

**Department of Economics & Finance**

# **The Macroeconomic Implications of Firm Selection**

Dudley Cooke and Tatiana Damjanovic

Working Paper No. 1, 2018

Department Economics and Finance  
Durham University Business School  
Mill Hill Lane  
Durham DH1 3LB, UK  
Tel: +44 (0)191 3345200  
<https://www.dur.ac.uk/business/research/economics/>

© Durham University, 2018

# The Macroeconomic Implications of Firm Selection\*

Dudley Cooke<sup>†</sup>

University of Exeter

Tatiana Damjanovic<sup>‡</sup>

Durham University

**Abstract:** This paper studies the macroeconomic implications of firm selection in a model with monopolistic competition and translog preferences. Firm selection magnifies the impact of aggregate technology shocks. Magnification is limited by diminishing returns to new varieties and misallocation. We provide analytical results linking selection, diminishing returns, and misallocation with measured total factor productivity (TFP) and the distribution of firm-level productivity. A calibrated version of our model suggests the contribution of firm selection to variations in TFP is over 20 percent.

**JEL Classification:** E32, L11

**Keywords:** Firm Selection, Translog Preferences, Productivity Distribution

---

\*We thank Hugo Hopenhayn, Omar Licandro, Julian Neira, and Rish Singhania for comments and suggestions.

<sup>†</sup>Department of Economics, University of Exeter, Streatham Court, Rennes Drive, Exeter EX4 4PU, United Kingdom. Email: d.cooke@exeter.ac.uk

<sup>‡</sup>Department of Economics, Durham University, Mill Hill Lane, Durham DH1 3LB, United Kingdom. Email: tatiana.damjanovic@durham.ac.uk.

## 1. Introduction

A key issue in macroeconomics is determining the relationship between firm heterogeneity and aggregate fluctuations.<sup>1</sup> In this paper, we address the following question: does firm selection along the extensive margin; namely, the decision of heterogeneous firms to produce or exit upon entry, play an important role in enhancing the effect of aggregate technology shocks. To provide an answer, we use a model of monopolistic competition and translog preferences in which variations in measured total factor productivity (hereafter, TFP) depend on the underlying distribution of firm-level productivity.

The novelty of our approach lies in expressing TFP as the combination of three effects: firm selection (extensive margin), which we measure by the productivity of the marginal, zero-profit firm; diminishing returns to new varieties, which is captured by the Herfindahl index; and misallocation, which we relate to the use of factors across firms.<sup>2</sup> We provide new analytical results which characterize how the elasticity of the density function of firm productivity relates to TFP. The diminishing returns and misallocation effects only operate when the elasticity of the density function is strictly increasing or when firm-level productivity is bounded from above.<sup>3</sup> Whilst firm selection magnifies the impact of aggregate technology shocks on TFP, diminishing returns and misallocation both act to mitigate this effect.

To provide intuition for our results, consider the case in which labor is the only productive input and firm productivity has a Pareto distribution. A positive change in technology

---

<sup>1</sup>It is well-documented that there is a considerable amount of between-firm and between-plant heterogeneity, even within narrowly defined sectors of the economy. For example, see Bartelsman and Doms (2000) and Syverson (2011).

<sup>2</sup>Diminishing returns to new varieties has also recently been emphasized by Feenstra and Weinstein (2017).

<sup>3</sup>For many distributions, such as the log normal, exponential, Fréchet, and Weibull, the elasticity of the density function is strictly increasing. For the Pareto distribution, the elasticity of the density function is constant.

encourages firm entry because it raises profitable opportunities, whilst an increase in entry makes it harder for the marginal, zero-profit firm to produce. Firm selection therefore leads to higher average productivity. If the Pareto distribution is unbounded, product variety is independent of aggregate technology because additional firm entry is exactly offset by a rise in the probability of exit. In this case, the magnification of technology shocks arises from reallocating labor to more productive firms.

A positive change in technology leads to a rise in product variety when firm productivity is bounded from above and Pareto. This is because placing an upper bound on productivity weakens the selection effect.<sup>4</sup> An increase in product variety leads to crowding in the product space, whereby the introduction of a new variety lowers the substitutability of all varieties. Crowding reduces TFP, which, in this special case, is proportional to output.<sup>5</sup> Thus, bounding the productivity distribution affects the magnification of aggregate technology shocks in two ways; it curtails the rise in the endogenous component of productivity, and it also generates diminishing returns.

