Micro-level Misallocation and Selection

By Mu-Jeung Yang*

How large are the aggregate productivity losses from the misallocation of resources across firms? With endogenous selection, micro-frictions can induce extensive-margin misallocation among firms: too many unproductive firms are active (Zombies) and too many productive firms are inactive (Shadows). Therefore, the same set of measured distortions potentially induces much larger aggregate productivity losses, as the composition of firms is shifted towards unproductive active firms. I develop and calibrate a model with plant-level micro-data for Indonesia to quantify aggregate welfare in the presence of extensive margin misallocation. My estimates show that selection can magnify aggregate TFP losses from micro-distortions by over 40%, compared to existing estimates. Realistic values of measurement error even increase the relative importance of extensive margin misallocation. *Keywords: Development accounting, firm heterogeneity, misallocation*

Micro-distortions that prevent optimal resource allocation across firms have been shown to contribute significantly to observed differences in total factor productivity (TFP) across countries, see (Chang-Tai Hsieh and Peter Klenow 2009). The current empirical literature has focused on intensive-margin misallocation: among active firms, some are too big and some too small relative to an equilibrium without frictions. At the same time, a parallel literature has emphasized the potential importance of selection for aggregate TFP, as in (R. Lucas 1978), (Marc J. Melitz 2003), (Nezih Guner, Gustavo Ventura and Yi Xu 2008).

This paper contributes to both literatures by quantitatively showing that even without direct selection barriers the presence of producer level frictions can distort selection indirectly and reduce aggregate productivity by shifting the composition of entering firms towards lower efficiency firms: too many low-efficiency firms are active (Zombies), while too many high-efficiency firms will be inactive (Shadows). This mechanism, of "extensive margin misallocation", can therefore also be understood as contributing to a literature, which endogenizes the observed distribution of level efficiency (TFPQ) as function firm level frictions (TFPR), as

* David Eccles School of Business, University of Uah. mjyang@eccles.utah.edu. I am indebted to Chang Hsieh, Pierre Gourinchas, Chad Jones and Sam Kortum for guidance, encouragement and continuous support. I am grateful to Christian Bayer, Dominick Bartelme, Thomas Chaney, Gabe Chodorow-Reich, Max Floetotto, Yuriy Gorodnichenko, Lorenz Kueng, Matthias Lux, Christian Opp, Ralph Ossa and Johannes Wieland and seminar participants at the Berkeley GEMS Conference and the Chicago International Economics workshop, the Cologne Macro Workshop, UC Berkeley, Brown University, the Enaudi Institute for Finance and Economics, the IIES Stockholm, Johns Hopkins SAIS, USC and the University of Washington, for helpful comments. I would also like to thank the editor, Richard Rogerson, who has been particularly helpful and constructive. The author declares that he has no relevant or material financial interests that relate to the research described in this paper

done in (D. Bhattacharya, N. Guner and G. Ventura 2013) and (A. Gabler and M. Poschke 2013)

The insight that aggregate productivity losses for the same set of measured frictions are much larger if selection is endogenous might at first seem counterintuitive: one might expect that adding an additional (selection) margin would lower aggregate productivivity losses from misallocation rather than increase them. However, this argument is correct only if applied to the full set of potential firms, which consist of active firms as well as inactive firms. Current studies focus on measuring the misallocation of resources among the selected sample of active firms. But this ignores the additional productivity losses from the fact that the productivity composition of this observed and selected sample would change if we remove the underlying micro distortions. This also reinforces the importance of trying calibrate the joint distribution of frictions and efficiency for potential firms, rather than just the selected sample of active firms.

While many of these qualitative insights have been previously understood in the literature, this study shows that extensive margin misallocation is also quantitatively important: overall TFP losses from misallocation could be larger by a factor of at least 47% under selection, as compared to a model without selection.

This paper makes at least two main contributions. First, it develops a calibration strategy combining the measurement of micro-distortions as in (Hsieh and Klenow 2009) with selection correction as in (James Heckman and Bo E. Honore 1990). To my knowledge, this study is the first to systematically address selection issues when measuring micro-frictions.[1] A key empirical challenge is that inactive Shadow firms are not observed in the data. I address this issue by imposing a strong functional form assumption of multi-variate log-normality and use equilibrium constraints to discipline the calibration: the equilibrium conditions of the model imply a connection between the underlying joint distribution of firm-level distortions and efficiency on the one hand and equilibrium selection on the other hand which I impose as additional non-linear restrictions. Since selection cut-offs and distributions of distortions and firm-efficiency are likely to differ by industry, I separately estimate parameters for narrow 4 digit industries.

The second contribution of this study is to show that selection effects are quantitatively important for aggregate productivity. I illustrate this point with data from Indonesia in 1990, where it has been argued that political connections and ethnic diversity strongly distorted business decisions, see (Raymond Fisman 2001). Consistent with this view, I find that removing micro-distortions could have increased macroeconomic TFP in Indonesia by about 80%.[2] Im-

[1]Existing studies typically calibrate either a stylized two-point distribution for distortions, as done by (Diego Restuccia and Richard Rogerson 2008), or compare model implications only roughly with specific moments of the data, as done by (Eric Bartelsman, John Haltiwanger and Stefano Scarpetta 2013).

[2] 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. Similar results were obtained for other countries: Argentina could increase its TFP by 50-80% ((P. Neumeyer and G. Sandleris 2010)), Bolivia by 60-70% ((Carlos G. Machicado and Juan C. Birbuet 2008)), Colombia by 50% ((Adriana Camacho and Emily Conover 2010)), Chile by 60-80% ((Ezra Oberfield 2011)) and

portantly, removing extensive margin misallocation could have increased aggregate TFP in Indonesia by an addititional 37%. In other words, the selection channel suggests that aggregate productivity losses **for the same set of measured micro-distortions** are over 47% larger than implied by considering only intensive-margin misallocation. Furthermore, I document that relative importance of extensive margin misallocation becomes even larger relative to intensive margin misallocation if I account for measurement error in the microdata.

The paper proceeds as follows. Section I. outlines a simplified initial model, defines the parameters of interest and draws implications for aggregate TFP. Section II. describes the empirical methodology including the full-blown model with multiple sectors and capital and describes how calibration works. Section III. describes the data and the empirical estimates. Section IV. uses the estimates of section III. to quantify and decompose aggregate misallocation losses. Section V. performs robustness exercises and section VI. concludes.

I. Theory

A. Economic Environment and Equilibrium Definition

Consider the following one-sector closed economy. Following (Lucas 1978), I assume that there is a continuum of agents who can either become workers or entrepreneurs, indexed by ω. Each agent ω, has an underlying managerial talent $A(\omega)$ as well as an associated output distortion $\tau(\omega)$, which might capture personal barganing power, political connectedness or expropriation risk. This distortion will be modeled as a revenue tax as in (Hsieh and Klenow 2009), so that tax revenues from distortions are compensated lump sum. If the agent decides to become an entrepreneur, the associated firm chooses labor optimally to maximize profits. Production is given by a decreasing returns to scale technology as in (Lucas 1978), parameterized with a "Span of Control" parameter $\gamma \in (0, 1)$

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

where $L(\omega)$ is the labor input used by firm ω in production, p is price of the homogeneous output, $y(\omega)$ is the quantity firm ω produces.

Selection into entrepreneurship is driven by a standard occupational choice problem: if an agent becomes a manager, entrepreneurial profit needs to be at least as large as the opportunity cost of entrepreneurship, which is forgone wage income. Agent ω will therefore decide to become a manager if

$$\Pi(\omega) \geq w \tag{2}$$

Uruguay by 50-60% ((Carlos Casacuberta and Nestor Gandelman 2009)).

This selection equation can be rewritten as

$$\ln\left(\frac{1}{1-\tau(\omega)}\right) - \ln(A(\omega)) \leq -\ln\left(\frac{w}{p}\right) + (1-\gamma)\cdot\ln(1-\gamma) + \gamma\cdot\ln(\gamma) = \bar{z}_J \tag{3}$$

I assume an empirically motivated functional form for the joint distribution of managerial ability and micro-frictions. In particular, I assume that $\ln(A(\omega))$ and $\ln\left(\frac{1}{1-\tau(\omega)}\right)$ are jointly normally distributed. These are strong assumptions needed to recover the underlying distribution of potential firms and will be discussed further in the application.

$$\begin{pmatrix} \ln A(\omega) \\ \ln\left(\frac{1}{1-\tau(\omega)}\right) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_A \\ \mu_\tau \end{bmatrix}, \begin{bmatrix} \sigma_A^2, \sigma_{A,\tau} \\ \sigma_{A,\tau}, \sigma_\tau^2 \end{bmatrix}\right) \tag{4}$$

with the the correlation of efficiency and distortion given by $\rho_{A,\tau} = \frac{\sigma_{A,\tau}}{\sigma_A \sigma_\tau}$.

DEFINITION: COMPETITIVE EQUILIBRIUM WITH DISTORTIONS. — In the Lucas Span-of-Control model with micro-distortions outlined above, a competitive equilibrium with distortions is defined as a set of allocations $(\{L(\omega)\}, s_e)$ and prices (p, w), such that

1) Firm level labor demand $L(\omega)$ is chosen optimally to maximize (1), taking prices (p, w) and individual draws $\{A(\omega), \tau(\omega)\}$ as given

2) A fraction $s_e = P(\Pi(\omega) \geq w)$ of workers become entrepreneurs, using criterion (2), taking prices (p, w) and individual draws $\{A(\omega), \tau(\omega)\}$ as given

3) The labor market clears: $(1 - s_e) \cdot L = \int_{\Pi(\omega)\geq w} L(\omega) d\omega$

4) The output price is set as numeraire: $p = 1$

B. *TFP with Selection*

To understand the aggregate welfare consequences of how micro-distortions and selection interact, note that aggregate TFP in the presence of micro-distortions can be summarized as follows[3]

$$Y \propto E\left[\left(A(\omega)\frac{1-\tau(\omega)}{1-\bar{\tau}}\right)^{\frac{1}{1-\gamma}}\middle|\Omega_S\right]^{1-\gamma} \cdot s_e^{1-\gamma}(1-s_e)^{\gamma} \tag{5}$$

where $s_e = P(\Omega_S)$ is the fraction of agents becoming managers with Ω_S as set of active firms.

There are two main differences in a model with selection when compared to a model with only an intensive margin as in (Hsieh and Klenow 2009). First, the number of firms is now endogenous and more active firms will tend to offset firm-level diseconomies of scale in the aggregate, as production is spread out across more producers.[4] I call this an "aggregate scale effect" and it is captured by the term $s_e^{1-\gamma}(1-s_e)^{\gamma}$. Second, micro-distortions change the set of active firms Ω_S and therefore the productivity composition of active firms. This is the "extensive margin misallocation effect".

C. Sample Selection and Welfare Effects

Figure 1 shows selection in this model for a simulated equilibrium. The x-axis shows managerial ability, with higher values for more efficient firms. The y-axis shows firm-level implicit micro-distortions, with higher values for more heavily distorted firms. Every dot captures a potential realization of $A(\omega), \tau(\omega)$ and therefore a potential firm. The solid line is the selection line, equation (3). Inactive firms are captured by the grey dots to the northwest of this selection line.

