade.

Figure 1.6 (b) compares the gain from trade under different σ_τ and other parameters are the same as in the benchmark example. The welfare gain (loss) from trade is always larger (smaller) when σ_τ is smaller.

Figure 1.7 (a) exhibits how TFP loss varies with ρ under an open economy with a fixed level of trade cost f_x and τ_x. The red line plots the total TFP loss— the difference between the levels of TFP in the case with distortions and without distortions. The blue line plots the TFP loss compared with HK efficiency, and the dashed black line is the TFP loss due to entry and selection margin, as a function of ρ. First, a higher ρ is associated with a

Figure 1.6: Gains/Loss from Trade

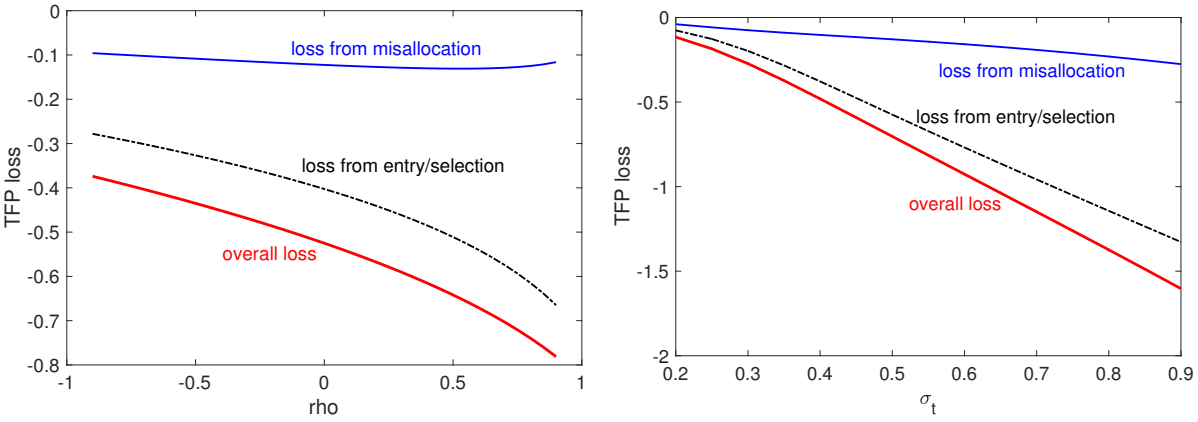


(a) Correlation Matters (b) Dispersion Matters

higher total efficiency loss. Second, the TFP losses induced by misallocation and entry and selection are both important; however, the TFP losses from entry and selections margin becomes more prominent as ρ increases.

Note that different from the standard HK analysis, the correlation between productivity and wedge affects TFP losses dramatically. The standard HK analysis has no entry margin and uses a joint log-normal distribution between productivity and wedge. In that analysis, TFP loss is fully captured by the dispersion of wedge, and the correlation of productivity and wedge does not matter at all. Our analysis is more general, featured with entry and exit into domestic and foreign market. The correlation between τ and φ is important as shown in Figure 1.7 (a).

Figure 1.7: TFP Loss Decomposition



(a) across ρ (b) across σ_τ

Figure 1.7 (b) exhibits how TFP loss varies with the standard deviation of τ. Again, we show three lines: the overall TFP loss, loss from misallocation, and loss from entry and selection. Higher dispersion of distortion leads to more misallocation and in turn higher loss in entry and selection. The overall losses increases with the dispersion of τ.

In summary, the size of TFP loss and welfare after opening up depends on the correlation of φ and τ and the dispersion of τ, σ_τ. The firm level data helps us identify these parameters. Specifically, in the next section, we will measure the firm-level output wedge and use its dispersion and its correlation with firm value added to identify ρ and σ_τ.

3 Empirical Results

We proceed to investigate whether key empirical implications of our model is supported by the data. Inferring firm-level wedges directly from the data has its challenges. Still, key properties and relationships of these measured wedges and productivity support core implications of our model. Though our focus is to estimate a structural model to infer welfare and productivity gains to trade, without directly using these empirical measures, the exercise that follows is of independent value, as well as providing a robustness check.

Data. Our data for Chinese firms are from an annual survey of manufacturing enterprises collected by the Chinese National Bureau of Statistics. The dataset includes non-state firms with sales over 5 million RMB (about 600,000 US dollars) and all of the state firms for the 1998-2007 period. We have information from the balance sheet, profit and loss statements, and cash flow statements, which incorporate more than 100 financial variables. The raw data consist of over 125,858 firms in 1998 and 306,298 firms by 2007.

Backing out Key Parameters. To back out factor and output distortions we adopt a Cobb-Douglas production function for a firm i in industry j, $y_{ji} = \varphi_{ji} k_{ji}^{\alpha} \ell_{ji}^{1-\alpha}$, where y_{ji}, ℓ_{ji}, k_{ji} represent the output, labor, and capital stock of firm i in sector j. The marginal revenue product of labor and capital is $\partial(p_{ji}y_{ji})/\partial(\ell_{ji})$ and $\partial(p_{ji}y_{ji})/\partial(k_{ji})$, and with firm profit

maximization, yields

$$MRPL_{ji} \equiv \frac{\sigma-1}{\sigma}(1-\alpha_j)\frac{p_{ji}y_{ji}}{\ell_{ji}} = \tau^{\ell}_{ji} w_j$$

$$MRPK_{ji} \equiv \frac{\sigma-1}{\sigma}\alpha_j\frac{p_{ji}y_{ji}}{k_{ji}} = \tau^{k}_{ji} r_j,$$

where w_j and r_j denote industry-level wages and interest rates. These marginal products are proportional to the average products, assuming common markups and capital elasticities, and no fixed cost as in HK. Firms equalize the after-tax marginal revenue products of factors. In the absence of distortions, revenue per person should be equalized across firms. In the presence of distortions, a firm that faces higher taxes will end up with a higher marginal revenue product and less capital/labor than an otherwise identical firm facing a subsidy.

Equilibrium allocations yield

$$p_{ji}y_{ji} \propto \left[\frac{\varphi_{ji}}{(\tau^{k}_{ji})^{\alpha_j}(\tau^{\ell}_{ji})^{1-\alpha_j}}\right]^{\sigma-1},$$

from which firm-level productivity can be inferred as

$$\varphi_{ji} = (P_j^{\sigma-1}X_j)^{\frac{1}{1-\sigma}}\frac{(p_{ji}y_{ji})^{\frac{\sigma}{\sigma-1}}}{k_{ji}^{\alpha_j}\ell_{ji}^{1-\alpha_j}}. \tag{17}$$

It is the relative marginal revenue and relative productivities–deviations from the industry mean—that matters. Thus,the measured relative $MRPK_{ji}$, or the relative average product ($ARPK_{ji}$), is calculated as $\log\left(p_{ji}y_{ij}/k_{ij}\right) - \log\left(\overline{p_jy_j/k_j}\right)$ where $\overline{p_jy_j/k_j}$ is the industry mean of the average product. The same holds for the measured marginal (average) revenue of labor. The elasticity of output with respect to capital in each industry is taken to be 1 minus the labor share in the corresponding industry in the U.S, following HK. The reason that labor shares are not computed from Chinese data is that the prevalence of distortions would affect these elasticities, and industry-level elasticities and distortions cannot be separately identified. The U.S. is taken to be the benchmark as the relatively undistorted economy. These labor share comes from the U.S. NBER productivity database, which is based on the Census and ASM. We take the benchmark elasticity of substitution

parameter σ to be 3, but experiment with other values within the conventional range. Different from HK, we take a firm's employment to measure ℓ_{ji} rather than the firm's wage bill. This addresses the problem that Chinese wage data implies too low of a labor share as measured by input-output tables and the national accounts. We define the capital stock as the book value of fixed capital net of depreciation.

Measured Distortions. We find large dispersions in measured distortions in China, similar to the levels in HK for the year 1998 and 2005. Measured distortions have come down over time, between 1998 and 2007, as evident in Table 1. There is also greater dispersion in the average product of capital than there is in the average product of labor.

We next turn to investigating further what factors are systematically related to measured distortions. Table 2 reports the regression results of the relative average product of capital of a firm (measured as value-added divided by total capital, deviated from industry mean) on a set of variables. The coefficient on firm-productivity is large and significant; 1 percent increase in relative productivity ($TFPQ$) is associated with a 0.7 percent increase in relative distortion. Moreover, more than half of the variation in distortions is explained by productivity alone. The positive relationship is consistent with the predictions of our model, as it predicts that firms that operate must have lower taxes on average, given the same level of productivity. The same is true for the results on exporters: given productivity, firms must have lower taxes on average in order to export. What these wedges are and where they might come from is suggested by its systematic relationship with firm characteristics: state-owned enterprises and foreign-owned firms are subject to lower taxes on average, given productivity. If we perform these regressions using the average product of labor instead, in Table 3, results are similar.