For the general version of our model we perform a simple quantitative exercise focusing on two commonly used productivity distributions; bounded Pareto and log normal.<sup>6</sup> For both productivity distributions we consider, there is endogenous pro-cyclical firm entry, product creation, and TFP. When we adopt a standard process for aggregate technology our analysis suggests that around 20 percent of the variation in TFP can be attributed to firm selection.

---

<sup>4</sup>Feenstra (2018) contains a discussion on the importance of bounded and unbounded Pareto distributions and our paper offers a natural extension of his analysis using the elasticity of the density function.

<sup>5</sup>In a model with homogenous firms, such as Bilbiie *et al.* (2012), when the number of varieties increases goods become closer substitutes. In this case, a higher elasticity of substitution leads to a lower aggregate markup. This result depends on the distribution of firm-level markups.

<sup>6</sup>By using a log normal distribution we allow for movements in the elasticity of the density function which are absent with a Pareto distribution. Combes *et al.* (2012) show that the productivity distribution of French firms is best approximated by a log normal distribution.

The response of output is more than 50 percent higher than when firm entry and exit are absent. Increased variation in output occurs because, in addition to variation in TFP, there is an increase in the sensitivity of factor inputs (capital and labor). Overall, our results suggest that firm heterogeneity and firm selection have quantitatively important implications for business cycle analysis.

Within our framework, we also consider fluctuations in the aggregate markup, which is the focus of numerous studies that seek to understand the link between firm entry and exit and the business cycle. When firm productivity is Pareto distributed, the initial response of the markup is around 0.3 percent, however, with a log normal distribution for firm productivity, the response of the markup is considerably stronger. To gauge the empirical relevance of our results, note that Rotemberg and Woodford (1991) estimate the elasticity of the markup with respect to output to be around 0.2, and more recently, Hong (2017) finds that firm-level markups have average elasticity of 0.9, with respect to real GDP. Our model implies an elasticity (upon impact) of between 0.19 and 0.44, for the Pareto and log normal distributions, respectively.

Translog preferences and the translog demand system have been successfully used in a number of different applications in macroeconomics. For example, Bilbiie *et al.* (2012) show that, in a flexible-price model with endogenous dynamic firm entry, translog preferences generate countercyclical markups relevant for business cycle analysis.<sup>7</sup> Lewis and Poilly (2015) estimate a monetary version of this model with nominal rigidities to evaluate the cyclical properties of the markup in the monetary transmission mechanism.<sup>8</sup> Finally, closer to our analysis, Rodriguez-Lopez (2011) develops an open economy monetary business cycle model

---

<sup>7</sup>Chatterjee and Cooper (1993) and Devereux *et al.* (1996) develop static models of firm entry with monopolistic competition.

<sup>8</sup>Bergin and Feenstra (2000) examine the persistence in output that results from the interaction of monetary shocks and nominal rigidities when the mass of firms is fixed and preferences are translog.

with heterogeneous firms to study exchange rate disconnect, a major empirical puzzle in international finance. Our focus is on selection and the distribution of firm-level productivity for the magnification of technology shocks.

There are alternative approaches to studying the macroeconomic implications of endogenous firm entry and exit with imperfect competition.<sup>9</sup> Jaimovich and Floettoto (2008) consider a model in which firms behave as oligopolistic competitors and show that endogenous markups magnify shocks over the business cycle. Minniti and Turino (2013) develop a model with multi-product firms and an intra-firm extensive margin along similar lines. Etro and Colciago (2010) suppose that firms compete under either Bertrand or Cournot competition, with endogenous dynamic firm entry, such that there are non-instantaneous zero profits. We consider a model in which firms differ in their productivity. Firm-level markups are endogenous and heterogeneous and the extent of (pro-cyclical) movements in the aggregate markup depend on the underlying distribution of firm-level productivity.<sup>10</sup>

Our analysis shares similarities with research that seeks to understand the impact of misallocation on aggregate productivity. As Restuccia and Rogerson (2017) emphasize, misallocation, which reflects the choice of how capital and labor are allocated among producers, and selection, which reflects the choice of which producers should operate, are not independent, and in our calibrated model both effects are present.<sup>11</sup> Since we focus on the cyclical implications of firm heterogeneity, our results also relate neoclassical models of the business

---

<sup>9</sup>Early contributions, such as Rotemberg and Woodford (1999), find that collusion can generate countercyclical markups, while Gali (1994) reaches a similar conclusion with a model in which there are variations in the composition of demand.