Note the presence of the key link between the selection equation (3) and the aggregate real wage level w/p. If average wages in the economy are high, a firm's relative competitive position is worse and hence being active is harder. On the other hand, selection influences the value of w/p, since the set of active firms impacts labor demand and therefore average wages paid in the economy. In equilibrium, both selection and real wages are driven by the latent heterogeneity in $A(\omega), \tau(\omega)$. In terms of figure 1, this means that the shape of distribution of grey and black dots determines both the position of the selection line – through its impact on the relative wage w/p – and the shape of the distribution of the

[3]To derive aggregate TFP, I first aggregate factor payments to production labor as $wL_P = \int_{\Pi(\omega)\geq w} wL(\omega)d\omega$ using optimal labor demand from (1), which is given by $L(\omega) = \left[\gamma(1-\tau(\omega))\frac{p}{w}A(\omega)\right]^{\frac{1}{1-\gamma}}$. Using the fact that $L_P = (1-s_e)\cdot L$, real aggregate income can be written as $Y = \frac{1}{\gamma}\frac{1}{1-\bar{\tau}}\frac{w}{p}(1-s_e)L$, where $1-\bar{\tau} = \int_{\Pi(\omega)\geq w}(1-\tau(\omega))\cdot\frac{py(\omega)}{pY}d\omega$, which is proportional to the average distortion. Then deriving equilibrium wages from labor market clearing, gives $\frac{w}{p} = \gamma[(1-s_e)L]^{-(1-\gamma)}\left(\int_{\Pi(\omega)\geq w}\left(\frac{1}{1-\tau(\omega)}\right)^{-\frac{1}{1-\gamma}}A(\omega)^{\frac{1}{1-\gamma}}d\omega\right)^{1-\gamma}$. Using this real wage in the real income formula and simplifying then leads to the expression for real TFP in equation (5).

[4]This is similar to the role of variety effects in monopolistic competition models, see (Roberto Fattal Jaef 2018)

black dots – the distribution of actually active firms. This is a key conceptual insight that I will exploit in the empirical section.

Moving to the welfare effects of selection, as per equation (3), the set of active firms in the distorted equilibrium is given by

$$\Omega_S = \left\{ \omega : \ln A(\omega) \geq \ln\left(\frac{1}{1-\tau(\omega)}\right) - \bar{z}_J \right\} \tag{6}$$

In figure 1, the set Ω_S captures the black dots. In contrast, only startups to the right of the vertical dashed line would enter in a frictionless equilibrium:

$$\Omega_S^* = \{\omega : \ln A(\omega) \geq -\bar{z}_J^*\} \tag{7}$$

For the set Ω_S^*, agents are becoming managers only based on their managerial productivity as there are no distortions in the frictionless equilibrium. Removing extensive-margin misallocation changes the set from Ω_S to Ω_S^* and therefore reweights the efficiency distribution of operating firms. Extensive margin misallocation highlights two groups of firms in particular: First, Shadows would be efficient enough to be active in a frictionless equilibrium but kept out by micro-frictions. Second, Zombies would not be active in a frictionless equilibrium, but are implicitly subsidized and can therefore be active in a distorted equilibrium.

To obtain the productivity composition of the frictionless equilibrium, two things need to happen: Zombie firms should be removed and Shadow firms should become active. Figure 2 illustrates how this would change the efficiency distribution of firms endogenously and shift the mass of firms from low-productivity Zombie firms toward high-productivity Shadow firms. This endogenous redistribution of the firm productivity distribution leads to additional gains from removing firm-level distortions relative to a model with intensive margin reallocation only.

Of course, the relative importance of Shadow and Zombie firms − and consequently extensive margin misallocation − will depend on both, the underlying distribution of distortions and firm efficiency and the relative position of the equilibrium cutoff lines in the distorted and frictionless equilibria. Therefore, I turn to the quantification of these objects next.

II. Empirical Methodology

This section gives an overview of the empirical 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 and evaluate the aggregate welfare losses from micro-distortions.

A. Extended Model

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

Extension 1: Capital enters the production function on the firm-level, and there will be a net wedge distorting the mix between capital and labor.
Extension 2: There will be multiple sectors, with an elasticity of substitution of 1 across sectors.

Each of these extensions serves a particular purpose when confronting the model with the data. Extension 1 introduces a net capital wedge since, within 4-digit industries, marginal revenue products of capital are distributed differently from marginal revenue products of labor.[5] 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 and selection behavior without changing anything important. Extension 2 addresses the fact that selection lines are likely to differ across industries as, for example, different industries have different distributions of efficiency and distortions.

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

$$Y = \Pi_{s=1}^{S} Y_s^{\xi_s} \quad \text{with} \quad \xi_s \in [0,1], \quad \sum_{s=1}^{S} \xi_s = 1 \tag{8}$$

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

(9)
$$\max_{\{K_s(\omega),L_s(\omega)\}} \Pi_s(\omega) = [1-\tau_{Y,s}(\omega)] \cdot p_s y_s(\omega) - w_s L_s(\omega) - [1+\tau_{K,s}(\omega)] \cdot R_s \cdot K_s(\omega)$$
subject to: $y_s(\omega) = A_s(\omega) \cdot \left[K_s(\omega)^{\alpha_s} L_s(\omega)^{1-\alpha_s}\right]^{\gamma}$

where R_s is the sector-specific rental rate of capital and the selection equation is given by

$$\Pi_s(\omega) \geq w_s \tag{10}$$

[5] Alternatively, this might suggest that technological factor shares are different across firms. I respond to this issue in two ways. First, in the validity check displayed in table 2, I show that measures of frictions, controlling for plant efficiency are negatively correlated with exit of age 0 firms. Such a systematic correlation is hard to explain if measured frictions only would reflect technological factor share differences. Second, in principle this issue could be resolved with corresponding establishment-level data from a wealthy country like the US, by not removing all distortions but moving the distortions in my data toward the US distribution of distortions as (Hsieh and Klenow 2009) have done. Results from such an analysis are available upon request.

Two remarks are in order to better understand this equilibrium. First, I assume that all workers can potentially become managers. However, since the data has both, production and non-production workers that differ by skill and therefore likelihood to become managers, I will convert production workers into efficiency units of non-production workers to preserve the property that all workers can potentially become managers. Second, to solve for equilibrium, I assume complete immobility of labor and capital across sectors, so that equilibrium for each industry is determined by factor prices w_s, R_s that ensure labor and capital market clearing for each industry. I choose complete factor immobility for two reasons. First, by shutting down cross-sectoral reallocation the welfare effects of removing distortions will be completely driven by within-sector reallocation across firms, which is the mechanism I focus on and which makes my estimates more comparable to (Hsieh and Klenow 2009).[6] Second, conceptually a natural model choice for agents that can become entrepreneurs is that each worker has a vector of managerial productivities with a differing managerial talent by sector. Instead, I assume that workers do not learn the value of their managerial talent until they have already worked in the sector. Hence, the equilibrium can be understood as the outcome of a two-stage process, in which workers are first randomly assigned to an industry. Then they find out their managerial talent for this industry only and select whether to become an entrepreneur.

B. Calibrated and Observed Data

This section describes additional assumptions that facilitate the empirical analysis, including calibrated parameters, data requirements and measurement assumptions.

An important calibrated parameter is the "Span of Control" parameter γ. I follow (Hsieh and Klenow 2009) and set $\gamma = 0.5$ to facilitate comparability. This value should just be seen as a starting point, as I will re-estimate all results for different values of γ.

Following (Hsieh and Klenow 2009), factor shares in the Cobb-Douglas production function are set to the corresponding factor shares of US 4-digit sectors from (Eric J. Bartelsman, Randy A. Becker and Wayne B. Gray 2000).[7] I choose to follow (Hsieh and Klenow 2009) in calibrating the factor shares to US levels, mainly to facilitate comparability. I also experimented with GMM estimation of factor shares as in (D. Ackerberg, K. Caves and G. Frazer 2015), which did not affect the results much[8]. For sectors that do not have any correspondence with a 4-digit sector in the US, I assume that the capital share is $\alpha_s = 1/3$.

[6] In a previous version of the model, I allowed for cross-sectoral reallocations, but quantitative results remained similar due to the fact that each industry is small relative to the whole manufacturing sector.

[7] 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.

[8] Results are available upon request.

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 $p_s y_s(\omega)$ to value added. As (Charles I. Jones 2011) points out, this basically ignores the effects of possible distortions 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 value added distortions. The method of this paper could easily be extended to accommodate distortions of intermediate input use.

Second, the selection in occupational choice used here states that each worker could potentially become an entrepreneur. This is likely to be the case for managerial or non-production labor, but less likely for production workers. I therefore aggregate labor into effective workers that might all select into entrepreneurship. I am using wage premia to accomplish this aggregation. In the Indonesian data non-production workers get paid approximately twice the average wage of production workers. Therefore, effective labor inputs are measured as

$$L_{E,s}(\omega) = L_{NP,s}(\omega) + \kappa \cdot L_{P,s}(\omega) \tag{11}$$

where $L_{NP,s}(\omega)$ is the number of non-production workers,$L_{P,s}(\omega)$ the number of production workers and the constant $\kappa = 0.5$ is used to convert production workers into efficiency units of non-production workers. Again, κ should be considered as an initial choice and I will vary this parameter in the robustness section to show how it affects the results.

C. *Recovery of underlying distribution of firm heterogeneity*

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 for estimation. The primary issue is that I do not directly observe the output and factor prices for each sector, yet strive to estimate the model for each sector separately, as underlying distributions are likely to differ across sectors. However, as noted before in the context of (3), these factor prices also influence firm selection. Under the current distributional assumptions, factor prices are non-linear fixed points that cannot be explicitly solved[9].

As discussed in section II.B., the main required data for this estimator are value added and factor inputs.

[9] Formally, the problem I face is similar to constrained MLE estimation of dynamic discrete choice problems, such as (John Rust 1987). Instead of using popular Nested Fixed Point algorithms to address this issue, I follow (Che-Lin Su and Kenneth Judd 2010) in formulating the estimation problem as a "Mathematical Program with Equilibrium Constraints" (MPEC). Like (Su and Judd 2010), and (J. Dube, Jeremy T Fox and Che-Lin Su 2009) I find in Monte Carlo test runs that the use of MPEC methods facilitates numerical stability and reliability of estimates.

The first observable variable is value added:

$$D_{1,s}(\omega) = p_s y_s(\omega) \tag{12}$$

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

$$D_{2,s}(\omega) = \left(\frac{R_s K_s(\omega)}{\alpha_s}\right)^{\alpha_s} \left(\frac{w_s L_s(\omega)}{1-\alpha_s}\right)^{1-\alpha_s} \tag{13}$$

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

$$D_{3,s}(\omega) = \left(\frac{\alpha_s}{1-\alpha_s}\right) \left[\frac{w L_s(\omega)}{R K_s(\omega)}\right] \tag{14}$$

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

$$\begin{pmatrix} \ln D_1(\omega) \\ \ln D_2(\omega) \\ \ln D_3(\omega) \end{pmatrix} \propto \begin{bmatrix} -\frac{\gamma}{1-\gamma} & -\alpha_s \frac{\gamma}{1-\gamma} & \frac{1}{1-\gamma} \\ -\frac{1}{1-\gamma} & -\alpha_s \frac{1}{1-\gamma} & \frac{1}{1-\gamma} \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} \ln\left(\frac{1}{1-\tau_{Y,s}(\omega)}\right) \\ \ln(1+\tau_{K,s}(\omega)) \\ \ln A_s(\omega) \end{pmatrix} \tag{15}$$

Therefore, it is straightforward to generate the distribution of these three data series as a function of the two wedges and firm efficiency.

