Title
Essays in Macroeconomics and International Economics

Author
Yang, Mu-Jeung

Publication Date
2012

Peer reviewed|Thesis/dissertation

Essays on Macroeconomics and International Economics

By

Mu-Jeung Yang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chang-Tai Hsieh, Co-Chair

Professor Pierre-Olivier Gourinchas, Co-Chair

Professor Charles I. Jones

Professor Marcus Opp

Spring 2012

Abstract

Essays on Macroeconomics and International Economics

by

Mu-Jeung Yang

Doctor of Philosophy in Economics

University of California, Berkeley

Chang-Tai Hsieh, Co-Chair

Pierre-Olivier Gourinchas, Co-Chair

This dissertation analyzes the interaction of firm level heterogeneity and selection. This selection can encompass either the survival of firms in the market, or their participation in international trade. In the first part I concentrate on the question of how large the aggregate productivity losses from the misallocation of resources across firms are. If firm exit is endogenous, micro-frictions can induce extensive margin misallocation: inefficient firms continue to survive (Zombies) and efficient firms are forced to exit (Shadows). I develop and estimate a fully specified model with plant-level micro-data to quantify extensive margin misallocation. This method allows me to identify Zombie firms, estimate the efficiency distribution of Shadow firms and calculate aggregate Total Factor Productivity (TFP) gains when Zombies are replaced by Shadows. Estimates for Indonesia show that extensive margin misallocation can increase aggregate TFP losses from micro-distortions by over 40%. Compared to existing estimates aggregate TFP losses from micro-distortions are 50-100% higher.

In the second part of the dissertation I trace out the implications of a simplified version of the framework I developed in the first part to questions of international trade. The cross-country comparisions in measured micro-distortions suggest that differences in firm heterogeneity could be potentially important to explain aggregate TFP and therefore also trade patterns. I consequently develop a tractable multi-country general equilibrium model of such differences in firm level heterogeneity across countries. I show how the model naturally links measured trade frictions to national firm-efficiency distributions and endogenously generates asymmetries in trade flows in the absence of asymmetric trade frictions. The model is able to generate key stylized facts, specifically the absence of a strong negative relationsship of firm-size dispersions and internal trade shares as predicted by the standard heterogeneous firm trade model with identical efficiency dispersions across countries.

To my mother,
an example in resilience and persistence

Contents

Acknowledgements

Academic research is comparable to the irreversible investment problem of an entrepreneur. Every research paper is like an investment project with uncertain returns and the time invested cannot be recouped. Navigating through these risky prospects is seldomly possible without the occasional stop at a quiet haven. I would like to thank my family, especially my mother, for providing this haven. My gratitute also belongs to my siblings for keeping me down to earth and inspiring my curiosity.

Intellectually, I am indebted to Helmut Booms for first sparking my interest in economics. My advisor Chang-Tai Hsieh was not only a role model for conducting "theory based empirical research", but without his generous support it would not have been possible for me to benefit from the Chicago culture alongside my Berkeley roots. In this context I would also like to express my deepest gratitude to Sam Kortum for the best education in quantitative trade one could hope for. I also owe an intellectual debt to Chad Jones for inspiring me to study "the field of interesting questions". Last but not least Pierre Gourinchas deserves special thanks for his support.

"Graduate students are not bad people, they only make terrible life choices", so Marge Simpson. I owe my sanity to the people with whom I jointly made the same "terrible" life choice. First and foremost among them is the gang of people who stuck together since the very first late nights in the Berkeley law library in the Fall of 2006. Lorenz Küng and Bryan Hong were kindred spirits and the best discussion partners about life and the economic absurdities of it I could have had. I am still waiting for your regression discontinuity paper to appear in AER. Seriously. Also, without Pedro Castro and Evgeny Yakovlev, life in grad school would have been at lot more tense. Second only in mentioning, but not in importance were my office mates Johannes Wieland, John Mondragon and Dominick Bartelme. Future generations of Berkeley macro students will sing songs about our office! And about John eating salad at Poker while Johannes is going all in. Other students who contributed to Berkeley's unique intellectual atmosphere were Nick Li, Gabe Chodorow-Reich and Josh Hausman. Finally, from the many people I met outside of Berkeley during or before my studies Matthias Lux, Raphael Schoenle, Christian Opp, Anna Wong and Louis Serranito deserve special mention.

Chapter 1

Extensive Margin Misallocation Estimation Methodology

1.1 Introduction

The dominant factor behind cross-country income differences is variation in aggregate total factor productivity (TFP) as documented by Hall and Jones (1999). According to a popular recent view exemplified by Hsieh and Klenow (2009), micro-distortions that prevent the optimal allocation of resources across firms within a country significantly contribute to cross-country differences in TFP. Existing empirical studies of micro-level misallocation and TFP do not allow for endogenous exit. As a consequence they can only capture intensive-margin misallocation: some firms are too big and some too small relative to perfectly competitive markets. However, the impact of micro-distortions on exit can have large effects, and potentially explains deep macroeconomic productivity slumps. A case in point is the study by Caballero, Hoshi and Kashyap (2008) on the Japanese Lost Decade. They focus on the retention of Zombie firms: inefficient firms that would otherwise exit under competitive conditions, but are artificially kept in business through implicit subsidies such as preferential access to credit. But, distortions of the exit margin also lead to firms exiting that would have been productive enough to survive in the absence of such distortions. I refer to these as Shadow firms. The key economic mechanism through which Zombies and Shadows impact aggregate TFP is misallocation along the extensive margin: under ideal conditions Zombies should exit and be replaced by Shadows.

In this chapter I develop a fully specified estimator of extensive margin misallocation. It is based on a model of firm heterogeneity under entry and exit. This model features both - micro-distortions as in Hsieh and Klenow (2009) and endogenous entry and exit of firms as in Melitz (2003). Micro-distortions are modeled as firm-specific implicit taxes and subsidies, standing-in for other widely-discussed frictions such as political connections, credit frictions, regulatory size restrictions, firm-specific bargaining power under incomplete contracts or insufficient enforcement of property rights. Monopolistically competitive firms differ in these wedges and in the firm's level of technology. Production involves fixed costs of operations, and, firms exit if their profits are less than these fixed costs. The estimation procedure used in this dissertation allows plant-level efficiency and distortions to be arbitrarily correlated, and is performed using data by narrow 4-digit sectors. What is more, I integrate all sectors into a multi-sector general equilibrium, so that the estimation methodology can be applied to typical establishment level data with industry identifiers.

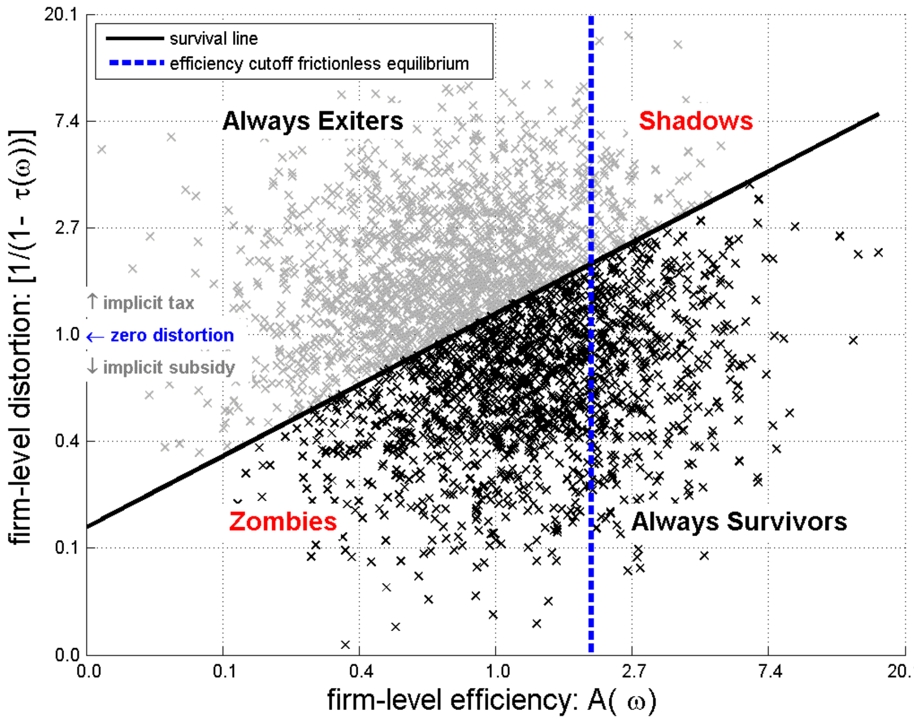

Figure 1.1: Parameters of Interest and Identification (Example simulation)

In evaluating aggregate TFP losses when micro-frictions and endogenous exit interact, I need to address two issues. First, which parameters are needed according to the theory in order to quantify the impact of micro-distortions on aggregate TFP when allowing for endogenous exit? Second, how can these parameters be identified with plant-level micro-data?

To understand what information is needed for evaluating extensive margin misallocation, consider first Figure 1. The x-axis shows firm-level productivity, with higher values for more efficient firms. The y-axis shows firm-level implicit micro-distortion,

with higher values for more heavily-distorted firms. The sloped line, which I call the survival line, captures exit in equilibrium with micro-frictions - firms that are either very productive or are not very distorted survive. All firms southeast of the survival line are therefore observed – in contrast, survival in the frictionless equilibrium is summarized by the dashed vertical line. Selection in an equilibrium without micro-distortions – referred to as "frictionless equilibrium" – is based only on firm efficiency, and therefore only firms to the east of the efficient cut-off survive. The area below the survival line and to the west of the frictionless cut-off contains firms whose productivity is too low to survive a frictionless equilibrium. However, they are nonetheless observed in the data – these are Zombie firms. The area above the survival line and to the east of the frictionless cut-off contains firms that would be productive enough to survive in a frictionless equilibrium, yet are unobserved in the data – these are Shadow firms.
Alleviating extensive margin misallocation would lead to the exit of Zombies and the entry of Shadows. Therefore, quantifying losses from extensive margin misallocation requires two objects. The first is the full underlying distribution of frictions and productivity, including the hypothetical firms northwest of the survival line. The second is the efficiency cut-off in the frictionless equilibrium, as shown by the vertical dashed line in Figure 1.

Identifying these objects works in three steps. In the first step, I separately identify both micro-frictions and firm productivity for the set of observed firms. Here I follow the insight of Hsieh and Klenow (2009) that in a frictionless market marginal revenue products should be equalized. Differences in marginal revenue products can therefore be used as a measure of firm-level frictions. Given an explicit demand system, I measure firm level productivity as the part of firm size that cannot be explained by differences in marginal revenue products.
In the second step, I recover the full underlying distribution of frictions and productivity from the observed data, using a structural model that allows me to account for selection. I assume that the joint distribution of frictions and productivity is log-normal; this assumption is supported by the empirical evidence on firm size distributions, and enhances tractability. The economic model of selection and the distributional assumption together imply that the observed data are generated by a truncated log-normal, subject to equilibrium constraints. Estimating the full underlying distribution therefore reduces to a constrained Maximum Likelihood problem. Since selection cut-offs and size distributions differ by sector, I obtain estimates for narrow 4-digit sectors .
In the third step, I apply the parameters that characterize the underlying distribution of plant-level micro-distortions and productivity by sector and integrate them into a multi-sector general equilibrium model, which is used to compute the counterfactual frictionless equilibrium. This frictionless equilibrium provides the information on the unique frictionless plant productivity cut-off – the vertical dashed line in Figure 1.

Related Literature

This study is part of a rapidly expanding recent literature on the importance of micro-distortions for aggregate productivity and welfare. It is closest in spirit to previous empirical studies that use micro-data to measure the extent of micro-level misallocation. Following the methodology of Hsieh and Klenow (2009), this literature consistently finds large potential aggregate TFP gains from removing intensive-margin misallocation: Argentina could increase its TFP by 50-80% (Neumeyer and Sandleris (2010)), Bolivia by 60-70% (Machicado and Birbuet (2008)), Colombia by 50% (Camacho and Conover (2010)), Chile by 60-80% (Oberfield (2011)) and Uruguay by 50-60% (Casacuberta and Gandelman (2009)). My research contributes to this literature by extending the estimation methodology to account for the endogenous entry and exit of firms.

This leads to two distinct innovations in the measurement and estimation exercise. First, to correct for sample selection effects I employ a constrained Maximum-Likelihood Estimator. This is necessary since factors influencing the survival of firms such as overall market demand and factor prices impact the observable degree of marginal revenue product dispersion.[1] Furthermore, since firms that face high implicit taxes only survive if they are very efficient, selection may bias the measured correlation of frictions and firm efficiency upwards. To my knowledge, this study is the first to systematically address selection issues when measuring micro-frictions. A related technical contribution is the estimation and solution of heterogeneous firm models using joint log-normality with multiple sources of heterogeneity.[2]

Second, the basic Melitz-type endogenous selection model implies that part of the observed factor demand is due to fixed costs of operation. As Bartelsman, Haltiwanger and Scarpetta (2009) point out, the presence of these fixed costs may induce an artificial correlation between measured distortions and plant productivity – suggesting that larger firms are more distorted than they really are. I demonstrate that the model

[1] For example, consider two industries with two firms each, assume firm productivity is equal across firms and industries. Within each industry, one firm is taxed and the other is subsidized. Suppose the tax and subsidy rates are the same for each industry, such that both allocations are distorted to the same degree. Within each industry, the subsidized firm will typically be able to offer its products at lower prices, and will therefore have higher revenues than the taxed firm. If firms have to pay a fixed cost of operation, they will decide to exit if their revenues are too low to cover this fixed cost. Now, suppose that due to a difference in preferences, industry # 2 has such a low market demand that only the subsidized firm can survive, while in industry # 1 both firms survive. An analyst ignoring market selection would observe a difference in marginal revenue products of firms only in industry # 1 where two firms survive, but not in industry # 2. Due to selection forces, the simple dispersion of marginal revenue products might be a misleading indicator about how distorted allocations are.

[2] In principle this innovation also allows me to incorporate differences in firm level dispersions into multi-country trade models with asymmetric trade frictions. This is part of ongoing research. See also chapter 3.

suggests a simple fixed-cost estimator that can be used to correct factor demands and recover marginal revenue products from average revenue products.

My research also builds on another, more theoretical strand of the literature on micro-frictions and aggregate TFP that does allow endogenous entry and exit. Examples are Restuccia and Rogerson (2008), Bartelsman et al. (2009) and Fattal Jaef (2011).[3] These papers typically calibrate either a stylized two-point distribution for distortions, or compare model implications only roughly with data. In contrast, I estimate the relevant parameters from micro-data. Furthermore, none of these theoretical papers analyzes the impact of distortions on the efficiency composition of firms, or decomposes the TFP effects of Zombies and Shadows. Instead, these papers typically focus on the net welfare impact.

This chapter is also close in spirit to the literature on the "Macroeconomics of Restructuring," and especially to Caballero et al. (2008).[4] They provide evidence for the hypothesis that the presence of Zombie firms in Japan could explain the dismal productivity performance during the Lost Decade. In contrast to Caballero et al. (2008), who used reduced-form sectoral TFP regressions, my work imposes more theory-driven structure on the data. I allow for industry and general equilibrium feedbacks of micro-frictions, utilize plant micro-data for the identification of Zombies, and concentrate on steady-state effects of micro-distortions on the level of aggregate TFP. [5]

1.2 Theory

This section develops the general equilibrium model in a simplified setup of a one-sector closed economy with one factor and one wedge only. I derive implications for measurement, and then discuss aggregate TFP effects in this model. In section 3 I derive the full model which is taken to the data.

1.2.1 Model Setup

Consider the following one-sector closed economy. Following standard practice, I assume that plants, indexed by ω, produce differentiated goods and sell them to a competitive final good sector. This final good sector can be characterized by the

[3] See also recent papers analyzing financial frictions as a particular source for micro-frictions, such as Buera, Kaboski and Shin (2011), Dong (2011), Greenwood, Sanchez and Wang (2010), Midrigan and Xu (2010), and Moll (2010).

[4] For a survey of this related strand of the literature, see Caballero (2006).

[5] The question of whether micro-level misallocation can quantitatively account for the deep Japanese slump, compatible with the Hayashi-Prescott hypothesis, is the topic of ongoing research.

following aggregate production function:

$$Y = \left(\int y(\omega)^{\frac{\eta-1}{\eta}} \cdot d\omega \right)^{\frac{\eta}{\eta-1}}$$

where $\eta > 1$ is the elasticity of substitution used to combine the differentiated goods in a CES aggregator. There is a continuum of potential producers. For this section only, I assume that the mass of producers indexed by ω is exogenously given.
Plants choose labor optimally to maximize their profits. Production is given by a simple linear technology:

$$\begin{aligned} &\max_{\{L(\omega)\}} \Pi(\omega) = (1 - \tau(\omega)) \cdot p(\omega) y(\omega) - wL(\omega) \\ &\text{subject to: } \; y(\omega) = A(\omega) \cdot L(\omega) \end{aligned} \tag{1.1}$$

where $L(\omega)$ is labor input used by firm ω in production, $p(\omega)$ is the price this firm charges, and $y(\omega)$ is the quantity it produces. Hence $p(\omega)y(\omega)$ is the revenue of the firm. Every producer of a particular variety ω is defined by a realization of his technology parameter $A(\omega)$. Following Hsieh and Klenow (2009), $\tau(\omega)$ is an output distortion which is modeled as a revenue tax. The simplest motivation for this revenue wedge is a model of expropriation or corruption. Suppose some firm owners are politically connected and receive implicit subsidies from the state – the output distortion of these firms will be negative $\tau(\omega) < 0$ reflecting the implicit subsidy. On the other hand, if other firms are regularly subject to expropriation due to insufficient protection of property rights, their output friction will be positive $\tau(\omega) > 0$ and act as implicit tax. Since my focus is the cross-sectional heterogeneity between firms, I ignore time-variation in both firm efficiency $A(\omega)$ and the wedge $\tau(\omega)$.[6]

The distribution of firm types $\{A(\omega), \tau(\omega)\}$ can be characterized by a measure $\mu\Big(A(\omega), \tau(\omega)\Big)$, which describes how many firms ω have productivity levels $A(\omega)$ and implicit taxes $\tau(\omega)$. With a slight abuse of notation, I adopt the convention that $\mu(\omega) \equiv \mu\Big(A(\omega), \tau(\omega)\Big)$, so that integration over $d\mu(\omega)$ should be understood as integrating over $\{A(\omega), \tau(\omega)\}$. This eases the notational burden, especially in the empirical sections where I integrate over three dimensions. Integration over the whole support of $d\mu(\omega)$

[6]The analysis of establishment dynamics since Baily, Hulten and Campbell (1992) suggests that a model with permanent level differences and temporary shocks explains the data very well. See Bartelsman and Doms (2000). Similarly Song and Wu (2011) and Midrigan and Xu (2009) find that adjustment frictions and measurement error impact primarily the time variation of plants. Neither factor can typically account for the cross-sectional heterogeneity of establishments. Following this line of research, I therefore concentrate on the cross-sectional variation in the data. Appendix 1.B provides additional empirical evidence on the viability of this approach.

then gives the number potential of firms.

The optimal price in this setup is given by

$$p(\omega) = \bar{m} \cdot \left(\frac{1}{1 - \tau(\omega)} \right) \cdot \left(\frac{w}{A(\omega)} \right) \tag{1.2}$$

where $\bar{m} = \left(\frac{\eta}{\eta - 1} \right)$ is the markup due to monopolistic competition, and $\left(\frac{1}{1-\tau(\omega)} \right)$ is the price markup due to the output distortion. Implicitly-taxed firms with positive frictions $\tau(\omega)$ will charge higher prices, and consequently produce sub-optimally low quantities. At identical prices, implicitly taxed firms would have lower marginal revenues than implicitly-subsidized firms with the same marginal cost. But since marginal costs are equal, profit maximization requires that marginal revenues should be equal as well. The optimal response of the implicitly taxed firm is therefore to cut back production to increase its marginal revenue and raise its price as is reflected in the price markup $\left(\frac{1}{1-\tau(\omega)} \right)$. A similar argument holds in reverse for firms receiving implicit subsidies. For future reference I define $\left(\frac{1}{1-\tau(\omega)} \right)$ as the measure of distortions, and will refer to this markup when talking about distortions.

Once the optimal pricing and production strategy is determined, firms can calculate their gross profits (1.1), and decide whether to continue operations or exit. If firms exit, they are assumed to get an outside value, normalized to zero. If they stay in business, they must hire an number of workers F at wage rate w. This fixed cost of operation can be thought of as administrative overhead cost. The firm exits if

$$\Pi(\omega) - w \cdot F < 0 \tag{1.3}$$

that is, if gross profits do not cover fixed cost wF.

To close the model, factor markets need to clear. One important assumption is that the rents generated by expropriation will not be "lost to the economy", but will be spent.[7] The aggregate resource constraint is therefore given by

$$PY = w \cdot L_P + \Pi + T$$

where PY is aggregate nominal expenditures, wL_p is labor income for production workers, Π is aggregate gross profits and T is net aggregate rents from micro-frictions. Profits and aggregate net rents from micro-frictions are assumed to be rebated lump-sum to households.[8] The fraction of overall spending due to these net rents can be

[7] Note that other forms of distortions – which may induce resources to disappear, or in the case of an implicit subsidy, to appear from nowhere – will typically be captured by the technology parameter $A(\omega)$.

[8] A number of plausible micro-frictions require this rebate of net rents, such as wedges related to

summarized by

$$\bar{\tau} = \frac{T}{PY} = \int_{\Pi(\omega)\geq f} \tau(\omega) \cdot \left(\frac{p(\omega)y(\omega)}{PY}\right) \cdot d\mu(\omega) \tag{1.4}$$

which is a market-share weighted average of the micro-frictions. Equation (1.4) implies that rent income is higher the higher distortions are on firms with large sales. Note that $\bar{\tau}$ can become negative if all firms are on net subsidized. In this case, the resources for these subsidies are assumed to come from lump-sum taxes on consumers.
The labor market clears

$$L = L_P + L_F$$

where L_p is aggregate labor used in production, and L_F is aggregate labor used for fixed costs of operation.

1.2.2 Equilibrium

An equilibrium is defined as combination of wages w, production labor allocation L_p and an average tax rate $\frac{1}{1-\bar{\tau}}$, such that (i) final goods firms and monopolistic competitors optimize, (ii) labor and goods markets clear, and (iii) net rents from micro-frictions are redistributed lump-sum to consumers.

> **Proposition 1:** Given that the labor force is normalized to $L = 1$, and choosing the ideal price level as numeraire $P = 1$, equilibrium output per worker is given by
>
> $$Y = \left[\int \left(A(\omega)\frac{1-\tau(\omega)}{1-\bar{\tau}}\right)^{\eta-1} \cdot d\mu(\omega)\right]^{\frac{1}{\eta-1}} = \frac{\bar{A}}{1-\bar{\tau}} = TFP_A \tag{1.5}$$
>
> where
>
> - $\bar{\tau}$ is given by (1.4)
> - $\bar{A} = \left[\int (A(\omega)[1-\tau(\omega)])^{\eta-1} \cdot d\mu(\omega)\right]^{\frac{1}{\eta-1}}$
>
> If exit is endogenous, the integrals above are conditioned on $\Pi(\omega) > wF$.

Proof: see Appendix 1.C.1
The next two sections explain the mechanisms through which micro-distortions lower aggregate productivity. In both, I start with the effect of frictions on the firm level and

information asymmetries, expropriation risk, and incomplete contracting models. Distortions that are not rebated could easily be accommodated.

the associated measurement implications. Then, I aggregate up and show the macro consequences of micro-level misallocation.

1.2.3 Allocation and aggregate TFP along the Intensive Margin

Without endogenous exit, micro-distortions appear only in differences of marginal revenue products across firms, as pointed out by Hsieh and Klenow (2009). To see this in the simple model, it is helpful to think about the labor demand of firms. The benefit to firm ω of hiring one more worker is $A(\omega)$ units of its goods. Selling these results in a revenue of $\frac{\eta-1}{\eta} \cdot p(\omega)$ per unit. The marginal revenue product is therefore $\frac{1}{\bar{m}} p(\omega) \cdot A(\omega)$. Equivalently, sales per worker adjusted for the CES markup are $\frac{p(\omega)y(\omega)}{\bar{m}L(\omega)}$. In a frictionless equilibrium these marginal revenue products should be equalized across firms; otherwise firms with higher marginal revenue products would have an incentive to hire more workers while firms with lower marginal revenue products would have an incentive to shed workers. It is straightforward to combine (1.2) with the production function to show that

$$\frac{p(\omega)y(\omega)}{L(\omega)} = \bar{m} \cdot \left(\frac{1}{1-\tau(\omega)}\right) \cdot w \tag{1.6}$$

Equation (1.6) states that differences in sales per worker across firms reflect differences in distortions, and not in firm productivity.[9] To recover distortion parameters, note that (1.6) can be solved for $\left(\frac{1}{1-\tau(\omega)}\right)$, and can be calculated from firm-level micro-data.

Once micro-level frictions are measured, I calculate the aggregate TFP implications. As in the case with distortionary taxation, two forces are at work. First, there is a price effect: higher distortions imply higher prices set by monopolistic competitors as in (1.2) so that buyers substitute out of more expensive products. Second, there is a rent (or revenue) effect: since the rents from higher taxes are not lost to the economy as a whole, consumers spend the higher tax revenue on goods. The net effect is a deadweight loss. This loss can be seen in equation (1.5), which shows how aggregate TFP is affected when individual distortions $\left(\frac{1}{1-\tau(\omega)}\right)$ deviate from their mean $\left(\frac{1}{1-\bar{\tau}}\right)$. Note that if $\tau(\omega) = \bar{\tau}$ micro-distortions drop out of equation (1.5).
Figure 2 displays the allocative consequences of removing micro-frictions for the sales of firms. The graph shows a distribution plot with firm level efficiency on the x-axis and firm level distortions on the y-axis. Firms with higher y-axis values face a higher implicit tax, while firms with higher x-axis values are more efficient. In this situation, setting all distortions to zero will let firms above the zero-distortion line win sales

[9] This relies on CES demand and monopolistic competition, since otherwise size impacts markups and therefore marginal revenue products. I consider the impact of this extension in appendix 1.D. But, given that the demand system is known it is straightforward to control for endogenous markups.

and firms below the line lose sales. This is the basic logic of calculating misallocation losses along the intensive margin. The removal of distortions will therefore lead all firms to line up along the zero distortion line, so that only efficiency is a source of heterogeneity across firms. The bottom panel of Figure 2 then sums up the number of firms for each efficiency type $A(\omega)$. Note that as the intensive-margin misallocation is removed, the efficiency distribution of firms is held fixed. From this perspective, studies that consider intensive-margin misallocation leave the efficiency distribution of firms unchanged. Introducing extensive margin misallocation effects partly endogenizes this efficiency distribution of firms.

1.2.4 Allocation and aggregate TFP with Extensive Margin

This section lays out how the presence of endogenous exit qualifies measurement arguments and aggregate TFP implications of the previous section. Again I start with measurement issues at the micro-level and then show macro implications for TFP.

Firms will exit if they do not generate sufficient gross profits to cover their fixed cost. Rewriting the zero cut-off profit condition (1.3), one obtains[10]

$$\log(A(\omega)) - \log(\bar{A}) - \left[\bar{m}\log\left(\frac{1}{1-\tau(\omega)}\right) - \frac{1}{\eta-1}\log\left(\frac{1}{1-\bar{\tau}}\right)\right] \geq \frac{1}{\eta-1}\log[(\eta-1)F] \tag{1.7}$$

Equation (1.7) shows that the zero profit condition in this simple model can be understood in terms of individual deviations from aggregate means. A firm will survive only if its efficiency is sufficiently above the aggregate mean or its distortion is sufficiently below the mean distortion. For future reference, I define components of the selection equation as follows:

$$\begin{aligned} Z(\omega) &= A(\omega)\cdot\left(\frac{1}{1-\tau(\omega)}\right)^{-\bar{m}} \\ \bar{Z}_J &\equiv \bar{m}\left(\frac{w}{P}\right)\cdot\left(\frac{\eta w F}{PY}\right)^{\frac{1}{\eta-1}} = [(\eta-1)\cdot F]^{\frac{1}{\eta-1}}\cdot\bar{A}\left(\frac{1}{1-\bar{\tau}}\right)^{\frac{1}{\eta-1}} \end{aligned} \tag{1.8}$$

with $z(\omega) = \log(Z(\omega))$. The $Z(\omega)$ variable captures the importance of idiosyncratic efficiency and firm distortions on net profits, while the $\bar{Z}_J$ variable captures the impact of both aggregate efficiency and fixed costs on survival.

To facilitate tractability in this section, I assume a specific functional form for joint distribution of firm-level efficiency and micro-frictions. In particular, I assume that

[10] See Appendix 1.C.2

$\log(A(\omega))$ and $\log\left(\frac{1}{1-\tau(\omega)}\right)$ are jointly normally distributed:[11]

$$\begin{pmatrix} \log A(\omega) \\ \log\left(\frac{1}{1-\tau(\omega)}\right) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_A \\ \mu_\tau \end{bmatrix}, \begin{bmatrix} \sigma_A^2, \sigma_{A,\tau} \\ \sigma_{A,\tau}, \sigma_\tau^2 \end{bmatrix} \right) \tag{1.9}$$

with the the correlation of efficiency and distortion given by $\rho_{A,\tau} = \frac{\sigma_{A,\tau}}{\sigma_A \sigma_\tau}$.
Note that given the distributional assumption in equation (1.9), $z(\omega) \equiv \log(A(\omega)) - \bar{m}\log\left(\frac{1}{1-\tau(\omega)}\right)$ is normally distributed with mean $\mu_A - \bar{m}\mu_\tau$ and dispersion $\sigma_z = \sqrt{\sigma_A^2 + \bar{m}^2\sigma_\tau^2 - 2\bar{m}\sigma_{A,\tau}}$. Figure 8 shows firm survival in this simple model.[12] The straight line is equation (1.7) solved for $\left(\frac{1}{1-\tau(\omega)}\right)$, and displays the exit cut-off. Only firms whose distortion is low enough or productivity high enough to be below this survival line will actually be operating. Aggregate variables such as the required fixed costs, aggregate efficiency $\bar{A}$, and aggregate rents $\bar{\tau}$ enter the picture through their impact on intercept of the survival line. For example, as aggregate efficiency $\bar{A}$ falls, the survival line shifts to the left – allowing more firms to survive for lower realizations of idiosyncratic efficiency or higher realizations of distortions.

The optimal exit decision induces sample selection. This is important for measurement, since the observable measures of marginal revenue products will only reflect firms that are not too heavily distorted to stay in business. To illustrate the sample selection issues, I exploit log-normality from equation (1.9) to derive the observable mean and dispersions of the distortion measures based marginal products. Since $z(\omega)$ is normally distributed, I rewrite equation (1.8) as

$$z(\omega) \equiv \log(A(\omega)) - \bar{m}\log\left(\frac{1}{1-\tau(\omega)}\right) = [\mu_A - \bar{m}\mu_\tau] + \sigma_z \cdot z_N(\omega)$$

where $z_N(\omega)$ is a standard normal random variable. Therefore, we can rewrite the selection equation (1.7) in terms of a standard normal random variable as

[11] For establishment-level micro-data such as the data used in this study, log-normality has been shown to approximate size-distributions very well. The often-used Pareto distribution for firm sizes is especially useful for the largest US firms, while even the US establishment-size distribution is better approximated by a log-normal – see Sutton (1997), Rossi-Hansberg and Wright (2007) and Luttmer (2010). Furthermore, a growing literature shows how power laws for the largest percentiles of firms could be generated by the fact that large firms are multi-establishment entities. Power laws of firm sizes could therefore result even as plant-level heterogeneity is characterized by log-normality. See Growiec, Pammolli, Riccaboni and Stanley (2008) and Bee, Riccaboni and Schiavo (2011).

[12] The formal setup of self-selection is well known in the labor literature on Roy models, such as Heckman and Honore (1990).

$$z_N(\omega) \geq \frac{\log(\bar{A}) - \frac{1}{\eta-1}\log\left(\frac{1}{1-\bar{\tau}}\right) + \frac{1}{\eta-1}\log\Big((\eta-1)F\Big) - (\mu_A - \bar{m}\mu_\tau)}{\sigma_z} \equiv \bar{z}_J \tag{1.10}$$

The analytical gain from this transformation is that truncated moments are especially tractable, as is well known from the applied econometrics literature on sample selection. To see the gain in tractability, let $\lambda_1(.)$ denote the Inverse Mill's Ratio and $\lambda_2(.)$ the variance of a standard normal conditional on (1.10).The observed mean and dispersion of the distortion are given by[13]

$$\begin{aligned} E\left[\log\left(\frac{1}{1-\tau(\omega)}\right)\Bigg| z_N(\omega) \geq \bar{z}_J\right] &= \mu_\tau + \left(\frac{\sigma_{A,\tau} - \bar{m}\sigma_\tau^2}{\sigma_z^2}\right) \cdot \lambda_1\Big(\bar{z}_J\Big) \\ Var\left[\log\left(\frac{1}{1-\tau(\omega)}\right)\Bigg| z_N(\omega) \geq \bar{z}_J\right] &= \sigma_\tau^2 + \Big[\lambda_2\Big(\bar{z}_J\Big) - 1\Big] \cdot \left(\frac{\sigma_{A,\tau} - \bar{m}\sigma_\tau^2}{\sigma_z}\right)^2 \end{aligned} \tag{1.11}$$

One noteworthy result is that the degree of sample selection depends on the underlying covariance of micro-distortions and firm productivity. When such productivity and distortions are uncorrelated, the sample mean of distortions will always underestimate the underlying mean distortion, while the sample mean of productivity will always overestimate the underlying mean productivity. Intuitively, both the lowest-productivity firms as well as the most highly-distorted firms tend to exit. This setup also allows us to perform simple direct comparative statics to support our intuition. For instance, it is straightforward to show that $\frac{\partial \lambda_1(\bar{z}_J)}{\partial \bar{z}_J} > 0$ so that sample selection biases get stronger as selection forces become more important.

Selection effects also have implications for the measurement of observable dispersions of micro-frictions. Consider the dispersion in equation (1.11): since $\lambda_2\Big(\bar{z}_J\Big) < 1$, in the case of uncorrelated micro-distortions and productivity, observed dispersions will underestimate underlying dispersions.[14] In general, it is misleading to generalize from the truncated observable distribution to the underlying distribution without taking selection into account.

The impact of endogenous exit on aggregate TFP and welfare can be summarized as follows:

$$Y \propto J^{\frac{1}{\eta-1}} \cdot TFP_A \tag{1.12}$$

where J is the number of firms operating, and TFP_A is as defined below in equation (1.13).

[13]The formulas for productivity mean and dispersion are analogous.

[14]These selection results continue to apply for as long as the density of the underlying type distribution exhibits log-concavity. See Heckman and Honore (1990).

There are two main differences in a model with an extensive margin when compared to a model with only an intensive-margin. As can be seen in equation (1.12), the number of firms is now endogenous and leads to variety effects. To illustrate these, I assume that the full population of draws is exogenously given by $\tilde{J}$.[15] In Figure 8, this would be the number of black and grey points together. The number of actual surviving firms is therefore:

$$J = \int_{\Pi(\omega) > wF} d\mu(\omega) = \tilde{J} \cdot Pr\Big(\Pi(\omega) > wF\Big) = \tilde{J} \cdot [1 - \Phi(\bar{z}_J)]$$

The number of operating firms is determined by the number of latent draws multiplied by the probability of survival. Whether variety effects magnify or dampen aggregate TFP losses due to distortions is not obvious *a priori* and will depend on the underlying distribution of efficiency and distortions. For example, suppose that efficiency and distortions are highly positively correlated. As the dispersion of frictions increases, mostly high-efficiency firms will be more heavily taxed. However, such firms are more likely to survive despite the distortion, so fewer firms will exit and the number of operating firms might actually increase as the dispersion of distortions increases. In this case, the gain in variety will work to to offset other misallocation losses. Contrast this with the case when the correlation of frictions and efficiency is negative – that is, primarily low-efficiency firms are more heavily taxed. These low-efficiency firms will exit very easily when the implicit tax on them is increased. Therefore, many firms will exit, leading to a large net loss of variety, which in turn reinforces misallocation losses.