<sup>10</sup>Ottaviano (2012) develops a model with Pareto-distributed productive heterogeneity and monopolistic competition which features pro- or countercyclical markups. de Blas and Russ (2015) focus on Bertrand competition and the distribution of markups in an open economy.

<sup>11</sup>This aspect of our analysis is similar to Barseghyan and DiCecio (2011) who study the role of entry costs in generating misallocation.

cycle with firm dynamics, which typically feature a rich set of ex-post firm-level shocks. In our model, heterogeneity in productivity is drawn ex-ante and is fixed, a simplification which allows us to generate new analytical results.<sup>12</sup>

The rest of the paper is as follows. In section two we develop the model economy. In section three we provide analytical results linking the distribution of firm-level productivity with selection, diminishing returns and misallocation. In section four we present a simplified model which we use to develop intuition. In section five we present the quantitative implications of our analysis. Section six concludes.

## 2. Model

This section outlines the model economy. The economy is populated by a continuum of households with mass normalized to one. There is a representative household which consumes a basket of goods and supplies labor. The household owns the capital stock which it rents to firms. There is a large number of ex-ante identical firms that have the option of paying  $f > 0$  units of labor to enter the market. Upon entry, each firm obtains a productivity level  $z \in (z_{\min}, z_{\max})$ , which is the realization of a random variable drawn independently across firms. If the domain is unbounded, we assume that  $z_{\max} = +\infty$ .

### 2.1. Translog Preferences

The representative household has a symmetric translog expenditure function over a set of differentiated goods,

$$\ln(e_t) = \ln y_t + \nu_t + \frac{1}{n_t} \int_{i \in \Delta} \ln p_t(i) di + \frac{\zeta}{2n_t} \int_{i \in \Delta} \int_{j \in \Delta} \ln p_t(i) [\ln p_t(j) - \ln p_t(i)] dj di \quad (1)$$

where  $e_t$  is the minimum expenditure required to obtain the basket of goods  $y_t$ . The set of differentiated goods available to the household is denoted  $\Delta$ , where  $n_t$  is the measure of  $\Delta$ .

---

<sup>12</sup>Our focus on ex-ante heterogeneity is consistent with Melitz (2003). The alternative approach, in which firms are hit by persistent shocks, ex-post, is developed in Hopenhayn and Rogerson (1993).

The term  $p_t(i)$  is the price of good  $i$  at time  $t$ . Finally,  $\nu_t \equiv \frac{1}{2\zeta n_t}$  captures the variety effect, where the parameter  $\zeta > 0$  determines the substitutability between goods.

**Proposition 1** Each firm produces a single good and faces a downward-sloped demand curve. The demand for good  $i$  is  $y_t(i) = [s_t(i) / \rho_t(i)] y_t$ , where,

$$s_t(i) \equiv \frac{1}{n_t} - \zeta \ln p_t(i) + \frac{\zeta}{n_t} \int_{j \in \Delta} \ln p_t(j) dj \quad (2)$$

is the expenditure share on good  $i$  and  $\rho_t(i) \equiv p_t(i) / P_t$  is the price of good  $i$  relative to the consumer-based price index, denoted  $P_t$ .

**Proof** Apply Shephard's Lemma to the equation (1). ■

To understand the implications of Proposition 1 we define the reservation price as the maximum price charged for a good. Such a good will have zero expenditure share,  $s_t(i^*) = 0$ , with reservation price equal to,  $\ln \rho_t(i^*) = \frac{1}{\zeta n_t} + \frac{1}{n_t} \int_{j \in \Delta} \ln \rho_t(j) dj$ , where the second term is the average price across all goods. Using this result, we re-write the expenditure share for good  $i$  as the ratio the price of good  $i$  to the reservation price,  $s_t(i) = -\zeta \ln [\rho_t(i) / \rho_t(i^*)]$ . Finally, the elasticity of demand for good  $i$  - defined as  $\epsilon_{y_t, p_t}(i) \equiv -\partial \ln y_t(i) / \partial \ln p_t(i)$  - is  $1 + \zeta / s_t(i)$ , which implies firms with lower market share face more elastic demand.<sup>13</sup>