Table 1: Dispersion of Distortions

	1998	2001	2004	2007
std(APK(deviation from industry mean))	1.348	1.306	1.241	1.185
std(APL(deviation from industry mean))	1.184	1.039	0.940	0.923

Relationship between productivity and distortion. There is no apriori reason to believe

Table 2: APK Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\ln(APK)$	$\ln(APK)$	$\ln(APK)$	$\ln(APK)$	$\ln(APK)$	$\ln(APK)$
$\ln(TFPQ)$	0.652***	0.697***	0.706***	0.705***	0.707***	0.711***
	(147.7)	(153.0)	(154.8)	(153.9)	(160.3)	(168.1)
age				-0.00178***	-0.00191***	-0.00174***
				(-8.772)	(-9.477)	(-9.386)
1.soe					-0.116***	-0.109***
					(-3.388)	(-3.313)
1.foreignown					-0.460***	-0.379***
					(-19.74)	(-20.60)
exporters						-0.233***
						(-13.82)
Constant	-3.617***	-3.280***	-3.204***	-3.173***	-3.049***	-3.042***
	(-134.6)	(-60.38)	(-54.16)	(-53.37)	(-44.45)	(-44.88)
Observations	1,616,507	1,616,507	1,506,572	1,505,657	1,505,657	1,505,657
R-squared	0.566	0.628	0.640	0.640	0.655	0.659
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes	Yes
Location FE	No	No	Yes	Yes	Yes	Yes

Robust t-statistics clustered at the four-digit industry level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: APL Regressions

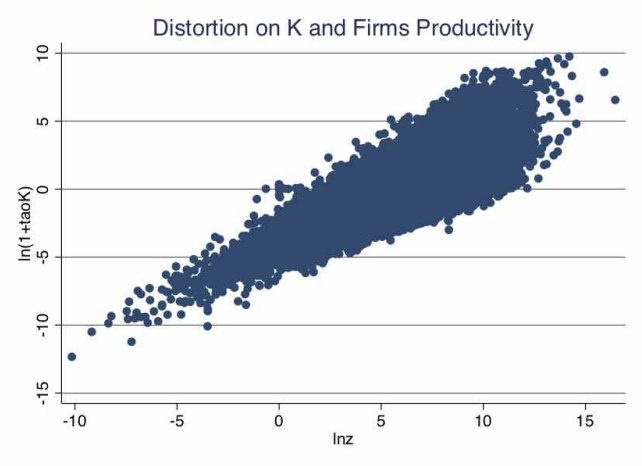
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	$\ln(APL)$	$\ln(APL)$	$\ln(APL)$	$\ln(APL)$	$\ln(APL)$	$\ln(APL)$
$\ln(TFPQ)$	0.530***	0.570***	0.569***	0.568***	0.565***	0.567***
	(110.7)	(228.5)	(222.5)	(224.2)	(228.4)	(229.4)
age				-0.00161***	-0.00140***	-0.00128***
				(-9.072)	(-8.783)	(-8.440)
1.soe					-0.0840***	-0.0787***
					(-7.136)	(-7.057)
1.foreignown					0.0615***	0.123***
					(3.925)	(8.317)
exporters						-0.175***
						(-27.08)
Constant	-3.593***	-3.274***	-3.229***	-3.201***	-3.172***	-3.167***
	(-123.2)	(-109.1)	(-103.2)	(-100.5)	(-95.80)	(-97.30)
Observations	1,616,507	1,616,507	1,506,572	1,505,657	1,505,657	1,505,657
R-squared	0.619	0.691	0.699	0.700	0.701	0.705
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes	Yes	Yes
Location FE	No	No	Yes	Yes	Yes	Yes

Robust t-statistics clustered at the four-digit industry level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

that more productive firms are associated with higher wedges, as is shown in the regression results. To show this graphically, we next plot the relationship between φ ($TFPQ$) and APK in Figure 1.8; a similar relationship holds for φ and APL. Though this relationship was not important in the special case of HK assuming a joint log normal distribution between the two variables—it does matter for more general cases. Moreover, it is vital for our quantitatively analysis.

Figure 1.8: Correlation Between Measured MPK and Measured TFPQ



The *observed* correlation cannot be treated to be the same as the *underlying* correlation between the two variables (ρ), as the observed relationship may be an outcome of a selection process, as our model demonstrates. Even if the underlying correlation is negative, the selection mechanism can induce the observed correlation to become positive, for the simple reason that high-taxed firms must be more productive in order to stay in the market. The selection mechanism will strengthen any underlying correlation between the two variables. In order to compute the impact of distortions on welfare and productivity gains, one would need to know the underlying correlation ρ, and therefore one would need micro data and a structural model to uncover it. What is also interesting about the relationship in the data is the line that seems to cut firms from above—reminiscent of the cutoff line endogenously generated in the model.

Costa-Scottini (2018) and Ho (2010) assume a perfect correlation of (log) productivity and (log) wedges in their analysis. We have shown why the assumption of an exogenous

correlation may be misplaced. Moreover, this assumption counters findings in the data in that 1) measured (log) productivity and (log) wedges are far from perfectly correlated, and 2) exporters have lower wedges. As long as these two variables are not perfectly correlated, selection will affect their measured correlation and one would need to estimate it from the model. For this reason, a careful estimation of the joint distribution of productivity and wedges based on a structural model and firm-level data is imperative to obtaining quantitative results on the impact of trade. We turn to this analysis in the subsequent section.

Fixed Cost. The positive relationship can also result from measuring marginal products using average products when there are fixed costs. The average products are $APL = p_{ji}y_{ji}/(\ell_{ji} + f_{ji})$, and productivities are $\varphi = y_{ji}/(\ell_{ji} + f_{ji})$. Thus, the presence of fixed cost will tend to induce a positive relationship between the two. In our quantitative analysis,

4 Quantitative Results

This section presents estimates of the quantitative effects of trade liberalization when accounting for domestic distortions. The two countries Home and Foreign, are calibrated to data corresponding to China and the U.S.

Table 4 reports the calibrated and chosen parameters. The Home labor L is normalized to 1 and Foreign labor L_f to 0.2 to match the relative labor force of US to China. Productivity levels are set to match the relative GDP of US to China. Given that Foreign affects Home only though aggregate variables, we can assume that Foreign is absent of distortions, while taking $f_e, f, f_x, \tau_x, \sigma_\phi$ to be the same as those in Home. We set the elasticity of substitution between varieties σ to be 3, the one taken in HK. This value is consistent with the estimates from plant-level US manufacturing data in Bernard et al. (2003).

The remaining 8 parameters are estimated jointly, to match the model moments with their data counterparts. Table 4 and 5 reports the estimated parameters and the moments in the data and model. The moments we choose are the ones that are most relevant and sensitive to variations in model parameters. Clearly, every parameter matters for the general equilibrium and affects other moments. However, there is by and

Table 4: Model Parameters

Parameter	*Value*	*Identification*
Elasticity of substitution σ	3	
Home labor L	1	normalization
Foreign Labor L^f	0.2	Relative labor size of US to China
	Internal Estimation	
Entry cost f_e	0.2	Fraction of firms producing (one year survive rate in the data)
Fixed cost of producing f	0.015	mean-lowest 5% $ln(K^\alpha L^{1-\alpha})$
Fixed cost of export f_x	0.12	fraction of firm exporting
Iceberg trade cost mean τ_x	1.5	export intensity
Std. productivity σ_φ	1.2	std of existing firms lnVA
Std. wedge σ_τ	0.9	std of existing firms $ln(K^\alpha L^{1-\alpha})$
Corr(wedge, productivity) ρ	0.86	Corr(lnVA, $ln(VA/K^\alpha L^{1-\alpha})$)
Mean foreign prod $\mu_{f\varphi}$	5.5	Relative GDP of US to China

Table 5: Data and Model Moments

Target Moments	Data(2005)	Model
Fraction of firms producing ω_e	0.85	0.85
Mean $-$ lowest 5% for $ln(K^\alpha L^{1-\alpha})$	1.82	1.53
Fraction of firm exporting	0.30	0.28
Export intensity	0.41	0.42
std of existing firms ln(VA)	1.20	1.26
std of existing firms $ln(VA/K^\alpha L^{1-\alpha})$	0.93	0.84
Corr(lnVA, $ln(VA/K^\alpha L^{1-\alpha})$)	0.41	0.35
Relative real GDP of US to China	1.79	1.77

large a clear correspondance between certain parameters and moments. The parameter most relevant for matching the fraction of surviving firms is the entry cost f_e, as $\omega_e E[\pi(\phi,\tau)] = wf_e$. Lower entry costs induces more entrants to pay the costs, and the result is a lower fraction of survivors. Next, to identify the fixed cost f, one needs only to turn to the smallest firms, which have their profit just about cover fixed cost. That is, the after-tax profit $\pi = wf$ and $wl_{min} = (\sigma-1)wf$, and the mean of firms' labor $wl_{mean} = (\sigma-1)(E[\pi(\phi,\tau)]+wf) = (\sigma-1)(\frac{wf_e}{\omega_e}+wf)$. Hence, the mean-lowest 5% $ln(K^\alpha L^{1-\alpha}) = \frac{\frac{f_e}{\omega_e}+f}{f}$ helps identify f.

We calibrate τ_x to match the fraction of exports in exporters' sales in Chinese manufacturing. The resulting parameter $\tau_x = 1.5$ is inline with the estimate of 1.7 in Anderson and Van Wincoop (2004), and the 1.83 in Melitz and Redding (2015). The dispersions in productivity and wedges, and their correlation are important for matching the observed joint distribution between value-added and inputs in the data. Table 5 shows that the discrepancy between our model and data moments is reasonably small, though we underestimate the dispersion in distortions and slightly overestimate the dispersion in size. An important variable is the correlation between size and distortions, Corr(lnVA, $ln(VA/K^\alpha L^{1-\alpha})$. This variable is more positive the higher is $\rho\frac{\sigma_\varphi}{\sigma_\tau}$, where ρ is the underlying correlation between wedges and productivity. A higher underlying correlation and a lower dispersion in wedges raise the observed correlation between value added and inputs.

4.1 Implied Gain from Trade and TFP Loss

Table 6 reports the gain from trade and efficiency losses for both Home and Foreign. The upper panel of the table compares welfare and TFP in the open economy to those in the closed economy. In the benchmark estimation, the gain from trade for Home is 4.4%. Without distortions, the gains from trade is more than doubled (9.8%). Foreign's gain from trade is about 8.2% when Home has domestic distortions. Eliminating Home distortions makes the foreign country benefit more from trading with Home—a 19% of welfare gain.