Additionally, aggregate TFP is now

$$TFP_A = E\left[\left(A(\omega)\frac{1-\tau(\omega)}{1-\bar{\tau}}\right)^{\eta-1}\middle|\Omega_S\right]^{\frac{1}{\eta-1}} \tag{1.13}$$

As per equation (1.8), the set of currently active firms in the distorted equilibrium is given by

$$\Omega_S = \left\{\omega : A(\omega) \geq \left(\frac{1}{1-\tau(\omega)}\right)^{\bar{m}} \bar{Z}_J\right\} \tag{1.14}$$

The presence of micro-frictions reduces aggregate TFP due to two channels. This is illustrated in Figure 6. The first channel is the intensive-margin misallocation mechanism of section 2.3. At the same time, holding the set Ω_S constant leaves the efficiency distribution of firms unaffected, as in Figure 7.

Removing extensive margin misallocation changes the set Ω_S and reweights the efficiency distribution of firms. Consider the example in Figure 1.1, which contrasts

[15] $\tilde{J}$ is endogenized later by modeling endogenous entry.

the exit decisions in a distorted equilibrium with the exit decision in a frictionless equilibrium. The set Ω_S captures the firms southeast of the survival line – these are the survivors in the distorted equilibrium. In contrast, only firms to the right of the vertical dashed line would survive in a frictionless equilibrium. The survivors in the frictionless equilibrium can be summarized by the set

$$\Omega_S^* = \left\{\omega : A(\omega) \geq \bar{Z}_J^*\right\} \tag{1.15}$$

For the set Ω_S^*, firms are selected based on their efficiency only. The difference between the two exit rules classifies firms into four categories. First, Always Survivors will be active in both the distorted and in the frictionless equilibrium. Second, always exiters will decide to leave both the distorted and the undistorted markets. Third, Shadows would be efficient enough to survive in a frictionless equilibrium, but are pushed out by micro-frictions. Fourth, Zombies would exit in a frictionless equilibrium, but are implicitly subsidized and can therefore survive in a distorted equilibrium.
To obtain the productivity composition of the frictionless equilibrium, two things need to happen: Zombie firms need to exit, and Shadow firms need to enter. Figure 8 illustrates how this would change the efficiency distribution of firms endogenously. The efficiency distribution in the frictionless equilibrium has more weight at higher values of firm productivity as in Figure 9.

How strongly micro-distortions affect the equilibrium efficiency composition depends on the underlying joint distribution of firm distortions and efficiency. The example equilibrium in Figure 4 exhibits a low covariance of distortions and efficienc,y and a low degree of dispersion of frictions. The sets of both Zombies and Shadows are mostly drawn from the middle of the efficiency distribution. Contrast this with Figure 5, in which the dispersion of efficiency is the same but both the correlation of frictions and efficiency at the firm level as well as the dispersion of distortions is high. The striking feature is that, in this situation, Shadow firms are the highest-efficiency firms, while Zombie firms are mostly the lowest efficiency firms. The extent of extensive margin misallocation will therefore be much higher in the allocation shown in Figure 5 than in Figure 4.

The composition effect of extensive margin misallocation on efficiency has implications for the measured distribution of firm-level efficiency. Specifically in countries in which extensive margin misallocation is less pronounced, the efficiency distribution of firms should be more left skewed – that is, there should be more mass concentrated at the higher values of efficiency. Figure 10, which is from Hsieh and Klenow (2009) is suggestive in this respect; it displays estimates of the efficiency distribution of firms relative to the mean efficiency in India, China, and the US. A major caveat when interpreting this graph is that the estimates are across all 4-digit sectors. This can be important if survival cut-offs differ across these sectors. Nevertheless the skewness of the US efficiency distribution compared to India and China is compatible with the view that extensive margin misallocation is less important in the US than in India or China. This stylized fact has been emphasized repeatedly in the literature – for instance in Banerjee and Duflo (2005), who refer to the "thick left tail" of Indian firms. Figure 11 shows that the qualitative features of this shift in the efficiency distribution can be replicated by contrasting the allocation of Figure 4 with the heavily-distorted allocation of Figure 5.

1.2.5 Conceptual Decomposition of Extensive Margin Misallocation Effects

In this section, I propose a decomposition that allows me to analyze how much of the extensive margin misallocation loss is due to the retention of Zombies versus the premature exit of Shadows. The overall aggregate TFP loss from removing distortions can be decomposed as follows:

$$\left(\frac{E\left[\left(A(\omega)\frac{1-\tau(\omega)}{1-\bar{\tau}}\right)^{\eta-1}\middle|\Omega_S\right]}{E\left[A(\omega)^{\eta-1}\middle|\Omega_S^*\right]}\right)^{\frac{1}{\eta-1}} =$$

$$\text{Intensive-Margin Misallocation}\left\{\left(\frac{E\left[\left(A(\omega)\frac{1-\tau(\omega)}{1-\bar{\tau}}\right)^{\eta-1}\middle|\Omega_S\right]}{E\left[A(\omega)^{\eta-1}\middle|\Omega_S\right]}\right)^{\frac{1}{\eta-1}}\right.$$

$$\text{Extensive Margin Misallocation}\left\{\begin{array}{l}\text{Zombies}\left\{\times\left(\frac{E\left[A(\omega)^{\eta-1}\middle|\Omega_S\bigcup\Omega_S^*\right]}{E\left[A(\omega)^{\eta-1}\middle|\Omega_S^*\right]}\right)^{\frac{1}{\eta-1}}\right.\\ \text{Res. EM}\left\{\times\left(\frac{E\left[A(\omega)^{\eta-1}\middle|\Omega_S\right]}{E\left[A(\omega)^{\eta-1}\middle|\Omega_S\bigcup\Omega_S^*\right]}\right)^{\frac{1}{\eta-1}}\right.\end{array}\right. \tag{1.16}$$

The first term in (1.16) removes all wedges between the set of firms that is given in the distorted allocation, formalizing the mechanism from section 2.2 that is displayed in Figure 2. The extensive margin misallocation effect is the combination of the second and third terms. The second term compares TFP at the efficient composition of firms shown in Figure 9 to aggregate TFP when firms of the frictionless equilibrium and

Zombie firms are active as in Figure 8. This ratio answers the question of how much aggregate TFP would fall if we would force the market to keep Zombie firms active at the production levels implied by their low efficiency. This term therefore summarizes how the reallocation of resources from firms active in the frictionless equilibrium toward the Zombie firms reduces TFP. The third term captures a residual effect conditional on the existence of Zombies – it compares aggregate TFP with the composition of currently active firms as in Figure 7 to the composition of firms in Figure 8. The latter is comprised of all firms in the frictionless equilibrium plus the Zombie firms. Then, given that Zombie firms with low productivity cannot be forced to exit, how much aggregate productivity would be recovered by letting Shadow firms exit and reallocate their resources to the Always Survivors and Zombies? Note that the exit of Shadow firms does not necessarily increase aggregate productivity, as the resources of Shadow firms are not only reallocated towards Always Survivors but also to Zombies as well. Which effect dominates, and if this Residual EM-Reallocation effect is increasing or reducing aggregate TFP, is ultimately an empirical question.

1.3 Empirical Methodology

This section gives an overview of the estimation strategy to correct for sample selection in order to estimate the full underlying distribution of firm efficiency and distortions. Based on these estimates, I calculate a counter-factual frictionless equilibrium to generate the efficiency cutoff in this frictionless equilibrium.

1.3.1 Extended Model

I start by extending the model to more credibly capture the main features of the data when quantifying the effects discussed in the theory section. The main extensions are the following.

<u>Extension 1</u>: Capital enters the production function on the firm level and there will be a net wedge distorting the mix between capital and labor

<u>Extension 2</u>: The number of potential firms is endogenized with free entry.

<u>Extension 3</u>: There will be multiple sectors, with an elasticity of substitution of 1 across sectors

Each of these extensions serves a particular function when confronting the model with the data. Extension 1 introduces a capital wedge, since within 4-digit industries marginal revenue products of capital are distributed differently from marginal revenue products of labor.[16] It should be noted that the capital wedge stands in for a *net* wedge between capital and labor – a separate labor wedge could be introduced, but its effects would map into the current size and net capital wedges. Extension 2 allows me to quantify the number of overall latent productivity and distortion draws, and relates them to resources allocated to each sector.[17] This enables me to estimate overall sectoral resources because it uncovers resources spent by entrants. Furthermore, this extension allows me to take into account welfare effects from expansion of variety due to changes in entry. Extension 3 is important for two reasons. First, exit cut-offs are likely to be differ across sectors as for example different distributions of efficiency and distortions will imply different ideal price levels. Second, when calculating counterfactuals, changes in factor prices could potentially affect entry and exit due to changes in the prices of fixed costs incurred by entrants and operating firms. Modeling a full multi-sector general equilibrium allows me to endogenize factor prices in counterfactual experiments.

The extended model is as follows. Total output is given by:

$$Y = \Pi_{s=1}^{S} Y_s^{\xi_s} \quad \text{with} \quad \xi_s \in [0,1], \quad \sum_{s=1}^{S} \xi_s = 1 \tag{1.17}$$

where ξ_s are assumed to be given by the sectoral shares in value added. Monopolistic competitors ω in sector s solve the static profit maximization problem:

$$\begin{aligned} &\max_{\{K_s(\omega), L_s(\omega)\}} \Pi_s(\omega) = [1 - \tau_{Y,s}(\omega)] \cdot p_s(\omega) y_s(\omega) - w L_s(\omega) - [1 + \tau_{K,s}(\omega)] \cdot R \cdot K_s(\omega) \\ &\text{subject to: } \ y_s(\omega) = A_s(\omega) \cdot K_s(\omega)^{\alpha_s} L_s(\omega)^{1-\alpha_s} \end{aligned} \tag{1.18}$$

where R is the economy-wide rental rate of capital and the endogenous exit decision given by

$$\Pi_s(\omega) \geq B_s \cdot [(1 + \tau_{K,s}(\omega))R]^{\alpha_s} w^{1-\alpha_s} F_s \tag{1.19}$$

Note that fixed costs here are firm-specific, since they are subject to the same factor frictions as variable costs. To facilitate exposition, define: $f_s(\omega) \equiv B_s \cdot [(1 + \tau_{K,s}(\omega))R]^{\alpha_s} w^{1-\alpha_s} F_s$ with $B_s = \alpha_s^{-\alpha_s}(1-\alpha_s)^{-(1-\alpha_s)}$

[16] Alternatively, this might suggest that technological factor shares are different across firms. While I cannot control for this issue with my current data, in principle this could be resolved with corresponding establishment-level data from a wealthy country like the US. It is with this caveat in mind that I also analyze in section 5 the TFP effects of only partially removing micro-distortions, allowing for much of the cross-sectional firm heterogeneity to reflect other factors than distortions.

[17] In other words this is endogenizing the number $\tilde{J}$ from section 2.5.

Finally the endogenous entry decision in steady state is given by:

$$\begin{aligned} V_{e,s} &= \sum_{t=0}^{\infty} \delta_s^t \cdot Pr\Big(\Pi_s(\omega) \geq f_s(\omega)\Big) \cdot E\left[\Pi_s(\omega) - f_s(\omega) \Big| \Pi_s(\omega) \geq f_s(\omega)\right] \\ &\geq B_s \cdot R^{\alpha_s} w^{1-\alpha_s} F_e \end{aligned} \tag{1.20}$$

where δ_s is an exogenous death shock for the firm, mostly capturing the age effects of firm exit behavior. The specification of the fixed cost of entry follows the way these are specified in earlier studies, such as Fattal Jaef (2011) or Hsieh and Klenow (2011).

1.3.2 Calibrated and Observed Data

This section describes additional assumptions that facilitate the empirical analysis. First, I calibrate a number of parameters of the model in accordance with evidence from other studies. Second, I describe the data requirements and measurement assumptions underlying the estimation.

The primary calibrated variables are the elasticity of substitution η and factor prices. I follow Hsieh and Klenow (2009) ans set $\eta = 3$ – in the middle range of existing estimates from trade data (Broda and Weinstein (2006)). Again, following Hsieh and Klenow (2009) I set the rental rate to 10%. The economy- wide wage is directly picked from evidence from the World Bank labor-market database, which is based on national establishment level surveys that are adjusted to enhance cross-country comparability.[18] Factor shares in the Cobb-Douglas production function of monopolistic competitors are set to the corresponding factor shares of US 4-digit sectors from Bartelsman, Becker and Gray (2000).[19] For sectors that do not have any correspondence with a 4-digit sector in the US, I assume that the capital share is $\alpha_s = 1/3$. Results seem to be unaffected by either calibrating this capital share or omitting these sectors altogether.

Mapping the available establishment-level micro-data to the quantities in the model is done according to the following principles. First, since my theory concentrates on output and capital wedges, I map the revenue measure $p_s(\omega)y_s(\omega)$ to value added.[20] As Jones (2011b) points out, this basically ignores the effects of possible distortions

[18] The mismeasurement of these aggregate prices will influence only the estimated means, and therefore plays virtually no role in the counterfactual exercises later, in which only the dispersions or correlations of distortions are changed.

[19] It is well-known that labor shares from these data underestimate actual compensations by excluding fringe benefits such as social security contributions. I therefore impute these following Hsieh and Klenow (2009) by inflating the reported wage-bill by a constant factor. This ensures that the aggregate capital share implied by the sectoral data for the US is consistent with the estimates from NIPA data.

[20] Note that this assumes the existence of a value-added aggregator that is Cobb-Douglas. See Basu and Fernald (2000) and Diewert (1978).

Parameter	Value	Explanation
η	3	Elasticity of Substitution between varieties Hsieh and Klenow (2009)
R	1.1	Aggregate rental rate Hsieh and Klenow (2009)
w	Country data	Aggregate wage rate World Bank Labormarket database
α_s	US data	Sectoral capital share Bartelsman et al. 2000
δ_s	from Turnover	Probability of death shock (see section 3.4)

Table 1.1: Calibrated Values

across firms to the use of intermediate goods. I do this mostly for reasons of comparability with Hsieh and Klenow (2009) and related studies that focus on intensive-margin misallocation. The method of this chapter could easily be extended to accommodate more sources of heterogeneity. Second, the observed data on factor demand are typically annual values that do not differentiate between fixed and variable cost factors. I therefore be maintain the assumption that measured labor is the sum of fixed-cost and variable-cost workers. More formally,

$$
\begin{aligned}
L_{M,s}(\omega) &= L_{P,s}(\omega) + L_{F,s}(\omega) \\
K_{M,s}(\omega) &= K_{P,s}(\omega) + K_{F,s}(\omega)
\end{aligned}
$$

i.e. measured labor for plant ω in sector s denoted by $L_{M,s}(\omega)$ is the sum of labor for production purposes $L_{P,s}(\omega)$ plus labor demand to cover fixed overhead costs $L_{F,s}(\omega)$. The same measurement assumption is applied to capital, which is measured as the reported book value of the establishment capital stock.

All variables calculated on the establishment-level are time averages for my seven year panel. This reflects the low persistence of temporary shocks I document in Appendix 1.B and will therefore capture the permanent plant level distortions and efficiency differences my theory focuses on. Song and Wu (2011) advocate this “Generalized Marginal Revenue Product”-Approach to focus on the cross-sectional dimension of establishment level frictions.

Model Quantity	Corresponding Data
$p_s(\omega)y_s(\omega)$	Establishment value added
$K_{M,s}(\omega)$	Book value of capital
$L_{M,s}(\omega)$	Establishment wage bill deflated by wage
$Pr\big(\Pi_s(\omega) > f_s(\omega)\big)$	Probability of survival of entrants

Table 1.2: Model variables and corresponding data

The last measurement assumption concerns the probability of survival of entrants. This probability is observed directly in the data, as I can measure how man establishments with a specific birth year in the survey are remain later.

1.3.3 Fixed-Cost Estimator

In this section, I derive from the model an estimator for fixed costs. To show why this might be necessary, I concentrate on the scale wedge $\frac{1}{1-\tau_Y(\omega)}$. The Melitz-type extensive margin model with fixed costs of operation would imply that as long as a firm produces, it must hire $L_P(\omega)$ workers for production and F workers for the overhead fixed costs. It follows that the measured amount of labor for a firm would be $L_M(\omega)$. It is straightforward to show how measurement of labor impacts distortion measurement. Measured distortions are given, similar to (1.6), by

$$\frac{p(\omega)y(\omega)}{L_M(\omega)} = \bar{m}\frac{w}{1-\tau(\omega)}\left(\frac{L_P(\omega)}{L_P(\omega)+F}\right)$$

If F is the same across firms, then this changes the correlation structure of micro-frictions and productivity. Note that as firms become larger, $L_P(\omega)$ is larger and hence the term $\left(\frac{L_P(\omega)}{L_P(\omega)+F}\right)$ will be close to 1. However, for very small firms this term will be less than 1, as $L_P(\omega)$ is small. As a consequence, larger firms will look more distorted than small firms, even if no such correlation is present. A simple approach for dealing with this issue in the current context is to utilize the optimal exit decision of firms to estimate the fixed cost of operation.

Proposition 2: Fixed-Cost Estimate

Let F_s be the amount of the Cobb-Douglas composite $K_M(\omega)^{\alpha_s}L_M(\omega)^{1-\alpha_s}$ needed to pay for fixed costs of production. Fixed costs of operation in the monopolistic competition model, with productivity heterogeneity and micro-distortions under endogenous entry and exit, are given by

$$F_s = \frac{1}{\eta} \min_{\omega} \left\{ K_{M,s}(\omega)^{\alpha_s} \cdot L_{M,s}(\omega)^{1-\alpha_s} \right\} \tag{1.21}$$

Proof: Appendix 1.C.3

The intuition behind this estimation approach is that the econometrician is focusing on the smallest firms in an industry. For these firms, production is only a very small fraction of factor demand, while fixed costs of operation are the bulk of factor demand. Hence, the factor demand of the smallest firms within the same narrow industry should be a viable proxy for the typical fixed cost for firms in the industry.[21]

In contrast to other attempts to deal with this issue in the literature, I do not need to assume a particular functional form for the productivity distribution or about the distribution of distortions to estimate fixed costs. These distributional assumptions are necessary only for dealing with selection effects. The cost of this is that the only heterogeneity in fixed costs of operation for which I allow, are the same wedges that distort production. However, under endogenous exit, even random variation in fixed costs across firms has a very limited impact on distortion measurement, as I show in Appendix 1.E.[22]

1.3.4 Maximum-Likelihood Estimation with Equilibrium Constraints

I now turn to the issue of sample selection. In this section, I derive a likelihood function based on the assumption of multivariate log-normality for efficiency and distortions. A novel feature compared to well-known selection estimators from the labor literature – such as Heckman and Honore (1990) – is the use of equilibrium constraints. The primary issue is that I do not directly observe the welfare-based CES price level. But, note that, as in (1.7) and (1.8), this ideal price level also influences the exit decision of firms. This can be an issue when estimating the underlying distributions of distortions and efficiency from sample-selected observed distributions. For example, as the underlying dispersion of distortions increases, there are more heavily distorted plants and more firms might exit. As a result, the ideal price index could rise as fewer varieties are offered, but with a higher price level more firms now might be able to

[21] I also experimented with using the first and the second percentile of factor demands, instead of the minimum. The estimation for distortions and efficiency seem not significantly affected by this variation, because I follow Hsieh and Klenow (2009) and drop the 1% outliers of the data.

[22] Note also that the bias in TFPR from overhead costs will be less important if fixed costs of production systematically increase with firm size.

survive. This connection between the underlying distributions and equilibrium sample selection needs to be taken into account. Moreover, under the current distributional assumptions, this price level is a non-linear fixed point that cannot be explicitly solved for the underlying parameter vector. Formally, the problem I face is similar to constrained MLE estimation of dynamic discrete choice problems, such as Rust (1987). Instead of using popular Nested Fixed Point algorithms to address this issue, I follow Su and Judd (2010) in formulating the estimation problem as a "Mathematical Program with Equilibrium Constraints" (MPEC).[23]

As discussed in section 3.2, the main required data for this estimator are value added, factor inputs, and the conditional probability of exit for entrants. To get an intuition for identification, remember that firm-heterogeneity in my model is driven by three sources: net output-wedges $\frac{1}{1-\tau_{Y,s}(\omega)}$, net capital wedges $\big(1+\tau_{K,s}(\omega)\big)$ and efficiency $A_s(\omega)$. In measuring these sources of heterogeneity I basically follow Hsieh and Klenow (2009). To identify the output wedge, recall that $L_{P,s}(\omega)$ is the labor input after the measured labor input $L_{M,s}(\omega)$ is corrected for fixed costs, as in the last section. Then, heterogeneity in output per worker on the firm level should reflect primarily the output wedge:

$$d_{1,s}(\omega) \equiv \frac{1-\alpha_s}{\bar{m}} \frac{p_s(\omega) y_s(\omega)}{w L_{P,s}(\omega)} = \frac{1}{1-\tau_{Y,s}(\omega)}$$

The identification of the net capital wedge relies on information on establishment-level labor intensity. After the calibration of sectoral capital shares α_s from US data and factor prices w and R, labor intensity should reflect net capital wedges:

$$d_{2,s}(\omega) \equiv \frac{\alpha_s}{1-\alpha_s} \frac{L_{M,s}(\omega)}{K_{M,s}(\omega)} \frac{w}{R} = (1+\tau_{K,s}(\omega))$$

In the discussions of estimates, I will sometimes follow Hsieh and Klenow (2009) and summarize both frictions into one joint distortion:[24]

$$\texttt{log TFPR}_s(\omega) = \frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)} \tag{1.22}$$

Note that the estimation procedure identifies both distortions separately. Later, the

[23]Like Su and Judd (2010), and Dube, Fox and Su (2009) I find in Monte Carlo test-runs that the use of MPEC methods facilitates numerical stability and reliability of estimates.

[24]Note that equation (1.22) applies in a strict sense only if factor shares are the same across firms within the same industry and Cobb-Douglas is a reasonable approximation to plant-level production functions. However, as Dong (2011) shows, generalizing the production function to CES and allowing for measurement error has quantitatively small implications for the measures of establishment-level frictions.

counterfactual experiments will analyze aggregate TFP changes as one friction is removed without changing the other.

Efficiency differences $A_s(\omega)$ are identified by using the CES demand system,[25]

$$d_{3,s}(\omega) \equiv \texttt{log TFPQ}_s(\omega) \equiv A_s(\omega) = \frac{y_s(\omega)}{K_{P,s}(\omega)^{\alpha_s} L_{P,s}(\omega)^{1-\alpha_s}}$$
$$\text{with} \quad y_s(\omega) = \Big[p_s(\omega) y_s(\omega)\Big]^{\frac{\eta}{\eta-1}} \frac{(P_s Y_s)^{-\frac{1}{\eta-1}}}{P_s}$$

so that efficiency can be recovered from inverting the size of the firm in terms of value added. The crucial point here is that the main data necessary are nominal value added $p_s(\omega)y_s(\omega)$, and factor inputs $K_{s,M}(\omega)$ and $L_{s,M}(\omega)$. The one parameter that cannot be readily identified is the value of the ideal price index P_s. Studies focusing on only intensive-margin misallocation typically normalize this value to one. Since the underlying distortions impact the sectoral ideal price level, which in turn affects selection, I do not follow this strategy here. Instead, I include the ideal price level as an additional estimation parameter but require that the price level satisfies a fixed point that describes industry equilibrium.[26]

This model also offers a natural method to utilize plant exit and survival data to get information about truncation. The probability of firms that survive given by

$$Pr\Big(z_N(\omega) \geq \bar{z}_{J,s}\Big) = 1 - \Phi\Big(\bar{z}_{J,s}\Big)$$

To get the necessary information, I note that, as in Melitz (2003), steady state turnover is given by

$$J_{e,s} \cdot Pr\Big(z_N(\omega) \geq \bar{z}_{J,s}\Big) = \delta_s \cdot J_s \tag{1.23}$$

where $J_{e,s}$ is the number of entering firms in industry s, δ_s is an exogenously-given exit probability, and J_s is the number of firms surviving in the industry. This steady-state turnover condition helps to calibrate two moments. First, $Pr\Big(z_N(\omega) \geq \bar{z}_{J,s}\Big)$ is given by the mass of firms surviving, conditional on entry. Second, given that we can observe the number of entering firms $J_{e,s}$ as well as the number of operating firms J_s, and have an estimate of $Pr\Big(z_N(\omega) \geq \bar{z}_{J,s}\Big)$, (1.23) indicates the value of δ_s that is compatible

[25]This strategy basically follows Klette and Grilliches (1996) and Hsieh and Klenow (2009) to infer efficiency from the contrast between firm size and factor demand.

[26]Note that estimating sequentially by sector does not mean that the estimation is "Partial Equilibrium only." I do not need to require anything about multi-sector GE in the estimation setup, as I can observe or calibrate aggregate factor prices w and R. I make sure later that these factor prices describe a general equilibrium in my model, when the factor demands in entry cost inputs are recovered from GE conditions.

with steady-state turnover.

Finally, I have to take account of the possibility that measurement error might lead to the observation that some plants are above the survival line. According to the steady-state model here, this would actually be impossible to observe. The simplest approach to deal with this issue is the addition of a zero-mean measurement error into the selection equation. The estimation section discusses how important this measurement error is in describing the observed data.

Proposition 3: MLE Estimation with Equilibrium Constraint

Let the parameter vector for each sector s be given by

$\theta_s = \left[\mu_{A,s}; \mu_{\tau_Y,s}; \mu_{\tau_K,s}; \sigma_A; \sigma_{\tau_Y,s}; \sigma_{\tau_K,s}; \rho_{A\tau_Y,s}; \rho_{A\tau_K,s}; \rho_{\tau_{YK},s}; \sigma_e\right]$

The Maximum-Likelihood Estimator for the model with productivity heterogeneity and micro-distortions under endogenous entry and exit, can be written as

$$\max_{\theta_s, \ln(P_s)} \sum_{\omega} \ln \left\{ \frac{\phi\left(d_{1,s}(\omega), d_{2,s}(\omega), d_{3,s}(\omega) \middle| \theta_s, \ln(P_s)\right)}{1 - \Phi\Big(\bar{z}_J\big(\theta_s, \ln(P_s)\big)\Big)} \right\}$$

subject to:

$$\begin{aligned} \ln(P_s) = \bar{p}_{IM}(\theta_s) + \frac{1}{\eta - 1} \ln\left(\frac{J_{e,s}}{\delta_s}\right) \\ + \frac{1}{\eta - 1} \ln\left[1 - \Phi\left(\bar{z}_J\big(\theta_s, \ln(P_s)\big) + \frac{\sigma_{xz}(\theta_s) - (\eta - 1)\sigma_z(\theta_s)^2}{\sigma_z(\theta_s)}\right)\right] \end{aligned}$$

where $\phi(.|\theta_s, \ln(P_s))$ is a normal density with parameter vector $\{\theta_s, \ln(P_s)\}$, and $\Phi(.)$ is the cdf of a Standard-Normal. Other implicitly-defined variables in this estimation exercise include the following:

$$u_s = \left(\frac{R}{\alpha_s}\right)^{\alpha_s} \left(\frac{w}{1 - \alpha_s}\right)^{1-\alpha_s}$$

$$\bar{Z}_J = \ln\left[\bar{m}\left(\frac{u_s}{P_s}\right)\left(\frac{P_s Y_s}{\eta \cdot u_s \cdot F_s}\right)^{-\frac{1}{\eta-1}}\right]$$

$$\begin{aligned} \sigma_{xz}(\theta_s) = \bar{m}\alpha_s(\alpha_s \sigma^2_{\tau_K,s} + \rho_{\tau_{YK},s}\sigma_{\tau_K,s}\sigma_{\tau_Y,s}) + \bar{m}(\alpha_s \rho_{\tau_{YK},s}\sigma_{\tau_K,s}\sigma_{\tau_Y,s} + \sigma^2_{\tau_Y,s}) \\ - (\alpha_s \rho_{A\tau_K,s}\sigma_{\tau_K,s}\sigma_{A,s} + \rho_{A\tau_Y,s}\sigma_{A,s}\sigma_{\tau_Y,s}) \end{aligned}$$

$$
\begin{aligned}
\sigma_z^2(\theta_s) = \big(\bar{m}\alpha_s\sigma_{\tau_K,s}\big)^2 + \big(\bar{m}\sigma_{\tau_Y,s}\big)^2 + \sigma_{A,s}^2 \\
+ 2\bar{m}\Big(\rho_{\tau_{YK},s}\sigma_{\tau_K,s}\sigma_{\tau_Y,s}) - \sigma_{A,s}(\rho_{A\tau_Y,s}\sigma_{\tau_Y,s} - \rho_{A\tau_K,s}\sigma_{\tau_K,s})\Big) + \sigma_e^2
\end{aligned}
$$

$$
\bar{z}_J\big(\theta_s, \ln(P_s)\big) = \frac{\bar{Z}_J + \bar{m}(\alpha_s\mu_{\tau_K,s} + \mu_{\tau_Y,s}) - \mu_{A,s}}{\sigma_z}
$$

$$
\begin{aligned}
\bar{p}_{IM}(\theta_s) = \mu_{\tau_Y,s} + \alpha_s\mu_{\tau_K,s} - \mu_{A,s} \\
- \frac{1}{2}(\eta - 1)\Big(\sigma_{A,s}^2 + \sigma_{\tau_Y,s}^2 + \sigma_{\tau_K,s}^2 + 2\rho_{\tau_{YK},s}\sigma_{\tau_K,s}\sigma_{\tau_Y,s}\Big) \\
+ (\eta - 1)\sigma_{A,s}\Big(\rho_{A\tau_Y,s}\sigma_{\tau_Y,s} - \rho_{A\tau_K,s}\sigma_{\tau_K,s}\Big)
\end{aligned}
$$

Proof: see Appendix 2.C.4

Note that the constraint in Proposition 3 is a nonlinear fixed point in the ideal price level P_s, which cannot be explicitly solved for the entering parameter vector θ_s.[27] Statistical inference is discussed in the next section.

To gain additional intuition behind the estimation procedure, consider Figure 12. The dark stars will typically be data for firms that stay in business. The ellipses constitute contour plots of the joint density of firm efficiency and distortions. Changes in the parameter vector θ_s adjust the shape of the underlying distribution. These changes in the underlying distribution also lead to equilibrium responses of the ideal price level $P_s(\theta_s)$ as shown in equation (1.8) and therefore shift the survival line. Hence, the MLE estimator will jointly adjust the shape of the distribution and the position of the survival line to optimally fit the data points of firms in operation.

1.3.5 Statistical Inference of MLE with Equilibrium Constraints

As discussed in Su and Judd (2010) the use of equilibrium constraints in estimation introduces additional considerations when computing standard errors. The equilibrium constraint is a restriction on the parameter space and therefore typically can increase the efficiency of estimates.

Proposition 4 (Su and Judd (2010)): MLE Inference with Constraints

Let $\psi = [\theta, \log(P)]$. Suppose $\theta_0 \in \Theta$ is the vector of true parameter values and denote the likelihood at the true values by $L(\psi_0)$ and the constraint

[27] Despite being highly non-linear, this MLE estimator is continuously differentiable in its ten parameters. I therefore proceed to use derivative-based quasi-Newton solvers linked to KNITRO to solve the corresponding constrained non-linear optimization problem. These solvers are especially well-equipped to deal with medium-scale non-linear optimization subject to non-linear constraints, as I encounter in this problem.

set by $h(\psi_0) = 0$. Given matrices $H_0 = \frac{\partial h(\psi_0)}{\partial \psi}$ and $B_0 = E\left[\frac{\partial L(\psi_0)}{\partial \psi}\frac{\partial L(\psi_0)}{\partial \psi'}\right]$. Assume the following conditions hold:

(i) Θ is compact

(ii) θ_0 is in the interior of Θ.

(iii) Observations are iid draws from the assumed distribution

(iv) The likelihood function is C^3 and the constraint is C^2

(v) The matrix B_0 is positive definite

Then,

$$\sqrt{J}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, S_0)$$

with $S_0 = B_0^{-1}[I - H_0(H_0'B_0^{-1}H_0)^{-1}H_0'B_0^{-1}]$

Proof: see Su and Judd (2010) and Silvey (1975)

1.3.6 Identification of Zombies and Shadows

To identify Zombies and Shadow firms exit rules, for the counter-factual frictionless equilibrium are required. After obtaining the estimation vector θ_s for each sector, I use these results to calculate aggregate sectoral factor demand L_s and K_s. Remember that the data report the annual factor demands of operating firms, but not the resources for the entry fixed costs prior to opening the business. To obtain aggregate factor demand, I need to infer resources spent by new entrants on fixed costs of entry. Combining the information on the latent mass of firms with the distribution of heterogeneity allows me to recover these parameters from aggregate resource constraints. Appendix 1.C.5 shows how this is done.

After inferring the values of L_s and K_s from the data, I define the relevant economy-wide factor endowments as $L = \sum_{s=1}^{S} L_s$ and $K = \sum_{s=1}^{S} K_s$. The counterfactual frictionless equilibrium is computed by taking the aggregate factor endowments for the whole economy as given. This is standard practice in studies that analyze intensive-margin misallocation only. Note that the welfare effects will be larger if capital is allowed to accumulate in response to removing frictions.

1.3.7 Aggregate Total Factor Productivity

For the purpose of evaluating the welfare gains of removing distortions, I summarize a key result in proposition 5 below.

Proposition 5 summarizes the aggregate sector-level TFP consequences of the extended setup and makes comparison with the simplified model in Section 2 possible.

The proposition displays TFP in real consumption units: sectoral TFP deflated by the ideal CES price index. As in section 2, the first term summarizes net variety effects, while the second term captures aggregate TFP. There are two main differences to the simple model in Section 2. First, the overall firm level distortion is a geometric composite of output and capital wedge. This is summarized in (i) of Proposition 5. Second, since the capital wedge distorts the overhead fixed costs the same way it distorts production costs, the capital wedge induces a reallocation of fixed cost resources across firms. I subsume this allocational effect under the aggregate TFP effects, as it is usually quantitatively small.

Note that varieties are normally not measured in industry-level price deflators, so that I will refer to sectoral TFP without any variety effects as TFP and to the TFP measure including these variety effects as "Overall Real TFP".

Proposition 5: Sectoral TFP

In equilibrium, sectoral aggregate real output per inputs is given by

$$\frac{Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}} =$$

$$\text{Variety}\left\{ J_s^{\frac{1}{\eta-1}} \right.$$

$$\text{TFP}\left\{ \begin{array}{l} \left(\int_{\pi_s(\omega) \geq f_s(\omega)} \left[\frac{1-\tau_s(\omega)}{1-\bar{\tau}_s} \right]^{\eta-1} A_s(\omega)^{\eta-1} \cdot \frac{d\mu(\omega)}{\mu(\pi_s(\omega) \geq f_s(\omega))} \right)^{\frac{1}{\eta-1}} \\ \times \left(1 - \frac{\bar{T}_{K\alpha,s}\left(\frac{1}{\eta} - l_{e,s}\right) + l_{e,s}}{\frac{1}{\bar{m}}\left(\frac{1}{1+\bar{\tau}_{K,s}}\right) + \bar{T}_{K\alpha,s}\left(\frac{1}{\eta} - l_{e,s}\right) + l_{e,s}} \right)^{\alpha_s} \left(1 - \frac{1}{\eta}\right)^{1-\alpha_s} \end{array} \right.$$

with

(i) Overall distortion: $\left(\frac{1}{1-\tau_s(\omega)}\right) = \frac{\left(1+\tau_{K,s}(\omega)\right)^{\alpha_s}}{1-\tau_{Y,s}(\omega)}$ and $\left(\frac{1}{1-\bar{\tau}_s}\right) = \frac{\left(1+\bar{\tau}_{K,s}\right)^{\alpha_s}}{1-\bar{\tau}_{Y,s}}$

(ii) Fixed cost allocation $\bar{T}_{K\alpha,s} = \left(\frac{\int_{\Pi_s(\omega) > f_s(\omega)} [1+\tau_{K,s}(\omega)]^{\alpha_s - 1} d\mu_s(\omega)}{\int_{\Pi_s(\omega) > f_s(\omega)} [1+\tau_{K,s}(\omega)]^{\alpha_s} d\mu_s(\omega)} \right)$

(iii) $l_{e,s}$ denoting the fraction of the sectoral labor force that are entry-cost workers

Proof: see Appendix 2.C.6

1.4 Conclusion

This chapter developed an estimation methodology to estimate extensive margin misallocation effects on aggregate TFP. The data requirements of this methodology are modest. At a minimum it requires one cross section of data on value added and factor inputs and industry-level estimates on entry and exit. In the next chapter I apply this methodology to establishment level microdata from a major developing economy to evaluate the quantitative importance of this extensive margin misallocation channel.