## 2.2. Production

Each firm produces a differentiated good under conditions of monopolistic competition by hiring labor and renting capital. Firms have access to a constant returns to scale technology. Given  $l_t(i)$  workers and  $k_t(i)$  units of capital, firm  $i$  produces,

$$y_t(i) = a_t z_t(i) [k_t(i)]^\alpha [l_t(i)]^{1-\alpha} \quad (3)$$

---

<sup>13</sup>Write the demand curve as,  $\ln y_t(i) = \ln s_t(i) - \ln p_t(i) + \ln P_t y_t$ , where,  $s_t(i) = \zeta \ln p_t(i^*) - \zeta \ln p_t(i)$ . Differentiating,  $-\partial \ln y_t(i) / \partial \ln p_t(i) = \zeta / [\zeta \ln p_t(i^*) - \zeta \ln p_t(i)] + 1$ , where the term in square bracket is the expenditure share. The expenditure share is declining with price and the elasticity is increasing with price.

where  $a_t$  is a technology common to all firms and  $z_t(i)$  is a firm-specific level of productivity. The profit function is,

$$\vartheta_t(i) = [\rho_t(i) - mc_t(i)] y_t(i) \quad (4)$$

where real marginal cost, denoted  $mc_t(i)$ , is determined by the following cost minimization problem;  $mc_t(i) = \min_{k_t(i), l_t(i)} [w_t l_t(i) + r_t k_t(i)]$ , s.t.  $y_t(i) \geq 1$ , where  $w_t$  is real wage and  $r_t$  is real cost of capital. We define  $mc_t \equiv w_t^{1-\alpha} r_t^\alpha / a_t (1-\alpha)^{1-\alpha} \alpha^\alpha$ , such that  $mc_t(i) = mc_t / z_t(i)$ .

Firms maximize profits, subject to the demand for their good, as defined in Proposition 1.

### Proposition 2

1. The optimal price chosen by firm  $i$ , with productivity  $z$ , is,

$$\rho_t(z) = \Omega \left( \frac{z}{z_t^*} \exp \right) \frac{mc_t}{z} \quad (5)$$

where  $\Omega_t \equiv \Omega \left( \frac{z}{z_t^*} \exp \right)$  denotes the Lambert-W function and  $z_t^*$  is the zero-profit level of firm productivity (cut-off), which indexes the reservation price,  $\rho_t^* = mc_t / z_t^*$ .

2. The market share and profit of firm  $i$  are,

$$s(z) = \zeta \left[ \Omega \left( \frac{z}{z_t^*} \exp \right) - 1 \right] \quad (6)$$

and,

$$\vartheta(z) = s(z) y_t - \phi(z) = s(z) \left[ 1 - \Omega^{-1} \left( \frac{z}{z_t^*} \exp \right) \right] y_t \quad (7)$$

where  $s(z) y_t$  is firm-level revenue (sales) and  $\phi(z) = \zeta \left[ 1 - \Omega^{-1} \left( \frac{z}{z_t^*} \exp \right) \right] y_t$  is the cost of production.

**Proof** See Appendix. ■

The term  $\Omega\left(\frac{z}{z_t^*} \exp\right)$  in Proposition 2 measures the gross firm-level markup. Firm-level markups depend positively on the firm-specific productivity draw,  $z$ , and negatively on the cut-off level of productivity,  $z_t^*$ , which is endogenous and depends on market conditions. The least productive firm, which is defined as the firm with  $z = z_t^*$ , has zero markup, since  $\Omega(\exp) = 1$ , and zero market share,  $s(z_t^*) = 0$ . Equation (5) also implies that more productive firm set higher markups and charge lower prices. Because markups are heterogeneous across firms we observe that price pass-through (from an exogenous change in marginal costs,  $mc_t$ ) is also heterogeneous. An exogenous change in marginal cost has two effects on prices; a direct effect and an indirect effect, via  $z_t^*$ . We follow Rodriguez-Lopez (2011) in deriving a simple expression for pass-through and confirm that smaller, less productive firms, which also have smaller market share, have smaller price pass-through.<sup>14</sup>

### 2.3. Firm Entry and Exit

There are a large number of ex-ante identical firms that have the option of hiring  $f > 0$  units of labor to enter the market. Each firm obtains a productivity level  $z \in (z_{\min}, z_{\max})$  which is the realization of a random variable drawn independently across firms from a distribution  $G(z)$ . Firm  $i$  enters if,

$$\vartheta_t \equiv \int_{z_t^*}^{z_{\max}} \vartheta_t [z(i)] dG(z) > (w_t/a_t) f \quad (8)$$