As mentioned before, since the latent distribution of firm level distortions and efficiency differs by sector, I estimate the model separately for each sector. Since in the model, the impact of latent heterogeneity on selection works through factor prices, I include equilibrium factor prices w_s and R_s as an additional parameters but require that the factor prices satisfies a fixed point that describes equilibrium.[10]

Proposition 1: Mathematical Program with Equilibrium Constraints

Let the parameter vector for each sector s be given by
$\theta_s = \left[\mu_{A,s}; \mu_{\tau_Y,s}; \mu_{\tau_K,s}; \sigma_A; \sigma_{\tau_Y,s}; \sigma_{\tau_K,s}; \rho_{A\tau_Y,s}; \rho_{A\tau_K,s}; \rho_{\tau_Y K,s}\right]$
The Maximum-Likelihood Estimator for the selection model with managerial productivity heterogeneity and micro-distortions, can be writ-

[10]Note that although I observe wages w_s, I do not observe industry specific prices p_s. I therefore normalize $p_s = 1$, so that estimated factor prices should be understood to be in units of industry prices.

ten as

$$\max_{\theta_s,w_s,R_s}\sum_{\omega}\ln\left\{\frac{\phi\Big(\ln D_{1,s}(\omega),\ln D_{2,s}(\omega),\ln D_{3,s}(\omega)\Big|\theta_s,w_s,R_s\Big)}{1-\Phi\Big(\kappa_Z\big(\theta_s,w_s,R_s\big)\Big)}\right\}$$

subject to:

$$w_s=\left(\frac{(1-\alpha_s)}{1-P(\Pi_s(\omega)\geq w)}\right)\gamma^{1-\gamma}\left[u_s(R_s,w_s)\right]^{-\frac{\gamma}{1-\gamma}}\Sigma_L(\theta_s,w_s,R_s)$$

$$R_s=\alpha_s\gamma^{1-\gamma}\left[u_s(R_s,w_s)\right]^{-\frac{\gamma}{1-\gamma}}\left(\frac{L_s}{K_s}\right)\Sigma_K(\theta_s,w_s,R_s)$$

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

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

$$\kappa_Z(\theta_s,w_s,R_s)=\ln\left(\frac{1-\gamma}{w_s}\right)+\left(\frac{\gamma}{1-\gamma}\right)\left[\ln(\gamma)-\ln(u_s(R_s,w_s))\right]$$

$$\Sigma_L(\theta_s,w_s,R_s)=\left(\int_{\Pi_s(\omega)\geq w_s}A(\omega)^{\frac{1}{1-\gamma}}\left(\frac{1}{1-\tau_Y(\omega)}\right)^{-\frac{1}{1-\gamma}}(1+\tau_K(\omega))^{-\alpha_s\frac{\gamma}{1-\gamma}}d\omega\right)/L_s$$

$$\Sigma_K(\theta_s,w_s,R_s)=\left(\int_{\Pi_s(\omega)\geq w_s}A(\omega)^{\frac{1}{1-\gamma}}\left(\frac{1}{1-\tau_Y(\omega)}\right)^{-\frac{1}{1-\gamma}}(1+\tau_K(\omega))^{-\left(1+\alpha_s\frac{\gamma}{1-\gamma}\right)}d\omega\right)/L_s$$

Proof: see Appendix

III. Data and Empirical Results

A. Overview of Data

For Indonesia, I use the *Statistik Industri*, an annual panel data set that collects information for all Indonesian establishments with more than 20 employees as well as a random set of establishments with less than 20 employees.[11] Note that this data is therefore capturing primarily medium sized and large establishments and that this is likely to influence the quantitative importance of extensive margin misallocation estimates. However, it is worthwhile to point out that the same is true for intensive margin misallocation losses in (Hsieh and Klenow 2009) so that

[11] As the data set has been intensively analyzed before by (Mary Amiti and Jozef Konings 2007) and (Michael Peters 2013), I refer the interested reader to the data discussions in these papers.

the relative importance of extensive margin effects to intensive margin effects is not likely to be strongly affected by missing data on very small producers. Additionally, it should be noted that altough small producers make up the majority of firms in both developed and developing economies, empirical studies such as (E. Hurst and B. Pugsley 2011) have shown that their employment, market shares and growth rates are low. In other words, if such very small firms are not growing much in relatively distortion-free environments, such as the US, they are unlikely to respond much to reforms that remove distortions in developing economies.

Among the surveyed variables are the wage bill and number of employees, capital stocks at book values, and value added. The sample contains approximately 20,000 plants each year. I focus on the cross section of 1990, but results are similar for other years from 1990 to 1996. Calibration is done at the 4-digit industry level, which, after some data cleaning, leaves me with 43 sectors capturing approximately 80% of manufacturing activity.[12]

B. Reduced Form Evidence on External Validity

Before turning to the actual results, I verify the key selection mechanism of the model and test the validity of my measures of distortions and efficiency. To check external validity, I deliberately use the part of the data that is not used in MPEC: I consider the effects of measured managerial productivity and disortions on exit of entrants. Note that my model has no explicit entry and exit dynamics, so a natural question is how this exercise is related to my selection model. A possible motivation is an extension of my model in which agents who decided to become entrepreneurs do have imprecise information about their managerial talent $A(\omega)$ and their distortion $\tau(\omega)$ and only learn about the true values of $A(\omega)$ and $\tau(\omega)$ after being in business for a year.

The panel regressions also allow me to evaluate whether adding further dimensions and state variables substantially alters the baseline predictions of the effects of TFPR and TFPQ on selection. [13]

[12] I restrict the exercise to sectors that have more than 100 establishments and include other sectors with more than 80 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 by sector to minimize the impact of outliers.

[13] For the following discussion, I follow Hsieh and Klenow and summarize all firm level discussion with TFPR, defined as

$$\texttt{TFPR}_s(\omega) = \frac{(1+\tau_{K,s}(\omega))^{\alpha_s}}{1-\tau_{Y,s}(\omega)} \tag{16}$$

The measure of firm level efficiency is given by TFPQ:

$$\texttt{TFPQ}_s(\omega) \equiv A_s(\omega) = \frac{y_s(\omega)}{K_{P,s}(\omega)^{\alpha_s} L_{P,s}(\omega)^{1-\alpha_s}}$$

$$\text{with} \quad y_s(\omega) = \frac{1}{p_s}\Big[p_s y_s(\omega)\Big]^{1-\gamma}\left[TFPR_s(\omega)\cdot\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\right]^{-\frac{\gamma}{1-\gamma}}$$

The key mechanism of this paper implies that distortions and efficiency should have predictive power for selection. Firms with higher values of TFPR, conditional on TFPQ should be more likely to exit, while firms with higher TFPQ, conditional on TFPR, should be less likely to exit. Table 2 reports estimates of the impact of TFPR and TFPQ on survival patterns of age zero firms in simple discrete choice regressions. As the third column in table 2 shows, higher TFPR plants are more likely to exit after year zero. Note that the literature surveyed in (Eric Bartelsman and Mark Doms 2000) and even more recent empirical studies such as (L. Foster, J. Haltiwanger and C. Syverson 2008) usually do not include both TFPR and TFPQ together in the exit regressions. From the perspective of this paper, this a potentially problematic practice as it conflates the direct impact of TFPR on exit with the indirect impact through the correlation of TFPR and TFPQ.

The simple regressions in table 2 are also instructive to evaluate competing hypotheses about the nature of distortions as measured by TFPR. While I follow Hsieh and Klenow and model TFPR as implicit taxes, TFPR differences could mainly reflect variations in firm-level markups. If TFPR captures primarily markups, then controlling for TFPQ, one should expect that firms with higher TFPR have higher profits and should therefore be more likely to survive their first year. This prediction contradicts the evidence in table 2.

C. Evaluation of Model Fit and the Role of Equilibrium Constraints

Before turning to a discussion of results, I review a number of ways to check whether my estimates are reasonable, concentrating on two aspects of my model. First, I check how well the estimated model captures patterns of distortions, plant efficiency, and selection. Second, I test the functional form assumption on the distribution of establishment-level frictions and efficiency.

Selection Patterns. — One way to analyze the role of firm distortions and efficiency is to compare graphically the estimated selection lines with the data. Figure 3 illustrates the selection of plants as function of TFPR and TFPQ in 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.[14] Both of these overlap fairly closely for most of the body of the distribution. Note especially the estimated black sloped selection line from the model. If TFPR would be driven mostly by markups, this would predict that higher-distortion firms are more profitable. In this case, one would expect high-TFPR firms to be more likely to enter despite low TFPQ. The estimated selection line is reassuring in this respect – the basic modeling of frictions as taxes seems to be compatible with the data.

I refer to (Hsieh and Klenow 2009) for a more detailed discussion of these measures.

[14]The predictions are generated based on MLE estimates and using MC draws.

Distributional Assumptions and Equilibrium Constraints. — Beyond these qualitative features of the model, the calibration is driven by the combination of the functional form of the distribution of firm heterogeneity and the equilibrium constraint. To evaluate the viability of the functional form assumptions I benchmark my estimation results with data on firm size distributions. To understand why the firm-size distribution is a useful comparison benchmark, it is useful to remember that, in my model, firm size differences are driven by TFPR and TFPQ and selection

$$\ln(py(\omega)) \propto -\left(\frac{1}{1-\gamma}\right)\left[\ln(TFPQ(\omega)) - \ln\left(TFPR(\omega)\right)\right]$$
$$\text{subject to selection:}\;\; \ln(TFPQ(\omega)) - \ln\left(TFPR(\omega)\right) \geq -\bar{z}_J$$

In other words, according to the model the shape of the firm-size distribution is driven by the estimated distribution of TFPR and TFPQ as well as selection. Firm size distributions are therefore a useful statistic to check the functional form of the overall distributions and the truncation assumed for the model. If either the functional form of the distribution of TFPR/TFPQ or the nature of the selection equation is misspecified, one would expect the data to generate very different firm-size distributions than those estimated by my model. I use a two-sample Kolmogorov-Smirnov test on firm-size distributions in the data versus the firm-size distribution generated by my model to formally test to what degree the data deviate from the model. I compare these two samples of firm-size distribution data under the null hypothesis that both samples are from the same continuous distribution. This null hypothesis can be rejected only for 13 out of 43 sectors. This implies that, for about 70% of the sectors, I cannot reject the hypothesis that the firm-size distribution in the data is generated from the same truncated log-normal that my model generates. Figure 4 illustrates the quality of the fit of my estimates compared to the data. It shows kernel density estimates with simple box kernels for the six largest manufacturing sectors by value added.

To benchmark the performance of my calibration procedure I compare the firm-size distribution implied by MPEC with three alternatives. First, given the recent popularity of the Pareto distribution to model firm heterogeneity, I fit a simple model of Pareto firm sizes on data for the largest 10% of firms and consider the implied size distribution vis-a-vis the data. The two-sample Kolmogorov-Smirnov test cannot be rejected in only 30% of sectors compared to the 70% of the MLE under equilibrium constraints. Second, a model without selection such as that used by (Hsieh and Klenow 2009) with the additional assumption of log-normality is rejected for all sectors by the Kolmogorov-Smirnov test. Third, a model with selection and log-normality of firm heterogeneity but without equilibrium constraints cannot be rejected in only about 50% of the cases. I conclude from this sequence of tests that the combination of functional form to model firm heterogeneity and the use of equilibrium constraints seem to perform reasonably well compared to common alternatives.