Chapter 2

Extensive Margin vs. Intensive Margin Misallocation in Indonesia

2.1 Introduction

This chapter contributes to the existing literature by showing that extensive margin misallocation can be a quantitiatively important determinant of aggregate TFP. I illustrate this point with data from Indonesia in the 1990s, where it has been argued that political connections and ethnic diversity strongly distorted business decisions. (Fisman (2001)). My empirical results demonstrate that removing micro-distortions would increase macroeconomic TFP in Indonesia by 67%. As such, by improving the allocation of resources along the intensive margin, aggregate TFP can rise by a factor of 1.7. This is in line with previous work by Hsieh and Klenow (2009) who show that the removal of intensive-margin misallocation could increase aggregate TFP by 80-100% for China and India. Additionally removing extensive margin misallocation could increase aggregate TFP by another 44%. This means that overall aggregate TFP gains from such adjustments are closer to a factor of 2.45, or 145%. The extensive margin misallocation channel suggests that misallocation losses are over 50% larger than implied by considering only intensive-margin misallocation. Furthermore, I document that the feature of the data that drives these results is the extent of covariance of efficiency and wedges at the firm level. Seen through the lens of my model, a higher covariance means that more high-efficiency firms face a high implicit tax, and therefore exit. At the same time, more low efficiency firms benefit from a high implicit subsidy and are artificially kept alive. Hence, the efficiency of Shadows will be particularly high and the efficiency of Zombies will particularly low so that replacing Zombies with Shadows would boost aggregate productivity more.

2.2 Data and Empirical Results

2.2.1 Overview of Data

For Indonesia, I use the *Statistik Industri*, an annual panel data set of medium and large plants.[1] The data set collects information for all Indonesian establishments with more than 20 employees. Among the surveyed variables are the wage-bill and number of employees, capital stocks at book values, birth year of the plant and value added. The sample contains approximately 20,000 plants each year. I use sample years 1990-1996, as these years have consistent industry classifications before the Asian Crisis. Estimation is done at the 4-digit industry level, which, after some data cleaning, leaves me with 40 sectors capturing approximately 80% of manufacturing activity.[2]

2.2.2 Reduced Form Evidence on External Validity

Before turning to the actual estimation, I verify the external validity of the constructed measures of distortions and efficiency. The key mechanism of this chapter implies that distortions and efficiency should have predictive power for exit decisions of plants. Plants with higher values of TFPR should be more likely to exit, while establishments with higher TFPQ should be less likely to exit. Table 2.1 documents this pattern in simple discrete choice regressions. A similar argument would suggest that participation in exporting is driven by firm-level distortions and efficiency. However, in contrast to the exit regressions, the signs on both efficiency and distortions should flip. As only the most efficient and the least distorted firms should decide to export, the prediction is that high-TFPR (low-TFPQ) plants should be less (more) likely to export. The final columns of Table 2.1 show how these patterns are observable in the data.

In the remainder of this chapter, I ignore international trade. Appendix 2.A documents simulation evidence from a multi-country trade model regarding why this is a reasonable approach in the context of measuring micro-level misallocation.[3]

[1] As the data set has been intensively analyzed before by Amiti and Konings (2007) and Peters (2011), I refer the interested reader to the data discussions in these papers.

[2] I restrict estimation to sectors that have more than 100 establishments and include other sectors with more than 70 plants only if their share in aggregate value added is at least 1%. Furthermore, I follow Hsieh and Klenow (2009) and remove the 1% tails of the data to minimize the impact of outliers.

[3] The topic of complementarities between domestic micro-distortions and frictions in international trade is research in progress.

Table 2.1: Exit and Trade Selection

	Depedent Variable			
	$\texttt{Exit}_{t+1}(\omega)$	$\texttt{Exit}_{t+1}(\omega)$	$\texttt{Export}_{t+1}(\omega)$	$\texttt{Export}_{t+1}(\omega)$
Independent Variable	Probit	Logit	Probit	Logit
$\texttt{log TFPR}_t(\omega)$	0.223	0.401	-0.992	-1.794
	[0.0146]	[0.0258]	[0.0134]	[0.0245]
$\texttt{log TFPQ}_t(\omega)$	-0.175	-0.314	0.766	1.382
	[0.00826]	[0.0147]	[0.00697]	[0.0130]
Industry FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: Pooled regressions of next period exit or export on current efficiency and micro-distortion.

2.2.3 Evaluation of Model Fit

Before turning to a discussion of estimation results, I review a number of ways to check whether my estimates are reasonable, concentrating on three key aspects of my model. First, I evaluate how well the estimated model captures patterns of distortions, plant efficiency, and survival. Second, I test the functional form assumption on the distribution of establishment-level frictions and efficiency. Third, I offer some checks on the estimates of fixed costs across sectors.

The core of the model relates distortions and efficiency to survival. The econometric specification allows a mean-zero error term to explain deviations from the selection mechanism of the model. If the variance of these estimated errors in the selection equation is large, then measured distortions and efficiency would be a poor explanation of observed survival. But, as can be seen from Table 2.3, the estimated error variance is usually negligible. An alternative way to analyze the role of firm distortions and efficiency is to graphically compare estimated survival lines with the data. Figure 23 illustrates the survival pattern of plants for the six largest sectors by value added. It shows plots of the bivariate data distribution of firm efficiency and micro-distortions in blue stars against predictions from estimates in red circles.[4] Both of these overlap fairly closely for most of the body of the distribution. Note especially the estimated black sloped survival line from the model. An important concern about the estimation procedure used here is that it assumes the measured wedges are work like taxes to firms, instead of markups. The latter would in fact predict that higher-distortion firms are more profitable, and hence more likely to survive.[5] In this case one would expect high-TFPR firms to be more likely to survive despite low TFPQ. The estimated survival line is reassuring in this respect – the basic modeling of frictions as taxes seems to be compatible with the data.

The next key ingredient under scrutiny is the assumed log-normality of the joint distribution of distortions and efficiency. To evaluate the viability of this approach, I use a two-sample Kolmogorov-Smirnov test on firm-size distributions in the data versus firm sizes generated by my estimates. The null hypothesis is that both samples are from the same continuous distribution. The second column Table 2.3 lists the asymptotic p-values from this test. The null hypothesis can be rejected for 10 out of 40 sectors. This implies that for 3/4 of sectors I cannot reject the hypothesis that the firm-size distributions are drawn from a truncated log-normal as given by my model. Figure 24 illustrates this result graphically by comparing kernel density plots of firm-

[4]The predictions are generated based on MLE estimates and using QMC draws from Niederreiter sequences.

[5]Another concern is the omission of continuation values that capture dynamic considerations in response to temporary shocks or option-value considerations when risk is time-varying. A fully-dynamic version of the current model specification the corresponding dynamic estimation method is work in progress.

size distributions from the data to the estimated firm-size distributions.

Finally, the estimates of common fixed costs can be cross-checked with other sectoral information. One would expect capital-intensive industries to have on average higher fixed cost of operation. Similarly, sectors that report a higher fraction of labor compensation towards "non production workers" might be expected to have higher fixed costs.[6] Both of these predictions hold in the data, as displayed in Figures 25 and 26. In general the estimated fixed costs are not large. The final column of Table 2.3 shows the value of fixed costs evaluated at unit costs $uc_s = \left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}$ relative to the median value added of establishments in this industry. Fixed costs are typically is between 2% and 3 % of the median annual value added.

2.2.4 Estimation Results from the Full Model

The baseline estimation results on the importance of selection are illustrated in Figures 27 to 29. These figures display estimated distributional parameters on the x-axis, while displaying on the y-axis the same parameter measures as in the data. Each dot is a parameter value for a particular sector. For example, in Figure 27 each dot gives a value of the dispersion of TFPR σ_τ for each sector. If a parameter is unaffected by selection, the data dots should line up along the 45-degree line. In Figures 27 and 28, most dots are significantly below the 45-degree line, implying that selection reduces the observable dispersions of both frictions and efficiency. Furthermore, the selection gets more severe as the underlying dispersions of frictions or efficiency increase. These estimation results foreshadow that extensive margin misallocation is important in determining aggregate TFP losses.

The correlation between firm productivity and distortions is another important dimension affected by selection. The parameter of interest is again the *underlying* correlation of firm efficiency and frictions, since this correlation determines whether high-efficiency firms are systematically selected out. As shown by Fattal Jaef (2011) and Hsieh and Klenow (2011), a strongly positive correlation *all other things equal*, might lead variety effects to offset misallocation effects from micro-distortions. In this case the aggregate welfare effects from micro-frictions might be surprisingly low. On the other hand, the *measured* correlation of firm productivity and frictions could reflect only the fact that highly-distorted and low-productivity firms exit. In other words, firms with high implicit taxes might survive only if they are also highly efficient.

[6]The data on "non production workers" are not directly used in my estimation, since such data are often considered to capture skilled workers. However, the skilled work force might expand with the scale of operation similar to production workers, so it is questionable whether this variable captures only fixed-cost workers.

Table 2.2: Validity checks and Fixed Costs by 4 digit sector, Indonesia

	Estimates					
ISIC Code, Name	σ_e	p_{KS}	F_s	α_s	γ_s	$\frac{uc_s \cdot F_s}{[p_s y_s]_{med}}$
3114, Canning/pres. fish	$0.06 \cdot 10^{-4}$	0.09	4.61	0.36	0.23	0.21%
3115, Vegetable/animal oils	$0.06 \cdot 10^{-4}$	0.08	165.05	0.63	0.36	1.43%
3116, Grain mill products	$0.10 \cdot 10^{-4}$	0.04	268.50	0.71	0.30	4.07%
3117, Bakery products	$0.12 \cdot 10^{-4}$	0.00	17.66	0.43	0.17	3.54%
3118, Sugar factories	$0.08 \cdot 10^{-4}$	0.00	38.17	0.46	0.34	0.21%
3119, Cocoa, chocolate	$0.06 \cdot 10^{-4}$	0.46	91.26	0.61	0.34	3.61%
3121, Food products n.e.c.	$0.24 \cdot 10^{-4}$	0.19	232.37	0.69	0.27	10.67%
3122, Processed tea/coffee	$0.06 \cdot 10^{-4}$	0.36	328.28	0.64	0.38	3.76%
3123, Ice manufacturing	$0.83 \cdot 10^{-4}$	0.04	41.81	0.33	0.38	8.03%
3124, Soy products	$321.08 \cdot 10^{-4}$	0.07	25.02	0.33	0.22	6.30%
3125, Food chips, animal	$0.12 \cdot 10^{-4}$	0.01	12.55	0.33	0.17	5.57%
3127, Pastry/cake/food	$0.25 \cdot 10^{-4}$	0.31	23.66	0.33	0.27	7.73%
3134, Soft drinks/carb.	$0.11 \cdot 10^{-4}$	0.32	116.51	0.53	0.41	6.94%
3141, Dried/proc. tobacco	$0.08 \cdot 10^{-4}$	0.00	2.07	0.33	0.18	2.83%
3142, Clove cigarettes	$0.22 \cdot 10^{-4}$	0.74	5.73	0.33	0.30	0.13%
3211, Spinning, weaving T.	$0.07 \cdot 10^{-4}$	0.46	3.64	0.19	0.22	0.64%
3212, Made-up textiles	$0.06 \cdot 10^{-4}$	0.91	7.63	0.20	0.21	2.24%
3213, Knitting mills	$0.08 \cdot 10^{-4}$	1.00	8.15	0.16	0.17	1.68%
3221, Wearing textile garm.	$0.06 \cdot 10^{-4}$	0.00	7.14	0.33	0.16	1.02%
3241, Footwear and shoes	$0.68 \cdot 10^{-4}$	0.96	29.88	0.33	0.13	1.77%

Notes: Columns display: (1) standard deviation of error term in selection equation, (2) asymptotic p-value of two sample Kolmogorov-Test of data vs. simulated data from the model, (3) estimated fixed cost of operation, (4) capital share, (5) non-production worker share, (6) fixed costs relative to median value added.

Table 2.3: Validity checks and Fixed Costs by 4 digit sector, Indonesia

ISIC Code, Name	Estimates					
	σ_e	p_{KS}	F_s	α_s	γ_s	$\frac{uc_s \cdot F_s}{[p_s y_s]_{med}}$
3311, Sawmills, planing	$0.76 \cdot 10^{-4}$	0.93	2.58	0.16	0.20	0.39%
3319, Wood/cork n.e.c.	$0.06 \cdot 10^{-4}$	0.77	8.88	0.19	0.23	2.40%
3321, Wood furniture/fixt.	$0.07 \cdot 10^{-4}$	0.86	19.01	0.33	0.17	1.96%
3411, Pulp, paper, paperb.	$0.43 \cdot 10^{-4}$	0.14	19.43	0.41	0.27	0.59%
3420, Printing/Publishing	$0.13 \cdot 10^{-4}$	0.16	29.65	0.32	0.25	3.86%
3511, Basic ind. chemicals	$0.08 \cdot 10^{-4}$	0.59	50.10	0.59	0.46	0.43%
3521, Paints/varnishes	$0.11 \cdot 10^{-4}$	1.00	324.81	0.56	0.40	5.26%
3522, Drugs/Medicines	$0.06 \cdot 10^{-4}$	0.26	232.77	0.66	0.54	0.59%
3523, Soap/cleaning prep.	$0.10 \cdot 10^{-4}$	0.64	314.15	0.73	0.37	2.29%
3529, Chem. prod n.e.c.	$0.06 \cdot 10^{-4}$	0.10	74.27	0.52	0.44	0.38%
3552, Smoked rubber	$0.06 \cdot 10^{-4}$	0.01	44.53	0.33	0.32	0.78%
3559, Rubber prod. n.e.c.	$0.07 \cdot 10^{-4}$	0.46	15.73	0.18	0.27	2.81%
3560, Plastic prod. n.e.c.	$0.09 \cdot 10^{-4}$	0.40	12.34	0.28	0.22	1.88%
3632, Cement prod.	$0.09 \cdot 10^{-4}$	0.01	11.37	0.33	0.22	4.38%
3642, Structural clay	$0.13 \cdot 10^{-4}$	0.00	7.95	0.33	0.12	4.48%
3691, Stone products	$0.07 \cdot 10^{-4}$	0.59	7.78	0.24	0.27	2.23%
3811, Cutlery/hardware	$0.88 \cdot 10^{-4}$	0.79	11.35	0.24	0.21	2.75%
3813, Structural metal	$0.51 \cdot 10^{-4}$	0.25	15.10	0.12	0.26	1.64%
3819, Fabricated metal	$0.07 \cdot 10^{-4}$	0.31	17.48	0.20	0.29	3.11%
3839, Electrical apparatus	$0.07 \cdot 10^{-4}$	0.42	65.70	0.32	0.34	1.40%
3841, Ship building/repair	$0.08 \cdot 10^{-4}$	0.06	18.41	0.40	0.33	0.42%
3844, Motorcycles/bicycles	$0.10 \cdot 10^{-4}$	0.84	31.00	0.26	0.25	3.28%

Notes: Columns display: (1) standard deviation of error term in selection equation, (2) asymptotic p-value of two sample Kolmogorov-Test of data vs. simulated data from the model, (3) estimated fixed cost of operation, (4) capital share, (5) non-production worker share, (6) fixed costs relative to median value added.

Figure 29 provides evidence that this selection bias is quantitatively important for the observed correlation of productivity and distortions. The sectoral dots line up mostly above the 45-degree line, implying that the observed correlation is higher than the underlying correlation. On average, 33% of the observed correlation between frictions and productivity can be explained by this selection bias.

To better understand what drives these selection effects, it is instructive to differentiate between output and capital wedges. Contrast Figure 31, which displays the estimated versus observed dispersions of output wedges, with Figure 32 for capital wedges. The output-wedge dispersion measures display significant sample selection, but selection is substantially more important for the capital-wedge. The difference in observed versus estimated standard deviations of the log output-wedges are typically not larger than 0.2, while these differences can easily be 0.5 for capital-wedge dispersions. A natural question is, which feature of the data is responsible for this contrast? The first crucial difference is that observed output-wedge dispersions are typically lower than dispersions of capital-wedges. Since selection effects tend to be more severe the larger the dispersion of the already-selected sample, this might explain part of the contrast between capital and output-wedge dispersions. A second answer can be found in Table 2.5: the correlation estimates of TFPQ and the output wedge are typically positive, while the correlations between TFPQ and the capital wedge are negative. The data show that low TFPQ plants are typically more labor intensive within the same narrow 4 digit industry. Seen though the lens of the model, less-efficient plants face higher net capital wedges. This correlation in turn is interpreted by the model as indicating that many small firms are highly distorted in their factor-mix choice, and so are forced to exit. This is also reflected in the comparison of observed and estimated correlations of firm efficiency and capital wedges, as in Figure 34. Again, the correlations of observable TFPQ and capital wedges are mostly positive, while estimates of the underlying correlations are mostly negative.

Finally, Figure 30 shows that capital and output wedges are mostly negatively correlated with each other, and not significantly affected by selection. Such a negative correlation means that within a sector firms with high output wedges tend to have low capital wedges and vice versa. This suggests that, for many sectors, the two wedges tend to offset one another. This observation plays a role when considering jointly removing wedges vis-a-vis focusing on removing either the output- or the capital-wedge margin only.

Table 2.4: Estimates by 4 digit sector, Indonesia

ISIC Code, Name	Estimates					
	$\sigma_{A,s}$	$\sigma_{\tau_Y,s}$	$\sigma_{\tau_K,s}$	$\rho_{A\tau_Y,s}$	$\rho_{A\tau_K,s}$	$\rho_{\tau_YK,s}$
3114, Canning/pres. fish	1.56	0.65	1.36	0.45	-0.47	-0.28
3115, Vegetable/animal oils	1.60	0.82	1.57	0.69	-0.44	-0.46
3116, Grain mill products	1.34	0.71	1.21	0.40	0.05	0.01
3117, Bakery products	1.24	0.49	1.51	0.63	-0.36	-0.37
3118, Sugar factories	2.43	0.61	0.84	0.77	0.19	-0.11
3119, Cocoa, chocolate	1.77	0.57	1.44	0.62	-0.32	-0.53
3121, Food products n.e.c.	1.28	0.72	1.32	0.80	-0.42	-0.73
3122, Processed tea/coffee	1.40	0.64	1.08	0.35	-0.00	-0.30
3123, Ice manufacturing	0.94	0.51	0.92	0.30	-0.15	-0.06
3124, Soy products	1.24	0.54	0.91	0.58	-0.12	-0.36
3125, Food chips, animal	0.82	0.72	1.25	0.17	-0.33	0.42
3127, Pastry/cake/food	0.80	0.45	1.04	0.47	-0.15	-0.17
3134, Soft drinks/carb.	1.79	0.63	1.03	0.80	-0.45	-0.64
3141, Dried/proc. tobacco	1.57	0.66	1.05	-0.21	0.63	-0.56
3142, Clove cigarettes	2.76	0.91	1.63	0.95	-0.37	-0.42
3211, Spinning, weaving T.	2.02	0.57	1.31	0.64	-0.55	-0.51
3212, Made-up textiles	1.40	0.56	1.08	0.33	-0.32	-0.28
3213, Knitting mills	1.49	0.57	1.21	0.39	-0.19	-0.42
3221, Wearing textile garm.	1.62	0.51	1.07	0.17	0.03	-0.24
3241, Footwear and shoes	1.46	0.64	1.15	0.19	0.08	-0.06

Notes: Columns display: (1) standard deviation of plant efficiency, (2) standard deviation of output distortions, (3) standard deviation of capital wedges, (4) correlation of plant efficiency and output wedges, (5) correlation of plant efficiency and capital friction, (6) correlation of output and capital distortions

Table 2.5: Estimates by 4 digit sector, Indonesia

ISIC Code, Name	Estimates					
	$\sigma_{A,s}$	$\sigma_{\tau_Y,s}$	$\sigma_{\tau_K,s}$	$\rho_{A\tau_Y,s}$	$\rho_{A\tau_K,s}$	$\rho_{\tau_{YK},s}$
3311, Sawmills, planing	1.91	0.60	0.96	0.40	-0.12	-0.33
3319, Wood/cork n.e.c.	1.25	0.84	1.02	0.00	0.11	-0.37
3321, Wood furniture/fixt.	1.39	0.48	1.15	0.34	-0.18	-0.15
3411, Pulp, paper, paperb.	1.93	0.68	2.08	0.77	-0.60	-0.45
3420, Printing/Publishing	1.54	0.55	0.93	0.55	-0.08	-0.34
3511, Basic ind. chemicals	1.45	0.82	2.11	0.73	-0.32	-0.43
3521, Paints/varnishes	1.70	0.76	1.22	0.67	0.05	-0.18
3522, Drugs/Medicines	1.34	0.81	1.95	0.32	0.13	0.44
3523, Soap/cleaning prep.	1.86	0.69	1.20	0.70	-0.17	-0.50
3529, Chem. prod n.e.c.	1.46	1.10	1.71	0.50	0.00	0.11
3552, Smoked rubber	1.38	0.84	0.98	0.27	0.14	-0.26
3559, Rubber prod. n.e.c.	1.68	0.47	0.97	0.32	0.09	-0.23
3560, Plastic prod. n.e.c.	1.77	0.53	1.07	0.61	-0.18	-0.29
3632, Cement prod.	1.61	0.54	0.96	0.59	-0.47	-0.43
3642, Structural clay	1.16	0.48	1.21	0.69	-0.71	-0.65
3691, Stone products	1.13	0.83	0.99	0.34	-0.03	-0.25
3811, Cutlery/hardware	1.68	0.66	1.01	0.06	0.11	-0.40
3813, Structural metal	1.76	0.73	0.99	0.67	-0.10	-0.15
3819, Fabricated metal	1.93	0.57	1.00	0.52	0.11	-0.33
3839, Electrical apparatus	1.82	0.65	0.96	0.57	-0.10	-0.31
3841, Ship building/repair	1.58	0.88	2.36	0.67	-0.13	0.25
3844, Motorcycles/bicycles	2.25	0.77	0.96	0.83	-0.41	-0.43

Notes: Columns display: (1) standard deviation of plant efficiency, (2) standard deviation of output distortions, (3) standard deviation of capital wedges, (4) correlation of plant efficiency and output wedges, (5) correlation of plant efficiency and capital friction, (6) correlation of output and capital distortions

2.3 Estimates of Aggregate TFP and Welfare Effects

The empirical estimates in the previous section suggest that extensive margin misallocation effects do matter. This section now quantifies the aggregate TFP and welfare effects, and puts them into the perspective of previous estimates and the data.

For the evaluation of aggregate TFP effects I proceed as follows. I remove the micro-distortions and calculate aggregate real TFP in the frictionless equilibrium for each sector individually. To be as close as possible to the exercises in the previous literature, I make two choices. First, a removal of distortions is defined as setting the dispersions of micro-frictions to zero but leaving the mean parameters μ_τ at their current levels. I do this to prevent confounding welfare improvements through better cross-sectoral allocation as mean distortions are equalized with gains from within-sectors reallocation. Because of this choice my estimates of the welfare effects will be on the conservative side. Second, I consider the welfare gains from equalizing marginal products in one sector at a time while leaving distortions in all other sectors in place. I calculate the welfare gains from removing distortions by sector, and then sum up these weighted by the sectoral value-added shares, to arrive at the manufacturing-wide welfare gains. I will then contrast the results of this benchmark with the full removal of all distortions across all sectors.

The benchmark results presented below also feature two specific choices that increase the robustness of results. First, I concentrate on sectors for which I cannot reject the hypothesis that firm size distributions are generated by a truncated log-normal. However, TFP and welfare results are very similar whether the other sectors are included or not. Second, I exclude sectors for which TFP gains exceed two log points. This will significantly reduce the estimates of extensive margin aggregate TFP effects for the economy as a whole. On the other hand, this also makes extensive margin TFP effects more representative of typical sectoral effects.[7]

2.3.1 Aggregate Productivity Effects of Removing Distortions

Table 2.6 summarizes the key welfare results. The first column displays welfare gains from equalizing marginal revenue products across the currently-existing set of firms, as done also by Hsieh and Klenow (2009). The gains from reallocating sales across existing firms can raise aggregate TFP by close to 70%. This is in line with

[7]There are three sectors that have TFP losses exceeding two log points. The largest is Clove Cigarettes which makes up 15% of Indonesian manufacturing value added. Including this sector alone with two log points would boost extensive margin TFP losses by 50%.

Table 2.6: Aggregate Gains from Removing Distortions

	Aggregate Effects			
Reform Experiment	IM	Variety	EM/Selection	Overall Real TFP
(1) Benchmark Reform (Remove τ_Y, $\tau_K$dispersion : $\sigma_{\tau_Y} = \sigma_{\tau_K} = 0$)	67.6%	6.08%	44.34%	156.64%
(2) Neutralize output wedges (Keep τ_K & $\sigma_{\tau_Y} = 0$)	20.90%	−10.06%	27.53%	38.66%
(3) Neutralize capital wedges (Keep τ_Y & $\sigma_{\tau_K} = 0$)	3.72%	23.09%	0.01%	27.69%
(4) Remove $\tau_Y - A$ corr. ($\rho_{A,\tau_Y} = 0$)	1.33%	20.06%	13.74%	38.41%
(5) Remove $\tau_K - A$ corr. ($\rho_{A,\tau_K} = 0$)	−2.72%	12.24%	−6.21%	2.39%
(6) Remove $\tau_Y - \tau_K$ corr. ($\rho_{\tau_Y,\tau_K} = 0$)	−16.83%	3.98%	−14.11%	−25.73%
(7) Partial Reform I (Reduce $\sigma_{\tau_Y}, \sigma_{\tau_K}, \rho_{A,\tau_Y}, \rho_{A,\tau_K}, \rho_{\tau_Y,\tau_K}$ by 33%)	25.05%	6.64%	28.89%	71.89%
(8) Partial Reform II (Reduce $\sigma_{\tau_Y}, \sigma_{\tau_K}$ by 33%)	31.24%	−0.19%	21.47%	59.10%
(9) Complete Reform ($\tau_Y = \tau_K = 0$,all sectors)	67.6%	116%	48%	433%

Notes: Percentage changes are expressed with current real TFP as base. Let $\texttt{TFP}_0$ be TFP before removal of distortions and $\texttt{TFP}_1$ TFP after the removal of distortions. The colums display then $100 \cdot \left(\frac{TFP_1}{TFP_0} - 1\right)$. As a consequence, the overall welfare effect is not the sum of the individual welfare gains but the product. For instance in row #1, the last column should be understood as being the result of: $1.67 \cdot 1.06 \cdot 1.44 = 2.56$

quantitative findings by Hsieh and Klenow (2009) for China and India, where such gains are around 80-100%.[8] Columns 2 and 3 display the new extensive margin misallocation estimates. Focusing first on the variety effects in column 2, my estimates suggest that micro-frictions lead to a net destruction of varieties that increases welfare losses. Removing these micro-frictions increases variety and implies a welfare gain of 6% relative to the current base. This estimate stands in contrast to previous calculations by Fattal Jaef (2011) that are primarily based on the correlation between firm efficiency and frictions in the data. As mentioned in Section 4.3, the selection correction implies that underlying correlations between micro-frictions and firm efficiency are much lower than in the data. This estimation result drives the difference between my results here and previous studies. The third column displays the welfare gains from removing extensive margin misallocation: eliminate Zombies and replace them with Shadows. The aggregate impact is sizable – reallocation along the extensive margin can raise aggregate TFP by 44%. These estimates suggest that the overall real TFP gains are huge, as documented by the last column.

The sectoral distribution of TFP losses in the benchmark case is shown in Figure 35. The y-axis displays the log TFP gains from removing micro-distortions for the set of industries employed in the aggregate calculations. There is a fair amount of heterogeneity of sectoral TFP gains, reflecting different distributions of TFPR and TFPQ across sectors.

Results from simultaneous and complete removal of all distortions in all sectors are displayed in the last row of Table 2.6. They suggest that due to the unit elasticity of substitution across sectors, the effect of joint reform mostly shows up in larger variety effects. Note however, that this joint reform does not significantly impact the size of intensive and extensive margin misallocation losses.[9]

2.3.2 Features of the Data that Drive Extensive Margin Misallocation Effects

An important question is, what feature of the data drives these large extensive margin misallocation effects? As discussed in section 2.4, one key determinant could be the covariance of micro-distortions and efficiency. Since the dispersion of micro-frictions itself has an impact on misallocation losses, even if frictions and efficiency are

[8]Note that Hsieh and Klenow also analyze the gains from moving to the TFPR distribution of the US, where they find gains of around 50%.

[9]The reason variety effects are so large in the case of joint reform is due to the fact that removing especially the mean distortions affects factor prices significantly. Hence, sunk entry costs and fixed costs of operations fall, letting more firms enter and making survival easier. Without the Cobb-Douglas assumption across sectors, one would expect that results from simultaneous reform across sectors are different from reforming one sector at a time, depending on the elasticity of substitution. See Jones (2011a)

uncorrelated I proceed as follows. I regress the model-implied extensive margin misallocation losses by sector on the observed within-sector dispersions of micro-frictions and efficiency calculated from the data. Figure 35 then plots the residual of this regression against the covariance between micro-distortions and firm efficiency from the data. There is a strong positive relationship, documenting that extensive margin misallocation losses are higher for sectors with a higher observed covariance of TFPR and TFPQ in the data. This relation should be contrasted with Figure 35, which shows that the same relation does not hold for model-implied intensive-margin misallocation losses.

2.3.3 Reform Complementarities

Here I compare the welfare effects that result from completely removing all frictions versus the effects from completely neutralizing one friction at a time. This analysis is in the spirit of identifying the "most important bottlenecks", as in Hausman, Rodrik and Velasco (2005). Additionally, my analysis identifies important complementarities between different types of reform.

Rows 2 and 3 of Table 2.6 illustrate the removal of frictions in isolation. First, note that variety and selection effects are different when removing output frictions as opposed to capital frictions. Removing only output frictions leads to a net loss of varieties because output distortions are typically positively correlated with efficiency while capital frictions are negatively correlated. This means that implicit net capital taxes are stronger for low-efficiency firms, and as such result in more firms exiting. The reverse is true when only capital frictions are shut down: as output frictions are typically positively correlated with efficiency, more firms will survive. As a result the second column Table 2.6 shows that net variety gains are large if only capital frictions are removed. The differences in correlation patterns of capital and output frictions with efficiency also explain the different responses of the extensive margin misallocation effect in the third column. Output frictions are positively correlated with efficiency – so the Zombie firms that stay alive are of particularily low-efficiency while the Shadow firms that exit are of particularily high-efficiency. Replacing Zombies with Shadows therefore increases allocational TFP gains substantially. For capital frictions, the picture is the reverse – as these frictions are negatively correlated with efficiency, the firms that are forced out by these frictions are not that efficient and the extensive margin gains from removing capital wedges are small.[10]

Second, there are important complementarities between the two types of reform.

[10] A possible explanation for the different correlations of output and capital wedges with efficiency might be that they these wedges reflect distortions from different markets. For instance, capital wedges might reflect financial frictions that primarily distort small firms. In contrast, output wedges might result from large and efficient firms facing extortionary demands by local public officials.

The sum of percentage welfare gains from removing only output frictions and from removing only capital frictions is significantly lower then the gains from removing both frictions together. This is true even for intensive-margin misallocation gains. This should be seen in the context of the aggregate TFP formula given in Proposition 5. If output wedges, capital wedges, and firm efficiency are mutually independent, then it follows that

$$
\begin{aligned}
&E\left[\left(\frac{1+\tau_{K,s}(\omega)}{1+\bar{\tau}_{K,s}}\right)^{\alpha_s(\eta-1)}\left(\frac{1-\tau_{Y,s}(\omega)}{1-\bar{\tau}_{Y,s}}\right)^{\eta-1}A_s(\omega)^{\eta-1}\middle|\Pi_s(\omega)\geq f_s(\omega)\right]^{\frac{1}{\eta-1}}\\
&=E\left[\left(\frac{1+\tau_{K,s}(\omega)}{1+\bar{\tau}_{K,s}}\right)^{\alpha_s(\eta-1)}\middle|\Pi_s(\omega)\geq f_s(\omega)\right]^{\frac{1}{\eta-1}}\\
&\times E\left[\left(\frac{1-\tau_{Y,s}(\omega)}{1-\bar{\tau}_{Y,s}}\right)^{\eta-1}\middle|\Pi_s(\omega)\geq f_s(\omega)\right]^{\frac{1}{\eta-1}}\\
&\times E\left[A_s(\omega)^{\eta-1}\middle|\Pi_s(\omega)\geq f_s(\omega)\right]^{\frac{1}{\eta-1}}
\end{aligned}
$$

If the set of firms is fixed, then under independence intensive-margin misallocation gains from complete removal of both wedges together are the same as the sum of removing one friction at a time. This is not the case here, as can be seen in row 1 of Table 2.6. The reason is that frictions are mutually correlated and also pairwise correlated with efficiency.

To analyze the nature of this complementarily, the following three rows of Table 2.6 show aggregate responses as different correlations are removed. Looking first at removing the correlation between wedges and plant efficiency, in both cases the intensive-margin effects are negligible, in contrast to sizable effects along the extensive margin. Both variety and extensive margin reallocation effects are typically an order of magnitude larger then intensive-margin effects. Still, in both cases removing the correlation between wedges and plant efficiency is typically welfare-enhancing.
However, this is not true of removing the correlation between wedges, displayed in row 6 of Table 2.6. Welfare falls as the correlation of capital and output wedges is removed. To understand this, remember that this correlation is negative both in the data as well as in the selection-corrected estimates. This means that firms with higher net capital frictions typically have lower output frictions. If this correlation is removed, firms with a high output friction will on net be more distorted. The table shows that this worsening misallocation shows up along both intensive- and extensive margins. Therefore, row 6 underlines the importance of taking into account the correlation among wedges as frictions are removed. This also explains why removing only output or

capital frictions yields considerably smaller gains than reducing both frictions together. Removing, for instance, the output distortion but leaving the capital wedge in place not only gets rid of this output friction, but also removes the offsetting effect og the output friction on the capital wedge. Without this offset, misallocation will be made *worse*. These types of surprising effects are well known in the theoretical literature on Second Best paths of reform, since Lipsey and Lancaster (1957).

2.3.4 Welfare Effects from Partial Reform

The last two columns of Table 2.6 address the issue of partial reform more directly. In these cases I reduce the dispersion of both wedges by one-third instead of fully removing all frictions. The reduction in the dispersion of micro-frictions is chosen conservatively, in order to leave a lot of room for the possibility that much of the cross sectional dispersion of marginal revenue products reflects unknown heterogeneity unrelated to policy distortions. Despite this, the welfare effects of partial reform are still sizable, and driven mostly by improved allocational efficiency along both the intensive and the extensive margin. Note that compared to TFP gains from the intensive-margin alone, extensive margin gains add another 60% to 100%. Returning to the topic of complementarity of reform, it is instructive to compare TFP gains from the partial reform experiments to reform cases in row 2 and 3. Partial reform that addresses both frictions at the same time still dominates fully removing only one friction at a time.