Firms endogenously enter the market until profits are zero net of the entry costs. Of the mass  $\mathbf{N}_t$  of firms entering the economy, the mass of products available to the household, denoted  $n_t$ , is equal to the mass of entrants, multiplied by the probability of successful entry,  $\int_{z_t^*}^{z_{\max}} dG(z) = 1 - G(z_t^*)$ .<sup>15</sup>

### 2.4. Aggregation

---

<sup>14</sup>We relegate the proof of this point to the Appendix.

<sup>15</sup>Our model of entry and exit is static, similar to, for example, Chatterjee and Cooper (1993) and Jaimovich and Floetotto (2008).

In this section, we present and discuss aggregate equations for the economy. Details of the derivations are presented in the Appendix. Recall that market share is related to productivity by,  $s_t(z) = \zeta(\Omega_t - 1)$ , where  $\Omega_t \equiv \Omega\left(\frac{z}{z_t^*} \exp\right)$ . Aggregating total demand (defined in Proposition 1) we can relate the mass of entrants,  $\mathbf{N}_t$ , to the cut-off level of productivity in the following way,

$$\pi_{1,t} = 1/\zeta\mathbf{N}_t \quad (9)$$

where we define  $\pi_{1,t} \equiv \int_{z_t^*}^{z^{\max}} (\Omega_t - 1) dG(z)$ . The parameter  $\pi_{1,t}$  is the average markup across all firms (that is, the net markup across successful and non-successful firms) and this falls with firm entry. A key point of note is that a change in  $\pi_{1,t}$  does not imply a change in the aggregate markup in our economy, which is inversely proportional to product variety,  $n_t$ . Recall that the mass of available products is given by,  $n_t = [1 - G(z_t^*)] \mathbf{N}_t$ , where  $G'(z_t^*) > 0$ . The aggregate markup (i.e., the markup across successful firms) is defined as,  $m(z_t^*) \equiv \pi_{1,t}/[1 - G(z_t^*)]$ , and its cyclical behavior depends on the strength of firm selection.

The mass of firms is determined by the free entry condition,

$$\pi_{2,t} = (w_t/a_t) \frac{f}{\zeta y_t} \quad (10)$$

where  $\pi_{2,t} \equiv \int_{z_t^*}^{z^{\max}} [(\Omega_t - 1)^2 / \Omega_t] dG(z)$  is derived from the aggregated profit function.

Given product demand and free entry, we use the optimal pricing decisions of firms (defined in Proposition 2) to relate the mass of firms with the cut-off productivity level. In doing so, we define the Herfindahl index in our economy - a commonly-used measure of market concentration - as,  $\mathbf{H}(z_t^*) \equiv \mathbf{N}_t \int_{z_t^*}^{z^{\max}} [s(z)]^2 g(z) dz$ , where  $s(z)$  is the market share of each firm. This leads to,

$$\begin{aligned} \rho(z_t^*) &= mc_t(z_t^*) \\ &= z_t^* \exp\left[\frac{\mathbf{H}(z_t^*)}{2\zeta}\right] \end{aligned} \quad (11)$$

where  $z_t^* > z_{\min}$  and  $\mathbf{N}_t \int_{z_t^*}^{z_{\max}} s(z) g(z) dz = 1$ . This is an important equation in our analysis. First, notice that the reservation price depends positively on the Herfindahl index, and moreover, that price is only a function of the productivity cut-off. Second, equation (11) implies  $\epsilon_{z_t^*, mc_t} > 0$ , which is discussed above in the context of price pass-through. It is now clear that the size of this elasticity depends on the Herfindahl index and that the specific relationship between the productivity cut-off and market concentration depends on the underlying distribution of firm-level productivity. The Herfindahl index is integral to the translog setting, since it will determine, along with the distribution of productivity, the extend of crowding in the product space, whereby the entry of new products raises substitutability of all products.