Note that the standard trade models, as in ACR, compute the welfare gain using aggregate import shares and abstracts from micro-level distributions. To see whether the

Table 6: Welfare and TFP

	Open relative to close			
	Welfare	TFP	Import Share	ACR gain
Home (%)				
Benchmark	4.4	-2.9	30.8	12.7
No-distortion	9.8	13.3	20.8	10.1
Foreign(%)				
Benchmark	8.2	12.9	17.9	6.9
No-distortion	18.9	13.3	35	17.6
	TFP loss: Distortion relative to no-distortion			
	Overall loss	Misallocation	Entry-selection	
Benchmark	140.4	119.2	21.2	
Home close	124.2	118.7	5.4	

micro structure matters or not, we compute the ACR gain moving from a closed to an open equilibrium as inferred from the import share and an average full trade elasticity ϵ. This elasticity is calculated by dividing the logarithmic percentage change in trade moving from autarky to the benchmark open economy by the logarithmic percentage reduction in variable trade costs. The ACR formula is

$$\text{ACR gain} = -\frac{1}{\epsilon} log(\text{domestic share}).$$

In our benchmark and in the data, the import share is 30.8%. The implied ACR gain from trade is therefore 12.7%, about 8.3 percentage points higher than in our model. Hence, only using aggregate data and ignoring firm-level distribution overestimates the gain from trade by 289%. Moreover, Foreign has a larger gain in the benchmark case—around doubling that of Home. But using the ACR formula, one would draw the opposite conclusion —that Foreign has a smaller gain than Home does, 6.9% versus 12.7% since the import share of Foreign is about half of that of Home.

In terms of TFP our benchmark result shows that opening up leads to a 3% loss in TFP. In contrast, without distortion, TFP increases by 13.3%. Hence, contrary to the standard predictions, trade liberalization can exacerbate rather than improve resource allocation, causing a decline rather than a rise in TFP. As foreign has no distortions, the TFP levels are basically the same between the two models.

The lower panel of Table 6 reports Home's TFP losses due to distortions both for open and closed economy cases. Not surprisingly, there are large TFP losses for Home with domestic distortions. Eliminating these distortions would increase China's TFP by 124% under a closed economy. The TFP losses are larger, 140%, in the benchmark case where China has an import share of more than 30%. To understand the efficiency loss, we decompose them into a misallocation effect and an entry and selection effect. The majority of the losses is still coming from misallocation. One reason is that the share of surviving firms is high as implied in the data, but this could overestimate the survival rate as the dataset includes firms of a certain scale. If smaller firms are observed, then it is possible that the losses coming from the entry and selection margin are smaller, though still quantitatively non-negligible.

4.2 Decomposing China's Growth from 1998-2005

The rapid growth in China over the last four decades has been one of the most remarkable phenomena the world has witnessed in recent history. In between 1998 and 2005, its real GDP increased by 57%. Accompanying this development was a combination of domestic reforms and opening up programs—policies that fostered trade and FDI inflows. As a result, both trade and technological progress increased over time, while measured domestic distortions concurrently fell. A natural question is how much of the growth is attributed to trade over this period. Other competing factors include technological improvement, factor accumulation, and domestic reforms—that is, the allocative gains associated with a reduction in distortions. In what follows, we perform a quantitative analysis to answer this question. Specifically, we recalibrate the model parameters for the year 1998 and compare the implied GDP and TFP levels to those in the benchmark year, 2005. Overall, our results attribute the majority of China's GDP and TFP growth to technological improvement, capital accumulation, and a mitigation of distortions. Trade alone contributes to only about 8% of GDP growth.

Table 7 reports the moments for both 1998 and 2005. The starting year is taken to be 1998, as it is the first year in which firm-level data is available, and three years before China joined the WTO. Compared to the year 2005, trade intensity was significantly lower

in 1998, both in terms of the fraction of firms that export, and also the export intensity of these firms. Distortions were large in the earlier years, as seen by the fact that the dispersion of measured distortion was about 20% higher in 1998 compared to 2005. This implies a higher trade cost τ_x and dispersion of distortion σ_τ in 1998— at about 43% and 20% higher than the level in 2005. The mean TFP in 2005 is about 45% higher than that in 1998, reflecting technological improvements and factor accumulation over time.

Table 7: Data, 1998 and 2005

Target Moments	Data (1998)	Data (2005)
Fraction of firms producing ω_e	0.77	0.85
Mean − lowest 5% for $ln(K^\alpha L^{1-\alpha})$	2.04	1.82
Fraction of firm exporting	0.25	0.30
Export intensity	0.30	0.41
std of existing firms ln(VA)	1.33	1.20
std of existing firms $ln(VA/K^\alpha L^{1-\alpha})$	1.12	0.93
Corr(lnVA, $ln(VA/K^\alpha L^{1-\alpha})$)	0.47	0.41
Relative real GDP of US to China	2.50	1.79
Change of China's real GDP		57%

These estimates are then used to run counterfactual experiments, in order to decompose China's growth in between 1998 and 2005. The factors considered include technological progress, input accumulation, and the reduction of trade costs and domestic distortions. In each experiment, the parameters for the year 1998 remain fixed, while each of the following parameters– mean TFP μ_φ, trade cost τ_x, or dispersion of distortion σ_τ—are allowed to vary to its 2005 level. Table 8 shows that the increase of technology and inputs alone lead to a 44% increase in GDP and a 46% increase in TFP. Reduction in trade costs would independently boost GDP by 8% and TFP by only 3%. In contrast, lowering the dispersion of distortions increases GDP by 66% and TFP by 69%.[14]

A notable point of comparison is with Tombe and Zhu (2018), which, despite adopting an altogether different approach, finds also small gains to trade. In their model that features migration across regions and sectors in China, international trade contributes to only 7% of productivity growth in between 2000 and 2005. In other words, international trade has

[14]Note that the contributions to GDP or TFP increase don't add up to 100% because the productivity distribution and fixed costs have also changed from 1998 to 2005. Furthermore, there are interacting effects on mean TFP, trade cost, and distortion dispersions.

led to very little allocative benefits of labor across regions and sectors—as compared to direct reforms that lower migration costs or reductions in internal trade costs. Their model does not feature distortions at the firm level that can render trade's allocative benefits even smaller. This leads us to find an even smalerl effect of trade on productivity in China over roughly the same period.

These quantitative experiment predicts a small welfare and TFP gains to trade. Although in the current framework the main efficiency gains to trade is through the positive effects of selection, along with other gains such as scale and increase in varieties as encapsulated in the ACR formula, there are still other benefits that may not be accounted for. One example is the pro-competition effect of trade. Another that is relevant to China is potentially transfers of technology (Ramondo and Rodriguez-Clare 2013). However, Arkolakis, Costinot, Donaldson, Rodríguez-Clare et al. (2015) points out that the standard gains to trade, encompassing old and new theories of trade, are fully inferred from observed trade flows. Given that trade shares are not that large in China, the ACR formula predicts small gains to trade, even though it embodies a range of other types of trade gains. Above and beyond these gains, the pro-competitive effect of trade is quantitatively small, and thus 'elusive'.

Our point is that these gains could be even smaller, so that the large commonly-perceived trade gains in China may be exaggerated— especially in a country known for its myriad of distortions. We do not consider the possibility that trade may also help reduce domestic distortions. If, say, the WTO requires certain kind of domestic reforms as a pre-condition to becoming a member, some of the technological improvement and reductions in the level of distortions could be partially induced by opening up policies. Still, our experiments point to the fact that the conventional notion of trade gains have been small for China, and even smaller when there are distortions —particularly for the more productive firms.

Table 8: Decomposition of China's Growth between 1998-2005

	Change of Real GDP	Change of TFP
Benchmark	57%	56%
Counterfactual Change from 1998-2005:		
Technology and inputs alone (Increase mean φ)	44%	46%
Trade alone (Decrease τ_x)	8%	3%
Distortion alone (Decrease σ_τ)	66%	69%

4.3 Selection through Export: an Out of Sample Test and Extension

To examine whether Chinese firm characteristics change when they become exporters, we examine the relationship between measured distortions and firm export status in both the cross-section and the time series dimension. According to our model, given productivity exporters face a lower distortion due to selection. This implication is broadly consistent with the data. Next, we use the time series data to check whether 'measured' distortions change when firms enter the export market. We find that the model implications are qualitatively consistent with the data. Finally, we consider model extensions with export rebate and allow for different distortions when exporting.

The first two columns of Table 9 reports the data and the model regressions of measured distortion on measured productivity $TFPQ$ and a dummy of exporters. Both the model and the data have the pattern that exporters face a lower marginal product. Note that we did not target the differences between exporters and non-exporters. What the model implies is a stronger selection effect than that implied by the data, as exporters' marginal product is about 64% lower than non-exporters in the model, compared to 26% in the data.

We further consider whether the measured marginal product or distortion vary with entering or exiting the foreign market. This will help us to understand the differences between exporters and non-exporters are from selection or there are further/different distortions when exporting. Through the sample periods, we define the following exporters categories: always exporters who are exporting through out the sample years from 1998 to 2007, starters who start to export after 1998, stoppers who stopped to export during the sample year. Entry effect measures the percentage difference of measured distortions for starters, between the post- and pre-exporting entry periods. Exit effect measures percentage difference of measured distortion for stoppers, between the post- and pre-exporting exit periods.

As shown in Column 3 of Table 9, in the data, always exporters have a lower marginal product. Firms' measured marginal product decrease when they start exporting and increase when they exiting. Note that although the firm level distortions are exogenous in the model, the measured distortions do change after firms start exporting due to fixed costs of exporting.