D. Empirical Results

Parameters for the underlying distribution of heterogeneity are estimated by 4 digit sector. To get an overall impression of these estimates table 3 contrasts specific percentiles of the estimated parameters with the same moment using the baseline Hsieh-Klenow methodology. The bottomline message from table 3 is that as discussed in section I.C., the selection correction increases the estimated dispersions and reduces the correlation of frictions and efficiency for nearly all 4 digit sectors.

As predicted by the theory, the latent dispersion of frictions is higher than the observed dispersion, while the underlying correlation of TFPR and TFPQ is lower than the observed correlation.

IV. Estimates of Aggregate TFP and Welfare Effects

This section quantifies the aggregate TFP and welfare effects. 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 get as close as possible to the exercises in the previous literature, I make three 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. Second, I consider the welfare gains from equalizing marginal products in one sector at a time not allowing inter-sectoral reallocation of resources as previously discussed. I then calculate the welfare gains from removing distortions by sector and then sum these up, weighted by the sectoral value-added shares, to arrive at the manufacturing-wide welfare gains. Third, I leave the aggregate factor endowments constant, so results are not confounded by the effect of capital or human capital accumulation.

Proposition 2 below summarizes the aggregate sector-level TFP consequences of the extended setup and makes comparisons with the simplified model in section I. possible. The proposition displays TFP in real consumption units. As in section I., the first term summarizes the aggregate scale effects, while the second term captures aggregate TFP. The main difference to the simple model in section I. is that the overall firm-level distortion is a geometric composite of output and capital wedges.

Proposition 2: Sectoral TFP

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

$$\frac{Y_s}{L_s} = \underbrace{s_{e,s}^{\gamma}(1-s_{e,s})^{1-\gamma}\left(\frac{K_s}{L_s}\right)^{\alpha_s\gamma}}_{\text{aggregate scale}} \underbrace{E\left[A_s(\omega)^{\frac{1}{1-\gamma}}\left[\frac{1-\bar{\tau}_{Y,s}}{1-\tau_{Y,s}(\omega)}\right]^{-\frac{1}{1-\gamma}}\left[\frac{1+\tau_{K,s}(\omega)}{1+\bar{\tau}_{K,s}}\right]^{-\alpha_s\frac{\gamma}{1-\gamma}}\middle|\Pi_s(\omega)\geq w_s\right]^{1-\gamma}}_{\text{TFP}}$$

Proof: see Appendix

The separation of intensive margin misallocation, extensive margin misallocation and aggregate scale effects is done according to the following principles. First, for intensive margin effects, I basically follow Hsieh and Klenow (2009) and calculate the aggregate TFP differences implied by removing firm level frictions but keeping the set of currently operating firms fixed. Second, for extensive margin misallocation effects, I contrast aggregate productivity under the TFPQ composition of firms in the distorted equilibrium with aggregate productivity under the TFPQ composition of firms in the frictionless equilibrium. In other words, I remove Zombie firms and make Shadow firms active. Note that this will also trigger market share reallocations, as TFPQ of Shadows is typically larger than TFPQ of Zombies. These reallocations are a key part of extensive margin misallocation as they are implied by a changing set of firms and therefore the TFPQ composition. Third, for aggregate scale effects, I compare the fraction of entrepreneurs in the distorted equilibrium with the fraction of entrepreneurs in the frictionless equilibrium.

A. *Benchmark Misallocation Effects*

Table 4 summarizes the key welfare results. The first entry in the first column displays the welfare gains from equalizing marginal revenue products across the currently-existing set of firms, as done also by (Hsieh and Klenow 2009). Removing intensive-margin misallocation can raise aggregate TFP by close to 80%. This is in line with quantitative findings by (Hsieh and Klenow 2009) for China and India, where such gains are around 80-100%.[15] Column 2 displays the implied TFP gain from removing extensive margin misallocation. The aggregate impact is sizable: under the counterfactual efficiency composition of firms implied by a frictionless equilibrium, aggregate productivity would be 37% higher. In other

[15]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%.

words, the same set of measured distortions implies that aggregate productivity losses from misallocation across firms are 47% larger than current estimates, due to efficiency composition shifts. These estimates suggest that the overall real TFP gains are huge with about 145%.

The sectoral distribution of TFP losses in the benchmark case is shown in Figure 5. The y-axis displays the percentage 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.

Implied aggregate scale effects are reported the last column of the first row of table 4. These effects are very modest compared to intensive margin or extensive margin misallocation effects. These aggregate scale effects tend to offset the misallocation losses, but only to a small degree.

B. Complementarities of Reform

Here I compare the welfare losses that result from a fully distorted market allocation with both output and capital frictions with market allocations that exhibit either only output or only capital wedges. This analysis is in the spirit of identifying the "most important bottlenecks", as in (Ricardo Hausman, Dani Rodrik and Andres Velasco 2005). Additionally, my analysis identifies important complementarities between different types of reform.

Rows 2 and 3 of Table 4 illustrate welfare costs of different wedge types in isolation. First, note that extensive margin misallocation losses are different when considering allocations with only output wedges with allocations that only exhibit capital wedges. Output frictions are more positively correlated with managerial productivity in the data – so the active Zombie firms are of particularly low-efficiency while inactive Shadow firms are of particularily high-efficiency.[16]

Second, there are interactions between the two types of distortions. Under independence of capital wedges, output wedges and managerial productivity, it follows that

$$E\left[\left(\frac{1+\tau_{K,s}(\omega)}{1+\bar{\tau}_{K,s}}\right)^{-\alpha_s\frac{\gamma}{1-\gamma}}\left(\frac{1-\tau_{Y,s}(\omega)}{1-\bar{\tau}_{Y,s}}\right)^{-\frac{\gamma}{1-\gamma}}A_s(\omega)^{\frac{\gamma}{1-\gamma}}\middle|\Pi_s(\omega)\geq w_s\right]^{1-\gamma}$$

$$=E\left[\left(\frac{1+\tau_{K,s}(\omega)}{1+\bar{\tau}_{K,s}}\right)^{-\alpha_s\frac{\gamma}{1-\gamma}}\middle|\Pi_s(\omega)\geq w_s\right]^{1-\gamma}\cdot E\left[\left(\frac{1-\tau_{Y,s}(\omega)}{1-\bar{\tau}_{Y,s}}\right)^{\frac{\gamma}{1-\gamma}}\middle|\Pi_s(\omega)\geq w_s\right]^{1-\gamma}$$

$$\times E\left[A_s(\omega)^{\frac{1}{1-\gamma}}\middle|\Pi_s(\omega)\geq w_s\right]^{1-\gamma}$$

[16]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 smaller firms. In contrast, output wedges might result from large and efficient firms facing exproporation risk by local public officials.

In other words, under independence, misallocation gains from complete removal of both wedges together are the same as the product of removing one friction at a time. This is not the case here, as can be seen in the first three rows of Table 4. The reason is that frictions are mutually correlated and also pairwise correlated with efficiency. In fact, output and captal distortions are negatively correlated in the data. 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. 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 of the output friction on the capital wedge. Without this offset, misallocation can be made *worse*. These types of surprising effects are well known in the theoretical literature on Second Best paths of reform, since (Richard G. Lipsey and Kelvin Lancaster 1957).

C. The Role of Correlations of Distortions and Productivity

The last row of table 4 reports the misallocation losses from a market allocation that exhibits the same dispersion of output and capital wedges as in the data, but removes all correlation between TFPQ and the frictions.[17] This counterfactual exercise therefore highlights the importance of the covariance of TFPR and TFPQ.

As can be seen, intensive margin misallocation losses are almost unaffected by the fact that distortions are completely random now. This is to be expected, as most of the intensive margin misallocation losses are likely driven by the dispersion of revenue products rather than the fact that more efficient firms are more distorted.[18] In contrast, extensive margin misallocation and aggregate scale effects are different in nature now. Starting with aggregate scale effects, without a correlation of frictions with productivity the removal of distortions will decrease the overall dispersion of firm heterogeneity, therefore implying that there are less very productive firms which are thougher competitors. This makes it easier for many less productive entrepreneurs to become active, thereby relaxing decreasing returns to scale from limited managerial span of control. For the same reason, the extensive magin allocation effect now becomes negative, as many low productivity entrepreneurs become active as distortions are removed. This effect is similar to the "productivity curvature" effect emphasized by (Costas Arkolakis, Svetlana Demidova, Andres Rodriguez-Claire and Peter J. Klenow 2008) in the context of trade liberalizations in models of monopolistic competition and product differentiation with heterogeneous firm productivity as (Melitz 2003).

[17]However, the correlations between output and capital frictions are held constant.

[18]Since the tax revenue from distortions is fully compensated, higher distortions on more efficient firms will generate more tax revenue, which is then lump-sum redistributed to households.

V. Robustness

The baseline counterfactual exercises of removing the full set of distortions from the last section are informative about the quantitative implications of my estimates and are helpful for clarifying the economic mechanisms through which these estimates work. However, the analysis of the last section also imposes several strong assumptions. This section therefore provides a discussion of the importance of these assumptions for the size of different welfare effects.

A. Measurement Error

Among the strongest assumptions imposed in the literature building on (Hsieh and Klenow 2009) is the premise that all measured disperions of TFPR in the data are the result of distortions instead of measurement error. Several studies such as (M. Rotemberg and K. White 2017) and (M. Bils, P. Klenow and C. Ruane 2020) shed doubt on the belief that TFPR dispersions are mostly driven by actual frictions. Indeed (Rotemberg and White 2017) argue that micro data from developing economies is likely to exhibit higher levels of measurement error than the US data, though (Bils, Klenow and Ruane 2020) find the opposite. To evaluate the importance of measurement error for my welfare calculations, I propose to reduce my estimated dispersions of TFPR and TFPQ under the assumption that the degree of measurement error equal across industries. To confirm the plausibility of this approach, I calculate the implied size dispersion of all active firms in the frictionless equilibrium across sectors and compare it with the reported firm size dispersion in the relatively undistorted US economy. For US data, there are not only several different published estimates of the firm size dispersion in manufacturing, but also already existing approaches to correct for measurement error. The first two rows in table 5 report estimates for establishment size dispersions, measured as the standard deviation of log employment in US manufacturing. These estimates are broadly similar, with the first estimates coming from (I. Kondo, L. Lewis and A. Stella 2018), who fit a log-normal distribution on establishment data from the Longitudinal Business Database (LBD) of the US Census. The second estimate builds on the reported standard deviation of log TFPQ by (Hsieh and Klenow 2009) and implies a broadly similar dispersion using their parameter of $\gamma = 0.5$. The next three rows then report different estimates of measurement error corrected firm size dispersions. First, (Bils, Klenow and Ruane 2020) use the relationsship between output and input growth to correct for measurement error in US Census of Manufacturing data and find that around 60% of estimated dispersions are due to measurement error. Second, (N. Bloom, E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten and J. Van Reenen 2019) use data from establishments that accidentally responded twice to the US Annual Survey of Manufacturing to calculate measurement error and find that around half the variation is driven by measurement error. Third, since the reported survey items in (Bloom et al. 2019) are rather novel, it is plausible that actual measurement error for regularly reported survey items such as value added, capital and labor is lower than 50%. I therefore also report the implied US firm size dispersion if measurement error is only 40% of the reported data.