2.3.5 Decomposing Selection-Effects: Zombies and Residual Extensive Margin Reallocation

As outlined in section 2.5, it is instructive to further decompose the selection effect estimated in the previous section into effects of the retention of Zombies versus the premature exit of Shadows. Table 2.7 shows that the TFP loss due to Zombie retention is many times larger than the extensive margin misallocation effect. The latter effect is smaller due to the residual extensive margin reallocation effect: as Shadows exit their resources are reallocated to Zombies but also to Always Survivors. In the estimated model, gains from reallocating production to these Always Survivors outweights the loss from more production at Zombie firms. Therefore the residual extensive margin reallocation effect actually increases aggregate TFP, although by less than the Zombie effect.

Figures 39 to 42 provide some insight into the relation between identified Zombies in the data and estimates of the sectoral TFP effects. The first three graphs show the firm share of Zombies on the x-axis. This is defined as the log of the fraction of all establishments in the sector, identified as Zombie plants. As Figure 39 shows, this measure is strongly positively correlated with the log market share of all Zombie firms

Table 2.7: Decomposition of Extensive Margin TFP Effects

	Aggregate Effects		
Reform Experiment	EM/Selection	Zombie Effects	Res. EM-Reallocation
Complete Reform	44.34%	134.34%	−38.40%

Notes: Percentage changes are expressed with current real TFP as base. Let $\texttt{TFP}_0$ be TFP before removal of distortions and $\texttt{TFP}_1$ TFP after the removal of distortions. The colums display then $100 \cdot \left(\frac{TFP_1}{TFP_0} - 1\right)$.

in the sector. Figures 40 and 41 display the relation of estimated TFP losses along specific margins versus the firm share of Zombies. Not surprisingly, the Zombie effect increases in the firm share of Zombies. As Zombie firms steal more business from non-Zombies, their drag on sectoral productivity becomes stronger. However, in contrast to this straightforward implication of the model, figure 41 shows that intensive-margin misallocation losses are typically lower for sectors with many Zombie firms. The reason for this seems to be related to Figure 42. This graph plots the firm share of Zombies against the dispersion of micro-frictions within a sector, and demonstrates that these typically are negatively related. In short, the dispersion of micro-frictions as a measure of misallocation is more useful for capturing intensive-margin than extensive margin misallocation.

2.4 Conclusion

This chapter has shown how quantitatively important extensive margin misallocation is for aggregate TFP. I applied the methodology developed in chapter 1 to the case of Indonesian manufacturing. The results suggest that the bottomline TFP losses are around 50% larger than was estimated before. If micro-distortions differ substantially across countries and are an important determinant of aggregate TFP, we should expect them to also play an important role for international trade. The standard specification of the Melitz model used in international trade precludes such an analysis, as firm efficiency dispersions are assumed to be identical across countries. The next chapter lines out how the results of a simplified version of the model developed in the last chapters can be used to analyze international trade.

Chapter 3

From Competitive Advantage of Firms to Comparative Advantage of Nations

3.1 Introduction

It has long been an article of faith that microeconomic factors of firm competitiveness and institutional structures are at the heart of national productivity and wealth. But with the expansion of data sources linking various dimensions of firm behavior to performance across countries, researchers have also empirically documented systematic relations between micro-patterns and macro productivity across countries. Bloom and Van Reenen (2007) document systematic differences in management practices that are correlated with firm-size distributions. Hsieh and Klenow (2009) show how measured firm-level frictions differ across China, India and the US. Kumar, Rajan and Zingales (1999) analyze how institutional factors shape firm-size distributions even within a set of rich European countries. If differences in firm-efficiency dispersions or micro-distortions impact national TFP, then we should expect them to impact international specialization and trade. The crucial link in this analysis is the interaction between firm-level heterogeneity and international trade. Building on insights of Melitz (2003), most of the focus in the trade literature was devoted to the impact of trade opening on firm size via reallocation effects. The novel contribution of this study is to analyze how differences in underlying efficiency dispersion drive both firm-size dispersions and trade patterns.

The starting point of my analysis is therefore the standard specification of the Melitz model. This specification implicitly assumes that the dispersion of firm-efficiency is identical across countries. Given empirical estimates of trade costs and aggregate productivity across countries, one can generate implied firm-size dispersions from this

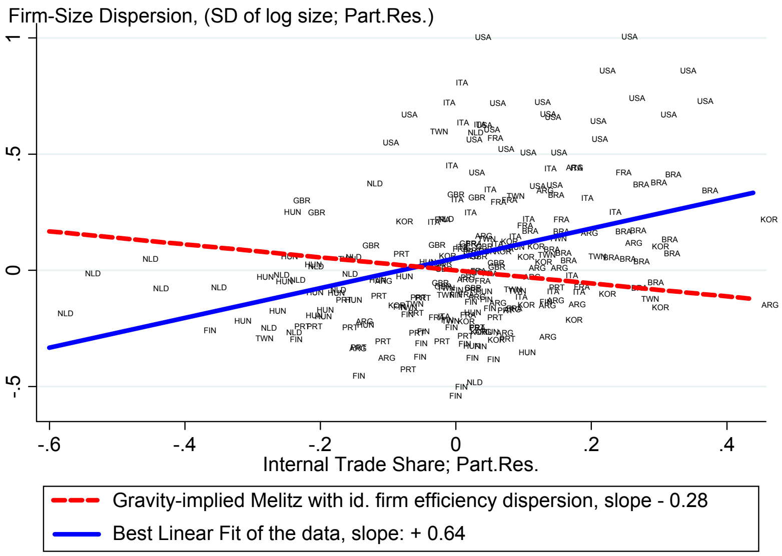

Figure 3.1: Basic Trade-Selection Puzzle. Figure displays cross-country firm-size dispersion within 2-digit sectors on the y-axis. The x-axis shows 2-digit internal trade shares, defined as value of shipments of the sector to itself, relative to all international shipments to this sector. Both variables are controlled for sector fixed effects to center graph at zero.

standard specification. The result is the dashed red line in figure 3.1: within-industry firm size dispersions should be narrower if the country is trading less. The mechanism behind this clear negative relation is an exporter-selection and reallocation effect. Countries whose firms are less productive on average will have only have a few very productive firms. These few productive firms will be larger due to export participation. On the other hand, there will be a lot of non-exporting firms, whose market shares are reduced due to exporters from other countries. Therefore, the lower the average productivity of a country, the larger the size difference between its few very productive firms and the many unproductive firms should be. The red line shows that this intra-industry reallocation effect which was qualitatively analyzed in Melitz (2003), can also be understood as a quantitative statement when combined with a standard Gravity equation.

In contrast, the solid blue line displays the relationship between trade openess and firm size dispersions in the data. To concentrate on intra-industry allocations, the data are dispersions of total firm size within 2-digit sectors for a dozend countries. In the data, country-sector pairs with a high internal trade share are datapoints with high firm size dispersions.

To rationalize this data, I extend the basic Melitz model to allow for international differences in efficiency dispersions. Similar to the data I analyze and in the spirit of the Melitz model, the theory is really a formalization of producer or establishment heterogeneity.[1] For establishments it is well known that log-normal distributions are an excellent alternative to the Pareto-distributions widely used in the trade literature. An innovation of this study will be to exploit the tractability of log-normal distributions to model differences in firm-efficiency dispersions across countries. Along the way I will develop a framework that can easily encompass multiple sources of heterogeneity and will be able to trace their impact on patterns of international specialization and trade. A country with a large underlying efficiency dispersion will exhibit large firm-size dispersion differences even taking selection of the least productive firms into account. What is more, a high efficiency dispersion will mean that there are a lot of very productive firms active. Combined with the fact that products are substitutable, this will imply that home consumers will prefer the product of these very efficient home producers to more expensive foreign firms. Hence economies with higher efficiency dispersion will trade more internally and look more closed to trade.

An important implication of the new model is that as one allows differences in efficiency dispersion of producers with a country, this impacts how strongly trade barriers influence exporter selection and the productivity distribution of firms conditional on survival. In other words, the underlying micro distributions of firm efficiency matter for macro outcomes such as aggregate productivity and trade. As an example of this connection I discuss asymmetries in international trade flows. In a standard constant-elasticity Gravity model, if country A has a large market share in country B, this implies that country B will have a low market share in country A. Deviations from this reciprocal relationship are known as "asymmetries". Recently it has been argued that asymmetries in trade flows may indicate asymmetric trade barriers. If this is correct eliminating these asymmetric trade barriers might be important for improving the international allocation of ressources. In the new model, asymmetries in trade flows are the natural result of differences in firm-level productivity distributions across countries. All other things equal, country A might be exporting systematically more to country B, because country A has more very efficient producers. On the flipside, country B might have nearly no efficient producers and will therefore systematically export less to country A. The model will formalize this intuition and show how asymmetries in measured trade frictions can be seen as consequence of productivity differences. In this sense the model highlights a potentially important factor related to the nature of measured "trade barriers".

[1] I follow the literature and will use the terms firm and establishment or plant interchangibly, but will really mean establishment.

3.2 Openess and Firm Sizes: Empirical Evidence

This section establishes the key stylized facts to be addressed by the theory. These facts center around trade patterns on the one hand and firm-size distributions on the other hand, where firm size is defined to include total sales of domestic firms to all markets including international sales. The natural starting point of an analysis of trade patterns and firm size distributions are the recently developed models of firm heterogeneity and trade, following Melitz (2003), Chaney (2008) and more recently Eaton, Kortum and Kramarz (2011). These have become the standard models to analyze firm heterogeneity and trade, as they are compatible with micro-level facts about export selection into trade and give rise to Gravity equations that are successful in matching aggregate trade flows empirically.

In order to organize the discussion of these facts, I start out with the standard specification of the the Melitz model under Pareto distributions of productivity draws. Implicit in this standard specification of the Melitz model is the assumption that the dispersion of firm level productivities is constant and identical across countries.[2] I will estimate the parameters of this Melitz-Pareto model with international trade data and will contrast its implications about firm sizes with the data.

The economic environment comprises CES preferences for the representative consumer of country i with real output defined by

$$Y_i = \left(\sum_{j=1}^{N} \int y_{ij}(A)^{\frac{\eta-1}{\eta}} \cdot d\mu_j(A) \right)^{\frac{\eta}{\eta-1}}$$

where P_i will denote the corresponding CES price index and P_iY_i overall nominal spending. Firms in these economies are comprised of monopolistic competitors who use a composite input good E_i to produce a differentiated output good:

$$\max_{\{E_{ij}(A)\}} \Pi_{ij}(A) = p_{ij}(A)y_{ij}(A) - c_jE_{ij}(A) - f_{ij}$$

$$\text{subject to: } y_{ij}(A) = \frac{1}{d_{ij}}AE_{ij}(A)$$

Monopolistic competitors differ by productivity draw A, which in this section is assumed to be generated by a Pareto-distribution:

$$\mu_i(A) = T_iA^{-\theta}, A > T_i^{\frac{1}{\theta}} \tag{3.1}$$

[2]See studies such as Chaney (2008), Eaton et al. (2011) and Hsieh and Ossa (2011)

CES preferences and Cobb Douglas technologies imply the following pricing rule

$$p_{ij}(A) = \bar{m} \cdot d_{ij} \left(\frac{w_j}{A} \right)$$

where $\bar{m} = \frac{\eta}{\eta-1}$ denotes the markup and $d_{ij} > 1$ as iceberg cost with normalization $d_{ii} = 1$. Let X_{ij} denote the nominal value of shipments from country j to country i.

I follow Chaney (2008) and Arkolakis et al. and derive trade shares as

$$\lambda_{ij} = \frac{X_{ij}}{\sum_j X_{ij}} = \frac{T_j [c_j d_{ij}]^{-\theta} L_j f_{ij}^{-\left(\frac{\theta}{\eta-1}-1\right)}}{\sum_{k=1}^{N} T_k [c_k d_{ik}]^{-\theta} L_k f_{ik}^{-\left(\frac{\theta}{\eta-1}-1\right)}} \tag{3.2}$$

where d_{ij} capture iceberg costs of transport, such that $d_{ii} = 1$ and $d_{ij} d_{jk} \geq d_{ik}$, f_{ij} capture fixed costs of shipping to market i from source country j with $f_{ij} > f_{jj}$ for $j \neq i$, $(c_j / T_j^{\frac{1}{\theta}})$ captures unit costs of production in country j and L_j denotes the labor force in country j. Equation (3.2) can be derived from a Melitz model where firm-level productivities are drawn from a Pareto distrubution and where entry is assumed to be proportional to the labor force of the economy.

These trade shares give rise to the following standard Gravity equation:

$$\log\left(\frac{X_{ij}}{X_{ii}}\right) = -\theta \log(d_{ij}) - \left(\frac{\theta}{\eta-1} - 1\right) \cdot \log\left(\frac{f_{ij}}{f_{ii}}\right) + \log\left(T_j L_j c_j^{-\theta}\right) - \log\left(T_i L_i c_i^{-\theta}\right) \tag{3.3}$$

It is worthwhile to interpret the factors in this equation to understand the patterns of data better. The structural Gravity equation describes the determinants of market shares in a given market i of exporters from country j vs. home producers from i. According to this formulation of the Gravity equation, there are three main factors determining bilateral trade. First, exporters from country j will have lower market shares if either iceberg costs of shipment d_{ij} are high, or if fixed costs of exporting from j to i are high, i.e. $f_{ij} > f_{ii}$. The major restriction of the theory with respect to these trade costs relates to the selection of exporters into trade. If as observed in the data, only a fraction of firms is exporting and these firms are typically larger and more productive than only domestically active firms, then $f_{ij} > f_{jj}$. Second, since the number of entrants is assumed to be proportional to the labor force countries with more firms tend to have higher market shares. This is the Love-of-Variety effect implied by the CES specification used for preferences. Third, country j's market share will be higher than country i's market share if its cost advantage is higher. To see this, note that marginal costs can be written as $mc_j = (c_j / T_j^{\frac{1}{\theta}})$, where $c_j \propto R_j^{\alpha} w_j^{1-\alpha}$ is the cost of a composite input bundle of capital and labor with R_i as the rental rate of capital and w_i the wage rate and $T_j^{\frac{1}{\theta}}$ is the mean productivity of firms in country j.

As emphasized by Arkolakis, Costinot and Rodriguez-Clare (n.d.), a key feature this specification shares with virtually all applied Gravity models, is that it generates a constant-elasticity import demand system.

Property 1: Constant Elasticity Import Demand

> Define the partial elasticity $\varepsilon_j^{ik} = \partial \log\left(\frac{X_{ij}}{X_{ii}}\right)/\partial \log(d_{ik})$. The import demand system is such that for any importer i and any pair of exporters $j \neq i$ and $k \neq i$, $\varepsilon_j^{ik} = \varepsilon$ if $j = k$ and zero otherwise

An implication of this class of demand systems is that country-level marginal costs mc_i and trade elasticities are strictly separable.

$$\partial\varepsilon_j^{ij}/\partial \log(mc_i) = \partial\varepsilon_j^{ij}/\partial \log(mc_j) = 0 \tag{3.4}$$

Equation (3.3) is even stronger than this condition, as it implies a form of symmetry of trade flows unless trade frictions are asymmetric. To see this, let me assume for the moment symmetric trade frictions $d_{ij} = d_{ji} = d$ and $\frac{f_{ij}}{f_{ii}} = \frac{f_{ji}}{f_{jj}} = f$. Define $\log(\tau) = -\theta \log(d) - \left(\frac{\theta}{\eta-1} - 1\right) \cdot \log(f)$.

Property 2: Symmetry

> Given trade frictions are symmetric as defined above, trade flows satisfy
>
> $$\log\left(\frac{X_{ij}}{X_{ii}}\right) + \log\left(\frac{X_{ji}}{X_{jj}}\right) = 2 \cdot \log(\tau) \tag{3.5}$$

In this setup, any country specific factor, such as technology T_i, factor costs c_i or labor force L_i that changes the trade flow $\log\left(\frac{X_{ij}}{X_{ii}}\right)$ impacts the reverse trade flow $\log\left(\frac{X_{ji}}{X_{jj}}\right)$ inverse proportionally. What is more, symmetric trade frictions can only reduce the sum of the flows $\log\left(\frac{X_{ij}}{X_{ii}}\right) + \log\left(\frac{X_{ji}}{X_{jj}}\right)$ but will not change one trade flow $\log\left(\frac{X_{ij}}{X_{ii}}\right)$ relative to the other $\log\left(\frac{X_{ji}}{X_{jj}}\right)$. The only way to generate asymmetries in these trade flows, is then to generate asymmetric trade frictions.

On the other hand, as argued by Waugh (2009) and Helpman, Melitz and Rubinstein (2007), asymmetries in trade flows seems to be an important stylized fact of the data. Rich countries like the US seem to have systematically higher market shares when exporting to other countries than implied by country-specific effects such as productivity alone. To document this point, let me follow Waugh (2009)'s modification of the Gravity equation:

$$\log\left(\frac{X_{ij}}{X_{ii}}\right) = \beta_\tau \log(\tau_{ij}) + \beta_{S,j} D_{S,j} - \beta_{S,j} D_{S,i} + \beta_{X,j} D_{X,j}$$
$$\text{s. th.} \quad \sum_{i=1}^{N} \beta_{S,i} = 1 \tag{3.6}$$
$$\text{s. th.} \quad \sum_{i=1}^{N} \beta_{X,i} = 1$$

where X_{ij} denotes the nominal value of shipments from country j to country i. As in traditional Gravity-type regressions, τ_{ij} includes observable factors such as distance or common borders, capturing bilateral trade barriers. On the other hand the fixed effects are used to recover unobservable factors influencing trade. These include dummies for country fixed effects

$$D_{S,i} = \begin{cases} 1 & \text{if country i is an exporter} \\ -1 & \text{if country i is an importer} \end{cases}$$

and dummies capturing systematic exporter effects

$$D_{X,i} = \begin{cases} 1 & \text{if country i is an exporter} \\ 0 & \text{if country i is an importer} \end{cases}$$

To be able to capture asymmetries in the trade flows, one can decompose the trade frictions into bilaterally observable components $d_{B,ij}$ and $f_{B,ij}$ and unobserved exporter-specific components $d_{X,ij}$ and $f_{X,ij}$:

$$\begin{aligned} d_{ij} &= d_{B,ij} d_{X,ij} \\ f_{ij} &= f_{B,ij} f_{X,ij} \end{aligned} \tag{3.7}$$

Therefore the coefficients of the Gravity regressions map into the structural model according to

$$\begin{aligned}\hat{\beta}_\tau \tau_{ij} &= \log d_{B,ij} + \left[\frac{1}{\eta-1} - \frac{1}{\theta}\right] \log f_{B,ij} \\ \hat{\beta}_{X,j} &= \log d_{X,j} + \left[\frac{1}{\eta-1} - \frac{1}{\theta}\right] \log f_{X,j}\end{aligned} \tag{3.8}$$

while the country effects summarize

$$\hat{\beta}_{S,i} = \log\left(T_i c_i^{-\theta} L_i\right) \tag{3.9}$$

Figure 45 summarizes the key result regarding asymmetric trade patterns. It displays on the x-axis the estimated coefficients of the country fixed effects, while showing coefficiencts of exporter fixed effects on the y-axis. There is a clear positive pattern, stating that countries with higher values of technology or lower values of costs, are also countries that systematically export more. This positive systematic relationship seems to be mostly driven by the difference between the group of very poor countries on the one hand and and the set of very rich countries on the other hand. On the other hand, for our purposes it makes sense to consider a subsample of rich OECD countries which correspond to countries for which I will have firm-size data later. Indeed, if the sample is conditioned on the 16 richest countries, the relationship turns negative as shown in 44.

Given that the coefficients on distance and the exporter fixed effects are proving estimates of the trade cost matrix, I follow Waugh (2009) to recover technology parameters. In particular, suppose factor endowments in country i are given by

$$E_i = K_i^\alpha L_i^{1-\alpha} \tag{3.10}$$

Then the overall spending of country i can be shown to satisfy

$$P_i Y_i \propto c_i E_i \tag{3.11}$$

Imposing Balanced Trade then gives

$$c_i = \sum_{j=1}^{N} \left(\frac{X_{ji}}{\sum_{i=1}^{N} X_{ji}}\right)\left(\frac{c_j E_j}{E_i}\right) \quad i = 1, ..., N \tag{3.12}$$

This is a system of equations that, given observable trade shares $\frac{X_{ji}}{\sum_{i=1}^{N} X_{ji}}$ and factor endowments E_i is solvable for c_i. With the c_i variables at hand and the endowments E_i observable, I can recover the technology parameters T_i. To summarize, the Gravity regression allows me to recover trade costs and technology parameters. Equipped with these parameters, I can generate now the implied firm-size dispersions within a country. Export participation is governed by the selection equation:

$$\Pi_{ij}(A) \geq f_{ij}$$

so that firms from country j only export to country i if gross profits from exporting cover the fixed costs of exports. The total size of a firm from country j with productivity draw A can therefore be written as

$$X_j(A) = \sum_{i=1}^{N} p_{ij}(A) y_{ij}(A) \cdot 1_{\{\Pi_{ij}(A) \geq f_{ij}\}}$$

The basic measure of dispersion that is reported in the data, is the standard deviation of log total sales of firms within a country, relative to the cross-country average. I therefore construct the same statistic by country with simulation draws that are based upon a solved equilibrium model with the parameters from the Gravity model.

$$\mathtt{SDN}_{x,j} = \frac{(Var[\log X_j(A)])^{1/2}}{\frac{1}{N}\sum_{k=1}^{N}(Var[\log X_k(A)])^{1/2}} \tag{3.13}$$

The results are displayed in figure X. The simulation model suggests that the baseline Melitz model with parameters from the Gravity equations should display a strong negative relationship between internal trade share and firm-size dispersions. This relation reflects the core prediction of the Melitz model about the connection of openess to trade and market share allocations across firms within industries. Countries that are more open to trade exhibit larger dispersions of firm sizes as the most efficient firms enter export markets and expand total sales. On the other hand, firms that are only domestically active loose market share to foreign competitors and contract. This exporter-selection and market share reallocation mechanism is a key novel insight from heterogeneous firm models following Melitz. The figure demonstrates that a model estimated to fit empirical trade patterns also delivers a quantitately significant negative relationship. The next question is how this implication compares to the data.

To shed light on this quantitative implication of the standard Melitz-Pareto gravity model, one requires data on firm-size distributions and trade patterns. Internationally comparable data on firm-size distributions within narrowly defined industries is difficult to obtain for a large sample of countries. A recent study by Bartelsmanetal08 compiles such comparable firm-size data for a dozend OCED countries within 2-digit industries. I merge this data with data ob internal trade shares across the same 2-digit industries from the OCED STAN database.

The baseline message from the table is that internal trade shares and firm-size dispersions exhibit a positive relationship. That is, sectors that are on average more

Table 3.1: Pooled Regressions

	Sample			
Dependent Variables	`all`	`all`	`ex.USA`	`ex.USA`
$\lambda_{ii,s}$	.64 [.083]	.633 [.106]	0.39 [.072]	.34 [.092]
$\texttt{log Y}_{i,s}$		.0007 [.005]		.004 [.004]
Sector FEs	Yes	Yes	Yes	Yes
Number of Obs.	246	246	222	222
R^2	0.16	0.16	0.15	0.16

Notes: Pooled regressions using all industry-country pairs. Dependent variable: $\texttt{SDN}_{x,j} = \frac{(Var[\log X_j(A)])^{1/2}}{\frac{1}{N}\sum_{k=1}^{N}(Var[\log X_k(A)])^{1/2}}$. Independent variables include $\lambda_{ii,s} = \frac{X_i i}{\sum_{j=1}^{N} X_{ij}}$, $\log Y_{i,s}$: gross output of industry s. Standard errors are clustered on the 2-digit industry level.

closed to trade exhibit systematically higher firm-size dispersions. This is true even as I control of sector size, as captured by overall gross output. This relationship is not driven by compositional effects across sectors, sector effects are controlled for. What is more, as I document in the appendix, the relationship holds within 2-digit sectors. The quantitative relationship between size and firm-size dispersions is clearly influenced by the US. Excluding the US cut the coefficient in half. But it still remains positive and significant, although dropping the US cuts out around 10% of the original sample.

To summarize, the data offers two key stylized facts. First, there seem to be significant asymmetries in trade patterns as captured by exporter fixed effects in Gravity regressions. If we take these frictions and the standard specification of the Melitz model with identical efficiency dispersions of firms across countries at face value, this implies a negative relation between internal trade shares and firm size dispersions. This implications seems to be at odds with the second stylized fact that the relationship between firm size dispersions and internal trade shares is positive. Motivated by these observations, I propose a simple deviation of the standard specification that can generate the second stylized fact and will also have implications for the asymmetric trade patterns observed in the data.

3.3 Theory

The theoretical innovation of this chapter is a tractable analysis of international differences in firm efficiency dispersions. In principle this could be achieved by using differences in the Pareto dispersion parameter across countries. It turns out that this is a numerically quite challenging task. In order to facilitate the computational analysis I will analyze a different assumption for efficiency distributions, which is Log-normality. In practice, both approaches to modeling size distributions have been popular in the empirical literature.

3.3.1 Closed Economy

I start out with the closed economy to discuss build up the basic intuition for the theory. As before, the aggregate production function is given by the CES aggregator

$$Y = \left(\int y(A)^{\frac{\eta-1}{\eta}} \cdot d\mu(A) \right)^{\frac{\eta}{\eta-1}}$$

Final composite good production is supplied by a contiuum of monpolistic competitors, each with an idiosyncratic productivity:

$$\max_{\{L(A)\}} \Pi(A) = p(A)y(A) - wL(A)$$
$$\text{subject to: } y(A) = A \cdot L(A)$$

I assume that micro-level productivity is drawn from a log-normal. The log-normal distribution is one of two major distributions that is used in the literature to model firm and plant-size distributions. It is typically thought to capture the body of the firm size distribution fairly well. The alternative distributional assumption is Pareto, which is often argued to capture the tail of the firm size distribution well. In fact, both distributions are can be considered close to each other. To illustrate this point, consider the following log density of the Pareto distribution:

$$\log(f(x)) = (-\theta - 1)\log(x) + \theta \cdot \log(T) + \log(\theta)$$

where T is the minimum value or location parameter of the Pareto and θ is the dispersion parameter. Plotting the log frequency of realizations against the log values, this distribution gives a straight line with a constant slope, governed by the dispersion paramter θ. In contrast, the log density of the log-normal distribution is given by:

$$\begin{aligned}\log(f(x)) &= -\log(x) - \log\left(\sqrt{2\pi\sigma_A^2}\right) - \frac{[\log(x) - \mu_A]^2}{2\sigma_A^2} \\ &= -\frac{[\log(x)]^2}{2\sigma_A^2} + \left(\frac{\mu_A}{\sigma_A^2} - 1\right)\log(x) - \log\left(\sqrt{2\pi\sigma_A^2}\right) - \frac{\mu_A^2}{2\sigma_A^2}\end{aligned} \tag{3.14}$$

where, μ_A, σ_A are the mean and standard deviation of the log values of the variable x. As can be seen from the comparision of both densities, the key difference between both distributions is the first term on the right hand side of equation (3.14). This term capture a quadratic curvature on values of $log(x)$[3].

Figure (46) contrasts data on plant size distributions for US establishments and enterprises in 2000. As is known by work of Axtell01, the distribution very large firms in the US economy seems to be well captured by a Pareto distribution with coefficient 1. The figures is from Rossi-Hansberg and Wright (2007) and illustrates how the distribution of plants and enterprises in US data seems to be well approximated by a log-normal distribution. Furthermore, a growing literature shows how power laws for the largest percentiles of firms could be generated by the fact that large firms are multi-establishment entities. Power laws of firm sizes could therefore result even as plant-level heterogeneity is characterized by log-normality. See Growiec et al. (2008) and Bee et al. (2011). Since the theory of this chapter is really a theory about efficiency differences across production units, one can justifiably consider the log-normal distribution as the better choice for efficiency distributions. What is more, the corresponding cross-country evidence presented in section 2 is mostly from national economic censuses that compile establishment level rather than firm level data.

After firms realized their productivity, they can calculate net profits and decide whether to stay in business or exit. The selection equation for this decision is given by

$$\Pi(A) \geq f$$

It is convenient to rewrite this equation in logs and exploit the fact that productivity is given by $\log A = \mu_A + \sigma_A \cdot D_N$, where D_N is a standard-normal random variable. Therefore the selection equation can be rewritten as

$$\log A = \mu_A + \sigma_A \cdot D_N \geq -c$$

[3]Note that in order to approximate the fat tail of the Pareto distribution better, the log-normal has to increase the dispersion parameter σ_A, but eventually for very large values of x, the quadratic term will dominate.

where the cutoff value is given by

$$c = \log\left(\frac{1}{\bar{m}}\frac{P}{w}\left(\frac{PY}{\eta f}\right)^{\frac{1}{\eta-1}}\right)$$

Higher values of this cutoff allow more producers with lower values of productivity to survive. That is, higher values of c correspond to more firms surviving. The cutoff shows which variables determine the level of survival ceteris paribus. The cutoff will tend to be higher with overall demand PY being higher, the average ideal price index P being higher, wages w being lower and fixed costs of production f being lower. To understand the mechanics of selection and trade elasticities better in the full trade model, it is instructive to analyze the selection mechanisms and its relation to micro-level productivity of survivors better. For this purpose, let me define the normalized cutoff as

$$c_{JX} = \frac{c + \mu_A}{\sigma_A}$$

Due to the log-normality of the productivity draws, the first to moments of the distribution of productivity, conditional on survival are given by

$$\begin{aligned} E\left[\log A \middle| D_N \geq -c_{JX}\right] &= \mu_A + \sigma_A \psi\left(c_{JX}\right) \\ Var\left[\log A \middle| D_N \geq -c_{JX}\right] &= \sigma_A^2 \end{aligned} \tag{3.15}$$

Therefore to characterize how the average productivity, conditional on survival and the dispersion of productivity among survivors responds to changes in the selection cutoff c, it is useful to directly calculate the conditional expectations of truncated standard normals. These are given by

$$\begin{aligned} E\left[D_N \middle| D_N \geq -c_{JX}\right] &= \psi\left(c_{JX}\right) = \frac{\phi(c_{JX})}{\Phi(c_{JX})} \\ Var\left[D_N \middle| D_N \geq -c_{JX}\right] &= 1 + \psi\left(c_{JX}\right) \cdot \left(-c_{JX}\right) - \psi^2\left(c_{JX}\right) \end{aligned} \tag{3.16}$$

These truncated moments summarize how selection impacts the corresponding moments of the productivity distribution among surviving firms as a function of the selection cutoff. Recall that increases in c correspond to easier survival and therefore reduced sample selection. One can show that indeed

$$\psi'(c_{JX}) < 0$$

This implies that the truncation factor $\psi()$ will be lower, the easier survivial is. As a consequence of easier survival, the average productivity of surviving firms falls, as (3.15) shows. Indeed, this is also reflected in the limiting values of the truncated expectation, as

$$\lim_{c_{JX,ij} \longrightarrow -\infty} \psi(c_{JX,ij}) = \infty$$

This means that as survival becomes infinitely hard, the only surviving firm will be the one with the highest productivity. Since the normal distribution as infinite support, the limiting productivity is not bounded and therefore average productivity of survivors will also not be bounded. On the flipside,

$$\lim_{c_{JX,ij} \longrightarrow \infty} \psi(c_{JX,ij}) = 0$$

which means that as survival become costless, the selection term disappears and average productivity of survivors just reflects the average productivity of the untruncated efficiency distribution of firms.

To further understand the implications of selection for a given selection cutoff, it is useful to consider how the moments (3.16) change in response to changes in the underlying productivity parameters μ_A and σ_A. This is shown in first four panels of figures 48 and 49. In general as the cutoff c rises and survival becomes easier, the truncated mean falls and the truncated dispersion rises. The dispersion effect reflects the fact that sample selection cuts the lower tail of productivity draws. As survival becomes easier with higher values of c, less of the lower tail of the productivity distribution is cut. The productivity dispersion of survivors will therefore reflect more and more the full dispersion of efficiencies from the underlying distribution. Both of these selected moments are affected by parameters of the underlying distribution. Figure 48 shows how truncated mean and truncated dispersions change as the underlying average productivity increases. Intuitively, higher values of μ_A imply that more firms have higher productivity and survive. As a consequence, the selection term $\psi(.)$ is lower and since more firms from the lower tail of the distribution survive, the efficiency dispersion of survivors is higher. A similar mechanism is a work when considering increases in the underlying efficiency dispersion of firms. As shown in the top left panel of figure 49, the increase in the variance of productivity distributes more mass into the tails of the distribution. But this implies that more mass will be concentrated at infra-marginal firms. There will be more firms are are either very productive and will therefore survive given a fixed value of c. Or there are more firms that are of such low productivity that they would have decided to exit even for every low values of c. With more inframarginal productivities, the wedge that selection drives between the underlying and the observed productivity distribution becomes smaller and smaller for any fixed value of c. This effect is reflected in the lower selection term $\psi(.)$ and the higher conditional dispersion in the middle panels of figure 49.

Let us now move over to the question of how the selected distribution of micro-level productivity is related to aggregate TFP. Explicit aggregation of establishment level productivity give the following expression for aggregate overall TFP in this economy

$$\begin{aligned}
\bar{A} &= \left(\int_{\Pi(A) \geq f} A^{\eta-1} d\mu(A) \right)^{\frac{1}{\eta-1}} \\
&= \underbrace{J^{\frac{1}{\eta-1}}}_{\text{Variety}} \cdot \underbrace{E\left[A^{\eta-1} \middle| D_N > -c_{JX} \right]^{\frac{1}{\eta-1}}}_{\text{TFP with Selection}}
\end{aligned}$$

Overall TFP consists of two elements. The first is the Dixit-Stiglitz variety effect. This is driven by the number of producers. As before, I assume that the number of entering firms J_e is exogenously given. Introducing entry would be straightforward but would not change the analysis much and would distract from the main core of the analysis, which is selection.

$$J = J_e \Phi(c_{JX}) \tag{3.17}$$

The number of operating firms is determined by the number of latent draws multiplied by the probability of survival. This probability of survival is directly related to the inverse Mill's ratio selection term $\psi(.)$ analyzed above. In particular, note that

$$\frac{\partial \log \Phi(c_{JX})}{\partial \log c_{JX}} = \psi(c_{JX}) \tag{3.18}$$

This is intuitive as at the same time higher selection effects impact the average productivity of survivors, it also determines how many firms survive. All other things equal, increases in c make survival easier, which implies more firms and therefore a higher variety effect on welfare.

The second component of overall TFP is average producer efficiency, taking substitutability of products and selection into account.

$$TFP_A = \exp\left\{ \mu_A + \frac{1}{2}(\eta - 1)\sigma_A^2 \right\} \cdot \left[\frac{\Phi(c_{PX})}{\Phi(c_{JX})} \right]^{\frac{1}{\eta-1}} \tag{3.19}$$

where the value for c_{PX} is defined by

$$c_{PX} = \frac{c + \mu_A + (\eta - 1)\sigma_A^2}{\sigma_A} = c_{JX} + (\eta - 1)\sigma_A$$

This expression is central for the economics of this study, as it relates the underlying dispersion of producer efficiency to the aggregate productivity, which in turn will be important to understand trade patterns. The key property in this context is the elasticity of substitution, since it translates strongly differences in producer efficiency translate into size differences and hence profitability and survival. Under substitutability of the differentiated products, $\eta > 1$, more dispersion in producer efficiency will lead to a concentration of resources at the most productive firms. What is more, the stronger the substitutability of products, i.e. the higher η, the more will the most productive firms dominate the economy. This effect is captured in the first term of (3.19). How does this effect interact with the selection to determine aggregate TFP? To answer this question, we need to analyze how TFP_A varies for fixed selection cutoffs c.