In our benchmark model, if trade cost decreases from 1998 level to 2005 level, some firms start to export, their measured distortion indeed decreases as shown in the entry-effect of Column 4. As in the data, always exporters and starters have lower measured marginal product as well. Our benchmark model shows a relative large selection effect since the model does not have large heterogeneity as in the data, and a relative small entry effect. In reality, the change of a firm's marginal product after exporting could be driven by multiple alternative reasons. For example, exporters face different distortions from non-exporters, or there are endogenous distortions that change with trade liberalization. We explore these in the following.

In Column 5, we introduce a 10% tax rebate after exporting. This generates an even larger selection effect and the measured wedge further decrease upon exporting. In this case, export subsidies from tax rebate benefit foreign country, and Home welfare gains from trade decreases to 1.6% and TFP loss increases to 4.5%.

Column 6 considers an extension that firms face different distribution of distortion when exporting. We use the standard deviation of export intensity to discipline the dispersion of the additional wedges when exporting. To be more precise, we estimate the mean and standard deviation of the extra export distortions to match the standard deviation of export intensity and the regression coefficient on always exporters. When the trade cost is reduced as in the benchmark, the coefficient on starter and the entry effect are similar to the data. Not surprisingly, the loss from trade is much larger in this extension than in the benchmark. The reason is because there are additional direct distortions on the selection of firms to foreign market when open up to trade.

Table 9: Measured Marginal Products, data vs model

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Data	Benchmark	Data	Benchmark	Export rebate	Different τ in export
entry effect			-0.104***	-0.050	-0.103	-0.09
			(-12.69)			
exit effect			0.0315***			
			(4.574)			
starter			-0.101***	-0.429	-0.400	-0.08
			(-21.74)			
stopper			-0.0891***			
			(-20.98)			
always exporters			-0.301***	-0.768	-0.791	-0.324
			(-23.47)			
log(TFPQ)	0.636***	0.652	0.638***	0.653	0.654	0.613
	(250.5)		(254.9)			
exporters	-0.264***	-0.637				
	(-24.14)					
Constant	-3.258***	0.401	-3.255***	0.414	0.425	0.714
	(-106.2)		(-107.0)			
Observations	1,587,629		1,584,242			
R-squared	0.823		0.826			
Time FE	Yes		Yes			
Industry FE	Yes		Yes			

Note: Robust t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is measured marginal product, which is $\log(TFPR)$ in the data and $\log(VA/(\ell+f))$ in the model.

5 Discussion

In this section, we explore a model with endogenous distortion arising from endogenous markup, which has been extensively studied in the standard trade literature. We show that the endogenous markup model runs counter with the data in that exporters in the model face a higher markup and distortion. We then address the issue that misallocation could be driven by mismeasurement of inputs or outputs. We show that even taking out the standard measurement errors, there are still large distortions remaining among Chinese firms.

5.1 Endogenous distortions

Here we build a model with endogenous markup as in Edmond, Midrigan, and Xu (2018). The consumer's problem is the same as before.

Final goods producer Final goods producers are competitive and produce with intermediate goods with a Kimball aggregator

$$\int_{\omega\in\Omega}\gamma\left(\frac{q}{Q}\right)d\omega=1,$$

where $\gamma(\cdot)$ follows Klenow and Willis (2016) specification as

$$\gamma\left(\frac{q}{Q}\right)=1+(\sigma-1)\exp\left(\frac{1}{\varepsilon}\right)\varepsilon^{\frac{\sigma}{\varepsilon}-1}\left[\Gamma\left(\frac{\sigma}{\varepsilon},\frac{1}{\varepsilon}\right)-\Gamma\left(\frac{\sigma}{\varepsilon},\frac{(q/Q)^{\frac{\varepsilon}{\sigma}}}{\varepsilon}\right)\right], \tag{18}$$

$\sigma>1,\varepsilon\geq 0$ and $\Gamma(s,x)$ denotes the upper incomplete Gamma function $\Gamma(s,x)=\int_x^\infty t^{s-1}e^{-t}dt$. The demand function for each intermediate good producer is therefore given by

$$p(\omega)=\gamma'\left(\frac{q(\omega)}{Q}\right)PD, \tag{19}$$

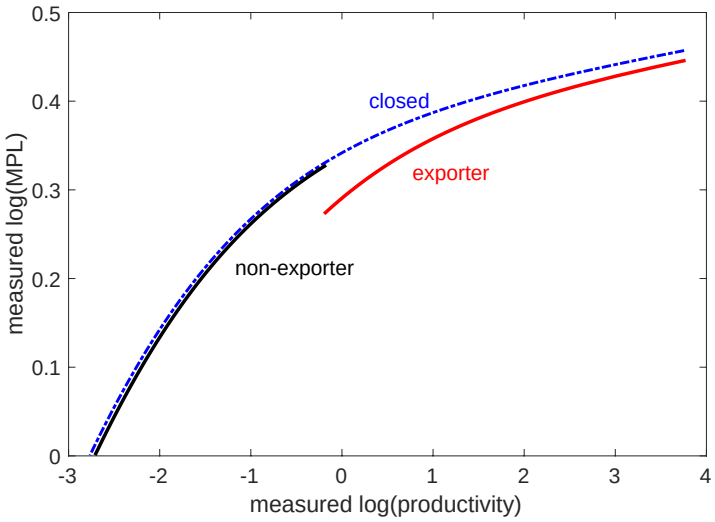
where D is a demand index, $D=\left[\int_{\omega\in\Omega}\gamma'\left(\frac{q(\omega)}{Q}\right)\frac{q(\omega)}{Q}d\omega\right]^{-1}$.

Intermediate good producer The problem of an intermediate good producer is similar as before except it faces a demand function as in equation (19). The firm will choose the price as a markup over the marginal cost,

$$p = \frac{\sigma}{\sigma - (q/Q)^{\frac{\varepsilon}{\sigma}}} \frac{w\tau}{\varphi}.$$

Note that the markup is endogenous and depends on the size of the firm, the higher the quantity a firm sells, the higher the markup it charges. The firm's optimal production and profit increase with φ and decrease with τ. Hence there exists a cutoff $\varphi^*(\tau)$, firms produce when $\varphi \geq \varphi^*(\tau)$.

Figure 1.9: Measured MPL and Productivity in an Endogenous Markup Model



To compare with the benchmark model, we choose ε as 0.08 to match the aggregate marginal product of labor of 1.45 as in Edmond, Midrigan, and Xu (2018), while keeping other parameters the same as in the benchmark. Figure 1.9 plots the relationship between the measured $\log(MPL)$ and the measured log productivity in this model. First, higher productivity firms produce more and end up with a higher endogenous markup. The measured MPL is therefore higher. Hence, we observe an upward sloping line for the closed economy. Second, this upward sloping patterns also show up in the open economy. Moreover, exporters are more productive and face a higher wedge. Non-exporters face a more competitive market after opening up and charge a lower markup, the measured wedge is smaller. Around the exporting cutoff, exporters face a lower MPL due to the

fixed cost of exporting. Overall, exporters face higher measured MPL.

In summary, if the observed wedges are purely driven by markups and they endogenously change with trade, we shall have 1) Exporters are on average have higher markup, hence higher instead of lower wedges, 2) Measured log (MPL) and log (VA) will be almost perfectly correlated. These implications are at odds with the regression results shown in section 4.3, where exporters face lower wedge. Moreover, Arkolakis et al. (2015) shows the gain from trade in a model with endogenous markup is similar to ACR. In the current model, the excessive welfare gain from trade when removing domestic distortions is 3.2%, while it is 5.4% (9.8%-4.4%) in the benchmark model. It is therefore in this extended model, similar exogenous distortions are needed to match the observed dispersion and correlation. This is consistent with Song and Wu (2015) and David and Venkateswaran (2017) that the heterogeneity in markup explains very limited MPK dispersion in China.

5.2 Detecting measurement error

Differences in measured average products need not imply differences in true marginal products. The presence of fixed costs in producing and exporting in our model, for instance, means that the average revenue products differ from the marginal revenue products. For this reason, a model estimation is used to back out the true dispersion in marginal products. Nevertheless, other types of mismeasurements in output and input may also generate a dispersion in the average revenue products, and thereby affect the measured TFPR—as shown in Bils, Klenow, and Ruane (2017) and Song and Wu (2015). In this section, we address the issues surrounding measurement error following the practices adopted in the literature.

The main approach involves using panel data to improve the estimates on the true marginal product dispersion, rather than simply employing cross-sectional data. With this method, we find that the measurement errors are small in China, accounting for only 18% of the variation in the average product.[15] This 18% includes the mismeasurement of production inputs in the presence of fixed cost, which is accounted for in our benchmark. We

[15] In constrast, Bils, Klenow, and Ruane (2017) finds measurement errors can explain about half of variation of average products in Indian, and about 80% of that in the U.S.

then reestimate our model and implied gains from trade using the adjusted dispersion of marginal product when measurement errors are corrected for. Our benchmark results are robust to this adjustment.

We exploit three alternative methods to detect measurement error: average annual observations within firms, first differences over years within firms, and covariance between first differences and average products. All three approaches point to the same conclusion: that 1) there is a large dispersion in marginal products in China; 2) measurement error only accounts for a small fraction of the dispersion in the measured marginal products (i.e. average products).

First, if measurement error were idiosyncratic across firms and over time, one can take the time average of annual observations within firms to wash out these errors, drastically reducing the dispersion of average products. The upper panel of Table 10 reports the statistics when we take the average within firms. The average standard deviation is 1.19 for the average marginal product of capital and 0.96 for the average marginal product of labor. The standard deviations of value added and the average product of inputs are 1.19 and 0.94, where the correlation between the two variables is 0.4. These results mimic our benchmark moments. In particular, the dispersions of average products of inputs are still high. This implies that measurement errors of the iid type cannot explain the observed dispersions in the average products.