For the Indonesian data, I strive to be conservative and assume that 50% of the standard deviations of TFPR or TFPQ are driven by classical measurement error. Given this assumption, I cut all standard deviations by 50% and then recalculate the counterfactual frictionless equilibria for each sector. To measure firm size, I use log number of workers, which is the firm size measure in (Kondo, Lewis and Stella 2018) and calculate firm size dispersions for all active firms in the frictionless equilibrium by sector. The last three colums show the implied standard deviation of firm sizes across sectors. It confirms that the median firm size dispersion of the frictionless Indonesian economy is broadly consistent with the measurement-error corrected firm size dispersions of the US economy. Therefore I proceed with this value for the remainder of this section.

Table 6 reports the welfare implications of removing micro-distortions under different assumptions on measurement error. The first row cuts the standard deviation of TFPR in half and then calculates the welfare losses from misallocation compared to a frictionless benchmark. Unsurprisingly, misallocation losses are are now only around 20%, a value much lower than in table 4. However, note that the relative size of extensive margin misallocation effects is now higher – exceeding 100% relative to intensive margin effects. Furthermore, it seems plausible that measurement error will not only affect the measurement of TFPR but also of TFPQ, since both variables rely on the same data items. Indeed my comparison of the frictionless firm size distributions used reduced dispersions in TFPQ only. I therefore report welfare losses of misallocation if dispersions of TFPR and TFPQ are cut, in the second row of table 6. This significantly reduces the size of extensive margin misallocation effects, since with lower underlying TFPQ dispersions, there are less high-productivity Shadows that can become active.

Up to this point, I have modeled measurement error as only biasing measured dispersions upwards, while leaving correlations constant. However, with classical measurement error, covariances should be unaffected by measurement error, so that the underlying correlations are actually higher.[19] I therefore repeat the calculations of welfare losses from misallocation while cutting TFPR/TFPQ dispersions but keeping covariances mostly unchanged, which implies that the underlying correlations will increase. To ensure that variance-covariance matrices of TFPQ and TFPR are still positive definite, I do adjust the correlation of TFPR and TFPQ downward if variance-covariance matrices become negative semi-definite. This can happen if the implied correlations exceed 1. This exercise has an additional conceptual advantage. Low correlations of frictions and managerial productivity imply very large size reversals of distorted firms in case

[19]Given two random variables X_1, X_2 and classical measurement error $\tilde{X}_k = X_k + \epsilon_k$ for $k = 1, 2$ with ϵ_i *i.i.d.*, it follows that $Cov(X_1, X_2) = Cov(\tilde{X}_1, \tilde{X}_2)$

of reform, as emphasized by (H. Hopenhayn 2014). Increasing the correlation of TFPQ and TFPR will limit such size reverals, which could be seen as an attractive feature from the perspective of (Hopenhayn 2014). The last two rows of table 6 present the results of full liberalizations under the assumption that half of the standard deviations of log TFPR and log TFPQ are driven by classical measurement error. In these cases, extensive margin misallocation effects become very large, as covariances are largely unchanged from the baseline exercises, implying the potential presence of very productive Shadow firms.

To summarize, throughout all of the re-calculations of misallocation losses to account for measurement error, the relative contribution of extensive margin misallocation compared to intensive margin misallocation either stayed similar or strongly increased. This implies that measurement error is unlikely to drive the quantitative importance of extensive margin misallocation effects.

B. *Span of Control*

I followed (Hsieh and Klenow 2009) and assumed that $\gamma = 0.5$ for the sake of comparability. However, it should be noted that this value for the managerial span of control parameter implies net profitability of 100%. In contrast, (A. Atkeson and P. Kehoe 2005) argue that curvature in utility and production implies a value of $\gamma = 0.8$, which corresponds to a profit rate of 25%. Table 7 reports the misallocation losses for different values of γ. For each of the rows, I re-estimate the distribution of TFPR and TFPQ with my MPEC procedure and then recalculate the counterfactual frictioness equilibrium using the new value of γ. As the second column of table 7 shows, a value of $\gamma = 0.5$ is conservative as compared to a value of $\gamma = 0.8$. In this case, higher values of γ (less decreasing returns to scale) imply more substitutability in production, which in turn increases the implied misallocation losses of measured frictions. Furthermore, the same is true of extensive margin misallocation and aggregate scale effects.

In contrast, a strong reduction in returns to scale to $\gamma = 0.3$ implies more complementarity in production. Each additional establishment helps to offset decreasing returns to scale in the aggregate, which is why with the removal of distortions, the number of entrepreneurs and therefore aggregate scale effects of welfare increase. In this case, extensive margin effects are again negative, as in the last row in table 4. The reason here is that a value of $\gamma = 0.3$ implies a profit rate of 200%, which makes it extremely attractive for entrpreneurs to become active, even if managerial productivity is low. As a result, the frictionless equilibrium features many very low productivity firms, which strongly turns the extensive margin effect negative.

C. *Effective Labor Force*

The third assumption needed to calibrate my model to the Indonesian microdata is the assumption on κ, the efficiency parameter of (unskilled) production labor relative to (skilled) non-production labor. My baseline value of $\kappa = 0.5$ used evidence from average wage premia in my data to calibrate this parameter. Here I explore how the welfare calculations change for different values of κ. For each of the different values, I re-estimate the distribution of frictions and managerial productivity and re-calculate the counterfactual frictionless equilibrium. Table 7 reports the results. As can be seen, estimates of the different misallocation effects are not strongly affected by different values of κ.

VI. Conclusion

This paper quantitatively explores the importance of selection in the presence of misallocation of resources across firms. I show that one does not need direct barriers for selection to have quantitatively important aggregate productivity effects. The same set of measured distortions implies a much larger welfare loss due to shifting the efficiency distribution of active firms from potentially high efficiency inactive Shadows to low-efficiency but active Zombies.

Extensive margin misallocation strongly magnifies the welfare effects of micro-level misallocation, when compared to estimates in (Hsieh and Klenow 2009). A full removal of distortions implies that extensive margin misallocation gains can boost TFP in Indonesian microdata by 37%. Given intensive margin welfare gains from removing distortions of almost 80%, this implies that extensive margin misallocation magnifies misallocation losses from micro-distortions by at least 47%. Additionally, the relative magnification of welfare effects of misallocation are even higher when taking measurement error into account. Measurement error corrected extensive margin misallocation losses are typically at least twice as high as intensive-margin misallocation losses.

REFERENCES

Ackerberg, D., K. Caves, and G. Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica.*

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

Arkolakis, Costas, Svetlana Demidova, Andres Rodriguez-Claire, and Peter J. Klenow. 2008. "Endogenous Variety and the Gains from Trade." *American Economic Review Papers and Proceedings.*

Atkeson, A., and P. Kehoe. 2005. "Modeling and Measuring Organizational Capital." *Journal of Political Economy.*

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

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review.*

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

Bhattacharya, D., N. Guner, and G. Ventura. 2013. "Distortions, Endogenous Managerial Skills and Productivity Differences." *Review of Economic Dynamics.*

Bils, M., P. Klenow, and C. Ruane. 2020. "Misallocation or Mismeasurement?" *Working Paper, Stanford University.*

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen. 2019. "What Drives Differences in Management Practices?" *American Economic Review.*

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

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

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

Fattal Jaef, Roberto. 2018. "Entry and Exit, Multi-Product Firms, and Allocative Distortions." *American Economic Journal: Macroeconomics.*

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

Foster, L., J. Haltiwanger, and C. Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review.*

Gabler, A., and M. Poschke. 2013. "Experimentation by Firms, Distortions and Aggregate Productivity." *Review of Economic Dynamics.*

Guner, Nezih, Gustavo Ventura, and Yi Xu. 2008. "Macroeconomic implications of size-dependent policies." *Review of Economic Dynamics.*

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

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

Hopenhayn, H. 2014. "On the Measure of Distortions." *NBER Working Paper.*

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

Hurst, E., and B. Pugsley. 2011. "What do small businesses do?" *Brookings Papers on Economic Activity.*

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

Kondo, I., L. Lewis, and A. Stella. 2018. "On the US Firm and Establishment Size Distributions." *Working Paper, Federal Reserve Board.*

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

Lucas, R. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics.*

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

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

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

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

Peters, Michael. 2013. "Heterogeneous Mark-Ups and Endogenous Misallocation." *mimeo London School of Economics.*

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

Rotemberg, M., and K. White. 2017. "Measuring Cross-Country Differences in Misallocation." *CES Working Paper.*

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

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

Parameter	Value	Explanation
γ	0.5	Span of control parameter Hsieh and Klenow (2009)
α_s	US data	Sectoral capital share Bartelsman et al. 2000
κ	0.5	Conversion factor of (unskilled) production labor into (skilled) non-production labor (from average wage premium)

Table 1—: Calibrated Values

Table 2—: External validity using exit in year after entry

	(1)	(2)	(3)	(4)	(5)
`log TFPR`	0.182**		0.454***	0.871***	1.080***
	(0.080)		(0.161)	(0.318)	(0.350)
`log TFPQ`		0.043	-0.190**	-0.446**	-0.529**
		(0.048)	(0.095)	(0.195)	(0.213)
`log Capital`				0.112	0.158**
				(0.073)	(0.079)
Obs.	622	622	622	622	596
Industry FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes

Notes: Sample includes entering (age zero) establishments only. Dependent variable is exit of establishment from the sample in the next year. Robust standard errors are provided in parentheses.

Table 3—: Estimates across sectors

Percentile	HK: σ_τ	MPEC: σ_τ	HK: σ_A	MPEC: σ_A	HK: $\rho_{A,\tau}$	MPEC: $\rho_{A,\tau}$
90^{th}	1.02	1.33	1.38	2.82	0.92	0.48
75^{th}	0.96	1.17	1.27	2.70	0.90	0.46
50^{th}	0.80	1.01	1.09	2.50	0.87	0.42
25^{th}	0.71	0.89	0.95	2.20	0.81	0.37
10^{th}	0.65	0.78	0.71	1.95	0.76	0.14

Notes: Table displays percentiles of estimated underlying parameters of the distribution of efficiency and distortions across sectors. Parameters are: σ_τ dispersion of distortions, σ_A: dispersion of efficiency, $\rho_{A,\tau}$: correlation of efficiency and distortions. "HK" denotes values of parameters using Hsieh-Klenow (2009) methodology and "MPEC" is the selection-corrected maximum likelihood estimator with equilibrium constraints.