As discussed above, increases in the the selection cutoff value c imply lower average producer productivity condtional on survival. This lower average producer productivity is reflected in TFP_A. For this consider the bottom right panel of figures 48 and 49. It plots different values of c against $\log TFP_A$ and shows how easier survival leads to lower TFP_A. The interaction between selection and differences in the dispersion of producer efficiency can be traced out in the differently colored lines in the bottom right panel of figure 49. As producers become more dispersed, TFP_A is less and less affected by an increase in c. The reason for this is related to the fact that with more efficiency dispersion, there will be less marginal firms. In the extreme case there are only extremely efficient firms that will survive even very low values of c and very unproductive firms that would have exited anyway. As a result marginally marking survival easier does not lead to many more firms staying in business so average productivity of survivors does not change much.

To understand the trade and specialization patterns, it will be important to see the net effect of changes in the selection cutoff c on overall aggregate TFP $\bar{A}$. Formally, one can show that since $\psi'(c) < 0$, this implies that

$$\frac{\partial \log \bar{A}}{\partial \log c} = \frac{1}{\sigma_A} \psi(c_{PX}) > 0 \tag{3.20}$$

This is the case since the variety effects tend to dominate the selection-effects in TFP_A discussed before. The effect is illustraed in the bottom left panel of figures 48 and 49. Both show how overall aggregate TFP $\bar{A}$ is increasing as surival becomes easier.

The analysis up to now assumes the selection cutoffs c as fixed and characterized partial elasticities to understand some of the basic mechanics. But the components in c are endogenous equilibrium objects. To see whether the baseline message of this analysis changes much, I calibrate and solve the closed economy version of the model, which is a system of two unknowns in two equations. Figure 50 plots out different equilibrium values for aggregate objects and micro-level moments for equilibria with different underlying dispersions σ_A. As one increases the dispersion of producer efficiency, TFP in the economy increases and the real wage rises. The right hand side top and middle panels in figure 50 trace out equilibrium variety and TFP_A compared to the initial equilibrium. As the underlying producer efficiency dispersion increases, variety falls and aggregate productivity increases. TFP_A increases strongly and dominates in welfare terms the loss due to a lower degree of variety. The bottom panels show the equilibrium selection effects in the micro-moments. Average productivity of survivors increases as was expected from stronger selection effects. On the other hand, the rise in the underlying dispersion of efficiency does lead to an increase in the efficiency dispersion of survivors, despite selection effects truncating more of the lower tail of the productivity distribution. This feature will help to reconcile the measured establishment-size dispersions and will tend to offset the strict implications of the Melitz model.

3.3.2 Open Economy

The open economy economic environment parallels the one analyzed in section 2. The major difference here is that I simplify the model further to facilitate the exposition of the model's properties. In particular, I will follow the closed economy framework and assume that labor is the only factor of production. Also, I will continue to assume that the number of entrants in each country is fixed and identical across countries.

The analog Gravity equation to (3.3) in this environment is given by

$$\log\left(\frac{X_{ij}}{X_{ii}}\right) = -(\eta-1)\log(d_{ij}) - (\eta-1)\log\left(\frac{w_j}{w_i}\right) + (\eta-1)\log\left(\frac{\bar{A}_{ij}}{\bar{A}_{ii}}\right) \tag{3.21}$$

where the overall aggregate TFP terms are the open economy analogs of the aggregate TFP factors analysed in the closed economy framework. These are given by

$$\bar{A}_{ij} = \exp\left\{\mu_{A,j} + \frac{1}{2}(\eta-1)\sigma^2_{A,j}\right\} \cdot \Phi\left[c_{PX,ij}\right]^{\frac{1}{\eta-1}}$$

The major difference to the closed economy version is that selection cutoffs of entering a foreign market are functions of the trading country pair.

$$c_{ij} = \frac{1}{\bar{m}} \frac{P_i}{w_j} \frac{1}{d_{ij}} \left(\frac{P_i Y_i}{\eta f_i} \right)^{\frac{1}{\eta-1}}$$

These cutoffs are independent of any assumption about the productivity distribution. For productivity moments, the normalized cutoff value is the feature of the model through which the efficiency distribution interacts with trade frictions:

$$c_{PX,ij} = \frac{c_{ij} + \mu_{A,j} + (\eta - 1)\sigma_{A,j}^2}{\sigma_{A,j}}$$

Similar to the case of the closed economy, the variable c_{PX} will govern the overall TFP response, including variety effects, to differences in the selection cutoffs c_{ij}.

To highlight the key differences of this Gravity equation with the standard Melitz-Pareto model with identical dispersions, it is helpful to contrast the implications of this model with the standard model of section 2. Remember that an important feature of the standard model is that trade elasticities are constant and identical across countries. In the Melitz model with Pareto distributions and identical dispersion parameter, the partial trade elasticity is

$$\frac{\partial \log \left(X_{ij}/X_{ii} \right)}{\partial \log(d_{ij})} = -\underbrace{(\eta - 1)}_{\text{IM}} - \underbrace{[\theta - (\eta - 1)]}_{\text{EM}}$$

The Pareto assumption implies that intensive margin effects of trade barriers and part of the extensive margin effect exactly offset each other. As a result, trade patterns generated by the Melitz-Pareto framework with identical dispersion parameters across countries will basially parallel a simple Armington constant-elasticity trade model. The same partial trade elasticity in the Melitz model with log-normality can be shown to be

$$\frac{\partial \log \left(X_{ij}/X_{ii} \right)}{\partial \log(d_{ij})} = -\underbrace{(\eta - 1)}_{\text{IM}} - \underbrace{\frac{1}{\sigma_{A,j}} \psi(c_{PX,ij})}_{\text{EM}} \tag{3.22}$$

The extensive margin effect in the new model is intimately connected to the selection effects discussed in the closed economy section. Note especially that the inverse mills ratio $\psi(.)$ reappears here, as trade barriers impact $\bar{A}$ along the extensive margin by through two channels. First, higher trade barriers reduce the number of exporters and therefore reduce the overall variety of exports from the exporting country. Second, trade barriers change the efficiency composition of exporting firms. The net extensive margin effect depends in the level of the normalized cutoff value $c_{PX,ij}$, which in turn depends in the underlying efficiency distribution. As trade barriers increase, exporting becomes harder and $c_{PX,ij}$ falls. This tends to increase the partial trade elasticity ceteris paribus, as the productivity of the marginal exporter rises. In other words, for a given productivity dispersion, a lower cutoff value c_{ij} means that the marginally non-exporting firm is of relatively high productivity. As trade barriers would be marginally lowered, the value of trade would expand a lot as the productive producer starts to generate sales from abroad.

A noticeable feature of these partial trade elasticities is the importance of the productivity distribution of the exporting country. The same change in a trade barrier can now generate very different responses depending on what the productivity distribution of the exporting country looks like.

3.4 Quantitative Analysis

This section developes the complete open economy implications in general equilibrium. The two features of the data I will focus most attention on are motivated by the stylized facts and the theory discussion of section 2. In particular, I ask whether realistic differences in underlying productivity distributions of production units can generate the relation of firm-size distributions and trade openess generated in the data. I then analyze the implications for trade asymmetries.

I start out by contrasting a baseline model where countries face simple symmetric trade frictions and only differ in underlying productivity. This helps to streamline the analysis to the key novel feature of the model in this study. The next subsection will then add a realistic geography and factor endowments from the data and will contrast the predictions of the model with the data.

3.4.1 Simple Geography and Identical Factor Endowments

Before turning to the solution of the model, a number of parameters need to be calibrated. These are listed in table 3.4.1. I mostly follow standard values of the literature. Since estimates on the fixed costs of exporting are more difficult to get, I calibrate a uniform value such that for the most productive country, around 20 percent of establishments export, see Bernard, Eaton, Jensen and Kortum (2003).

Parameter	Value	Explanation
η	3	Elasticity of Substitution Hsieh and Klenow (2009)
F_{ii}	1	Domestic Fixed Cost Workers
F_{ij}	F	Export Fixed Cost Workers Match 20% of firms export
$\tilde{J}$	20000	Number of Latent Plants
d_{ij}	3	Iceberg trade friction Anderson and van Wincoop(2004)
$\mu_{A,j}$	[-0.2200, 1.4050]	mean country-level efficiency
$\sigma_{A,j}$	[1,2.05]	dispersion of firm efficiencies

The productivity differences are chosen so that the mean underlying productivities are in the same range. The underlying mean productivity for log-normal idiosyncratic productivity is

$$\tilde{A} = \exp\left\{\mu_A + \frac{1}{2}\sigma_A^2\right\}$$

I assume that $\tilde{A}$ varies in the range [0.5, 2.12], which generates differences in aggregate technology of a factor of around 5. In contrast two possible world economys. In the first case, I assume that efficiency dispersion across producers within a country are identical and equal to $\sigma_A = 1.2$. I then pick μ_A to generate the given range for $\tilde{A}$. In the second case, I assume that only these dispersions vary such that $\sigma_A \in [0.5, 2.12]$. Quantitatively similar results can be obtained if both dispersions and mean productivities vary across countries.

The results of contrasting both equilibria can be seen in figure 51 and 52. The top and middle panels summarize the key differences in welfare and trade patterns among the two equilibria. To understand these, note that the same difference in mean productivity $\tilde{A}$ translates into very different differences in overall TFP $\bar{A}$. The reason is that if a country has a higher μ_A or a higher σ_A matters. The reason is with substituability of products more dispersion generates more very productive firms that attract overproportional market shares. Therefore aggregate TFP differences and welfare differences are quite different. The ratio of real wages in the richest to the poorest economy is around 4.8 with only μ_A differing. But the same ratio is 12.2 in the allocation where only σ_A differs. The middle panel corroborates that both economies display the same relationship of productivity and openess. Countries that have higher productivity levels will substitute external trade for internal trade and will therefore tend to be endogenously more closed to trade than countries have low levels of technology. The bottom panel contrasts how the micro-level implications differ across the two allocations. The world equilibrium with only σ_A differences will have countries that are at the same time very closed to trade but will exhbit higher dispersion in firm sizes. The slope of the relationship between internal trade shares and firm-size dispersions is $+0.3269$. In contrast the allocation where only μ_A differs internationally exhibits a strong negative relation between firm-size dispersions and internal trade shares. The log-normal model without differences in dispersions across countries generates therefore similar predictions as the Melitz-Pareto Model of section 2. The implied slope of internal trade shares and firm-size dispersions across countries is -0.6246.

These two different ways technology differences across countries can take, have also different implications for trade patterns. These are displayed in figure 52. Let me first investigate how the the differences in technology across country impact the partial trade elasticities from (3.22). For this purpose I calculate the average extensive margin trade elasticity of an exporter, defined by

$$ATE_{EM,j} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_{A,j}} \psi(c_{PX,ij}) \tag{3.23}$$

Plotting these average extensive margin elasticities for exporters against technology differences is done in the top panel of figure 52. extensive margin parts of trade elasticities are lower for more productive countries, as more firms from these countries export. Therefore a given trade barrier will mostly only reduce trade along the intensive margin. This is in contrast to countries with a lot of unproductive firms. A given trade barrier impacts exporter selection into trade and has therefore a larger impact on trade flows.

Let us now move over to the question of asymmetries in trade flows. For this purpose, I proceed as before and run Waugh's estimator (3.6) on the simulated data from the two allocations. Similar to section 2, country-factors like technology differences should be captured by the country fixed effects and asymmetries in trade flows are captured by the exporter fixed effects. In contrast to the standard setup of section 2, recall that I assume that all trade frictions are symmetric here. So any difference in exporter fixed effects are solely driven by technology differences here. The middle panel of figure 52 reports the results of the Waugh Gravity Estimator on the simulated data. Qualitatively the model can generate differences in exporter fixed effects. Note especially that even without any differences in dispersions, there can be systematic differences in exporter fixed effects. What international differences in σ_A offer is a potentially non-monotonic relation between country effects and exporter fixed effects.

These asymmetries generated by technology differences in these examples are typically orders of magnitude smaller than those documented in section 2. Consider for instance the rich country subsample of 44. The 90-10 difference of exporter fixed effects even in this restricted sample is 1.5701. In contrast, the corresponding 90-10 difference in both simple calibrations is 0.0251. One reason for these quantitatively small effects could be that the geography of these examples is exceedingly simple. Specifically, every country has exactly the same distance to all other countries. This might matter, as export cutoffs c_{ij} could more stronly differ with more realistic geographies and therefore induce larger variations in $\psi(c_{PX,ij})$. I analyze the possibility for this in the next section.

3.4.2 Realistic Geography and Factor Endowments

The simple model of the last section already generated realistic relations between firm size dispersions and internal trade share, comparable in magnitute to the data. It achieved this by only relying on differences in technology across countries. But it could not generate systematic asymmetries in trade flows that come close to quantitatively match the data. I therefore add endowment differences and a realistic geography that correspond to what we observe in the data. Endowment differences are given by the aggregate factor endowments given in section 2. I will basically replace labor forces by "equipped labor E_i". Second, to add a realistic geography I include distance and border-related iceberg trade frictions from standard Gravity estimates. I will exclude the estimated exporter fixed effects from the trade cost matrix, as the aim is to see how much of the asymmetric trade flows can be matched without asymmetric trade frictions.

I use the same technology differences in σ_A as in the last section to generate the positive relation between firm-sizes and internal trade shares. To match the geography in the data, I order the estimated country effects and assign different values of σ_A according to the order in the data. This will mean the geographic positions of countries of different productivities are roughly the same as in the data.

Figure 53 shows how adding the realistic geography alters some of the main conclusions. Welfare differences are much more pronounced compared to before, partly due to differences in capital across countries. The key result is that the variation in geography will already generate substantially larger exporter fixed effect variation. Compare the variation in exporter fixed effects as displayed in the bottom right panel of figure 53 to a comparable variation in the data, such as 44. First, both plots generate the same negative relation between country fixed effects and exporter fixed effects. Second, the differences of the exporter fixed effects are now of comparable magnitude. Recall that in the restricted sample had a 90-10 difference of around 1.5701. The same 90-10 difference is now around 0.8 in the model with realistic geography. The model can therefore roughly generate half of the variation in trade flow asymmetries as observed in the data without assuming asymmetries in trade frictions.

3.5 Conclusion

This chapter has argued that differences in efficiency dispersions are important to understand firm-size and trade patterns in the data. It has developed a framework that can easily encompass multiple sources of heterogeneity and can intuitively trace their impact on trade and international specialization.

There are three main areas the framework of this dissertation might be of immediate use. First, combining the ideas and data of this study with industry-level trade and production data it is possible to structurally estimate the contribution of efficiency-dispersion differences versus fixed costs in generating asymmetric trade patterns. The framework here can readily incorporate ideas of Helpman et al. (2007) who use Heckman-type two stage selection equations in a Pareto-framework.

Second, the framework can incorporate multiple sources of heterogeneity such as establishment-level frictions as in Hsieh and Klenow (2009), but also technology differences in factor shares. In particular, such a model with multiple sources of heterogeneity will be able to analyze how micro-distortions as in Hsieh and Klenow (2009) impact within-industry Heckscher-Ohlin specialization and therefore wage-inequality.

Third, the log-normal framework predicts that trade elasticities are potentially non-linear and related to firm heterogeneity of exporting and importing countries. An analysis of trade liberalization episodes with establishment level production data from both the liberalizing importer and an exporter country could help shed light on the quantitative importance of this prediction.

Bibliography

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 2007.

Anderson, James and Eric Van Wincoop, "Trade Costs," *Journal of Economic Literature*, 2004.

Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, Forthcoming.

_ , _ , and _ , "New Trade Models, Same Old Gains?," *American Economic Review.*

Atkenson, Andy and Ariel Burstein, "Pricing to Market, Trade Costs, and International Relative Prices," *American Economic Review*, 2008.

Baily, M., C. Hulten, and D. Campbell, "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity: Microeconomics*, 1992.

Banerjee, Abhijit and Esther Duflo, "Growth Theory through the Lens of Development Economics," in "Handbook of Economic Growth," Elsevier, 2005.

Bartelsman, Eric and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 2000.

Bartelsman, Eric J., Randy A. Becker, and Wayne B. Gray, "NBER-CES Manufacturing Industry Database," *NBER Working Paper*, 2000.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross Country Differences in Productivity: The Role of Allocative Efficiency," *NBER Working Paper*, 2009.

Bee, Macro, Massimo Riccaboni, and Stefano Schiavo, "Pareo versus Lognormal: A Maximum Entropy Test," *mimeo, University of Trento*, 2011.

Bernard, Andrew.B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93*, 1268–1290.

Bloom, Nicolas and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Jounral of Economics*, 2007.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006.

Buera, Francisco, Joseph P. Kaboski, and Yongseok Shin, "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 2011.

Caballero, R. J., Takeo Hoshi, and Anil K. Kashyap, "Zombie Lending and Depressed Restructuring in Japan," *American Economic Review*, 2008.

Caballero, Ricardo, *Specificity and the Macroeconomics of Restructuring*, MIT Press, 2006.

Camacho, Adriana and Emily Conover, "Misallocation and Productivity in Colombia's Manufacturing Industries," *Inter-American Development Bank Working Paper*, 2010.

Casacuberta, Carlos and Nestor Gandelman, "Productivity, exit and crisis in Uruguayan manufacturing and services sectors," *mimeo Universidad ORT Uruguay*, 2009.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review (forthcoming)*, 2008.

Diewert, Erwin, *Hicks' Aggregation Theorem and the Existence of a Real Value-Added Function*, Amsterdam: North-Holland, 1978.

Dong, Baiyu, "Misallocation or Mismeasurement? Evidence from Plant-Level Data," *mimeo New York University*, 2011.

Dube, J., Jeremy T Fox, and Che-Lin Su, "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," *mimeo University of Chicago, Booth*, 2009.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011.

Fisman, Raymond, "Estimating the Value of Political Connections," *American Economic Review*, 2001.

Greenwood, Jeremy, Juan M. Sanchez, and Cheng Wang, "Quantifying the Impact of Financial Development on Economic Development," *Working Paper, Federal Reserve Bank of St. Louis*, 2010.

Growiec, Jacub, Fabio Pammolli, Massimo Riccaboni, and H. Eugene Stanley, "On the size distribution of business firms," *Economics Letters*, 2008.

Hall, Robert E. and Charles I. Jones, "Why Do Some Countries Produce So Much More Output per Worker than Others?," *Quarterly Journal of Economics*, 1999, *114*, 83–116.

Hausman, Ricardo, Dani Rodrik, and Andres Velasco, "Growth Diagnostics," *Kennedy School of Government working paper*, 2005.

Heckman, James and Bo E. Honore, "The Empirical Content of the Roy Model," *Econometrica*, 1990.

Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *mimeo, Harvard University*, 2007.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen, "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 1988.

Hsieh, C. and Peter J. Klenow, "The Life Cycle of Plants in India and Mexico," *mimeo Stanford University*, 2011.

_ and R. Ossa, "A Global View of Productivity Growth in China," *Working Paper, Booth School of Business*, 2011.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009.

Jaef, Roberto Fattal, "Entry, Exit and Misallocation Frictions," *mimeo IMF research department*, 2011.

Jones, Charles I., "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macro*, 2011.

_ , "Misallocation, Economic Growth and Input-Output Economics," *Proceedings of the 10th World Congress of the Econometric Society*, 2011.

Klette, Tor Jacob and Arvid Raknerud, "Heterogeneity, Productivity and Selection: An Empirical Study of Norwegian Manufacturing Firms," *mimeo Statistics Norway Research Department*, 2005.

— and Zvi Griliches, "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, 1996.

Kumar, Krishna B., R. Rajan, and L. Zingales, "What Determines Firm Size?," *NBER Working Paper*, 1999.

Lien, Donald and N. Balakrishnan, "Moments and properties of multiplicatively constrained bivariate lognormal distributions with applications to futures hedging," *Journal of Statistical Planning and Inference*, 2006.

Lipsey, Richard G. and Kelvin Lancaster, "The General Theory of Second Best," *Review of Economic Studies*, 1957.

Luttmer, Erzo, "Models of Growth and Firm Heterogeneity," *Annual Reviews of Economics*, 2010.

Machicado, Carlos G. and Juan C. Birbuet, "Misallocation and Manufacturing TFP in the Market Liberalization Period of Bolivia," *mimeo Institute for Advanced Development Studies*, 2008.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695 – 1725.

Midrigan, Virgiliu and Daniel Xu, "Accounting for Plant-Level Misallocation," 2009.

— and — , "Finance and Misallocation: Evidence from Plant-Level Data," *NBER Working Paper*, 2010.

Moll, Benjamin, "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," *mimeo Princeton University*, 2010.

Neumeyer, P. and G. Sandleris, "Understanding Productivity During the Argentine Crisis," *Working Paper*, 2010.

Oberfield, Ezra, "Productivity and Misallocation During Crisis," *mimeo Federal Reserve Bank of Chicago*, 2011.

Peters, Michael, "Heterogeneous Mark-Ups and Endogenous Misallocation," *mimeo MIT*, 2011.

Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Journal of Economic Dynamics*, 2008.

Rossi-Hansberg, Esteban and Mark L. J. Wright, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, 2007.

Rust, John, "Optimal Replacment of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987.

Song, Zheng and Guiying Wu, "A Structural Estimation of Capital Market Distortions in UK and Chinese Manufacturing Firms," *mimeo, University of Chicago - Booth*, 2011.

Su, Che-Lin and Kenneth Judd, "Constrained Optimization Approaches to Estimation of Structural Models," *mimeo University of Chicago*, 2010.

Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, 1997.

Waugh, Michael, "International Trade and Income Differences," *American Economic Review*, 2009.

Appendix Chapter 1

Appendix 1.A: Figures

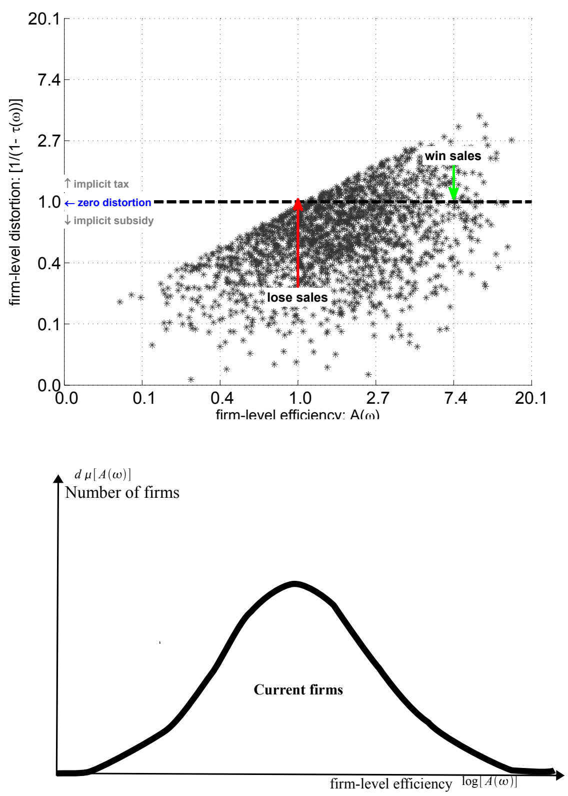

Figure 2: Reallocation with a given set of firms

Notes: Top diagram: Distribution and allocation plot of micro-frictions and firm productivity. Simulated data. Firms with values to the to the south or to the east have higher levels of sales. Reallocation implied when removing distortions shown by arrows. Bottom diagram: efficiency distribution of firms. Simulated example.

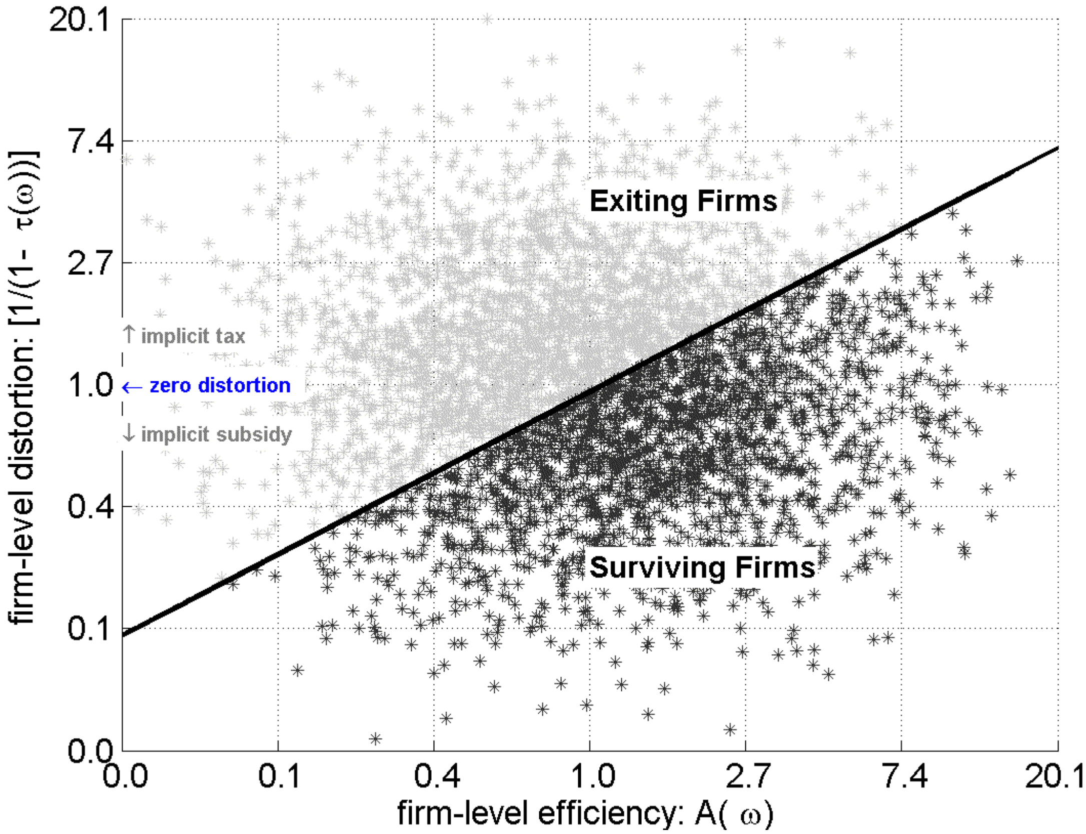

Figure 3: Simulated example of micro-frictions and firm efficiency with endogenous exit.

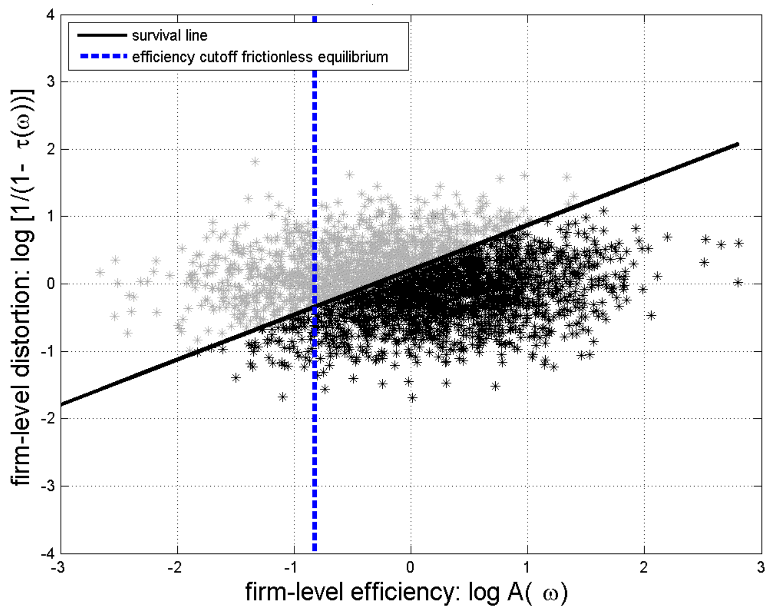

Figure 4: Selection in equilibrium with low covariance of frictions and efficiency. Equilibrium exit rate is 16%.

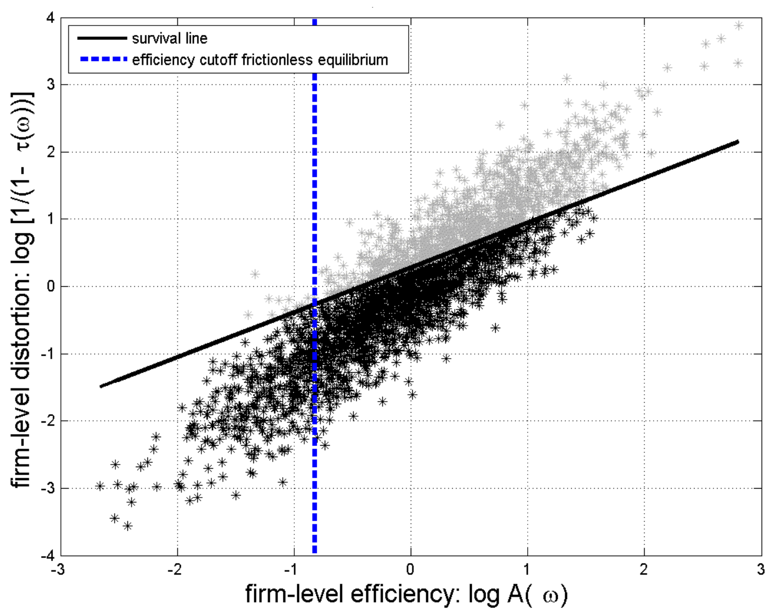

Figure 5: Selection in equilibrium with high covariance of frictions and efficiency. Equilibrium exit rate is 15%.

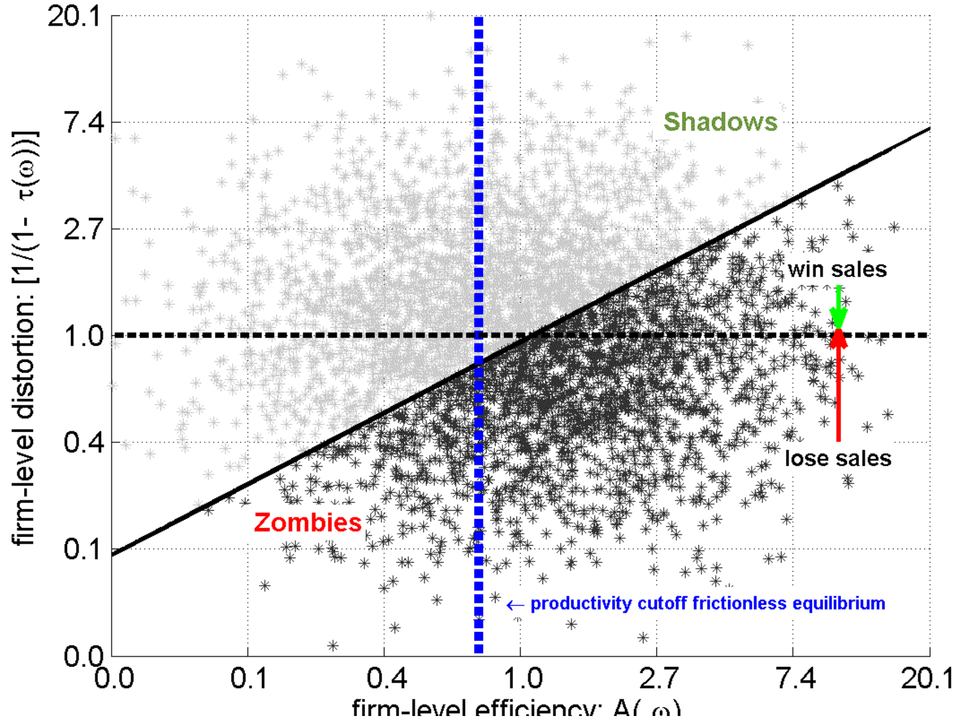

Figure 6: Intensive-margin gains: reallocation with fixed set of firms.

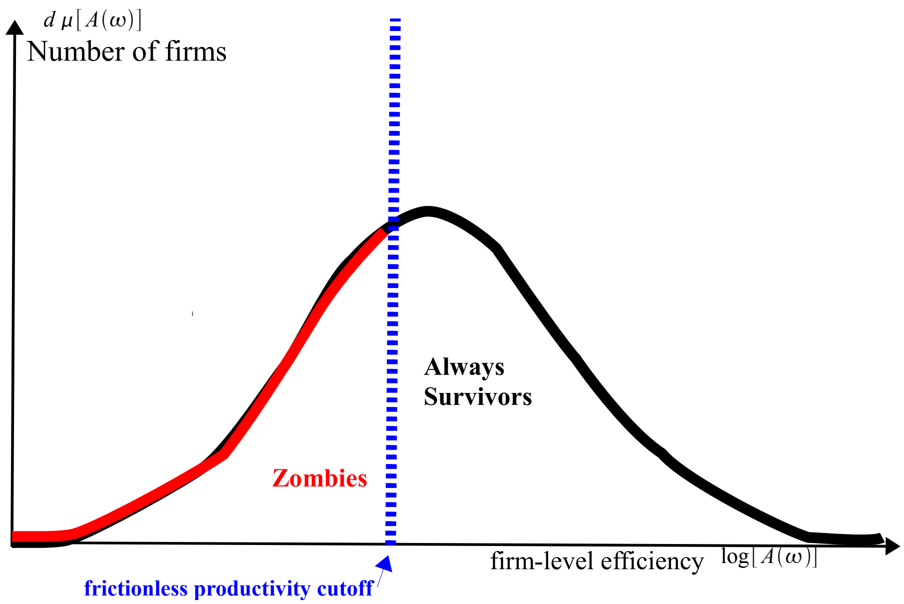

Figure 7: Efficiency composition of currently active firms

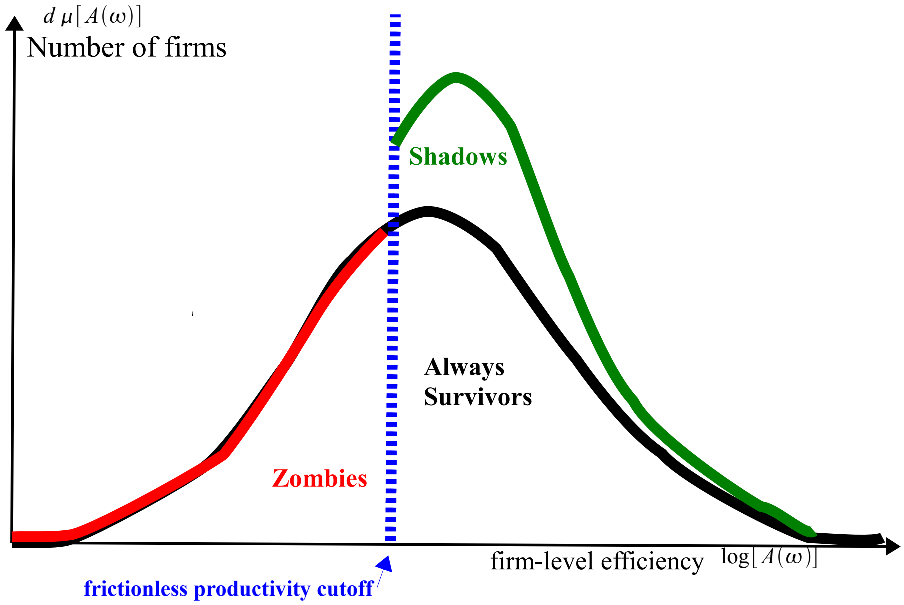

Figure 8: Efficiency compositions of Zombies and Shadows

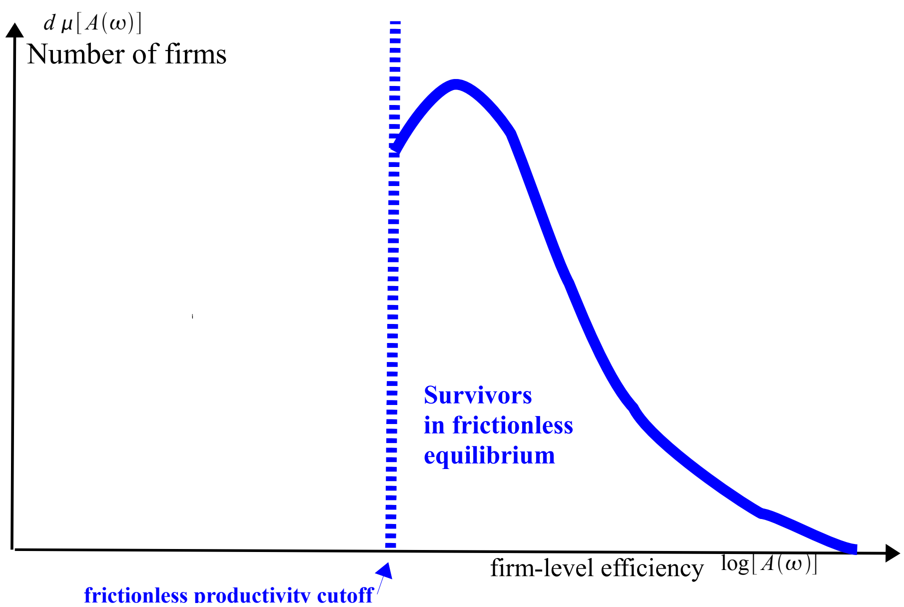

Figure 9: Efficiency composition of firms active in frictionless equilibrium.