Second, as pointed out by Bils, Klenow, and Ruane (2017), the dispersion of first differences reflect the true distortion if marginal products are constant over time. Calculating the first differences of value added $\triangle VA$, capital $\triangle K$, and labor $\triangle L$, and then taking the ratio $\triangle VA/\triangle K$ and $\triangle VA/\triangle L$ gives us an alternative measure of marginal products. The 1% tails of both ratios are trimmed, and the results are displayed in the middle panel of Table 10 for the year of 2001, 2004, and 2007. The dispersions are even higher than those in Table 1 for the standard measured marginal product of inputs using this measure.

Moreover, the alternative measured marginal products are highly correlated with our average products. Figure 1.10 plots the $\ln(\triangle VA/\triangle I)$ against the benchmark average product of input $\ln(VA/I)$ where I is the composite of inputs, $I = K^{\alpha}L^{1-\alpha}$, where each dot corresponds to one of 100 percentiles of $\ln(VA/I)$. The regression coefficient at the firm

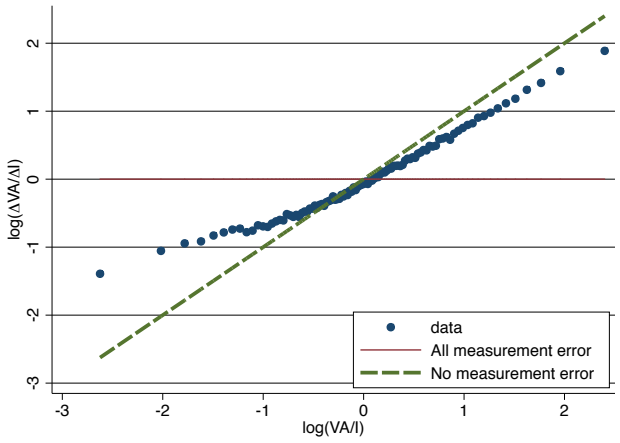
Table 10: Detecting Measurement Errors

Average annual observation within firm				
std(ln(APK))	std(ln(APL))	$std(lnVA)$	$std(ln(VA/I))$	$corr(lnVA, ln(VA/I))$
1.19	0.96	1.19	0.94	0.4
		First level differences		
		2001	2004	2007
$std(ln(\triangle VA/\triangle K))$		1.82	1.78	1.76
$std(ln(\triangle VA/\triangle L))$		1.68	1.60	1.61
		Regression		
		Ψ	$\Psi(1-\lambda)$	
		0.53^{***}	-0.0997^{***}	
		(34.58)	(−20.65)	

Note: This table reports three ways to detect measurement errors.
The upper panel reports the average annual levels within firms.
The middle panel reports the ratio of first differences as another measure of marginal product, where $\triangle VA$ denotes the first difference of value added.
The lower panel reports regression coefficient as in equation (20).
Robust t-statistics in parentheses.

level is 0.72, see Table 11. Note that without measurement errors, the two measures are perfectly correlated. For the case with only measurement error, the two measures have no correlation. Hence, the high correlation between the alternative measure and the average products suggest small measurement errors and a large distortion-induced misallocation.

Figure 1.10: Measured Marginal Product using First Differences vs TFPR



Lastly, we follow Bils, Klenow, and Ruane (2017) and run the following regression to further quantify the extent to which measured average products reflect true marginal products:

$$\triangle \widehat{VA}_i = \Phi \cdot \log(TFPR_i) + \Psi \cdot \triangle \hat{I}_i - \Psi(1-\lambda) \cdot \log(TFPR_i) \cdot \triangle \hat{I}_i + D_s + \xi_i \qquad (20)$$

where $\triangle \widehat{VA}_i$ and $\triangle \hat{I}_i$ are the growth rate of measured value added and inputs respectively, and $\log(TFPR_i)$ is the measured average products. The underlying assumption here is that the measurement errors are additive. The variable of interest in the regression is λ, the variance of distortions relative to that of $TFPR$:

$$\lambda = \frac{\sigma^2_{\ln \tau}}{\sigma^2_{\ln(TFPR)}}.$$

The regression coefficient for Ψ is 0.53 and for the interaction of $\log(TFPR_i)$ and $\triangle \hat{I}_i$ is -0.0997. Both are significant, and the robust t-statistics are reported in Table 10. The implied λ is therefore 0.81. Hence, 81% of variation in $TFPR$ or average products is accounted for by distortion τ and 19% is due to measurement errors. The results are robust if we weight the observations with their share of aggregate value added or if we control for higher orders of $\ln(TFPR)$ to allow for stationary shocks to firms productivity and distortions. See the Appendix for details.[16]

In summary, the three alternative ways of sifting out measurement errors using panel data all point to the result that the dispersion in the average product of inputs are mainly driven by distortions rather than measurement error typically conceived.

6 Conclusion

This paper evaluates efficiency and welfare gains to trade in an economy with firm-level distortions. We find that the existence of distortions that drive differences in marginal products across firms substantially reduces the welfare gains to trade and may even bring

[16] Bils, Klenow, and Ruane (2017) also considers the following extension to allow for stationary shocks to firms productivity and distortions:

$$\begin{aligned}\triangle \widehat{VA}_i = {} & \Phi \cdot \log(TFPR_i) + \Psi \cdot \triangle \hat{I}_i - \Psi(1-\lambda) \cdot \log(TFPR_i) \\ & + \Gamma \cdot [\log(TFPR_i)]^2 + \Lambda(1-\lambda) \cdot [\log(TFPR_i)]^2 \triangle \hat{I}_i \\ & + \Upsilon \cdot [\log(TFPR_i)]^3 + \Lambda(1-\lambda) \cdot [\log(TFPR_i)]^3 \triangle \hat{I}_i.\end{aligned}$$

about welfare and efficiency losses. For the case of China, trade gains are much lower than standard predictions based on observed trade flows and trade elasticities. There is also a TFP loss rather than a significant TFP gain–which is at the core of the new trade theory models in which trade engenders a flow of resources towards the more productive firms. ACR formula suggests that one can abstract from micro-level distributions to estimate macro-level gains, fully inferred from observed variables on trade flows and trade elasticities. We show that firm-level distortions have large impact on trade gains. Further work can be done on evaluating how these wedges respond to trade liberalization. Our paper points to the fact that distortions may be an important factor determining countries' experiences with trade liberalization.

References

Anderson, James E and Eric Van Wincoop. 2004. "Trade costs." *Journal of Economic literature* 42 (3):691–751.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, Andrés Rodríguez-Clare et al. 2015. "The elusive pro-competitive effects of trade." Tech. rep., National Bureau of Economic Research.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New trade models, same old gains?" *American Economic Review* 102 (1):94–130.

Bils, Mark, Peter J. Klenow, and Cian Ruane. 2017. "Misallocation or Mismeasurement?" Working paper, University of Rochester.

Costa-Scottini, Lucas. 2018. "Firm-Level Distortions, Trade, and International Productivity Differences." Working paper, Brown University.

David, Joel M. and Venky Venkateswaran. 2017. "The Sources of Capital Misallocation." Tech. rep., Working paper.

Ho, Giang T. 2010. "Trade Liberalization with Idiosyncratic Distortions: Theory and Evidence from India." Working paper.

Hsieh, Chang-Tai and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly journal of economics* 124 (4):1403–1448.

Melitz, Marc J and Stephen J Redding. 2015. "New trade models, new welfare implications." *American Economic Review* 105 (3):1105–46.

Song, Zheng and Guiying Laura Wu. 2015. "Identifying capital misallocation." Tech. rep., Working paper.

Alessandria G., Choi H,. Lu, D. (2016)."Trade Integration and the Trade Balance in China". IMF review.

Anzoategui, D., Chivakul, M., Maliszewski, W. (2015)."Financial Distortions in China: A General Equilibrium Approach". IMF Working Papers 15/274.

Arkolakis C., Costinot A. and Donaldson D. (2017)."The Elusive Pro-Competitive Effects of Trade".

J. Asturias, S. Hur, T. J. Kehoe, and K. J. Ruhl (2016), ?The Interaction and Sequencing of Policy Reforms,? Journal of Economic Dynamics and Control, 72, 45?66. Bai, Y., Lu, D., and Tian, X. (2016)."Do Financial Frictions Explain Chinese Firms Saving and Misallocation?"

Bai, C., Hsieh, C., and Song, Z. (2019)."Special Deals with Chinese Characteristics. "

Badlwin, R. E., and F. ROBERT-NICOUD (2008): ?Trade and Growth with Heterogeneous Firms,? Journal of International Economics, 74(1), 21?34.

BENHABIB, J., J. PERLA, AND C. TONETTI (2015): ?The Growth Dynamics of Innovation, Diffusion, and the Technology Frontier,? working paper.

Brandt, Loren, Chang-tai Hsieh, and Xiaodong Zhu. (2007)."Growth and Structural Transformation in China". In China"s Great Economic Transformation. Cambridge University Press.

L. Brandt, J. Van Biesebroeck, and Y. Zhang (2012), ?Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing,? Journal of Development Economics, 97, 339?351.

Caliendo L., Parro F. and Tsyvinki A. (2017)."Distortions and the Structure of the World Economy".

Chang C,. Chen K,. Waggoner Daniel F,. Zha T,. (2015).“Trends and Cycles in China?s Macroeconomy”. Working paper 21244. National Bureau of Economic Research.

Chen, T,. and J. Kung. (2019).“Busting the ?Princelings?: The Campaign against Corruption in China?s Primary Land Market”. Working paper, HKUST.

Chang, R., Kaltani, L., and Loayza, N. V. (2009). Openness can be Good for Growth: The Role of Policy Complementarities. Journal of Development Economics, 90(1):33? 49.

Cheremukhin, Anton, Mikhail Golosov, Sergei Guriev, and Aleh Tsyvinski. (2015).“The Economy of People’s Republic of China from 1953”, National Bureau of Economic Research Working paper 21397.