Table 4—: Productivity Effects from Removing Distortions

	Decomposition			
Full liberalization of..	Intensive-Margin	Selection	aggregate Scale	Welfare
(1) Fully distorted economy	79.12%	37.20%	−1.72%	141.52%
(2) Output distortions only	114.25%	19.85%	−3.11%	148.79%
(3) Capital distortions only	18.29%	5.89%	−0.69%	24.39%
(4) Random distortions only	91.61%	−10.00%	41.01%	143.17%

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 overall effect is calculated as $2.54 = 1.94 \cdot 1.39 \cdot 0.94$

Table 5—: Calibration of Measurement Error

Moment	Value
Firm (establishment) size dispersions (US data)	
Kondo et al., 2018[1]	1.76
Hsieh and Klenow 2009[2]	1.7
Measurement-error corrected size dispersions	
Bils et al. 2020[3]	0.64
Bloom et al. 2019[4]	0.8
Low measurement error[5]	1.03
Frictionless equilibrium size dispersion across sectors (50% measurement error of TFPQ)	
25th Percentile	0.76
Median	0.83
75th Percentile	0.93

Note: [1] Based on log employment data of establishments in manufacturing in the Longitudinal Business Database (LBD), table 10 in Kondo et al. 2018.
[2] Based on reported standard deviation of log TFPQ from the US Census of Manufacturing (table I in Hsieh and Klenow 2009) and assuming $\gamma = 0.5$.
[3] Measurement error correction based on correlation of revenue and inputs, which implies for the US that 60% of TFPQ is measurement error.
[4] Measurement error correction based on double reporting of US Annual Survey of Manufacturing, which implies 50% measurement error.
[5] Assumed value of 40% measurement error for US data.

Table 6—: Analysis of Measurement Error

	Decomposition			
	Intensive-Margin	Selection	aggregate Scale	Welfare
(1) σ_Y, σ_K mismeasured	18.23%	22.73%	−13.90%	24.93%
(2) $\sigma_Y, \sigma_K, \sigma_A$ mismeasured	22.19%	9.12%	−3.07%	29.24%
(3) σ_Y, σ_K mismeasured and limited size reverals	9.94%	59.49%	−31.52%	20.07%
(4) $\sigma_Y, \sigma_K, \sigma_A$ mismeasured and limited size reversals	14.25%	26.78%	−18.88%	17.49%

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.

Table 7—: Analysis of Span of Control

	Decomposition			
	Intensive-Margin	Selection	Aggregate Scale	Welfare
(1) .. $\gamma = 0.5$ [1]	79.12%	37.20%	−1.72%	141.52%
(2) ..$\gamma = 0.8$ [2]	194.57%	61.03%	−24.47%	258.27%
(3) ..$\gamma = 0.3$	43.30%	−55.43%	161.18%	66.81%

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.
[1] Calibrated value from Hsieh and Klenow 2009
[2] Calibrated value from Atkeson and Kehoe 2005

Table 8—: Analysis of Effective Labor Force

	Decomposition			
	Intensive-Margin	Selection	Aggregate Scale	Welfare
(1) $\kappa = 0.5$	79.12%	39.20%	−1.72%	141.52%
(2) $\kappa = 0.75$	81.86%	37.87%	−1.47%	130.18%
(3) $\kappa = 0.25$	83.26%	44.01%	−14.98%	124.37%

Notes: Effective labor units are measured as $L_{E,s}(\omega) = L_{NP,s}(\omega) + \kappa \cdot L_{P,s}(\omega)$, where $L_{NP,s}$ is the number of non-production workers and $L_{P,s}$ the number of production workers. 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.

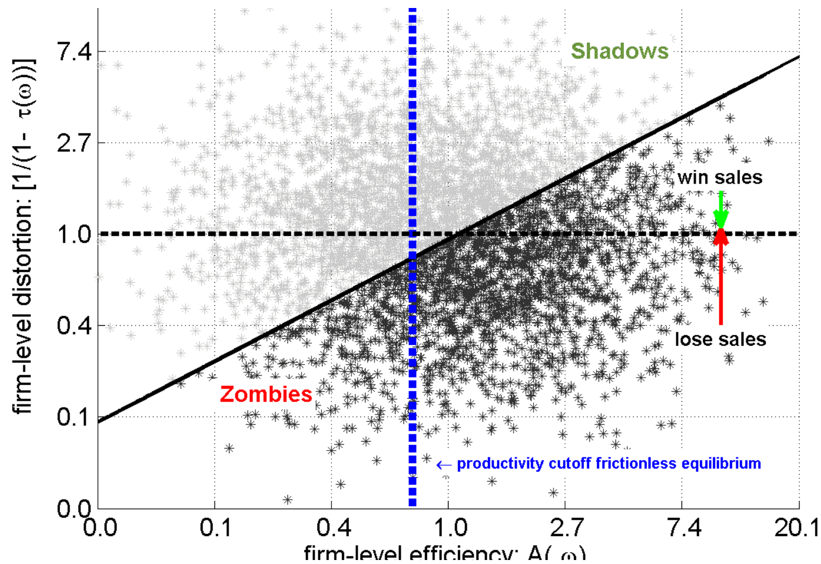

Figure 1. : Firm types and survival, based on example simulation. Solid line captures selection line in distorted equilibrium. Grey dots capture inactive firms, black dots active firms in the distorted equilibrium. Vertical dashed line captures selection cutoff in the frictionless equilibrium.

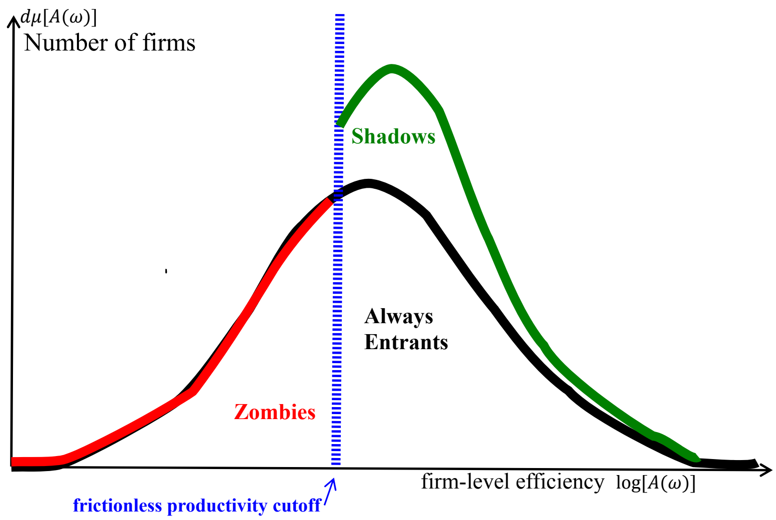

Figure 2. : Change in the efficiency composition of firms from distorted equilibrium (Always Entrants + Zombies) to frictionless equilibrium (Always Entrants + Shadows)

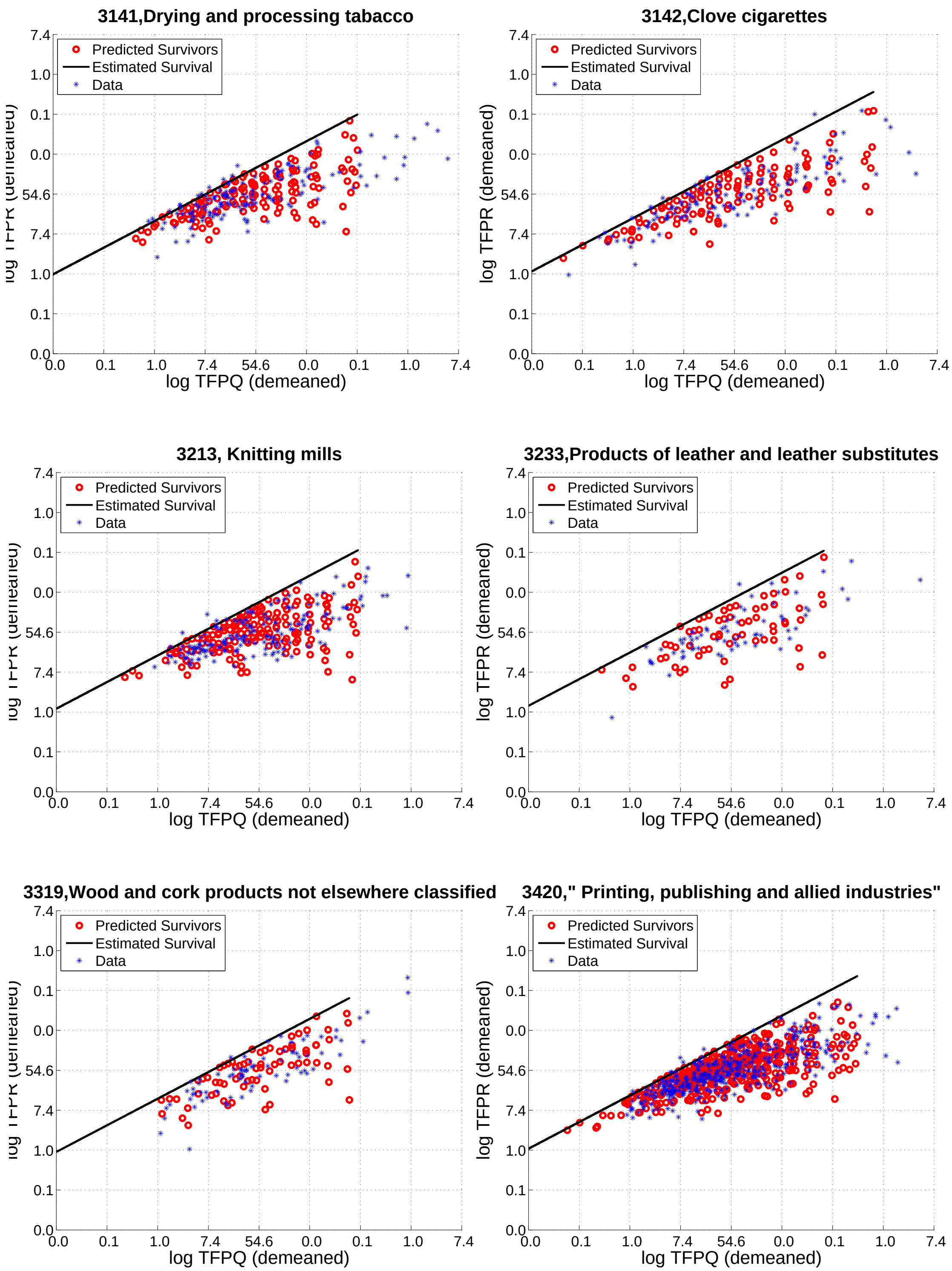

Figure 3. : Illustration of estimated survival pattern versus data for six largest manufacturing sectors.