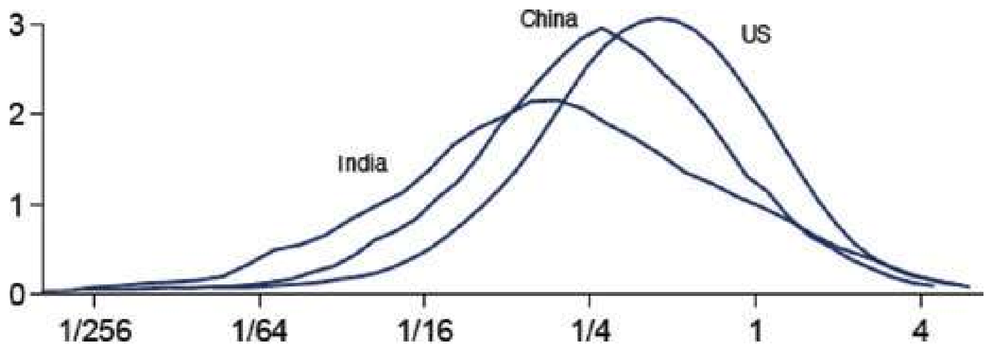

Figure 10: Comparision of log TFPQ = $\log A(\omega)$ distributions for India, China and the US. Estimates are over all 4 digit manufacturing sectors. Source: Hsieh and Klenow (2009)

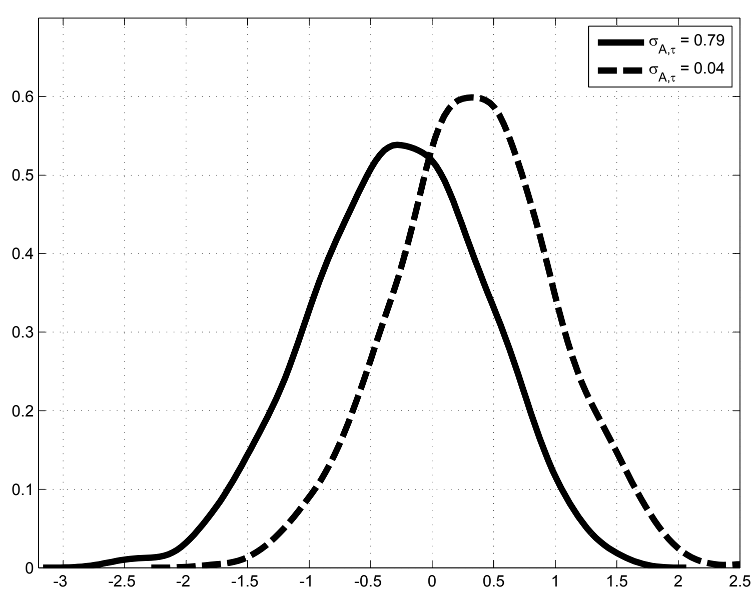

Figure 11: Shift in efficiency distribution from allocation with low dispersion of frictions and low covariance of frictions and efficiency to allocation with high dispersion of frictions and high covariance of frictions with efficiency. Dispersion of efficiency across firms is held constant.

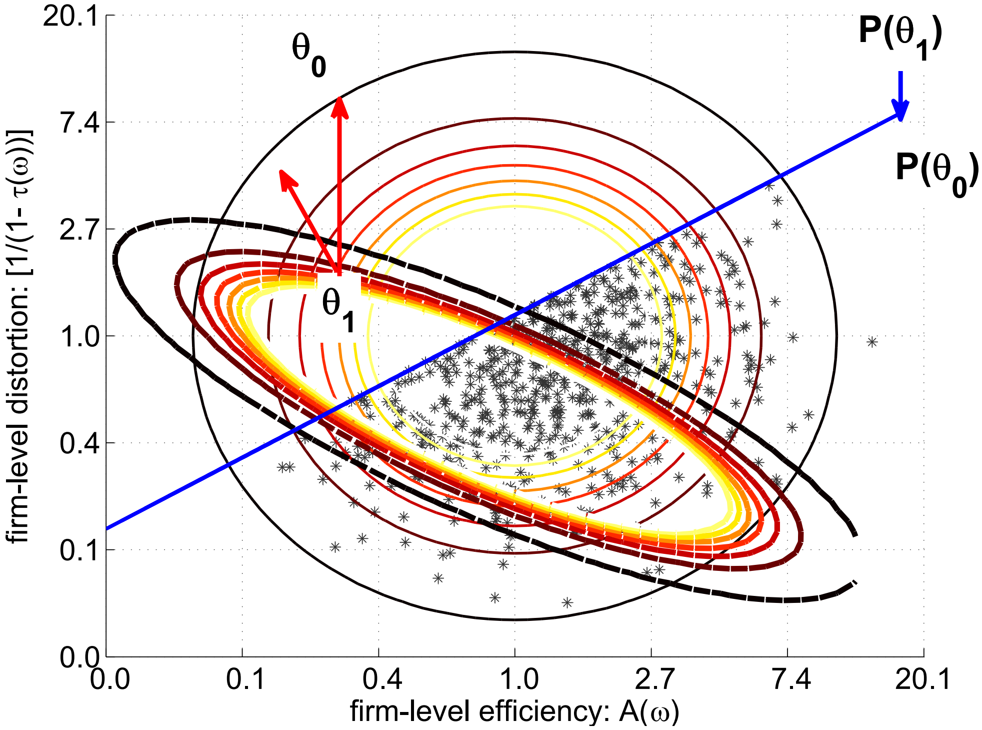

Figure 12: Graphical Illustration of MLE estimation with Equilibrium constraints

Appendix 1.B: Dynamics

The estimation strategy in the main text follows the assumption that establishment level heterogeneity is well characterized by a model of permanent level differences. I am motivated to pursue this approach by the literature on establishment size dynamics such as Baily et al. (1992) and more recently Klette and Raknerud (2005), who find that permanent level differences are crucial to understand the cross sectional heterogeneity of firm level TFP measures. In the literature on measurement of micro-distortions, studies such as Song and Wu (2011) and Midrigan and Xu (2009) suggest that the cross sectional dispersion of TFPR is not well captured by adjustment frictions, temporary shocks or measurement error.

To evaluate whether their findings hold up in my data as well, I estimate a Panel VAR of TFPR and TFPQ. The crucial feature of this approach is that I can remove time invariant plant fixed effects and estimate the degree of persistence once these fixed effects are taken out. As is well known simple time-differencing leads to inconsistent estimates due to the mixture of AR and fixed effects components. I therefore follow Holtz-Eakin, Newey and Rosen (1988) and exploit the fact that lagged values of the dependent variables qualify as instruments. Table 2 reports the results from pooled regressions concentrating on firms that do not decide to exit. Controlling for fixed effects, TFPR and TFPQ are not very persistent. The half lives of TFPR and TFPQ shocks are below one year. This is illustrated in figures 13 to 16, which show impulse responses to a one-standard deviation shock. The Cholesky decomposition of the shocks assumes that TFPQ is ordered first so that TFPR does not have any contemporaneous impact on TFPQ. As one can see temporary shocks usually have zero impact after 3 years. Note however that the size of temporary TFPQ shocks is large. The size of a one standard deviation shock is 80% to the level of TFPQ. TFPR shocks are significantly smaller, barely larger than 12%.

I conclude from this evidence that permanent level differences will most likely dominate time-averages of TFPR and TFPQ over the 6 year horizon of my data, due to the low persistence of TFPR and TFPQ once fixed effects are controlled for. Large temporary shocks might be an important additional margin of dynamic effects and is left for analysis in related research.

Table 2: Panel VAR

	Contemporaneous Variable	
Lagged Variable	$\texttt{log TFPR}_t(\omega)$	$\texttt{log TFPQ}_t(\omega)$
$\texttt{log TFPR}_{t-1}(\omega)$	.16842341 [.02711359]	−.39260905 [.03944105]
$\texttt{log TFPQ}_{t-1}(\omega)$	.03228435 [.01645172]	.52007713 [.02496301]
Number of Obs.	28,087	

Notes: Observations pooled by all 4 digit sectors and conditioned on non-exiters.

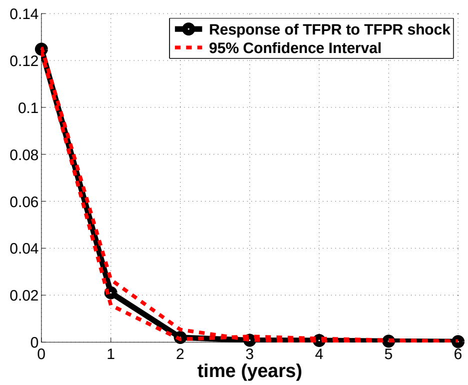

Figure 13: Impulse Response of `log TFPR` to a shock in `log TFPR`. Bootstrapped confidence intervals, 500 replications

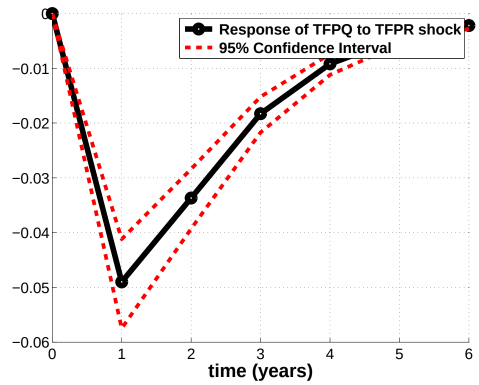

Figure 14: Impulse Response of `log TFPQ` to a shock in `log TFPR`. Bootstrapped confidence intervals, 500 replications.

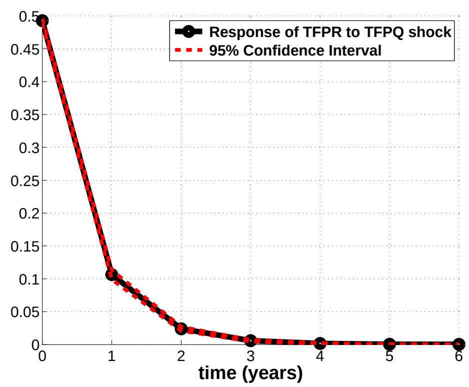

Figure 15: Impulse Response of `log TFPR` to a shock in `log TFPQ`. Bootstrapped confidence intervals, 500 replications.

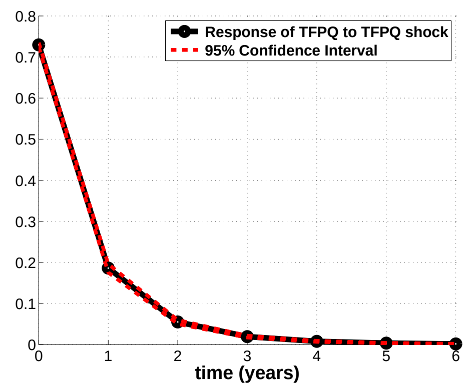

Figure 16: Impulse Response of `log TFPQ` to a shock in `log TFPQ`. Bootstrapped confidence intervals, 500 replications.

Appendix 1.C: Proofs

1.C.1: Proof of Proposition 1

Under free entry, the resource constraint implies

$$PY = \frac{1}{1-\bar{\tau}} wL \tag{24}$$

with $L = 1$, one only needs to calculate the CES ideal price index

$$P = \bar{m}\frac{w}{\bar{A}} \tag{25}$$

where $\bar{A} = \left[\int \left(A(\omega)[1-\tau(\omega)]\right)^{\eta-1} \cdot d\mu(\omega)\right]^{\frac{1}{\eta-1}}$. Substituting the CES ideal price index in equation (24) and setting $P = 1$ gives the result.

1.C.2: Derivation of (1.7)

Substituting for $p(\omega)y(\omega)$ in $\Pi(\omega)$ gives

$$\Pi(\omega) = \frac{1}{\eta}\left(\frac{p(\omega)}{P}\right)^{1-\eta} PY \tag{26}$$

This is the profit used in (1.3). Further substitution of (1.2) lets me solve for $A(\omega)$ and $\frac{1}{1-\tau(\omega)}$.

$$A(\omega) \cdot \left(\frac{1}{1-\tau(\omega)}\right)^{-\bar{m}} \geq \bar{m}\left(\frac{w}{P}\right) \cdot \left(\frac{\eta w F}{PY}\right)^{\frac{1}{\eta-1}} \tag{27}$$

Substitute from equation (25) for P and from equation (24) for PY to arrive at (1.7)

1.C.3: Proof of Proposition 2

To derive the fix cost estimate I need to analyze factor demands of the monopolistic competitors. Production-related factor demands are

$$R \cdot K_{P,s}(\omega) = \alpha_s \frac{1}{\bar{m}} \frac{1-\tau_{Y,s}(\omega)}{1+\tau_{K,s}(\omega)} p_s(\omega) y_s(\omega)$$
$$w \cdot L_{P,s}(\omega) = (1-\alpha_s) \frac{1}{\bar{m}} (1-\tau_{Y,s}(\omega)) p_s(\omega) y_s(\omega)$$

Fixed cost of operations require factor demand of

$$R \cdot K_{F,s}(\omega) = \alpha_s \frac{1}{1+\tau_{K,s}(\omega)} \left[\frac{w}{1-\alpha_s}\right]^{1-\alpha_s} \left[\frac{(1+\tau_{K,s}(\omega))R}{\alpha_s}\right]^{\alpha_s} \cdot F_s$$
$$w \cdot L_{F,s}(\omega) = (1-\alpha_s) \left[\frac{w}{1-\alpha_s}\right]^{1-\alpha_s} \left[\frac{(1+\tau_{K,s}(\omega))R}{\alpha_s}\right]^{\alpha_s} \cdot F_s$$

Measured factor demand is

$$\begin{aligned} K_{M,s}(\omega) &= K_{P,s}(\omega) + K_{F,s}(\omega) \\ L_{M,s}(\omega) &= L_{P,s}(\omega) + L_{F,s}(\omega) \end{aligned} \tag{28}$$

which implies for measured factor intensities

$$\frac{\alpha_s}{1-\alpha_s} \left(\frac{wL_{M,s}(\omega)}{RK_{M,s}(\omega)}\right) = [1+\tau_{K,s}(\omega)] \tag{29}$$

Survival condition $\Pi_s(\omega) \geq f_s(\omega)$ combined with $wL_{M,s}(\omega)$ implies

$$wL_{M,s}(\omega) \geq (1-\alpha_s) \left(\eta \cdot \left[\frac{w}{1-\alpha_s}\right]^{1-\alpha_s} \left[\frac{(1+\tau_{K,s}(\omega))R}{\alpha_s}\right]^{\alpha_s} \cdot F_s\right) \tag{30}$$

Combining (29) and (30) gives the bounds

$$\hat{F}_s = \frac{1}{\eta} \min_{\omega} \left\{L_{M,s}(\omega)^{1-\alpha_s} K_{M,s}^{\alpha_s}\right\}$$

1.C.4: Proof of Proposition 3

The estimation problem consists of two parts. First, a maximum likelihood problem of fitting a truncated multivariate distribution to the data from surviving firms. Second, this MLE problem is subject to an equilibrium constraint. In this derivation I proceed in three steps. The first two steps will derive the likelihood function. The third will derive the equilibrium constraint.

Step 1: Derivation of the log-likelihood

I slightly deviate from the exposition in the text and show the likelihood function in terms of observables, rather then the calculated wedges. The first observable are sales or value added:

$$D_1(\omega) = p_s(\omega)y_s(\omega) \tag{31}$$

In terms of underlying heterogeneity in the model, nominal output is determined by

$$\log D_1(\omega) \propto -(\eta-1)\left[\alpha_s \log\left(1+\tau_{K,s}(\omega)\right) + \log\left(\frac{1}{1-\tau_{Y,s}(\omega)}\right)\right] + (\eta-1)\log(A_s(\omega)) \tag{32}$$

The next observable is the composite input used by every firm. This is constructed from the underlying factor demand for capital and labor.

$$D_2(\omega) = \left(\frac{RK_s(\omega)}{\alpha_s}\right)^{\alpha_s}\left(\frac{wL_s(\omega)}{1-\alpha_s}\right)^{1-\alpha_s} \tag{33}$$

In terms of underlying heterogeneity from the model, this is

$$\log D_2(\omega) \propto (\eta-1)\log A_s(\omega) - \eta\alpha_s \log(1+\tau_{K,s}(\omega)) - \eta \log\left(\frac{1}{1-\tau_{Y,s}(\omega)}\right) \tag{34}$$

The third data source are factor intensities across firms within an industry

$$D_3(\omega) = \left(\frac{\alpha_s}{1-\alpha_s}\right)\left[\frac{wL_{M,s}(\omega)}{RK_{M,s}(\omega)}\right] \tag{35}$$

this factor intensity should directly map into the net capital wedge

$$\log D_3(\omega) = \log(1+\tau_{K,s}(\omega)) \tag{36}$$

These three data sources map into the three sources of heterogeneity through the mapping

$$\begin{pmatrix}\log D_1(\omega)\\ \log D_2(\omega)\\ \log D_3(\omega)\end{pmatrix} \propto \begin{bmatrix}(\eta-1) & -(\eta-1) & -(\eta-1)\alpha_s\\ (\eta-1) & -\eta & -\eta\alpha_s\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}\log A_s(\omega)\\ \log\left(\frac{1}{1-\tau_{Y,s}(\omega)}\right)\\ \log(1+\tau_{K,s}(\omega))\end{pmatrix} \tag{37}$$

Therefore, it is straightforward to generate the distribution of these three data series as a function of the two wedges and firm efficiency. Under trivariate log-normality, one gets

$$
\begin{aligned}
&\mathscr{L}\left(\theta \middle| \log D_1(\omega), \log D_2(\omega), \log D_3(\omega)\right) \\
&= \sum_{\omega} \left\{ \frac{3}{2}\log(2\pi) + \frac{1}{2}\log(|\Sigma|) - \frac{1}{2} \begin{pmatrix} \log D_1(\omega) - \mu_1 \\ \log D_2(\omega) - \mu_2 \\ \log D_3(\omega) - \mu_3 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} \log D_1(\omega) - \mu_1 \\ \log D_2(\omega) - \mu_2 \\ \log D_3(\omega) - \mu_3 \end{pmatrix} \right\}
\end{aligned}
\tag{38}
$$

where μ_1, μ_2, μ_3 are the means of $\log D_1(\omega), \log D_2(\omega), \log D_3(\omega)$ as a function of the parameter vector θ, while Σ is the variance covariance matrix of these three data series in terms of the theoretical parameters.

Step 2: Adjustment for noisy truncation

The key selection variable is defined in the trivariate case, similar to the bivariate case from section 2:

$$
\log Z(\omega) = \log A(\omega) - \bar{m} \cdot \left[\alpha_s \log\big(1 + \tau_{K,s}(\omega)\big) + \log\left(\frac{1}{1 - \tau_{Y,s}(\omega)}\right)\right] \tag{39}
$$

Now account for truncation noise, I add a zero mean iid error term that is also assumed to be log-normal

$$
\log \hat{Z}(\omega) = \log Z(\omega) + e(\omega)
$$

with $e(\omega) \sim N(0, \sigma_e)$. The selection equation with this truncation noise can therefore be written as

$$
\log \hat{Z}(\omega) \geq \log \bar{Z}_J
$$

with the selection cutoff defined by

$$
\begin{aligned}
\log \bar{Z}_J &= \log \bar{m} - \log P_s \\
&\quad + \bar{m}\left[\alpha_s \log\left(\frac{R}{\alpha_s}\right) + (1 - \alpha_s)\log\left(\frac{w}{1 - \alpha_s}\right)\right] - \left(\frac{1}{\eta - 1}\right)\log\left(\frac{P_s Y_s}{\eta F_s}\right)
\end{aligned}
$$

The probability of survival then can consequently be written as

$$\begin{aligned} & Pr\left(\Pi_s(\omega) \geq f_s(\omega)\right) \\ & = Pr\left(\log \hat{Z}(\omega) \geq \log \bar{Z}_J\right) \\ & = 1 - \Phi\left(\bar{z}_J(\theta, \log P_s)\right) \end{aligned}$$

Hence, the log likelihood should be adjusted by adding the term $\log\left[1 - \Phi\left(\bar{z}_J\right)\right]$ to adjust for truncation. Furthermore the estimation problem will also maximize over the parameter σ_e to account for noise in the truncation threshold.

Step 3: Derivation of the Equilibrium Constraint

The equilibrium constraint is basically the equilibrium CES price index, which in the full model can be written as

$$P_s = \bar{m}\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s} \bar{A}_s^{-1} J_s^{-\frac{1}{\eta-1}} \tag{40}$$

where

$$\bar{A}_s = \left(\int_{\Pi_s(\omega) \geq f_s(\omega)} \left[\frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)}\right]^{1-\eta} A_s(\omega)^{\eta-1} \frac{d\mu_s(\omega)}{\mu_s\left(\Pi_s(\omega) \geq f_s(\omega)\right)}\right)^{\frac{1}{\eta-1}}$$

The key integral to evaluate this expression is

$$\int_{\Pi_s(\omega) \geq f_s(\omega)} \left[\frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)}\right]^{1-\eta} A_s(\omega)^{\eta-1} d\mu_s(\omega) \tag{41}$$

In order to avoid MC-integration and be able to use derivative based optimization methods, I exploit the following result that will generate a smooth and differentiable form for integral (41).

Lemma 1 (Lien and Balakrishnan (2006))

Let X and Z be two jointly log-normally distributed random variables. Define the multiplicative constraint by the set

$$1_{\{a,b,K\}} = \begin{cases} 1 \text{ if } X^a \cdot Z^b \leq K \\ 0 \text{ if else} \end{cases} \tag{42}$$

Then it follows that

$$\begin{aligned}
&\mathbb{E}\left[X^m Z^n \cdot 1_{\{a,b,K\}}\right] = \\
&\exp\left\{m\mu_X + n\mu_Z + \frac{1}{2}\left(m^2\sigma_m^2 + n^2\sigma_n^2 + 2mn\sigma_{X,Z}\right)\right\} \\
&\times\Phi\left(\frac{\log K - (a\mu_X + b\mu_Z) - [am\sigma_X^2 + (bm+an)\sigma_{X,Z} + bn\sigma_Z^2]}{\sqrt{a^2\sigma_X^2 + b^2\sigma_Z^2 + 2ab\sigma_{X,Z}}}\right)
\end{aligned} \tag{43}$$

where $\Phi(.)$ is the cdf of a standard normal.

Lemma 2

Let X_1, X_2, X_3 be three jointly log-normally distributed random variables. Define the multiplicative constraint by the set

$$1_{\{\alpha,\beta,\gamma,K\}} = \begin{cases} 1 & \text{if } X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \leq K \\ 0 & \text{if else} \end{cases} \tag{44}$$

Then it follows that

$$\begin{aligned}
&\mathbb{E}\left[X_1^m X_2^n X_3^l \cdot 1_{\{\beta_1,\beta_2,\beta_3,K\}}\right] = \mathbb{E}\left[X \cdot Z^c \cdot 1_{\{0,0,1,K\}}\right] \\
&= \exp\left\{\mu_X + c\mu_Z + \frac{1}{2}\left(\sigma_X^2 + c^2\sigma_Z^2 + c\sigma_{X,Z}\right)\right\} \\
&\times \Phi\left(\frac{\log K - \mu_Z - [\sigma_{X,Z} + c\sigma_Z^2]}{\sigma_Z}\right)
\end{aligned} \tag{45}$$

where $\Phi(.)$ is the cdf of a standard normal and X and Z are defined by

$$\begin{aligned}
\log X &= a \log X_1 + b \log X_2 \\
\log Z &= \beta_1 \log X_1 + \beta_2 \log X_2 + \beta_3 \log X_3
\end{aligned} \tag{46}$$

and the coefficients a, b, c are given by

$$a = m - \beta_1 \frac{l}{\beta_3}, b = n - \beta_2 \frac{l}{\beta_3}, c = \frac{l}{\beta_3} \tag{47}$$

Proof: apply mapping (46) and (47) to reduce the trivariate problem to the bivariate problem of Lemma 1.

Lemma 2 can now directly be used to evaluate integral (41). For this note that we can map this into the form required by the Lemma, by noting that

$$\log X = a \log(1 + \tau_{K,s}(\omega)) + b \log \left(\frac{1}{1 - \tau_{Y,s}(\omega)} \right)$$

$$\log \tilde{Z} = -\log Z = \alpha_s \bar{m} \log(1 + \tau_{K,s}(\omega)) + \bar{m} \log \left(\frac{1}{1 - \tau_{Y,s}(\omega)} \right) - \log A_s(\omega)$$

with

$$a = \alpha_s, b = 1, c = -(\eta - 1) \tag{48}$$

Evaluating the integral (41) and substituting into the price level equation (40) implies the constraint displayed in the main text.

1.C.5: Recovering aggregate factor endowments

This section shows how endogenous entry and equilibrium conditions are used together to recover sectoral factor endowments. I derive aggregate sectoral factor demands and show how these are related to estimated parameters.

Production-related aggregate factor demands are given by

$$\begin{aligned} R \int_{\Pi_s(\omega) \geq f_s(\omega)} K_{P,s}(\omega) d\mu_s(\omega) &= \frac{\alpha_s}{\bar{m}} \left(\frac{1 - \bar{\tau}_{Y,s}}{1 + \bar{\tau}_{K,s}} \right) P_s Y_s \\ w \int_{\Pi_s(\omega) \geq f_s(\omega)} L_{P,s}(\omega) d\mu_s(\omega) &= \frac{1 - \alpha_s}{\bar{m}} \left(1 - \bar{\tau}_{Y,s} \right) P_s Y_s \end{aligned} \tag{49}$$

where the aggregate wedges are given by

$$(1 - \bar{\tau}_{Y,s}) = \frac{\int_{\Pi_s(\omega) \geq f_s(\omega)} A_s(\omega)^{\eta-1} \cdot [1 - \tau_{Y,s}(\omega)]^{\eta} \cdot [1 + \tau_{K,s}(\omega)]^{-\alpha_s(\eta-1)} d\mu_s(\omega)}{\int_{\Pi_s(\omega) \geq f_s(\omega)} \left[A_s(\omega) \frac{1 - \tau_{Y,s}(\omega)}{[1 + \tau_{K,s}(\omega)]^{\alpha_s}} \right]^{\eta-1} d\mu_s(\omega)} \tag{50}$$

$$(1 + \bar{\tau}_{K,s}) = \frac{\int_{\Pi_s(\omega) \geq f_s(\omega)} A_s(\omega)^{\eta-1} \cdot [1 - \tau_{Y,s}(\omega)]^{\eta} \cdot [1 + \tau_{K,s}(\omega)]^{-\alpha_s(\eta-1)} d\mu_s(\omega)}{\int_{\Pi_s(\omega) \geq f_s(\omega)} A_s(\omega)^{\eta-1} \cdot [1 - \tau_{Y,s}(\omega)]^{\eta} \cdot [1 + \tau_{K,s}(\omega)]^{-(1+\alpha_s(\eta-1))} d\mu_s(\omega)} \tag{51}$$

Plant-level factor demand for fixed costs of operations is given by

$$
\begin{aligned}
wL_{f,s}(\omega) &= (1-\alpha_s)\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\left(\frac{[1+\tau_{K,s}(\omega)]R}{\alpha_s}\right)^{\alpha_s}F_s \\
RK_{f,s}(\omega) &= \alpha_s\frac{1}{1+\tau_{K,s}(\omega)}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\left(\frac{[1+\tau_{K,s}(\omega)]R}{\alpha_s}\right)^{\alpha_s}F_s
\end{aligned} \tag{52}
$$

Aggregation of these factor demands across establishments implies

$$
\begin{aligned}
wL_{F,s} &= (1-\alpha_s)J_s\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\frac{\bar{T}_{\kappa 1,s}}{\mu_s\left(\Pi_s(\omega)\geq f_s(\omega)\right)}F_s \\
RK_{F,s} &= \alpha_s J_s\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\frac{\bar{T}_{\kappa 2,s}}{\mu_s\left(\Pi_s(\omega)\geq f_s(\omega)\right)}F_s
\end{aligned} \tag{53}
$$

where aggregate wedges related to fixed costs of operation are given by

$$
\begin{aligned}
\bar{T}_{\kappa 1,s} &= \int_{\Pi_s(\omega)\geq f_s(\omega)}[1+\tau_{K,s}(\omega)]^{\alpha_s}d\mu_s(\omega) \\
\bar{T}_{\kappa 2,s} &= \int_{\Pi_s(\omega)\geq f_s(\omega)}[1+\tau_{K,s}(\omega)]^{\alpha_s-1}d\mu_s(\omega)
\end{aligned} \tag{54}
$$

Note that there is a simple relationship between aggregate capital and labor used for fixed costs of operation.

$$
RK_{F,s} = \frac{\alpha_s}{1-\alpha_s}\bar{T}_{\kappa\alpha,s}wL_{F,s} \tag{55}
$$

where the aggregate distortions between capital and labor for fixed costs of operation can be written as

$$
\bar{T}_{\kappa\alpha,s} = \left(\frac{\bar{T}_{\kappa 2,s}}{\bar{T}_{\kappa 1,s}}\right) \tag{56}
$$

This term summarizes the direct effect of capital wedge distortions of the exit margin on aggregate relative factor demand.

Finally I need to recover the resources spent on entry. The net present value of a firm in Steady State is given by

$$
V_{e,s} = \frac{1}{\delta_s}Pr\Big\{\Pi_s(\omega)\geq f_s(\omega)\Big\}\mathbb{E}\left[\Pi_s(\omega)-f_s(\omega)\middle|\Pi_s(\omega)\geq f_s(\omega)\right] \tag{57}
$$

Aggregate valuations of entrants are

$$
\begin{aligned}
J_{e,s}V_{e,s} &= \frac{1-\bar{\tau}_{Y,s}}{\eta} P_s Y_s - J_s \mathbb{E}\left[f_s(\omega) \middle| \Pi_s(\omega) \geq f_s(\omega) \right] \\
&= \frac{1-\bar{\tau}_{Y,s}}{\eta} P_s Y_s - J_s \left(\frac{R}{\alpha_s}\right)^{\alpha_s} \left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s} \frac{\bar{T}_{\kappa 1,s}}{\mu_s(\Pi_s(\omega)) \geq f_s(\omega)}
\end{aligned}
\tag{58}
$$

The free entry condition then implies:

$$
V_{e,s} = \left(\frac{R}{\alpha_s}\right)^{\alpha_s} \left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s} F_{e,s} \tag{59}
$$

which is used to calculate the factor demand by entrants. Therefore aggregate labor demand for the purpose of sunk costs of entry is:

$$
\begin{aligned}
wL_{e,s} &= (1-\alpha_s) \left(\frac{R}{\alpha_s}\right)^{\alpha_s} \left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s} F_{e,s} J_{e,s} \\
&= (1-\alpha_s) J_{e,s} V_{e,s}
\end{aligned}
\tag{60}
$$

which recovers resources spent on entry cost labor. Similarily, capital demand to cover sunk costs of entry is given by:

$$
RK_{e,s} = \frac{\alpha_s}{1-\alpha_s} wL_{e,s} \tag{61}
$$

The aggregate factor demands for each sector are therefore given by

$$
\begin{aligned}
wL_s &= wL_{P,s} + wL_{F,s} + wL_{e,s} \\
RK_s &= RK_{P,s} + RK_{F,s} + RK_{e,s}
\end{aligned}
\tag{62}
$$

Aggregate factor endowments are then calculated as $L = \sum_{s=1}^{S} L_s$ and $K = \sum_{s=1}^{S} K_s$.

1.C.6: Multi-Sector General Equilibrium

This section defines the multi-sector general equilibrium. In the counterfactuals I assume that aggregate factor endowments L, K are fixed.

A multi-sector equilibrium in the economy with micro-distortions consists of a set of variables for each sector $\left\{P_s, l_{e,s}, (1-\bar{\tau}_{Y,s}), (1+\bar{\tau}_{K,s}), \bar{T}_{\kappa\alpha,s}, P_s Y_s, l_s\right\}_{s=1}^{S}$ and the economy wide relative factor price $\frac{R}{w}$ and the numeraire $w = 1$ such that

1. The sectoral ideal CES price index is given by (40) for each sector $s = 1, ..., S$.
2. Steady state turnover of firms is given by (1.23) for each sector $s = 1, ..., S$.
3. The sectoral aggregate output wedge is given by (50) for each sector $s = 1, ..., S$.
4. The sectoral aggregate capital wedge is given by (51) for each sector $s = 1, ..., S$.
5. The sectoral exit capital wedge is given by (56) for each sector $s = 1, ..., S$.
6. Sectoral share of labor in total employment is given by

$$l_s = \frac{(1-\alpha_s)(1-\bar{\tau}_{Y,s})\xi_s}{\sum_{s=1}^{S}(1-\alpha_s)(1-\bar{\tau}_{Y,s})\xi_s} \tag{63}$$

for each sector $s = 1, ..., S$.

7. Sectoral spending is given by

$$P_s Y_s = \frac{\xi_s w L}{\sum_{s=1}^{S}(1-\alpha_s)(1-\bar{\tau}_{Y,s})\xi_s} \tag{64}$$

for each sector $s = 1, ..., S$.

8. Economy-wide factor markets clear

$$\frac{R}{w} = \frac{\sum_{s=1}^{S}(1-\bar{\tau}_{Y,s})\xi_s\alpha_s\left\{\frac{1}{\bar{m}}\left(\frac{1}{1+\bar{\tau}_{K,s}}\right)+\bar{T}_{\kappa\alpha,s}(1-l_{e,s})+l_{e,s}\right\}}{\sum_{s=1}^{S}(1-\alpha_s)(1-\bar{\tau}_{Y,s})\xi_s} \tag{65}$$

where aggregate endowments L, K are given. It is a system of $S \times 7 + 1$ variables in as many equations.

1.C.6: Proof of Proposition 5

Here I derive the aggregate TFP formula. Combine both aggregate factor demand for production in equation (49) to get

$$P_s Y_s = \bar{m}\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\left[\frac{(1+\bar{\tau}_{K,s})^{\alpha_s}}{1-\bar{\tau}_{Y,s}}\right]K_{P,s}^{\alpha_s}L_{P,s}^{1-\alpha_s} \tag{66}$$

Now divide both sides by $P_s K_s^{\alpha_s} L_s^{1-\alpha_s}$ and substitute for P_s with equation (40). What is left is to substitute $\frac{K_{P,s}}{K_s}$ as well as $\frac{L_{P,s}}{L_s}$. For this, combine (49) and (62).