Costinot A., Rodrigues-Claire A., Werning I. (2016).“Micro and Macro: Optimal Trade Policy with Firm Heterogeneity”. Working paper 21989. National Bureau Of Economic Research.

Costinot A. and Rodriguez-Claire A.(2014).“Trade Theory with Numbers: Quantifying the Consequences or Globalization”. Handbook of international Economics, Volume 4. ISSN 1573-4404.

Demidova S. and Rodriguez-Clare A. (2009).“Trade Policy Under Firm-Level Heterogeneity in a small economy”. Journal of International Economics 78 100-112.

Defever, F. and A. Riaño (2012).“Protectionism through Exporting: Subsidies with Export Share Requirements in China” CEP Discussion Paper 1182.

Dhingra S. and Morrow J. (2012).“The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition. CEP Discussion Paper No. 1130.

Edmond C., Midrigan V. and Yi Xu D., (2015).“Competition, Markups and the Gains from International Trade”. American Economic Review 105(10):3183-3221.

Ge, S., and Yang, D. T. (2014).“Changes in China’s Wage Structure”. Journal of the European Economic Association, 300-336.

Guo D., Jiang, K., Kim B., and Xu, C. (2014). “The Political Economy of Private Firms in China”. Journal of Comparative Economics Volume 42, Issue 2, May 2014, Pp 286?303

Haley, U. C., and Haley, G. T. (2013)." Subsidies to Chinese Industry: State Capitalism, Business Strategy and Trade Policy". USA: Oxford University Press.

Haltiwanger J. (2008)."Top Ten Questions for Understanding Firm Dynamics and Productivity Growth". University of Maryland, NBER and IZA.

Harrison, A. E. and Rodr??guez-Clare, A. (2010). Trade, Foreign Investment, and Industrial Policy for Developing Countries, volume 5 of Handbook of Development Economics. Elsevier.

Headd B,. Nucci A,. and Boden R,."What Matters More: Business Exit Rates or Business Survival Rates?" (2005)

Hsieh, Chang-Tai, and Peter J. Klenow. (2013). "Misallocation and Manufacturing TFP in China and India". Correction Appendix.

Huang, Y., Pagano, M., and Panizza, U. (2015)."Public Debt and Private Firm Funding: Evidence from Chinese Cities".

Itskhoki, O., and Moll, B. (2015)."Optimal Development Policies with Financial Frictions". Mimeo.

Johansson, A. C. (2012)."Financial Repression and China's Economic Imbalances". In Rebalancing and Sustaining Growth in China (pp. 45 - 65).

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

Melitz Marc J. (2003)."The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". Econometrica, Vol 71, No. 6. Pp 1695-1725.

Melitz, Marc J., and Stephen Redding. 2015. "New Trade Models, New Welfare Implications." American Economic Review 105 (3): 1105-46.

Rodrigues-Clare A. (2018)."Globalization, And the Gains from Trade in Rich and Poor Countries", in *Meeting Globalization's Challenges*, Princeton University Press.

SAMPSON, T. (2015): ?Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth,? Quarterly Journal of Economics, forthcoming.

Scott, R. E. (2008).“The China Trade Toll”. Economic Policy Institute .

Tombe Trevor and Xiaodong Zhu (2015). “Trade, Migration and Productivity: A Quantitative Analysis of China”, University of Toronto Working Paper No. 542, 2015.

Wu Laura G,. (2015).“Capital Misallocation in China: Financial Frictions or Policy Distortions?”

Wu, Harry X. (2013).“Measuring and Interpreting Total Factor Productivity in Chinese Industry”. Comparative Studies, 69.

7 Appendix

7.1 Model Derivation

Closed Economy Equilibrium. In a closed economy, the free entry condition requires that the present value of producing equals the entry cost. The probability of entry ω_e is given by

$$\omega_e = \int_0^\infty \int_{\varphi^*(t)}^\infty g\left(\varphi, t\right) d\varphi dt, \tag{21}$$

and the distribution of operating firms $\mu(\varphi, t)$

$$\mu(\varphi, t) = \frac{g(\varphi, t)}{\int_0^\infty \int_{\varphi^*(t)}^\infty g\left(\varphi, t\right) d\varphi dt} = \frac{g(\varphi, t)}{\omega_e}$$

if $\varphi \geq \varphi^*(t)$; and 0 otherwise. Let the per-period expected profit conditional on producing be

$$E[\pi(\varphi, t)] = \int_0^\infty \int_{\varphi^*(t)}^\infty \pi(\varphi, t)\mu\left(\varphi, t\right) d\varphi dt.$$

The free entry condition is given by

$$\omega_e \sum_{s=0}^{\infty} (1-\delta)^s E[\pi(\varphi, t)] = \omega_e \frac{E[\pi(\varphi, t)]}{\delta} = w f_e. \tag{22}$$

Normalizing domestic wage such that $w = 1$, a closed-economy equilibrium consists of aggregate price P, the measure of producing firms M, and individual firms' decisions $\{p(\varphi, t), q(\varphi, t)\}$.

The expected profit of entry upon entry is

$$\begin{aligned} \omega_e E[\pi(\varphi, t)] &= \left[\int_0^\infty \int_{\varphi^*(t)}^\infty g\left(\varphi, t\right) d\varphi dt\right] \int_0^\infty \int_{\varphi^*(t)}^\infty \pi(\varphi, t)\mu\left(\varphi, t\right) d\varphi dt \\ &= \int_0^\infty \int_{\varphi^*(t)}^\infty \pi(\varphi, t) g\left(\varphi, t\right) d\varphi dt. \end{aligned}$$

The optimal profit function (3) combined with the above expression in the free entry con-

dition yields

$$\int \int_{\varphi^*(t)} \left[\sigma^{-\sigma} (\sigma - 1)^{\sigma-1} P^{\sigma} Q w^{1-\sigma} \varphi^{\sigma-1} t^{-\sigma} - wf \right] g(\varphi, t) d\varphi dt = w\delta f_e,$$

which yields an equation for P, in

$$\frac{PQ}{\sigma} \left(P \frac{\sigma - 1}{\sigma} \right)^{\sigma-1} w^{1-\sigma} \int \int_{\varphi^*(t)} \left[\varphi^{\sigma-1} t^{-\sigma} \right] g(\varphi, t) d\varphi dt - wf \int \int_{\varphi^*(t)} g(\varphi, t) d\varphi dt = w\delta f_e. \tag{23}$$

Let M_e be the measure of new entrant. The labor market clearing condition requires

$$L = ME \left[\frac{q}{\varphi} + f \right] + M_e f_e,$$

where the average labor demanded by firms $E\left[\frac{q}{\varphi} + f\right]$ is given by

$$E \left[\frac{q}{\varphi} + f \right] = \int_0^{\infty} \int_{\varphi^*(t)}^{\infty} \left[\frac{q}{\varphi} + f \right] \mu(\varphi, t) \, d\varphi dt.$$

In a stationary equilibrium, the number of entrants equals number of exits, such that $\omega_e M_e = \delta M$. Replacing M_e with $\delta M / \omega_e$ yields

$$M = \frac{L}{E\left[\frac{q}{\varphi} + f\right] + \delta f_e / \omega_e}. \tag{24}$$

The firm's expected optimal profit can be written as

$$E[\pi(\varphi, t)] = E\left[\frac{1}{\sigma - 1} \frac{wq}{\varphi} - wf \right],$$

which together with the free-entry condition (22) implies

$$\omega_e \left[E\left(\frac{1}{\sigma - 1} \frac{wq}{\varphi} \right) - wf \right] = w\delta f_e.$$

Normalizing w, and plugging the above into (24), we have

$$M = \frac{L}{\sigma\left(\frac{\delta f_e}{\omega_e} + f\right)}. \tag{25}$$

Open Economy Equilibrium. Optimal prices and cutoff functions are straightforward analogues of the closed economy case. The free entry condition for home implies

$$\int\int_{\varphi^*(t)} \pi(\varphi,t)g(\varphi,t)d\varphi dt + \int\int_{\varphi_x^*(t)} \pi_x(\varphi,t)g(\varphi,t)d\varphi dt = w\delta f_e,$$

or,

$$\begin{aligned}
&\frac{PQ}{\sigma}\left(P\frac{\sigma-1}{\sigma}\right)^{\sigma-1} w^{1-\sigma}\int_0^\infty\int_{\varphi^*(t)}^\infty \left[\varphi^{\sigma-1}t^{-\sigma}\right] g(\varphi,t)\,d\varphi dt - wf\int_0^\infty\int_{\varphi^*(t)}^\infty g(\varphi,t)\,d\varphi dt \\
&+\left[\frac{P_fQ_f}{\sigma}\left(P_f\frac{\sigma-1}{\sigma}\right)^{\sigma-1}(\tau_x w)^{1-\sigma}\int_0^\infty\int_{\varphi_x^*(t)}^\infty \left[\varphi^{\sigma-1}t^{-\sigma}\right] g(\varphi,t)\,d\varphi dt - wf_x\int_0^\infty\int_{\varphi_x^*(t)}^\infty g(\varphi,t)\,d\varphi dt\right] \\
&\qquad = \delta w f_e. \quad (26)
\end{aligned}$$

Similarly, the free entry condition for foreign country is

$$\begin{aligned}
&\frac{P_fQ_f}{\sigma}\left(P_f\frac{\sigma-1}{\sigma}\right)^{\sigma-1} w_f^{1-\sigma}\int_0^\infty\int_{\varphi_f^*(t)}^\infty \left[\varphi^{\sigma-1}t^{-\sigma}\right] g_f(\varphi,t)\,d\varphi dt - w_f f\int_0^\infty\int_{\varphi_f^*(t)}^\infty g_f(\varphi,t)\,d\varphi dt \\
&+\left[\frac{PQ}{\sigma}\left(P\frac{\sigma-1}{\sigma}\right)^{\sigma-1}(\tau_x w_f)^{1-\sigma}\int_0^\infty\int_{\varphi_{xf}^*(t)}^\infty \varphi^{\sigma-1}t^{-\sigma} g_f(\varphi,t)\,d\varphi dt - w_f f_x\int_0^\infty\int_{\varphi_{xf}^*(t)}^\infty g_f(\varphi,t)\,d\varphi dt\right] \\
&\qquad = \delta w_f f_e. \quad (27)
\end{aligned}$$