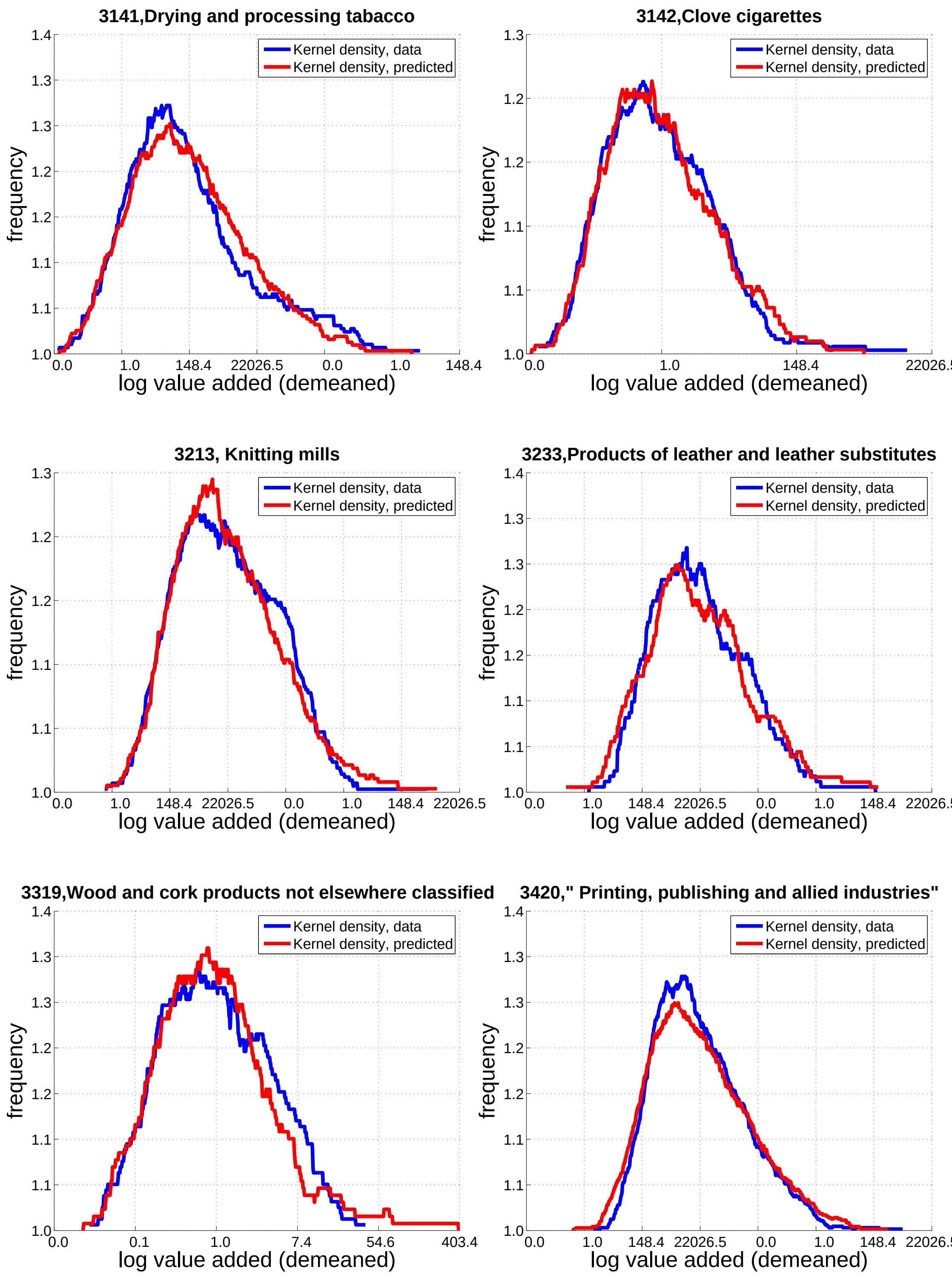

Figure 4. : Illustration of predicted firm size distribution versus data. Six largest manufactuing sectors by value added. Kernel density estimates.

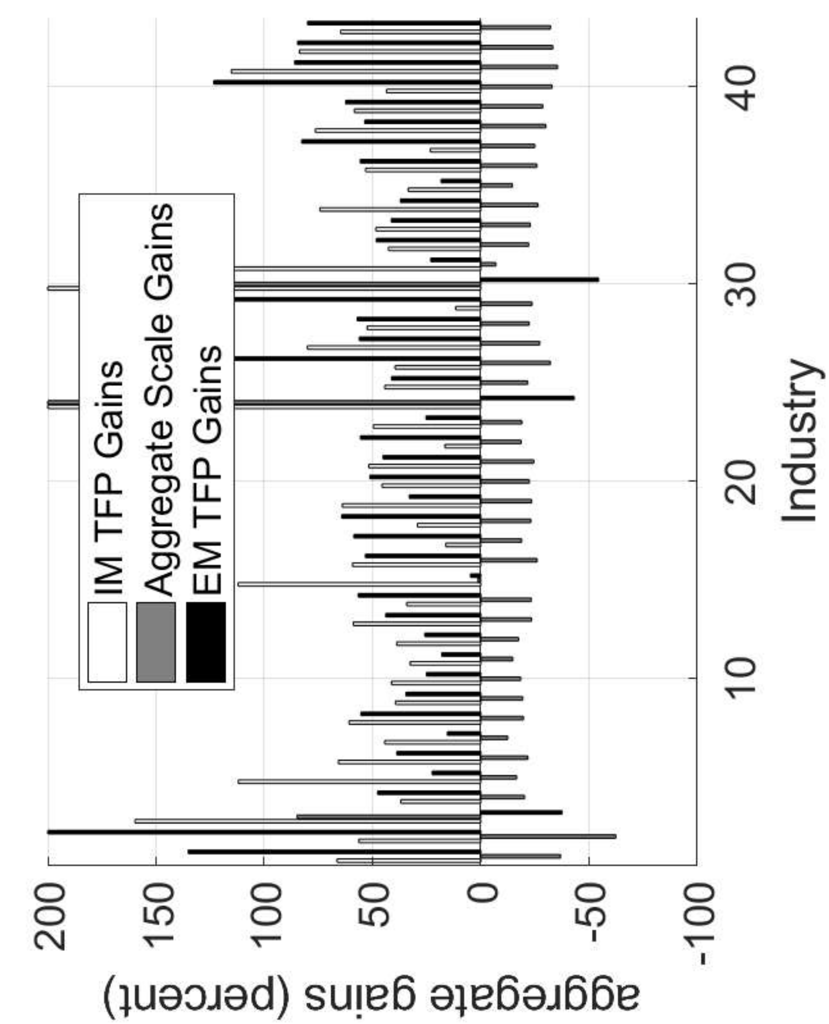

Figure 5. : Distribution of aggregate welfare gains by sector

Online Appendix

Micro-level Misallocation and Selection

Mu-Jeung Yang

To save on notation, I supress the sectoral index s in all following calculations.

1 Derivation of Aggregate Productivity

Profit maximization and Optimal size

Profits are given by

$$\begin{aligned} \max_{\{K(\omega),L(\omega)\}} \Pi(\omega) &= [1-\tau_Y(\omega)]\cdot py(\omega) - wL(\omega) - [1+\tau_K(\omega)]\cdot R\cdot K(\omega) \\ \text{subject to: } & y(\omega) = A(\omega)\cdot\left[K(\omega)^{\alpha}L(\omega)^{1-\alpha}\right]^{\gamma} \end{aligned} \tag{1}$$

Taking first order conditions and solving for optimal size gives

$$\begin{aligned} \frac{wL(\omega)}{1-\alpha} &= [\gamma p(1-\tau_Y(\omega))A(\omega)]^{\frac{1}{1-\gamma}}\left[\left(\frac{(1+\tau_K(\omega))R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right]^{-\frac{\gamma}{1-\gamma}} \\ \frac{(1+\tau_K(\omega))RK(\omega)}{\alpha} &= [\gamma p(1-\tau_Y(\omega))A(\omega)]^{\frac{1}{1-\gamma}}\left[\left(\frac{(1+\tau_K(\omega))R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right]^{-\frac{\gamma}{1-\gamma}} \\ py(\omega) &= \left(\frac{1}{\gamma}(pA(\omega))^{\frac{1}{1-\gamma}}\right)\left[\left(\frac{(1+\tau_K(\omega))R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right]^{-\frac{\gamma}{1-\gamma}} \end{aligned} \tag{2}$$

Aggregation I: factor payments

Labor market clearing can be written as

$$(1-s_e)wL = \int_{\Pi(\omega)\geq w} wL(\omega)d\omega \tag{3}$$

where L is a given labor supply. Equation can also be rewritten as

$$pY = \frac{1}{\gamma(1-\alpha)}\frac{1}{1-\bar{\tau}_Y}w(1-s_e)L \tag{4}$$

with the average output distortion is defined by

$$(1-\bar{\tau}_Y) = \int_{\Pi(\omega)\geq w} (1-\tau_Y(\omega)) \left(\frac{py(\omega)}{pY}\right) d\omega \tag{5}$$

and aggregate output in (4) is given by

$$Y = \int_{\Pi(\omega)\geq w} y(\omega) d\omega \tag{6}$$

Using (2) in (1) one obtains

$$\frac{w}{1-\alpha} = (1-s_e)^{-\frac{1-\gamma}{1-\alpha\gamma}} (\gamma p)^{\frac{1}{1-\alpha\gamma}} \left(\frac{R}{\alpha}\right)^{-\frac{\alpha\gamma}{1-\alpha\gamma}} \Sigma_L^{\frac{1-\gamma}{1-\alpha\gamma}} \tag{7}$$

with

$$\Sigma_L = E\left[A(\omega)^{\frac{1}{1-\gamma}} \left(\frac{1}{1-\tau_Y(\omega)}\right)^{-\frac{1}{1-\gamma}} (1+\tau_K(\omega))^{-\alpha\frac{\gamma}{1-\gamma}} \middle| \Pi(\omega) \geq w\right] s_e \tag{8}$$

Similarily, capital market clear is given by

$$RK = \int_{\Pi(\omega)\geq w} RK(\omega) d\omega \tag{9}$$

which can be rewritten as

$$pY = \frac{1}{\gamma\alpha}\left(\frac{1+\bar{\tau}_K}{1-\bar{\tau}_Y}\right) RK \tag{10}$$

where

$$\frac{1-\bar{\tau}_Y}{1+\bar{\tau}_K} = \int_{\Pi(\omega)\geq w} \left(\frac{1-\tau_Y(\omega)}{1+\tau_K(\omega)}\right)\left(\frac{py(\omega)}{pY}\right) d\omega \tag{11}$$

Combining (2) in (9) implies

$$\frac{R}{\alpha} = (\gamma p)^{\frac{1}{1-\gamma+\alpha\gamma}} \left(\frac{w}{1-\alpha}\right)^{-\frac{(1-\alpha)\gamma}{1-\gamma-\alpha\gamma}} \left(\frac{L}{K}\right)^{\frac{1-\gamma}{1-\alpha+\alpha\gamma}} \Sigma_K^{\frac{1-\gamma}{1-\alpha+\alpha\gamma}} \tag{12}$$

with

$$\Sigma_K = E\left[A(\omega)^{\frac{1}{1-\gamma}} \left(\frac{1}{1-\tau_Y(\omega)}\right)^{-\frac{1}{1-\gamma}} (1+\tau_K(\omega))^{-\frac{1-\gamma+\alpha\gamma}{1-\gamma}} \middle| \Pi(\omega) \geq w\right] s_e \tag{13}$$

To obtain aggregate production, note that $pY = (pY)^{\alpha}(pY)^{1-\alpha}$ and use (4) and (18) to obatain

$$\frac{Y}{L} = \left(E\left[\Sigma_K \middle| \Pi(\omega) \geq w\right]^{\alpha} E\left[\Sigma_L \middle| \Pi(\omega) \geq w\right]^{1-\alpha} \right)^{1-\gamma} \left(\frac{(1+\bar{\tau}_K)^{\alpha}}{1-\bar{\tau}_Y} \right) \times \left(\frac{K}{L} \right)^{\alpha\gamma} s_e^{1-\gamma}(1-s_e)^{\gamma(1-\alpha)} \tag{14}$$

Aggregation II: output

Combine equation (6) and (2) gives

$$Y = \left[\frac{1}{\gamma p} \left(\frac{R}{\alpha} \right)^{\alpha} \left(\frac{w}{1-\alpha} \right)^{1-\alpha} \right]^{-\frac{\gamma}{1-\gamma}} E\left[\Sigma_Y \middle| \Pi(\omega) \geq w\right] s_e L \tag{15}$$

with

$$E\left[\Sigma_Y \middle| \Pi(\omega) \geq w\right] = E\left[A(\omega)^{\frac{1}{1-\gamma}} \left[\frac{(1+\tau_K(\omega))^{\alpha}}{1-\tau_Y(\omega)} \right]^{-\frac{\gamma}{1-\gamma}} \middle| \Pi(\omega) \geq w \right] \tag{16}$$