Appendix 1.D: Endogenous Markups

The main part of the paper follows Hsieh and Klenow (2009) in modeling plant-level distortions as taxes as opposed to markups. But markup differences potentially show up in TFPR dispersion. To quantify the degree in which endogenous markups could drive results I follow Atkenson and Burstein (2008). They point out that if one drops the assumption that monopolistic competitors within sectors take the CES price index as given, it is possible to generate endogenous markups with CES preferences. Atkenson and Burstein (2008) show that the pricing decision of firms can be written as

$$p_s(\omega) = \frac{\varepsilon_s(\omega)}{\varepsilon_s(\omega) - 1} \cdot mc_s(\omega)$$

where $mc_s(\omega) = \frac{1}{A(\omega)} \cdot B_s R^{\alpha_s} w^{1-\alpha_s}$ is the marginal cost of a firm and the elasticity of substitution is given by

$$\varepsilon_s(\omega) = \left[\frac{1}{\eta}(1 - s_s(\omega)) + s_s(\omega)\right]^{-1}$$

where $s_s(\omega)$ is the market share of firm ω in its 4-digit sector s. I keep the assumptions that the elasticity of substitution across sectors is one and the within-sector elasticity of substitution is 3. Note that these choices can generate large markup differences. At one extreme, as a firm dominates its industry and $s_s(\omega) \to 1$ the markup will diverge toward infinity. Intuitively as one firm dominates an industry, the relevant elasticity of substitution will be the cross-sectoral elasticity of substitution of one. The monopolist realizes that his product is essential and charges an infinite markup. On the other extreme, if the market share in sector s $s_s(\omega) \to 0$ the within sectoral elasticity of substitution η is relevant. So the lower bound for the markup is 1.5.

With these assumptions in place it is possible to quantitatively evaluate to which degree endogenous markup differences can drive TFPR dispersion in the data. Suppose endogenous markups would be the only source of micro-distortions. Then measured TFPR in a model without fixed costs is

$$TFPR_s(\omega) = \frac{p_s(\omega) y_s(\omega)}{K_s(\omega)^{\alpha_s} L_s(\omega)^{1-\alpha_s}} = \frac{\varepsilon_s(\omega)}{\varepsilon_s(\omega) - 1} \cdot B_s R^{\alpha_s} w^{1-\alpha_s}$$

For each 4 digit sector I calculate the market shares $s_s(\omega)$ and generate markups according to the Atkenson-Burstein formula. Additionally I calculate Herfindahl-Hirshman concentration indices to contrast markup dispersion with industry concentration. Figure 17 shows the results of this exercise. As one would expect endogenous markups would lead markup dispersions to increase the more concentrated the industry is. However, endogenous markup dispersion is not able to generate any quantitatively significant amount of TFPR dispersion. The maximum dispersion of TFPR generated by endogenous markups is below 0.1 and most sectors are well below 0.05. This point reinforced by figure 18, which plots the observed TFPR dispersion against the log Herfindahl index, by 4 digit sector. Note that the actual TFPR dispersion is typically an order of magnitude larger than suggested by the Aktenson-Burstein calculations. There is indeed a positive relation between industry concentration and TFPR dispersion in the data, but it seems very weak at best.

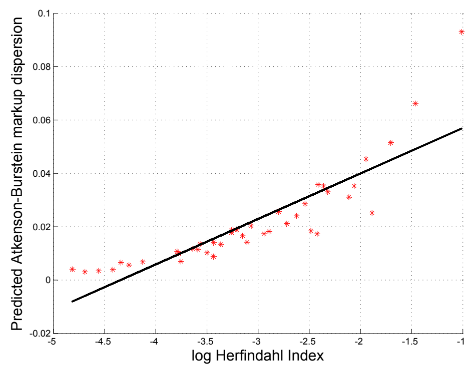

Figure 17: Standard deviation of log-markups predicted by Atkenson-Burstein framework versus log Herfindahl indices. Each data point is one ISIC Rev.2, 4 digit sector.

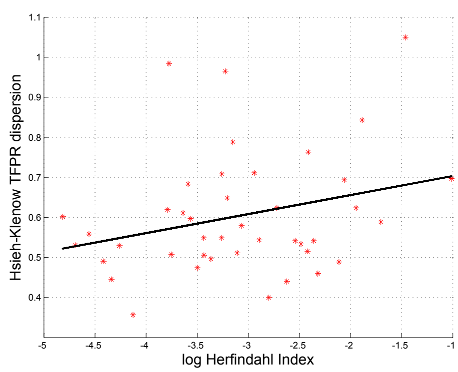

Figure 18: Standard deviation of log TFPR vs log Herfindahl indices. Each data point is one ISIC Rev.2, 4 digit sector.

Appendix 1.E: Heterogeneity in Fixed Costs

In this appendix I analyze how the presence of heterogeneity in fixed costs across firms could impact the measurement of TFPR and TFPQ. I will proceed in three steps. First, I derive bounds on the maximum possible differences in measured TFPR with heterogeneity in fixed costs. Second I quantitatively illustrate how much TFPR dispersion can be generated in a model that only has firm heterogeneity in efficiency and fixed costs. Third, I re-introduce heterogeneity in micro-distortions and analyze the quantitative impact of fixed cost heterogeneity on TFPR measurement.

The model of firm level heterogeneity in productivity is as follows. As in the main text, monopolistic competitors solve

$$\begin{aligned} \max_{L(\omega)} \Pi(\omega) &= p(\omega)y(\omega) - wL(\omega) \\ \text{s.th. } \; y(\omega) &= A(\omega)L(\omega) \\ \text{and } \; p(\omega) &= \left[\frac{y(\omega)}{Y}\right]^{-\frac{1}{\eta}} P \end{aligned}$$

Optimal exit decisions are given by

$$\begin{aligned} \Pi(\omega) &\geq w \cdot F(\omega) \\ F(\omega) &= F \cdot \exp\{\sigma_F \cdot \varepsilon(\omega)\} \\ &\text{with } \; \varepsilon(\omega) \; \text{Standard-Normal} \end{aligned}$$

Then rewriting this optimal survival condition implies that

$$\frac{1}{\eta} p(\omega)y(\omega) \geq wF(\omega)$$

This imposes a bound on maximum fixed costs: $wF_{max} = \frac{1}{\eta}p(\omega)y(\omega)$. The intuition for this result is straightforward. If fixed costs are too large relative to profits, firms optimally decide to exit. Firms with very large fixed costs will therefore not be observed in the data. Measured TFPR with this maximum possible fixed cost is:

$$TFPR_1(\omega) = \frac{p(\omega)y(\omega)}{w(L_p(\omega) + F_{max})} = \frac{p(\omega)y(\omega)}{\frac{p(\omega)y(\omega)}{\bar{m}} + \frac{p(\omega)y(\omega)}{\eta}} = 1$$

On the other hand TFPR with minimum possible fixed costs is:

$$TFPR_0(\omega) = \frac{p(\omega)y(\omega)}{wL_p(\omega) + 0} = \frac{p(\omega)y(\omega)}{\frac{p(\omega)y(\omega)}{\bar{m}}} = \bar{m}$$

Hence the maximum possible variation in measured TFPR from fixed costs is between 1 and $\frac{\eta}{\eta-1}$ or 1.5 with $\eta = 3$. Now to quantify how strong these bounds are, take $\sigma_F = 4$. This implies that the 95-5 percentile ratio is $8.8861 \cdot 10^5$! Furthermore let me assume that $\sigma_A = 1.4$ which is comparable with the mean TFPQ dispersion I estimate from the data. The observable standard deviation of log TFPR is 0.12, while the covariance of TFPR and TFPQ is 0.04. A model with only fixed cost heterogeneity does not seem to be able to generate TFPR dispersion quantitatively comparable to the data. What is more, Figure 19 shows that the correlation patterns of TFPR and TFPQ implied by this type of model are very different from the ones in the data.

How strongly does heterogeneity in fixed cost affect the measurement of TFPR and TFPQ? To investigate this question I set $\sigma_F = 1$, $\sigma_A = 1.4$ and $\sigma_\tau = 0.9$. Figure 20 shows that this introduces a lot of noise around the survival line. However, the impact on measured TFPR and TFPQ is limited. The standard deviation of log TFPR ideally measured is 0.71, the covariance of TFPR and TFPQ is 0.59 and the standard deviation of log TFPQ is 1.08. In contrast measures of TFPR and TFPQ ignoring fixed cost heterogeneity are: 0.69 for the standard deviation of log TFPR, 0.62 for the covariance of TFPR and TFPQ and 1.14 for the standard deviation of log TFPQ.

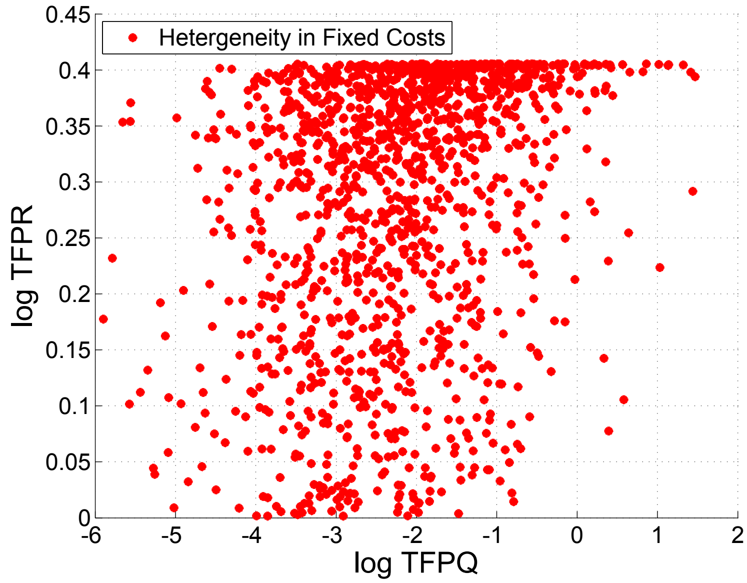

Figure 19: Measured TFPR and TFPQ with $\sigma_A = 1.4$ and $\sigma_F = 4$

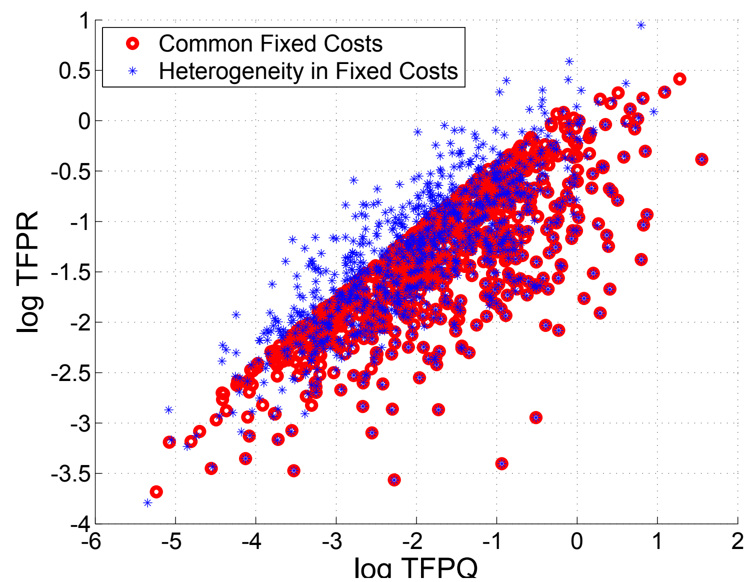

Figure 20: Measured TFPR and TFPQ with $\sigma_A = 1.4$, $\sigma_\tau = 0.9$ and $\sigma_F = 1$.

Appendix Chapter 2

Appendix 2.A: Impact of International Trade on Distortion and Efficiency Measures

This appendix quantitatively illustrates the impact of international trade on the measurement of micro level frictions and efficiency in a multi-country general equilibrium model. There are two reasons why one might be interested in the impact of international trade on measurement of frictions. First, as pointed out by Melitz (2003), changes in international trade induce re-allocation effects across firms. As trade frictions fall, the most productive firms should be expected to expand, while the least productive firms should contract. This impacts measured dispersions of firm sizes. Second, if the export decision involves a fixed cost of operation then very productive firms will have higher fixed cost of operations. These will show up in measured factor demands. Since the measurement of firm level distortions and efficiency rely on sales and factor demand, variations in aggregate trade frictions can impact the distortions measurement conducted in this paper.

Let there be N countries with $i = 1, ...N$ indexing the source country and $j = 1, .., N$ the destination country. Each country's aggregate production is given by

$$\min_{y_{ij}(\omega)} P_i Y_i - \sum_{j=1}^{N} \int p_{ij}(\omega) y_{ij}(\omega) \cdot d\mu_j(\omega)$$

$$\text{subject to: } Y_i = \left(\sum_{j=1}^{N} \int y_{ij}(\omega)^{\frac{\eta-1}{\eta}} \cdot d\mu_j(\omega) \right)^{\frac{\eta}{\eta-1}}$$

Within each country, there exists a continuum of monopolistically competitive firms. These firms differ by their level of technology and micro-distortions as in the main part of the paper.

$$\max_{\{L_{ij}(\omega)\}} \sum_{i=1}^{N} \Pi_{ij}(\omega) = \sum_{i=1}^{N} [(1-\tau(\omega)) p_{ij}(\omega) y_{ij}(\omega) - w_j L_{ij}(\omega)]$$

$$\text{subject to: } y_{ij}(\omega) = \frac{1}{d_{ij}} A(\omega) \cdot L_{ij}(\omega)$$

The joint distribution of firm level distortions and efficiency are allowed to differ by country and assumed to be log-normal. Producers shipping goods from country j to country i are assumed to face an iceberg trade frictions $d_{ij} > 1$. A firm from country j decides to operate in market i if gross profits are sufficient to cover fixed costs of operating in this market $f_{ij} = w_i \cdot F_{ij}$:

$$\Pi_{ij}(\omega) \geq f_{ij}$$

In typical establishment level micro-data, value added and factor demand are not differentiated by destination:

$$X^j(\omega) = \sum_{i=1}^{N} p_{ij}(\omega) y_{ij}(\omega) \cdot 1_{\{\Pi_{ij}(\omega) \geq f_{ij}\}}$$

$$L^j(\omega) = \sum_{i=1}^{N} [L_{ij}(\omega) + F_{ij}] \cdot 1_{\{\Pi_{ij}(\omega) \geq f_{ij}\}}$$

Exporters will have higher sales as they enter more markets, but also have higher fixed costs of operation.

Parameter	Value	Explanation
η	3	Elasticity of Substitution Hsieh and Klenow (2009)
F_{ii}	1	Domestic Fixed Cost Workers
F_{ij}	18	Export Fixed Cost Workers
F_e	0	Entry Cost Workers HMR (2008), Melitz-Ghironi (2008)
$\tilde{J}$	20000	Number of Latent Plants
d_{ij}	3	Iceberg trade friction Anderson and van Wincoop(2004)
$\mu_{A,j}$	[0.01, 1]	mean country-level efficiency
$\sigma_{A,j}$	1.2	dispersion of firm efficiencies
$\sigma_{\tau,j}$	1	dispersion of micro-distortions

I calibrate this model with ten countries, which I assume differ by their level of efficiency. Aggregate efficiency differences are assumed to be a factor of 2.5. Other parameters are explained in the table. Iceberg trading frictions are assumed to add a factor of three to prices, which is consistent with evidence from Anderson and Van Wincoop (2004). To quantify the impact of trade on distortion measurement I proceed as follows. For a specific country with an average technology level across countries I vary the trade friction. Specifically I start at 60% below the average trade friction and increase iceberg costs to 160% of the average trade friction. This has a dramatic impact on trade. I follow the trade literature and characterize the degree of openness to trade using the internal trade share. It is defined as the value of domestic nominal sales to all international sales to the country. Recently Arkolakis, Costinot and Rodriguez-Clare (Forthcoming) have argued that this statistic is key in a class of constant-elasticity trade models to evaluate the welfare gains from trade. The internal trade share in this exercise starts at 25% and increases to a value of 80%. Thus, the country starts as a relatively open economy and converges to a relatively closed one as I increase iceberg trading frictions.

As figures 21 and 22 document, the impact on measures of TFPR and TFPQ dispersion is modest. As one would expect, lower trade frictions increase the dispersion of TFPQ due to Melitz-type reallocation effects. This is compatible with the conjecture of Hsieh and Klenow (2009) that international trade would primarily show up in TFPQ. One notable feature is that measured TFPR dispersions fall as there is more trade.

The results of this exercise are not much changed if instead of a middle income country I choose the same calibration with a low income or a high income country. In each case measured TFPR and TFPQ dispersions are affected by international trade. But the quantitative impact is never more than 0.1 log points in the standard deviation of either log TFPR or log TFPQ.

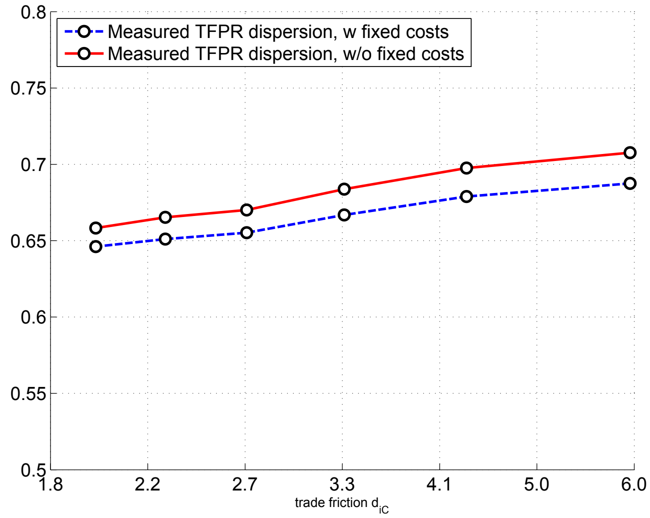

Figure 21: Impact of changes in trade friction on measured TFPR dispersion. Measured TFPR dispersion is defined as standard deviation of log TFPR. Plot is based on a ten country trade model with trade frictions and differences in mean productivity across countries.

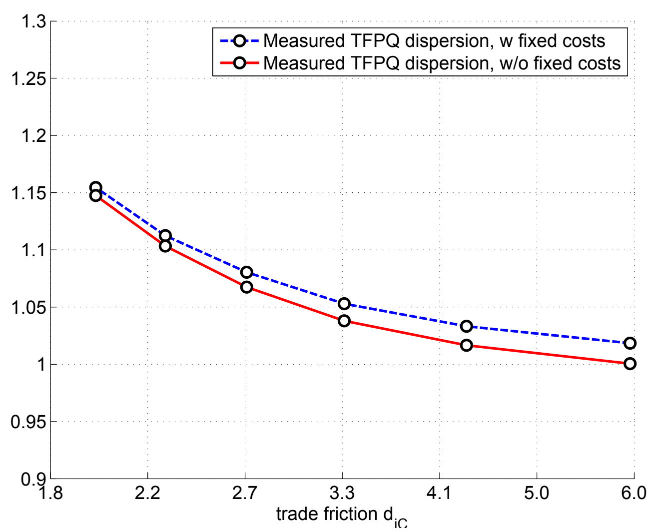

Figure 22: Impact of changes in trade friction on measured TFPQ dispersion. Measured TFPQ dispersion is defined as standard deviation of log TFPQ. Plot is based on a ten country trade model with trade frictions and differences in mean productivity across countries.

Appendix 2.B

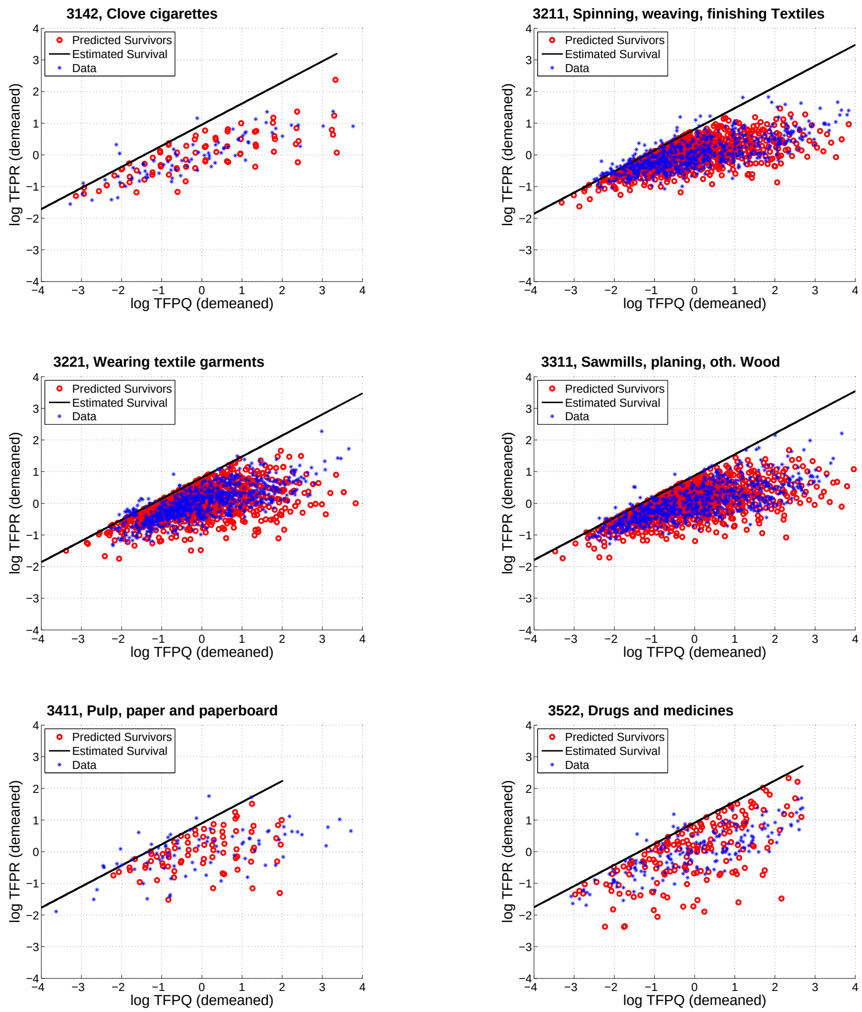

Figure 23: Illustration of estimated survival pattern versus data for six largest manufacturing sectors.

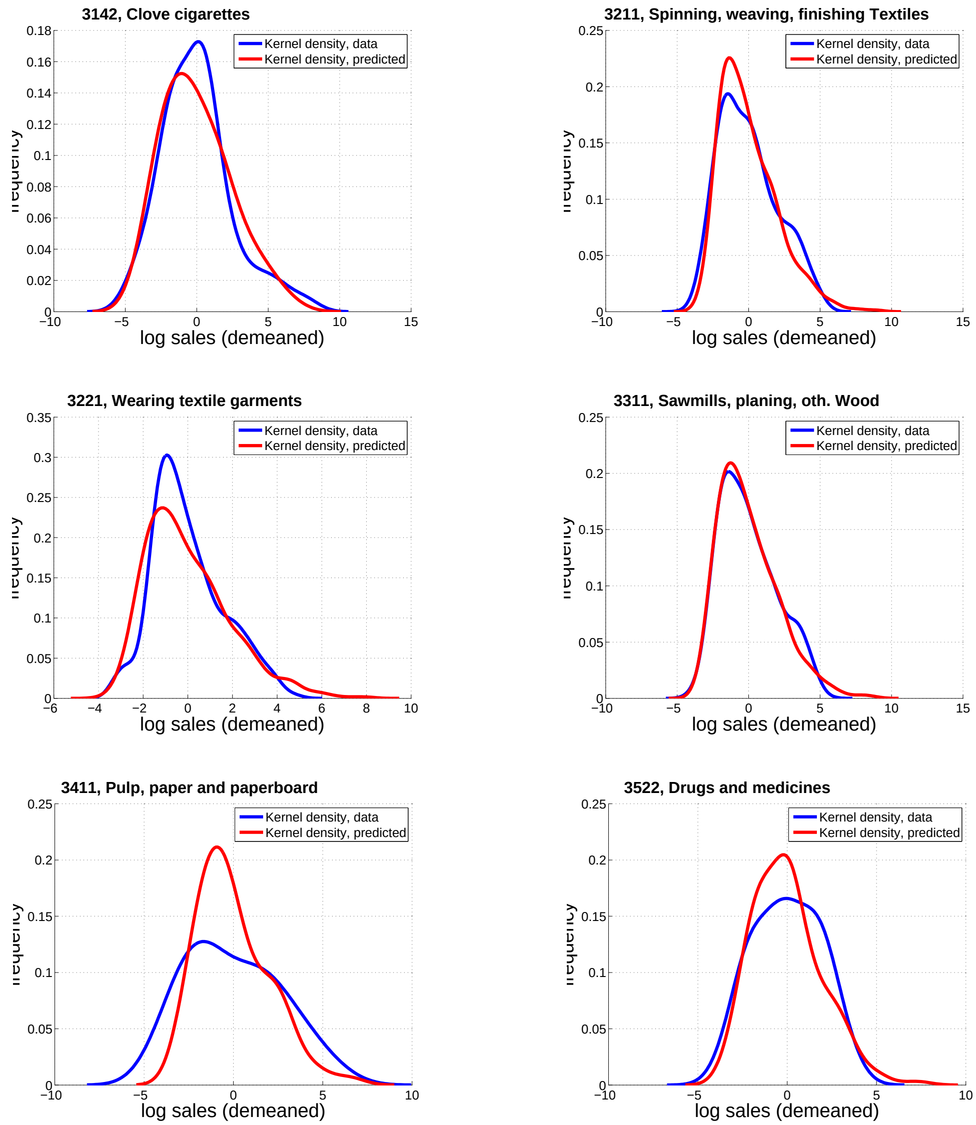

Figure 24: Illustration of predicted firm size distribution versus data. Six largest manufactuing sectors by value added. Kernel density estimates.

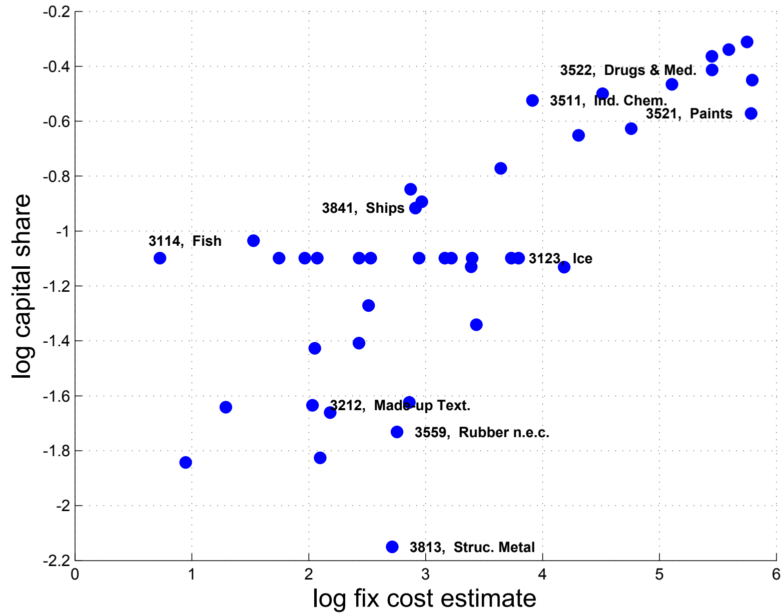

Figure 25: Display of log sectoral capital intensity $\log(\alpha_s)$ from US data versus estimate of log fixed costs $\log(F_s)$. Each data point is one ISIC Rev.2, 4 digit sector.

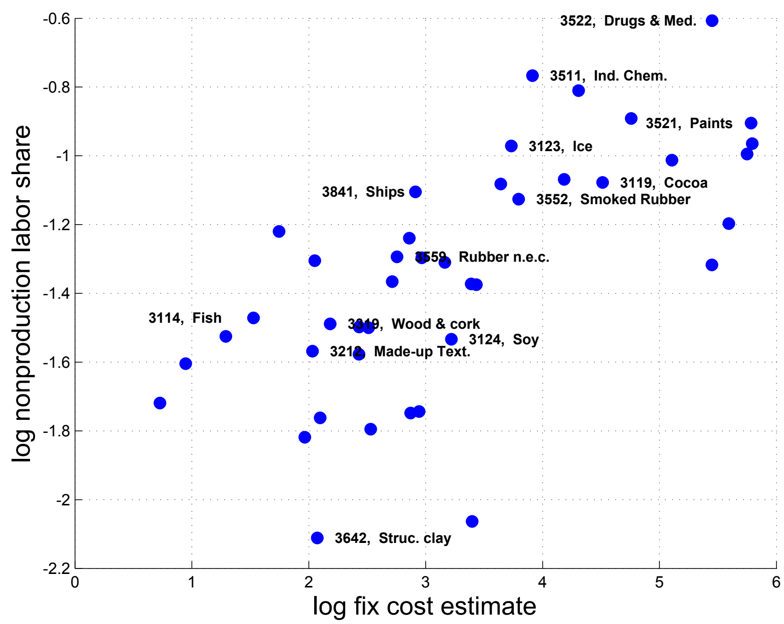

Figure 26: Display of log non-production worker share $\log(\gamma_s)$ from Indonesian data versus estimate of log fixed costs $\log(F_s)$. Each data point is one ISIC Rev.2, 4 digit sector.

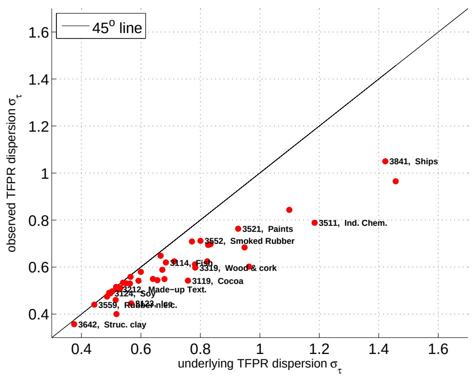

Figure 27: Standard deviation of `log TFPR` $= \log\left[\frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)}\right]$ (combined firm level wedge) against selection corrected estimates from MLE.

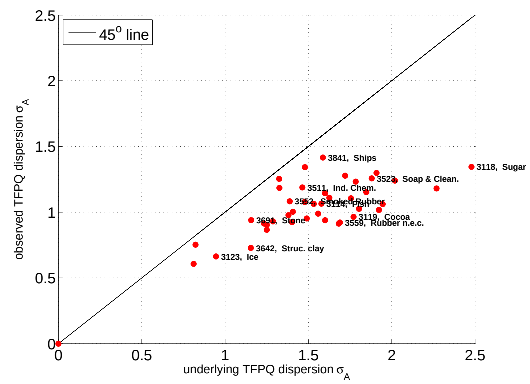

Figure 28: Standard deviation of `log TFPQ` $= \log A_s(\omega)$ (firm level efficiency) against selection corrected estimates from MLE.

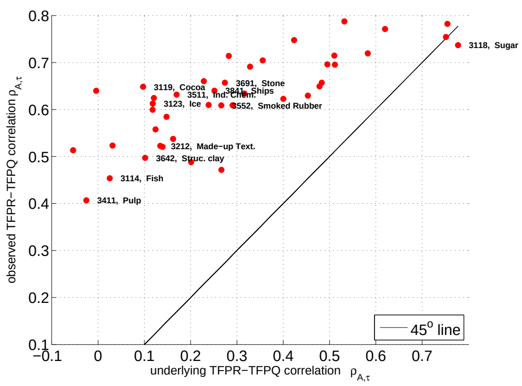

Figure 29: Correlation of `log TFPR` $= \log\left[\frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)}\right]$ and `log TFPQ` $= \log A_s(\omega)$ against selection corrected estimates from MLE.

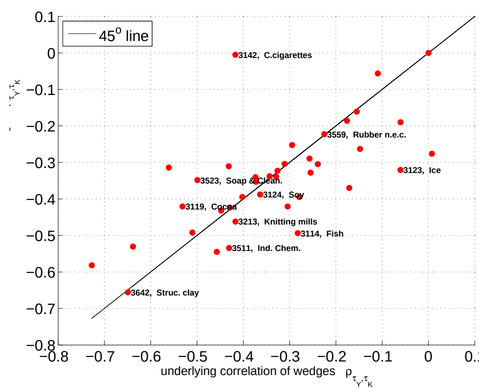

Figure 30: Correlation of output wedge $\log\left[\frac{1}{1-\tau_{Y,s}(\omega)}\right]$ and capital wedge $\log[1+\tau_{K,s}(\omega)]$ against selection corrected estimates from MLE.

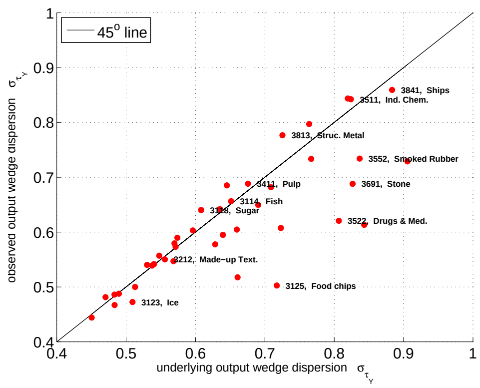

Figure 31: Standard deviation of $\log\left[\frac{1}{1-\tau_{Y,s}(\omega)}\right]$ (output wedges) against selection corrected estimates from MLE.

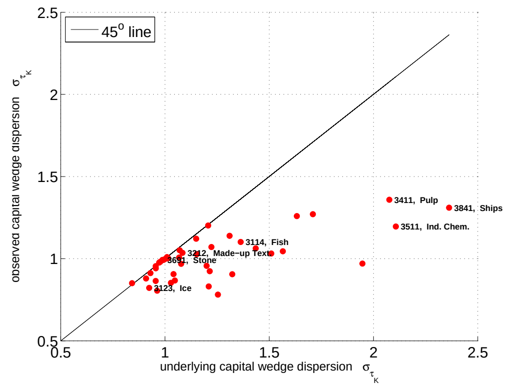

Figure 32: Standard deviation of $\log[1+\tau_{K,s}(\omega)]$ (capital wedges) against selection corrected estimates from MLE.

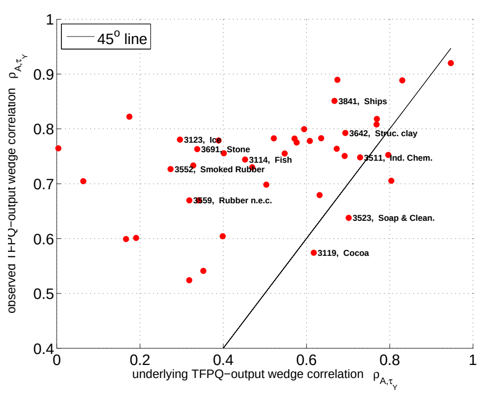

Figure 33: Correlation of $\log\left[\frac{1}{1-\tau_{Y,s}(\omega)}\right]$ and `log TFPQ` $= \log A_s(\omega)$ against selection corrected estimates from MLE.

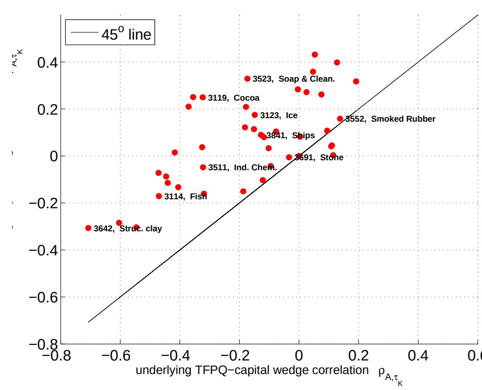

Figure 34: Correlation of capital wedge $\log[1+\tau_{K,s}(\omega)]$ nd `log TFPQ` $= \log A_s(\omega)$ against selection corrected estimates from MLE.