It follows that the aggregate prices of Home and Foreign are

$$P^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left[Mw^{1-\sigma}\frac{\int_0^\infty\int_{\varphi^*(t)}^\infty(\frac{\varphi}{t})^{\sigma-1}g(\varphi,t)\,d\varphi dt}{\int_0^\infty\int_{\varphi^*(t)}^\infty g(\varphi,t)\,d\varphi dt} + M_f(\tau_x w_f)^{1-\sigma}\frac{\int_{\varphi_{xf}^*(t)}^\infty(\frac{\varphi}{t})^{\sigma-1}g_f(\varphi,t)\,d\varphi dt}{\int_0^\infty\int_{\varphi_f^*(t)}^\infty g_f(\varphi,t)\,d\varphi dt}\right]$$

$$P_f^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left[M_f w_f^{1-\sigma}\frac{\int_0^\infty\int_{\varphi^*(t)_f}^\infty(\frac{\varphi}{t})^{\sigma-1}g_f(\varphi,t)\,d\varphi dt}{\int_0^\infty\int_{\varphi_f^*(t)}^\infty g_f(\varphi,t)\,d\varphi dt} + Mw^{1-\sigma}\tau_x^{1-\sigma}\frac{\int_{\varphi^*(t)_x}^\infty(\frac{\varphi}{t})^{\sigma-1}g(\varphi,t)\,d\varphi dt}{\int_0^\infty\int_{\varphi^*(t)}^\infty g(\varphi,t)\,d\varphi dt}\right]$$

Next we derive the measure of operating firms at home and foreign M and M^*. Let M_e be the measure of new entrant, ω_e be the entry probability given by (21), and ω_x be the export probability conditional on entry

$$\omega_x = \int_0^\infty \int_{\varphi_x^*(t)}^\infty \mu\left(\varphi, t\right) d\varphi dt = \frac{\int_0^\infty \int_{\varphi_x^*(t)}^\infty g\left(\varphi, t\right) d\varphi dt}{\int_0^\infty \int_{\varphi^*(t)}^\infty g\left(\varphi, t\right) d\varphi dt}.$$

The expected per-period profit includes the profit from both domestic production and international production,

$$\begin{aligned}
&\int \int_{\varphi^*(t)} \pi(\varphi, t)\mu(\varphi, t)d\varphi dt + \int \int_{\varphi_x^*(t)} \pi_x(\varphi, t)\mu(\varphi, t)d\varphi dt \\
&= \int \int_{\varphi^*(t)} \pi(\varphi, t)\mu(\varphi, t)d\varphi dt + \omega_x \int \int_{\varphi_x^*(t)} \pi_x(\varphi, t)\mu_x(\varphi, t)d\varphi dt \\
&\equiv E\pi + \omega_x E\pi_x
\end{aligned}$$

where $\mu_x(\varphi, t) = \mu(\varphi, t)/\omega_x$.

Average profits are

$$E\pi = E\left[\frac{1}{\sigma - 1}\frac{wq}{\varphi} + wf\right], \quad E\pi_x = E\left[\frac{1}{\sigma - 1}\frac{wq_x}{\varphi} + wf_x\right].$$

Free entry implies that

$$E\left[\frac{1}{\sigma - 1}\frac{wq}{\varphi} + wf\right] + \omega_x E\left[\frac{1}{\sigma - 1}\frac{wq_x}{\varphi} + wf_x\right] = \frac{w\delta f_e}{\omega_e}.$$

In a stationary equilibrium, $\pi_e M_e = \delta M$. The labor market condition is

$$\begin{aligned}
L &= M\left(E\frac{q}{\varphi} + f\right) + M\left(E\frac{q_x}{\varphi} + \omega_x f_x\right) + M_e f_e \\
&= M\sigma\left(\delta\frac{f_e}{\omega_e} + f + \omega_x f_x\right)
\end{aligned}$$

Using the free entry condition above, and using the definition for ω_e and ω_x, we have

$$M = \frac{L}{\sigma\left(\frac{\delta f_e}{\int_0^\infty \int_{\varphi^*(t)}^\infty g(\varphi,t)d\varphi dt} + f + \frac{\int_0^\infty \int_{\varphi_x^*(t)}^\infty g(\varphi,t)d\varphi dt}{\int_0^\infty \int_{\varphi^*(t)}^\infty g(\varphi,t)d\varphi dt} f_x\right)} \tag{28}$$

and

$$M_f = \frac{L_f}{\sigma\left(\frac{\delta f_e}{\int_0^\infty \int_{\varphi_f^*(t)}^\infty g_f(\varphi,t)d\varphi dt} + f + \frac{\int_0^\infty \int_{\varphi_{xf}^*(t)}^\infty g_f(\varphi,t)d\varphi dt}{\int_0^\infty \int_{\varphi_f^*(t)}^\infty g_f(\varphi,t)d\varphi dt} f_x\right)}. \tag{29}$$

A. Proof for Proposition 2

Consider two symmetric countries with domestic and trade frictions. To derive Proposition 2, we use the following four equilibrium equations,

(a) Free entry condition:

$$P^\sigma Q \frac{1}{\sigma}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} w^{1-\sigma}\left[\int\int_{\varphi^*(\tau)} \varphi^{\sigma-1}\tau^{-\sigma} g(\varphi,\tau)d\varphi d\tau + \tau_x^{1-\sigma}\int\int_{\varphi_x^*(\tau)} \varphi^{\sigma-1}\tau^{-\sigma} g(\varphi,\tau)d\varphi d\tau\right]$$
$$= \delta w f_e + wfH + wf_x H_x$$

(b) The labor market clearing condition:

$$M_e = \frac{L}{\sigma\left(\delta f_e + f + H_x f_x\right)}.$$

(c) Price index:

$$P^{1-\sigma} = con_p \times \left[M_e \int\int_{\varphi^*(\tau)} \varphi^{\sigma-1}\tau^{1-\sigma} g(\varphi,\tau)d\varphi d\tau + M_e \tau_x^{1-\sigma}\int_{\varphi_x^*(\tau)} \varphi^{\sigma-1}\tau^{1-\sigma} g(\varphi,\tau)d\varphi d\tau\right]$$

where con_p is a constant depends on the model parameters.

(d) Cutoff of producing:

$$\varphi^* = con_v \times P^{-1}\left(PQ\right)^{\frac{1}{1-\sigma}} \tau^{\frac{\sigma}{\sigma-1}}$$

where con_v is a constant depends on the model parameters.

(e) Definition of domestic share λ:

$$\frac{1-\lambda}{\lambda} = \frac{\tau_x^{1-\sigma}\left[\int\int_{\varphi_x^*}\varphi^{\sigma-1}\tau^{1-\sigma}g(\varphi,\tau)d\varphi d\tau\right]}{\left[\int\int_{\varphi^*}\varphi^{\sigma-1}\tau^{1-\sigma}g(\varphi,\tau)d\varphi d\tau\right]}$$

(f) Definition of the share of labor used in producing domestic goodsS_d:

$$S_d = \frac{\int\int_{\varphi^*(t)}\varphi^{\sigma-1}t^{-\sigma}g(\varphi,t)d\varphi dt}{\int\int_{\varphi^*(t)}\varphi^{\sigma-1}t^{-\sigma}g(\varphi,t)d\varphi dt + \tau_x^{1-\sigma}\int\int_{\varphi_x^*(t)}\varphi^{\sigma-1}t^{-\sigma}g(\varphi,t)d\varphi dt}$$

Differentiating the above system of equations and using the relationship $d\ln W = d\ln(wL+T) - d\ln P$, we get Proposition 2 with the following definitions:

$$\begin{aligned} d\ln Q = \frac{1}{\gamma_{sd}+\sigma-1}[&-d\ln\lambda \qquad (ACR)\\ &+ d\ln M_e \qquad (MR)\\ &+ \frac{\sigma}{\sigma-1}(\gamma_{sd}-\gamma_d)d\ln M_e\\ &- \left(\sigma-1+\frac{\sigma\gamma_{sd}}{\sigma-1}\right)d\ln\lambda + \left(\sigma-1+\frac{\sigma\gamma_d}{\sigma-1}\right)d\ln S_d] \end{aligned}$$

where

$$\gamma_d = \frac{\int(\varphi^*(\tau))^{\sigma-1}\tau^{1-\sigma}g(\varphi^*(\tau),\tau)\varphi^*(\tau)d\tau}{\int\int_{\varphi^*(\tau)}\varphi^{\sigma-1}\tau^{1-\sigma}g(\varphi,\tau)d\varphi d\tau}$$

$$\gamma_{sd} = \frac{\int(\varphi^*(\tau))^{\sigma-1}\tau^{-\sigma}g(\varphi^*(\tau),\tau)\varphi^*(\tau)d\tau}{\int\int_{\varphi^*(\tau)}\varphi^{\sigma-1}\tau^{-\sigma}g(\varphi,\tau)d\varphi d\tau}$$

With homogenous productivity and if $1/\tau$ follows a Pareto distribution with parameter θ, $\gamma_d = \frac{\sigma-1}{\sigma}(\theta-\sigma+1)$ and $\gamma_{sd} = \frac{\sigma-1}{\sigma}(\theta-\sigma)$, hence $dlnW = \frac{\sigma}{\sigma-1}[dlnS_d - dln\lambda]$. Changes in the share of expenditure on domestic goods and the share of labor used in producing domestic goods are sufficient statistics for welfare change. In addition, the open economy has an unambiguously lower welfare than in the closed economy. Because with distortion,

in open economy, the input share using to produce for exports exceeds the export share,

$$\frac{S_x}{1-\lambda} = \frac{\tau_x^{\frac{(\sigma-1)(1+\gamma)}{\sigma}}(f/f_x)^{\frac{\sigma-\gamma-1}{\sigma}}+1}{\tau_x^{\frac{(\sigma-1)(\gamma)}{\sigma}}(f/f_x)^{\frac{\sigma-\gamma}{\sigma}}+1} \geq 1,$$

and $lnS_d \leq ln\lambda$.