Using (7) and (12) in (15) to get

$$\frac{Y}{L} = \frac{E\left[\Sigma_Y \middle| \Pi(\omega) \geq w\right]}{\left(E\left[\Sigma_K \middle| \Pi(\omega) \geq w\right]^{\alpha} E\left[\Sigma_L \middle| \Pi(\omega) \geq w\right]^{1-\alpha} \right)^{\gamma}} \left(\frac{K}{L} \right)^{\alpha\gamma} s_e^{1-\gamma}(1-s_e)^{\gamma(1-\alpha)} \tag{17}$$

Matching coefficients of (14) and (17) gives

$$\left(\frac{(1+\bar{\tau}_K)^{\alpha}}{1-\bar{\tau}_Y} \right) = \frac{\Sigma_Y}{\Sigma_K^{\alpha}\Sigma_L^{1-\alpha}} \tag{18}$$

Then, using (18) in (17) gives

$$\frac{Y}{L} = E\left[A(\omega)^{\frac{1}{1-\gamma}} \left[\frac{1-\bar{\tau}_Y}{1-\tau_Y(\omega)} \right]^{-\frac{1}{1-\gamma}} \left[\frac{1+\tau_K(\omega)}{1+\bar{\tau}_K} \right]^{-\alpha\frac{\gamma}{1-\gamma}} \middle| \Pi(\omega) \geq w \right]^{1-\gamma} \left(\frac{K}{L} \right)^{\alpha\gamma} s_e^{1-\gamma}(1-s_e)^{\gamma(1-\alpha)} \tag{19}$$

2 Derivation of MPEC estimator

This section builds on the previous section to derive the MPEC estimator used in the paper. As before, I supress sector subscripts s to simplify notation.

MLE Objective

From (2) it follows that

$$\begin{aligned} D_1(\omega) &= py(\omega) \\ &= \left(\frac{1}{\gamma}(pA(\omega))^{\frac{1}{1-\gamma}}\right)\left[\left(\frac{(1+\tau_K(\omega))R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right]^{-\frac{\gamma}{1-\gamma}} \end{aligned} \tag{20}$$

Combining the expressions for factor demands in (2) it also follows that

$$\begin{aligned} D_2(\omega) &= \left[\left(\frac{RK(\omega)}{\alpha}\right)^{\alpha}\left(\frac{wL(\omega)}{1-\alpha}\right)^{1-\alpha}\right] \\ &= (p\gamma)^{\frac{1}{1-\gamma}}\left[\left(\frac{R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right]^{-\frac{\gamma}{1-\gamma}} A(\omega)^{\frac{1}{1-\gamma}}\left(\frac{1}{1-\tau_Y(\omega)}\right)^{-\frac{1}{1-\gamma}}(1+\tau_K(\omega))^{-\frac{\alpha}{1-\gamma}} \end{aligned} \tag{21}$$

as well as

$$\begin{aligned} D_3(\omega) &= \ln\left(\frac{wL(\omega)/(1-\alpha)}{RK(\omega)/\alpha}\right) \\ &= (1+\tau_K(\omega)) \end{aligned} \tag{22}$$

Equation (20), (21), (22) can be rewritten to yield

$$\begin{pmatrix} \ln D_1(\omega) \\ \ln D_2(\omega) \\ \ln D_3(\omega) \end{pmatrix} \propto \begin{bmatrix} -\frac{\gamma}{1-\gamma} & -\alpha\frac{\gamma}{1-\gamma} & \frac{1}{1-\gamma} \\ -\frac{1}{1-\gamma} & -\alpha\frac{1}{1-\gamma} & \frac{1}{1-\gamma} \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} \ln\left(\frac{1}{1-\tau_Y(\omega)}\right) \\ \ln(1+\tau_K(\omega)) \\ \ln A(\omega) \end{pmatrix} \tag{23}$$

which I assume is distributed according to a tri-variate normal distribution with parameters $\mu_i = E[\ln D_i(\omega)], \sigma_{ii} = Var[\ln D_i(\omega)]$ for $i = 1,2,3$ and $\sigma_{ij} = Cov(\ln D_i(\omega), \ln D_j(\omega))$ for $i,j = 1,2,3$ and $i \neq j$. As equation (23), shows, these parameters in turn are functions

of the underlying heterogeneity parameters $\mu_A = E[\ln A(\omega)], \mu_{\tau_Y} = E\left[\ln\left(\frac{1}{1-\tau_Y(\omega)}\right)\right], \mu_{\tau_K} = E[\ln(1+\tau_K(\omega)], \sigma_A = Var[\ln A(\omega)], \sigma_{\tau_Y} = Var\left[\ln\left(\frac{1}{1-\tau_Y(\omega)}\right)\right], \sigma_{\tau_K} = Var[\ln(1+\tau_K(\omega)], \rho_{A\tau_Y} = Corr\left(\ln A(\omega), \ln\left(\frac{1}{1-\tau_Y(\omega)}\right)\right), \rho_{A\tau_K} = Corr\left(\ln A(\omega), \ln(1+\tau_K(\omega))\right)$
$\rho_{\tau_{YK}} = Corr\left(\ln\left(\frac{1}{1-\tau_Y(\omega)}\right), \ln(1+\tau_K(\omega)\right)$

Selection is given by

$$\Pi_s(\omega) \geq w_s \tag{24}$$

which after plugging in (2) and taking the log, gives

$$\left(\frac{1+\gamma}{1-\gamma}\right)\ln\left(\frac{1}{1-\tau_Y(\omega)}\right) + \alpha\left(\frac{\gamma}{1-\gamma}\right)\ln(1+\tau_K(\omega) - \left(\frac{1}{1-\gamma}\right)\ln A(\omega) \leq \ln\kappa_Z \tag{25}$$

where the log truncation threshold $\ln\kappa_Z$ is given by

$$\ln\kappa_Z = \ln\left(\frac{1-\gamma}{w}\right) + \left(\frac{1}{1-\gamma}\right)[\ln p + \gamma\ln\gamma] - \frac{\gamma}{1-\gamma}\ln\left[\left(\frac{R}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\right] \tag{26}$$

The MLE objective with parameters $\theta = \left[\mu_A; \mu_{\tau_Y}; \mu_{\tau_K}; \sigma_A; \sigma_{\tau_Y}; \sigma_{\tau_K}; \rho_{A\tau_Y}; \rho_{A\tau_K}; \rho_{\tau_{YK}}\right]$ under truncation for a single observation can then be written as

$$\ln\left\{\frac{\phi\left(\ln D_1(\omega), \ln D_2(\omega), \ln D_3(\omega)\middle|\theta, w, R\right)}{1-\Phi\left(\kappa_Z\right)}\right\} =$$
$$\left\{\frac{3}{2}\ln(2\pi) + \frac{1}{2}\ln(|\bar{\sigma}|) - \ln\Phi\left(\frac{\ln\kappa_Z - \mu_Z}{\sigma_Z}\right) - \frac{1}{2}\begin{pmatrix}\ln D_1(\omega) - \mu_1\\ \ln D_2(\omega) - \mu_2\\ \ln D_3(\omega) - \mu_3\end{pmatrix}'\bar{\sigma}^{-1}\begin{pmatrix}\ln D_1(\omega) - \mu_1\\ \ln D_2(\omega) - \mu_2\\ \ln D_3(\omega) - \mu_3\end{pmatrix}\right\} \tag{27}$$

with

$$\begin{aligned}\mu_Z &= g\cdot\mu_{\tau_Y} + h\cdot\mu_{\tau_K} + k\cdot\mu_A\\ \sigma_Z^2 &= g^2\sigma_{\tau_Y}^2 + h^2\sigma_{\tau_K}^2 + k^2\sigma_A^2 + 2(g\cdot h\cdot\sigma_{\tau_Y,\tau_K} + h\cdot k\cdot\sigma_{\tau_K,A} + g\cdot k\cdot\sigma_{\tau_Y,A})\\ g &= \frac{1+\gamma}{1-\gamma}, h = \frac{\alpha\gamma}{1-\gamma}, k = \frac{1}{1-\gamma}\end{aligned} \tag{28}$$

Furthermore, $\bar{\sigma}$ is the variance-covariance matrix of $\ln D_1(\omega), \ln D_2(\omega), \ln D_3(\omega)$ and the term

$|\bar{\sigma}|$ the determinant of that variance-covariance matrix. $\Phi()$ denotes the cdf of a standard normal distribution.

Equilibrium Constraints

Equilibrium constraints are given by the terms (8) and (13). To evaluate the truncated power means in these expressions, I use the following Lemmas.

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

$$1_{\{a,b,K\}} = \begin{cases} 1 \text{ if } X^a \cdot Z^b \leq K \\ 0 \text{ if } \text{else} \end{cases} \tag{29}$$

Then it follows that

$$\begin{aligned} E\left[X^m Z^n \cdot 1_{\{a,b,K\}}\right] &= \exp\left\{m\mu_X + n\mu_Z + \frac{1}{2}\left(m^2\sigma_m^2 + n^2\sigma_n^2 + 2mn\sigma_{X,Z}\right)\right\} \\ &\times\Phi\left(\frac{\log K - (a\mu_X + b\mu_Z) - [am\sigma_X^2 + (bm+an)\sigma_{X,Z} + bn\sigma_Z^2]}{\sqrt{a^2\sigma_X^2 + b^2\sigma_Z^2 + 2ab\sigma_{X,Z}}}\right) \end{aligned} \tag{30}$$

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

To apply the Lien and Balakrishnan result to the trivariate lognormal truncated moments in (8) and (13), I use the following result, based in the fact that sums of normal random variables are themselves normally distributed.

Lemma 2

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

$$1_{\{\alpha,\beta,\gamma,K\}} = \begin{cases} 1 \text{ if } X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \leq K \\ 0 \text{ if } \text{else} \end{cases} \tag{31}$$

Then it follows that

$$E\left[X_1^m X_2^n X_3^l \cdot 1_{\{\beta_1,\beta_2,\beta_3,K\}}\right] = E\left[X \cdot Z^c \cdot 1_{\{0,0,1,K\}}\right]$$
$$= \exp\left\{\mu_X + c\mu_Z + \frac{1}{2}\left(\sigma_X^2 + c^2\sigma_Z^2 + c\sigma_{X,Z}\right)\right\} \cdot \Phi\left(\frac{\log K - \mu_Z - [\sigma_{X,Z} + c\sigma_Z^2]}{\sigma_Z}\right) \tag{32}$$

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

$$\log X = a \log X_1 + b \log X_2$$
$$\log Z = \beta_1 \log X_1 + \beta_2 \log X_2 + \beta_3 \log X_3 \tag{33}$$

and the coefficients a, b, c are given by

$$a = m - \beta_1 \frac{l}{\beta_3}, b = n - \beta_2 \frac{l}{\beta_3}, c = \frac{l}{\beta_3} \tag{34}$$

Proof: apply mapping (33) and (34) to reduce the trivariate problem to the bivariate problem of Lemma 1.