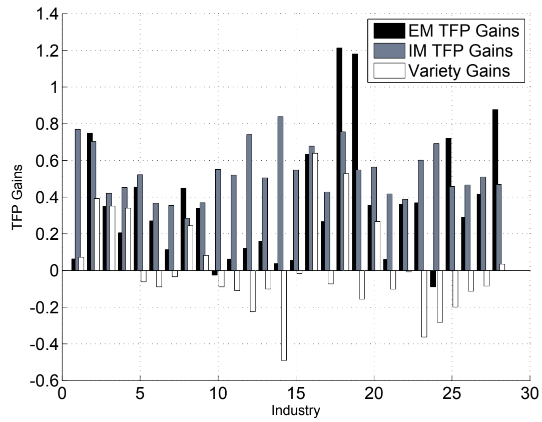

Figure 35: Distribution of TFP Gains across sectors as misallocation is removed.

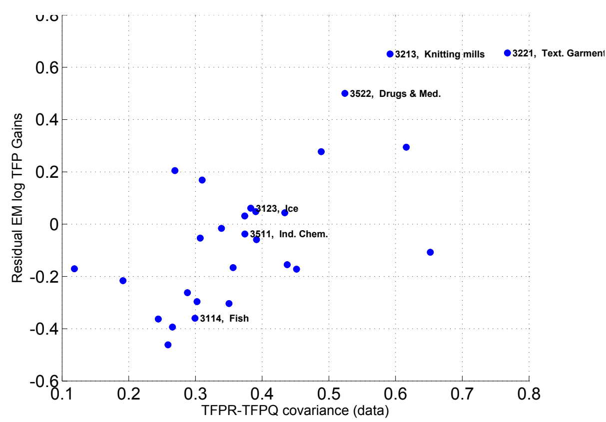

Figure 36: Entensive margin misallocation losses and covariance of efficiency and distortions across sectors. Each data point is a 4 digit sector.

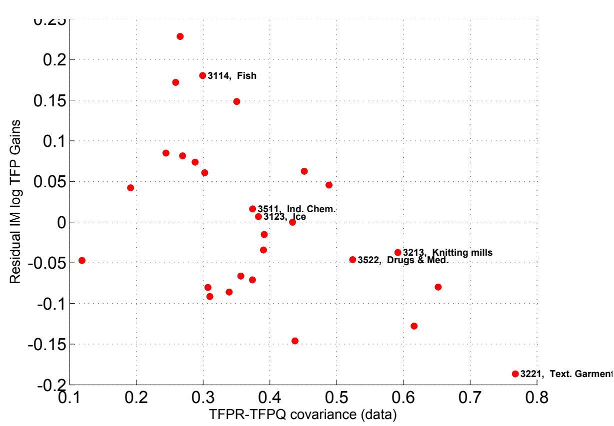

Figure 37: Intensive margin misallocation losses and covariance of efficiency and distortions across sectors. Each data point is a 4 digit sector.

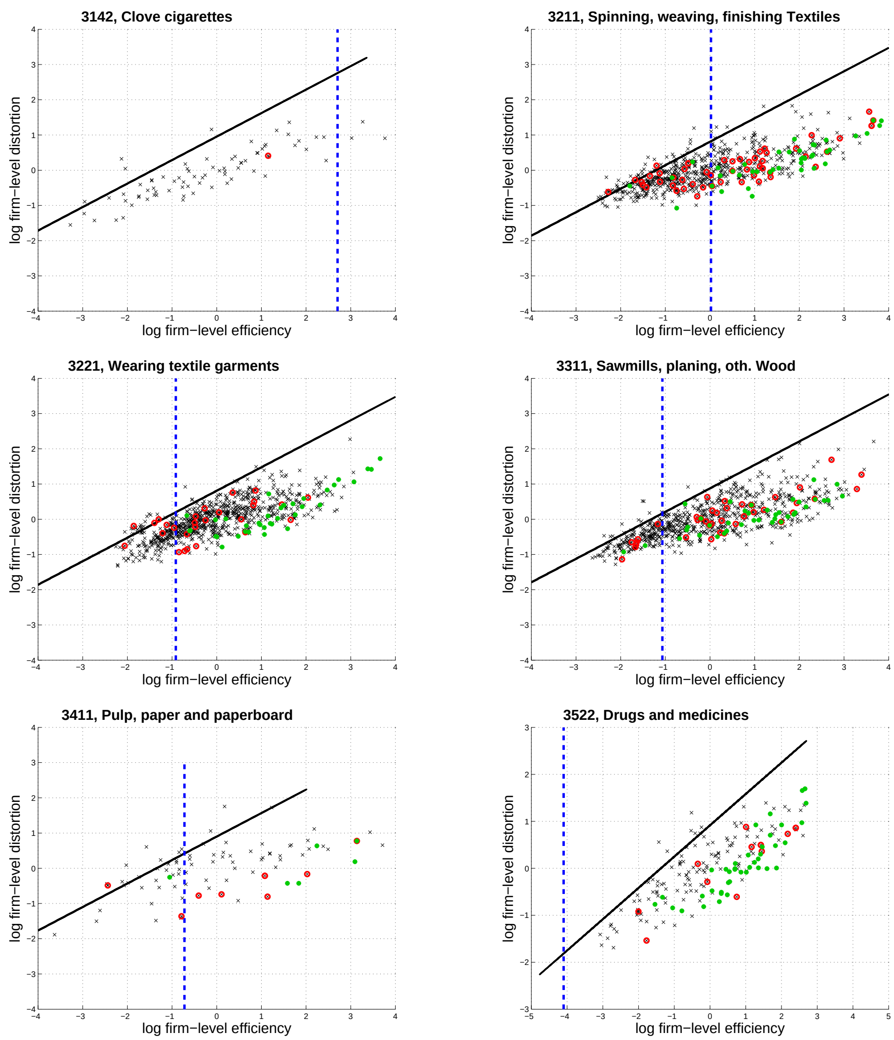

Figure 38: Estimates of Zombie firms for six largest 4-digit sectors. Solid sloped line denotes estimated survival line, while vertical dashed line shows estimated frictionless efficiency cut-off. Color coding denotes plant ownership patterns. Green plants are partially owned by foreigners while red plants are partially owned by central or local government.

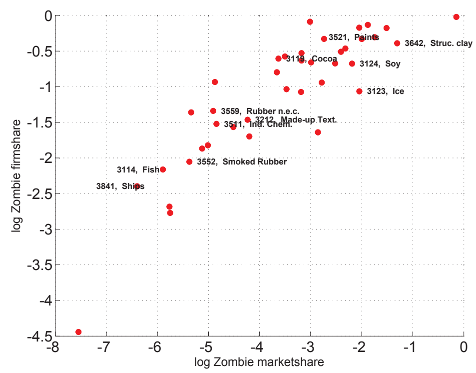

Figure 39: Market shares and firm shares of Zombies across ISIC 4 digit sectors

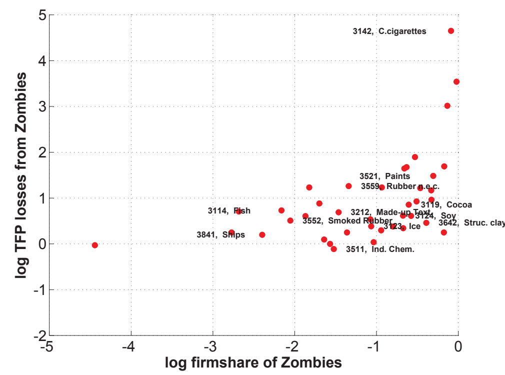

Figure 40: Zombie TFP losses and Zombie firm shares across ISIC 4 digit sectors

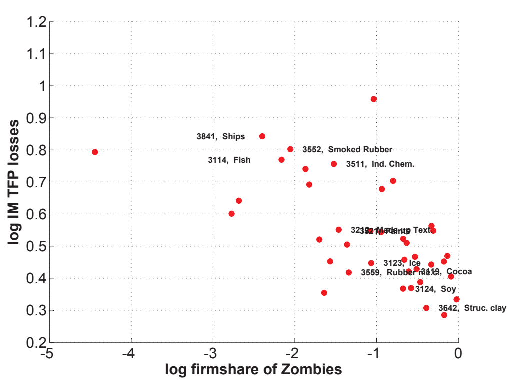

Figure 41: Intensive-margin TFP losses and firm shares of Zombies across ISIC 4 digit sectors

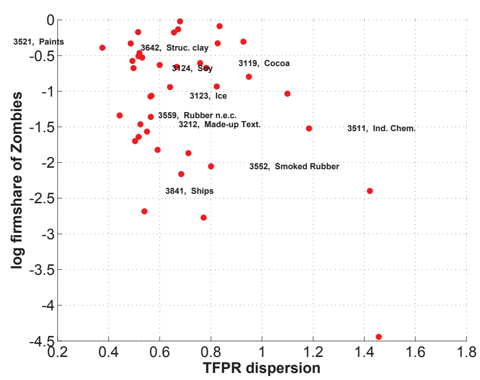

Figure 42: Zombie firm shares versus dispersion of micro-frictions across ISIC 4 digit sectors

Appendix Chapter 3

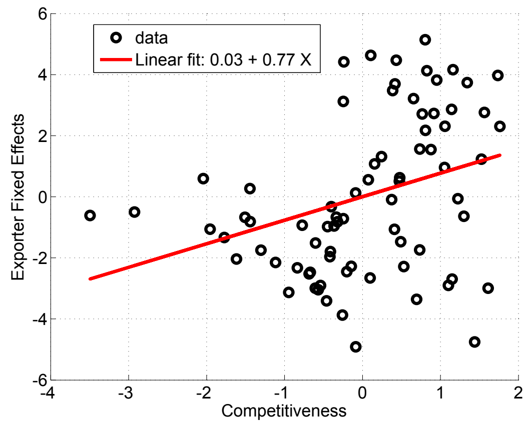

Figure 43: Competitiveness and export advantage, full sample. Data: Waugh (2009).

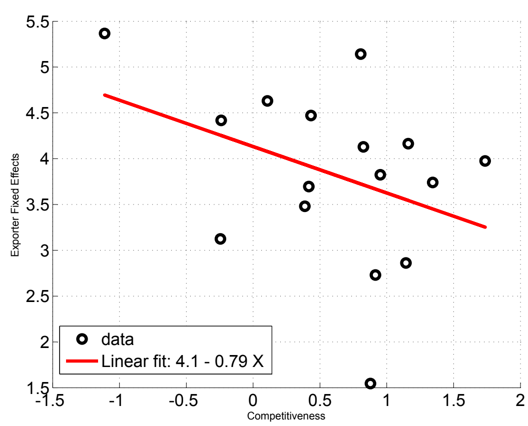

Figure 44: Competitiveness and export advantage, rich country subsample. Data: Waugh (2009) and Author's calculations.

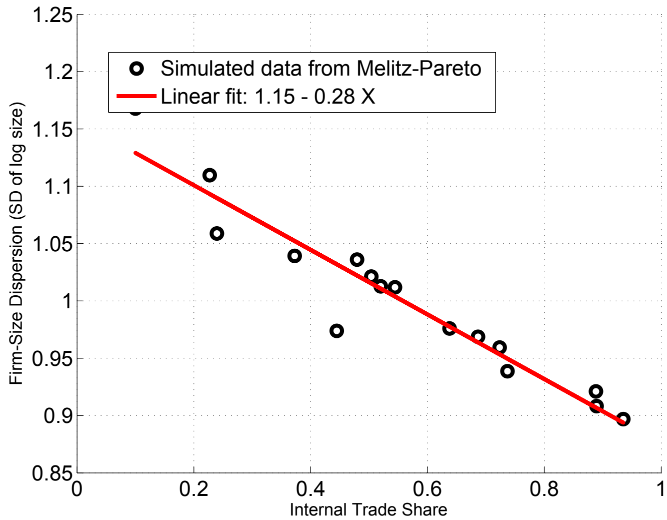

Figure 45: Implications from Melitz-Pareto model with parameters implied by Gravity equation.

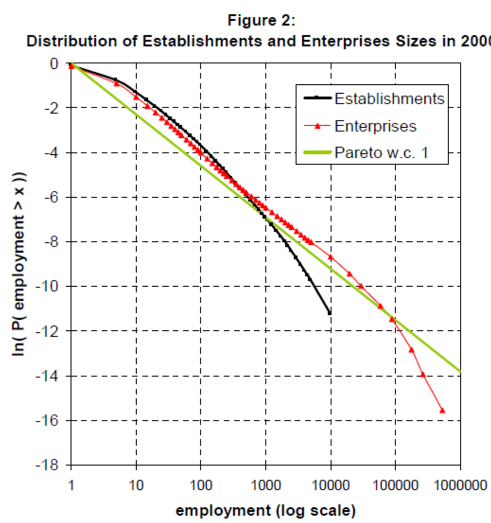

Figure 46: Establishment vs. Firm size distributions. Source: Rossi-Hansberg and Wright (2007)

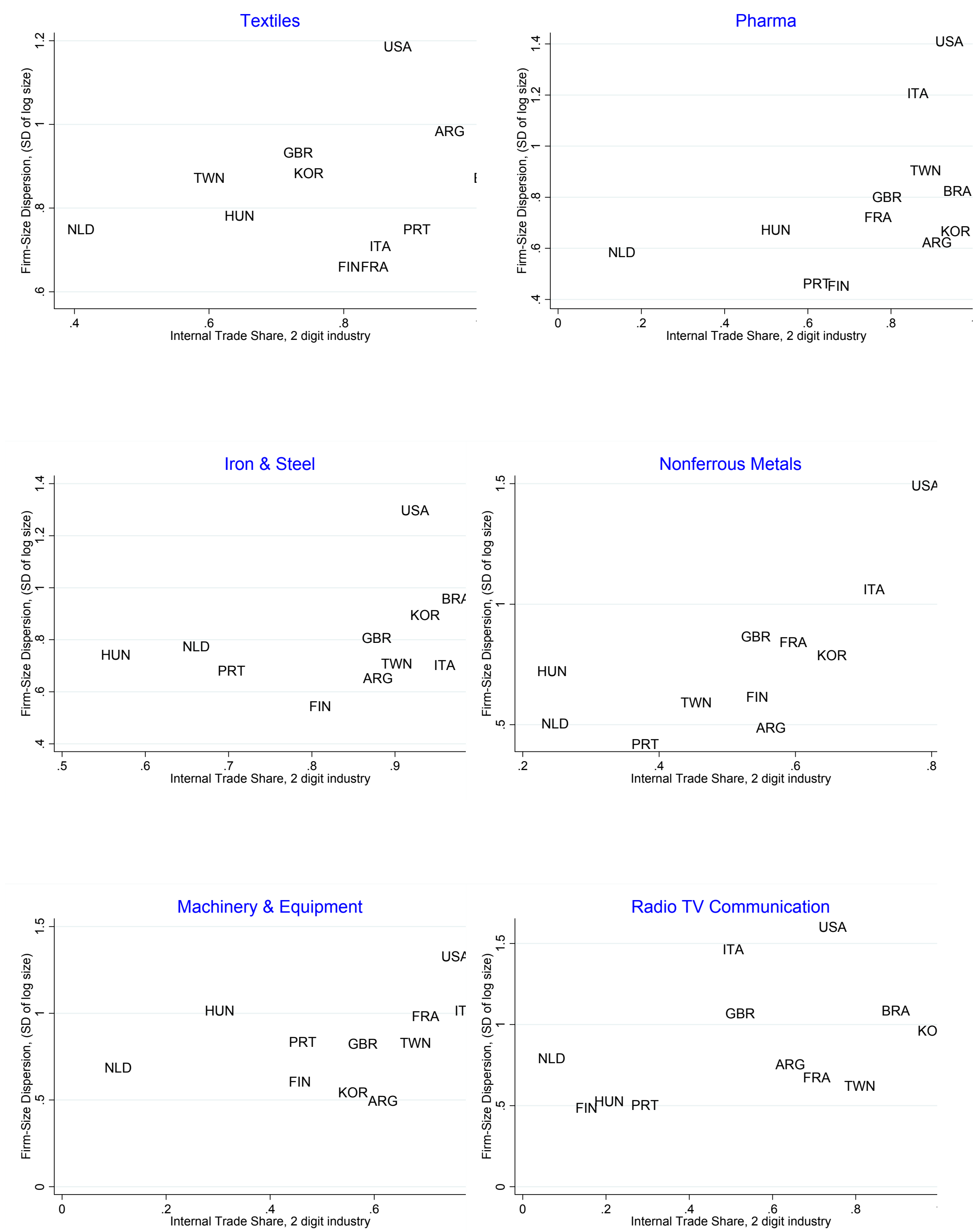

Figure 47: Within industry relation of firm size dispersion and internal trade share.

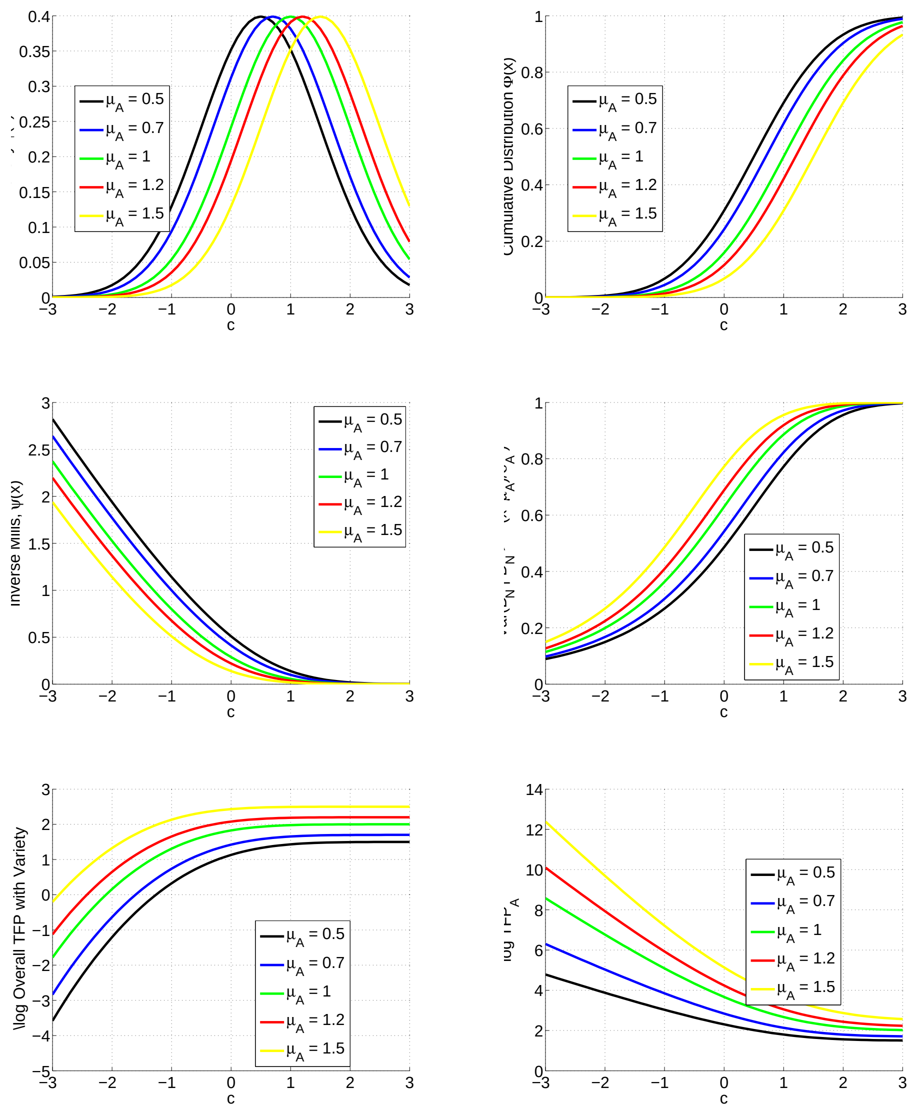

Figure 48: Firm level and aggregate productivity responses to changes in selection cutoffs, for different values of parameters.

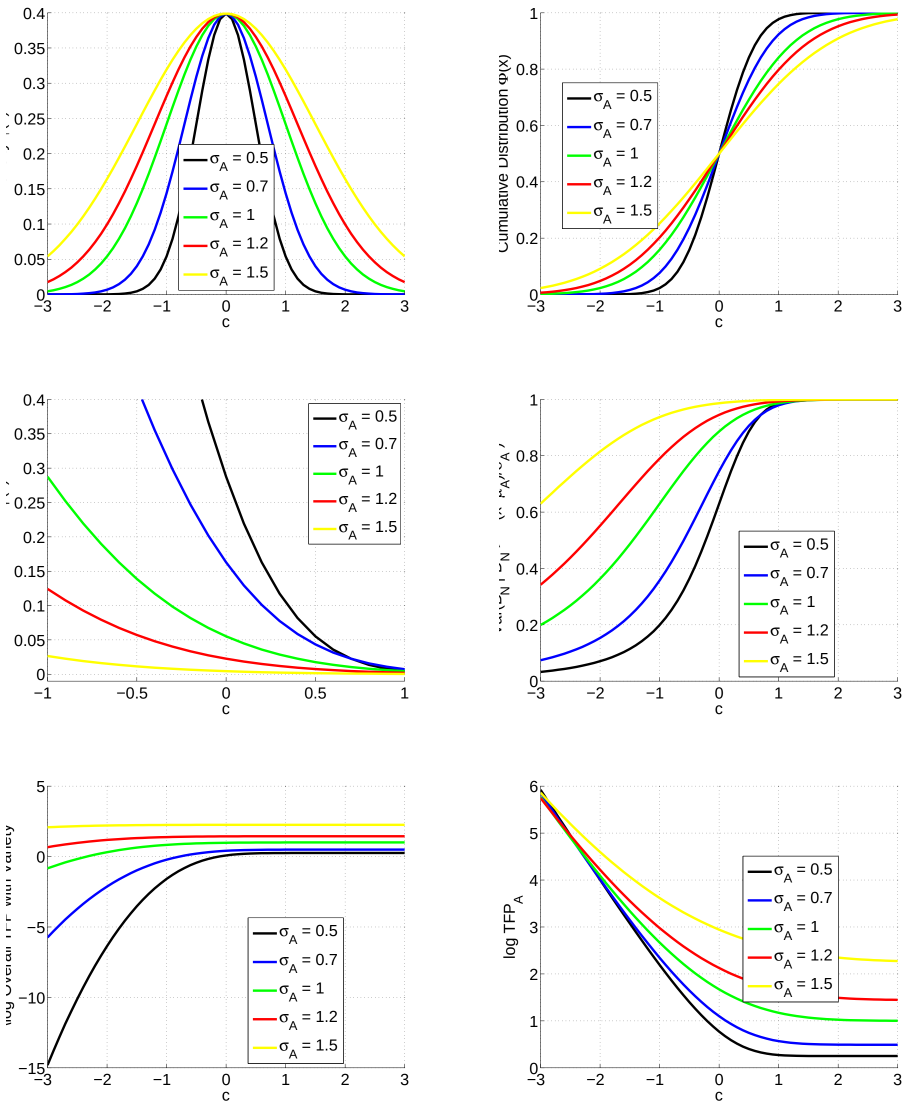

Figure 49: Firm level and aggregate productivity responses to changes in selection cutoffs, for different values of parameters.

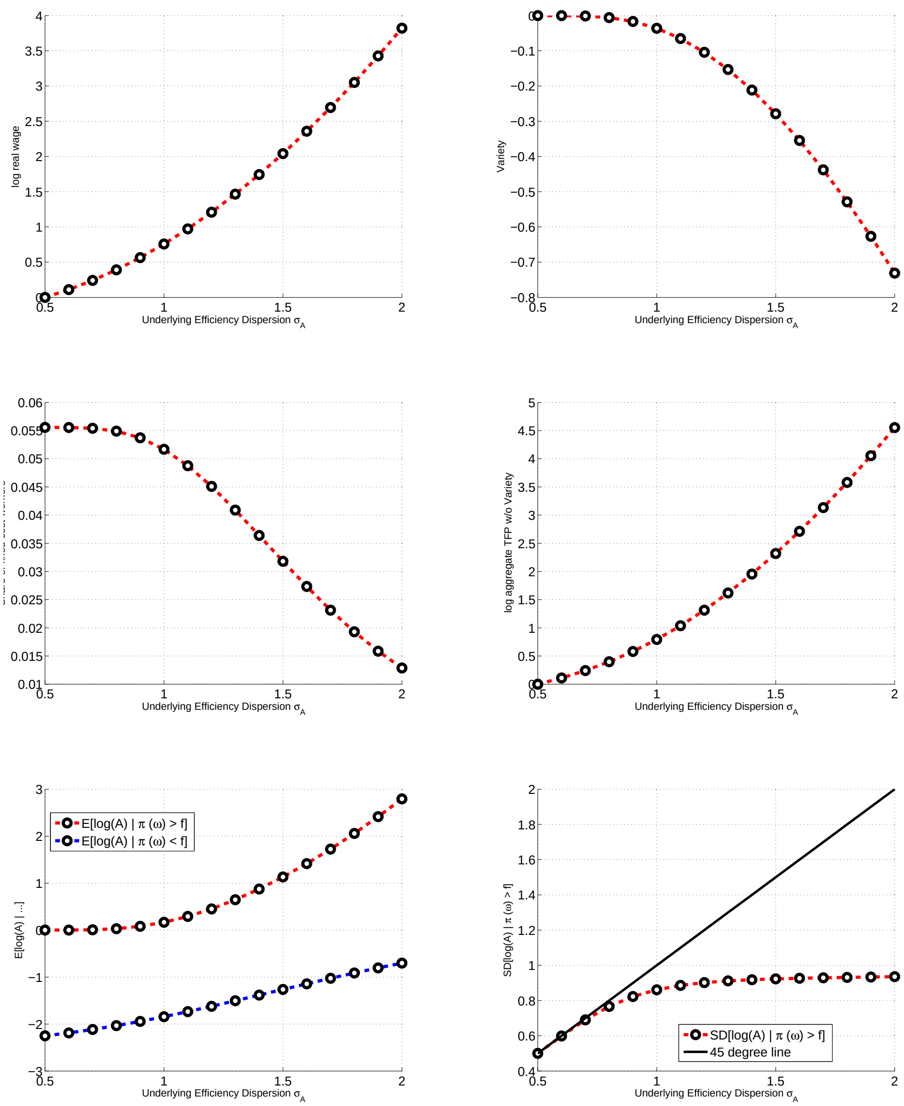

Figure 50: Comparative statics of the closed economy.

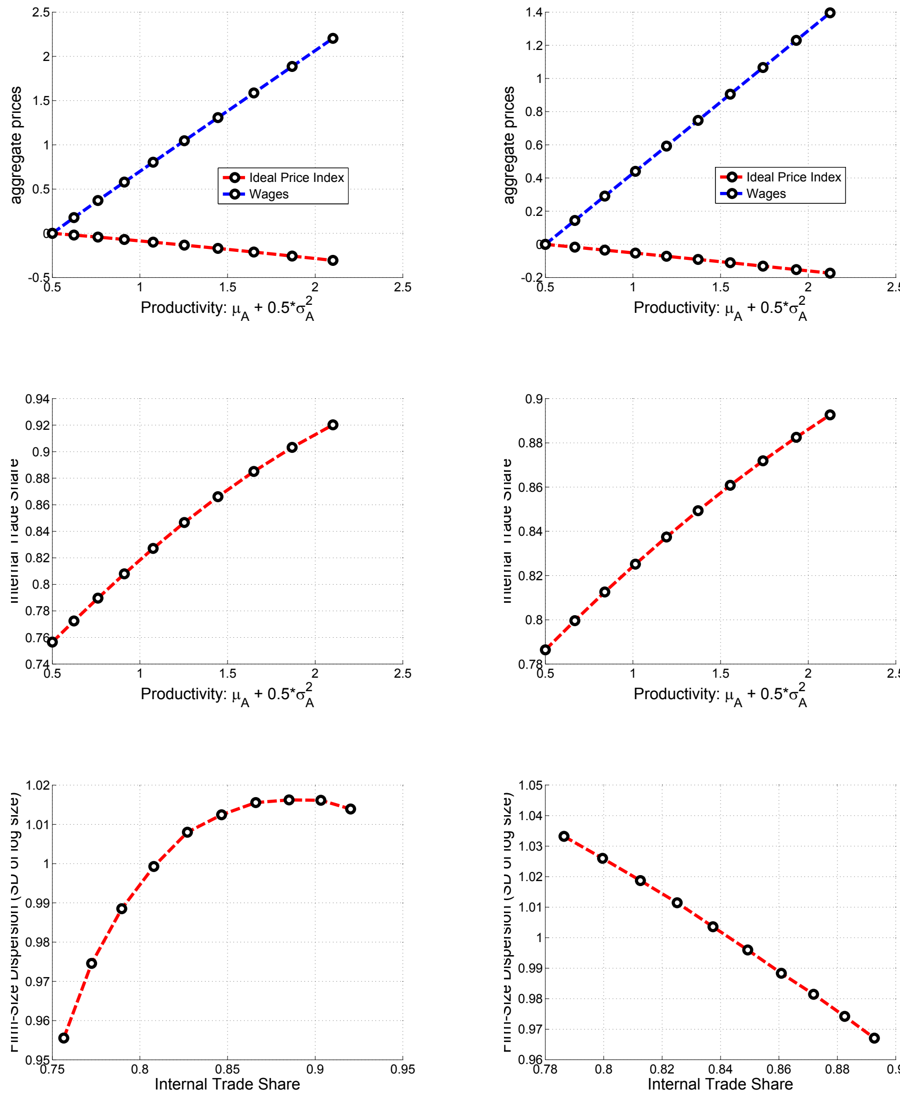

Figure 51: Plots of equilibrium objects in the simulated world economy. Left panels: differences in dispersions. Right panels: differences in mean productivity

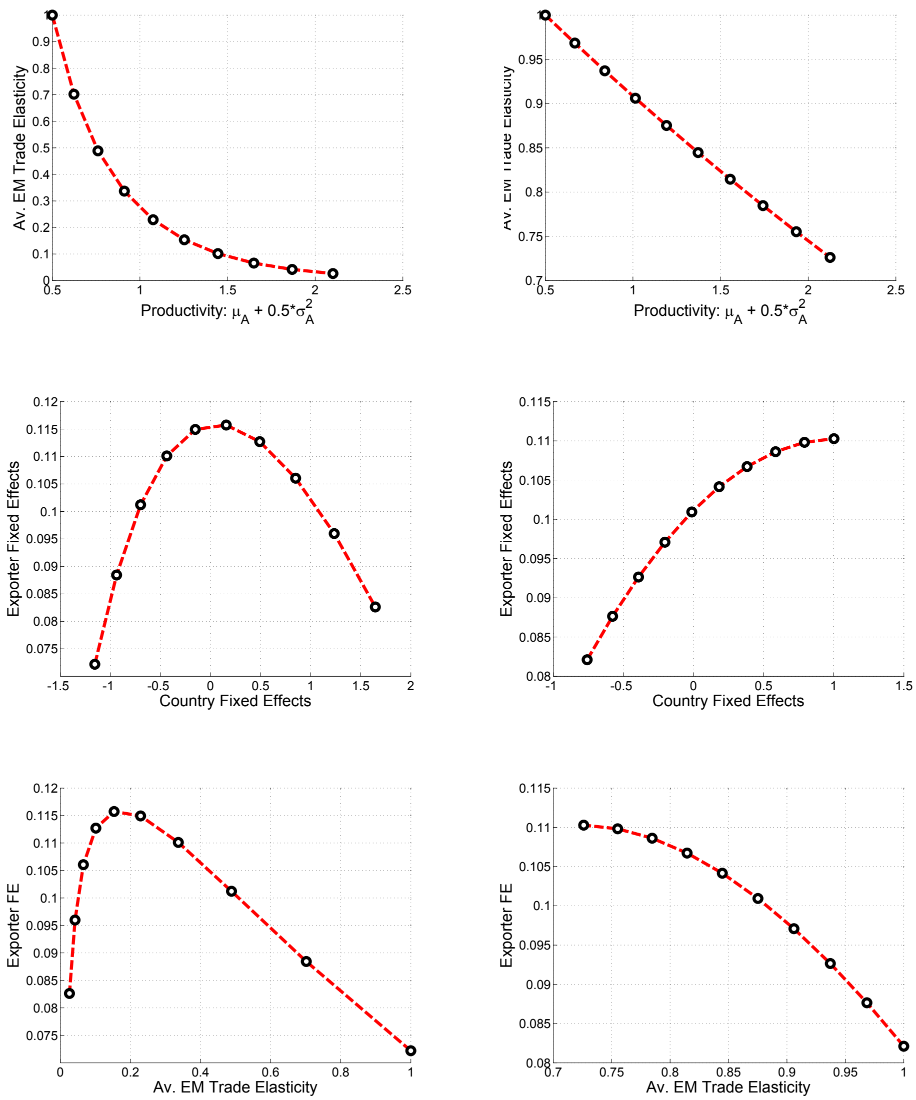

Figure 52: Plots of equilibrium objects in the simulated world economy. Left panels: differences in dispersions. Right panels: differences in mean productivity

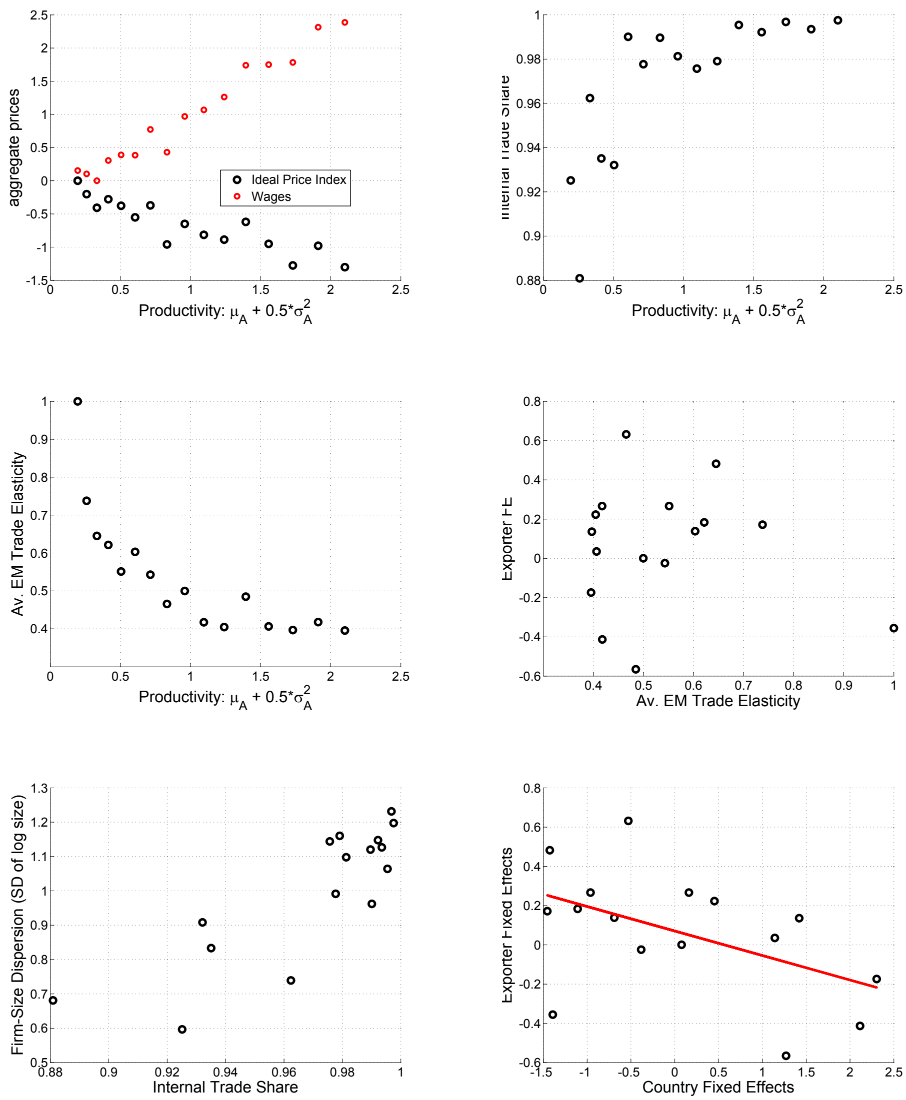

Figure 53: Results from model with realistic geography and differences in σ_A