B. Equilibrium under endogenous markup

A closed-economy equilibrium consists of aggregate (P, Q, M) that satisfy:

$$M = \frac{\omega_e L}{Q\int_0^{\sigma^{\sigma/\varepsilon}}\int^{\hat{\tau}(\hat{q})}\left[\frac{\sigma}{\sigma-\hat{q}^{\frac{\varepsilon}{\sigma}}}\frac{\hat{q}}{\varphi}\right]g(\varphi(\tau,\hat{q}),\tau)\frac{d\varphi(\tau,\hat{q})}{dq}d\tau d\hat{q}}$$

$$Q\int_0^{\sigma^{\sigma/\varepsilon}}\int^{\hat{\tau}(\hat{q})}\left[\frac{\hat{q}^{\frac{\varepsilon}{\sigma}}}{\sigma-\hat{q}^{\frac{\varepsilon}{\sigma}}}\frac{\hat{q}}{\varphi}\right]g(\varphi(\tau,\hat{q}),\tau)\frac{d\varphi(\tau,\hat{q})}{dq}d\tau d\hat{q} = \omega_e f + \delta f_e$$

$$\frac{M}{\omega_e}\int_0^{\sigma^{\sigma/\varepsilon}}\int^{\hat{\tau}(\hat{q})}\gamma(\hat{q})\,g(\varphi(\tau,\hat{q}),\tau)\frac{d\varphi(\tau,\hat{q})}{dq}d\tau d\hat{q} = 1,$$

where

$$\omega_e = \int_0^{\sigma^{\sigma/\varepsilon}}\int^{\hat{\tau}(\hat{q})} g(\varphi(\tau,\hat{q}),\tau)d\tau d\hat{q}$$

and

$$\gamma'\left(\frac{q}{Q}\right) = \frac{\sigma-1}{\sigma}\exp\left(\frac{1-(q/Q)^{\frac{\varepsilon}{\sigma}}}{\varepsilon}\right).$$

The open equilibrium consists of unknowns $(P, Q, M, P_f, Q_f, M_f, w_f)$ that satisfy:

$$\frac{\sigma}{\sigma-\hat{q}_x^{\frac{\varepsilon}{\sigma}}}\frac{w\tau_x\tau}{\varphi} = \gamma'(\hat{q}_x)P_f D_f$$

$$\pi_x = \left[\frac{\hat{q}_x^{\frac{\varepsilon}{\sigma}}}{\sigma-\hat{q}_x^{\frac{\varepsilon}{\sigma}}}\frac{\tau_x\hat{q_x}}{\varphi}Q_f - f_x\right]w,$$

where we get the zero profit cutoff. The free entry condition becomes:

$$\int\int_{\varphi^*(\tau)} \left[\frac{\hat{q}^{\frac{\varepsilon}{\sigma}}}{\sigma - \hat{q}^{\frac{\varepsilon}{\sigma}}} \frac{\hat{q}}{\varphi} Q - f \right] g(\varphi,\tau) d\tau d\varphi \tag{30}$$
$$+ \int\int_{\varphi_x^*(\tau)} \left[\frac{\hat{q}_x^{\frac{\varepsilon}{\sigma}}}{\sigma - \hat{q}_x^{\frac{\varepsilon}{\sigma}}} \frac{\tau_x \hat{q}_x}{\varphi} Q_f - f_x \right] g(\varphi_x,\tau) d\tau d\varphi = \delta f_e$$

The labor market clearing condition is:

$$M = \frac{\omega_e L}{\int\int_{\varphi^*(\tau)} \left(\frac{\sigma}{\sigma - \hat{q}^{\frac{\varepsilon}{\sigma}}} \frac{\hat{q}}{\varphi} Q \right) g(\varphi,\tau) d\tau d\varphi + \int\int_{\varphi_x^*(\tau)} \left(\frac{\sigma}{\sigma - \hat{q}_x^{\frac{\varepsilon}{\sigma}}} \frac{\tau_x \hat{q}_x}{\varphi} Q_f \right) g(\varphi,\tau) d\tau d\varphi} \tag{31}$$

$$\left[\frac{M}{\omega_e} \int\int_{\varphi^*(\tau)} \gamma\left(\hat{q}\right) g(\varphi,\tau) d\tau d\hat{q} + \frac{M_f}{\omega_{ef}} \int\int_{\varphi_{xf}^*(\tau)} \gamma\left(\hat{q}_{xf}\right) g_f(\varphi,\tau) d\tau d\varphi \right] = 1 \tag{32}$$

For the foreign,

$$\int\int_{\varphi_f^*(\tau)} \left[\frac{\hat{q}_f^{\frac{\varepsilon}{\sigma}}}{\sigma - \hat{q}_f^{\frac{\varepsilon}{\sigma}}} \frac{\hat{q}_f}{\varphi} Q_f - f \right] g_f(\varphi,\tau) d\tau d\varphi \tag{33}$$
$$+ \int\int_{\varphi_{xf}^*(\tau)} \left[\frac{\hat{q}_{xf}^{\frac{\varepsilon}{\sigma}}}{\sigma - \hat{q}_{xf}^{\frac{\varepsilon}{\sigma}}} \frac{\tau_x \hat{q}_{xf}}{\varphi} Q - f_x \right] g_f(\varphi,\tau) d\tau d\varphi = \delta f_{ef}$$

$$M_f = \frac{\omega_e L_f}{\int\int_{\varphi_f(\tau)} \left(\frac{\sigma}{\sigma - \hat{q}_f^{\frac{\varepsilon}{\sigma}}} \frac{\hat{q}}{\varphi} Q_f \right) g_f(\varphi,\tau) d\tau d\varphi + \int\int_{\varphi_{xf}^*(\hat{q})} \left(\frac{\sigma}{\sigma - \hat{q}_{xf}^{\frac{\varepsilon}{\sigma}}} \frac{\tau_x \hat{q}_{xf}}{\varphi} Q \right) g_f(\varphi,\tau) d\tau d\hat{q}} \tag{34}$$

$$\left[\frac{M_f}{\omega_{ef}} \int\int_{\varphi_f^*(\tau)} \gamma\left(\hat{q_f}\right) g_f(\varphi,\tau) d\tau d\hat{q} + \frac{M}{\omega_e} \int\int_{\varphi_x^*(\tau)} \gamma\left(\hat{q}_x\right) g(\varphi,\tau) d\tau d\varphi \right] = 1 \tag{35}$$

Finally, the goods market clearing condition is:

$$\frac{M}{\omega_e} \int\int_{\varphi_x^*(\tau)} \left[\frac{\sigma}{\sigma - \hat{q}^{\frac{\varepsilon}{\sigma}}} \frac{w\hat{q}}{\varphi} Q_f \right] g(\varphi,\tau) d\tau d\varphi = \frac{M_f}{\omega_{ef}} \int\int_{\varphi_{xf}^*(\tau)} \left[\frac{\sigma}{\sigma - \hat{q}_{xf}^{\frac{\varepsilon}{\sigma}}} \frac{w_f \hat{q}_{xf}}{\varphi} Q \right] g_f(\varphi,\tau) d\tau d\varphi \tag{36}$$

C. Regressions for measurement errors

Table 11: Measured Marginal Products using First Differences vs TFPR

VARIABLES	(1) $\log(\frac{\triangle VA}{\triangle I})$	(2) $\log(\frac{\triangle VA}{\triangle I})$	(3) $\log(\frac{\triangle VA}{\triangle I})$
$\log(TFPR)$	0.718***	0.715***	0.718***
	(135.3)	(158.6)	(135.3)
Constant	1.410***	0.331***	1.410***
	(78.31)	(17.49)	(78.31)
Observations	624,659	624,699	624,659
R-squared	0.173	0.269	0.173
Time FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Specification (2) weights all the observations with the absolute value of composite input growth.
Specification (3) weights all the observations with the share of aggregate value added.

Table 12: Estimate Measurement Error

VARIABLES	(1) $\widehat{\triangle VA}$	(2) $\widehat{\triangle VA}$	(3) $\widehat{\triangle VA}$
$\log(TFPR)$	0.0376***	0.0144***	0.0616***
	(22.62)	(9.170)	(16.07)
$[\log(TFPR)]^2$			-0.0128***
			(-6.110)
$[\log(TFPR)]^3$			0.00152***
			(4.008)
$\widehat{\triangle input}$	0.530***	0.523***	0.524***
	(34.58)	(33.03)	(31.13)
$\log(TFPR) \times \widehat{\triangle input}$	-0.0997***	-0.0954***	-0.0893***
	(-20.65)	(-19.16)	(-6.420)
$[\log(TFPR)]^2 \times \widehat{\triangle input}$			-0.00611
			(-0.919)
$[\log(TFPR)]^3 \times \widehat{\triangle input}$			0.00108
			(1.040)
Constant	-0.0207***	0.0551***	-0.0241***
	(-3.125)	(8.231)	(-3.592)
Observations	1,106,982	1,106,914	1,106,982
R-squared	0.044	0.042	0.044
Time FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Specification (2) weights all the observations with the share of aggregate value added.