Labour is used for production and the creation of new firms, whereas capital is only used for production. The total use of labor is,  $L_t = \mathbf{N}_t \int_{z_t^*}^{z_{\max}} l_t(z) dG(z) + \mathbf{N}_t (f/a_t)$ , which we express as,

$$w_t = \left( \frac{\pi_{3,t}}{\pi_{1,t}} \right) \frac{y_t}{L_t} \quad (12)$$

where  $\pi_{3,t} \equiv \int_{z_t^*}^{z_{\max}} [(\Omega_t - 1)(\Omega_t - \alpha) / \Omega_t] dG(z)$ . Similarly, the aggregate stock of capital is,  $K_t = \mathbf{N}_t \int_{z_t^*}^{z_{\max}} k_t(z) dG(z)$ , which we express as,

$$r_t = \alpha \left( \frac{\pi_{4,t}}{\pi_{1,t}} \right) \frac{y_t}{K_t} \quad (13)$$

where  $\pi_{4,t} \equiv \int_{z_t^*}^{z_{\max}} (1 - 1/\Omega_t) dG(z)$ . We note that the two-sector structure of our economy introduces a simple asymmetry which is also present in the analysis of, for example, Barseghyan and DiCecio (2011).<sup>16</sup> In our case, there is an endogenous wedge that depends on the value we assign to  $\alpha < 1$ , which determines the fraction of capital used in the production of goods for consumption - see equation (3).

---

<sup>16</sup>In Barseghyan and DiCecio (2011), labor is used for production and overhead costs, and overheads determine the cut-off level of productivity. Our model does not feature a fixed production cost and the productivity cut-off is determined by the existence of a choke-price.

### 2.5. Representative Household

The representative household has the following lifetime utility function,

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, L_t) \quad (14)$$

where  $c_t$  is consumption (see equation (1)) and  $L_t$  is labor supply. Households maximize lifetime utility subject to the following constraints,

$$c_t + i_t = w_t L_t + r_t K_t \quad \text{and} \quad i_t = K_{t+1} - (1 - \delta) K_t \quad (15)$$

where  $k_t$  ( $r_t$ ) is the stock (rental rate) of capital and  $w_t$  is the wage rate. This leads to standard conditions which characterize savings and labor supply decisions,

$$u_c(t) = \beta \mathbb{E}_t u_c(t+1) [r_{t+1} + (1 - \delta)] \quad \text{and} \quad w_t = -u_L(t) / u_c(t) \quad (16)$$

where  $u_c(t)$  is the period  $t$  marginal utility with respect to the consumption and similarly for  $u_L(t)$ .

### 2.6. Model Summary

Table 1 presents a summary of the equations that solve the model economy.

===== **Table 1 Here** =====

The nine equations in the middle column of Table 1 solve for the variables  $\{i_t, K_{t+1}, c_t, y_t, L_t\}$  and  $\{w_t, r_t\}$  and  $\{z_t^*, \mathbf{N}_t\}$ . The first 5 equations are associated with auxiliary variables;  $\pi_{i,t}$ -definitions, the Herfindahl index,  $\mathbf{H}_t = \mathbf{H}(z_t^*)$ , and the markup,  $\mathbf{m}_t = \mathbf{m}(z_t^*)$ . The final 4 equations are standard conditions describing the optimal choice over savings and labor supply along with resource constraints. It is worth noting that the mass of available products,  $n_t$ , does not appear explicitly in our system of equations. If we know the mass of entrants ( $\mathbf{N}_t$ ,

from the free entry condition) and the productivity cut-off (defined by the reservation price), product variety is determined. This feature of our model is associated with the translog expenditure system.

### 3. Analysis

In this section we do two things. First, we prove that there is a solution for the productivity cut-off,  $z_t^* > z_{\min}$ , for a general class of truncated distributions. Second, we discuss the properties of average productivity, the Herfindahl index, the aggregate markup, and measured TFP, all of which feature in our discussion below.

#### 3.1. Distributional Properties and Productivity Cut-Off

Lemma 1 and Lemma 2 establish important properties of aggregate variables in our economy.

**Lemma 1** Let  $G(z)$  be the CDF of  $z$ , bounded by  $z_{\max} \leq +\infty$ . For any function,  $J(z_t^*) = \int_{z_t^*}^{z_{\max}} j\left(\frac{z}{z_t^*}\right) dG(z)$ , where  $j(1) \geq 0$  and  $j' \geq 0$ , then, (i),  $J'(z_t^*) < 0$ , and (ii),  $\lim_{z_t^* \rightarrow z_{\max}} J(z_t^*) = 0$ .

**Proof** See Appendix. ■

Lemma 1 implies that auxiliary variable  $\pi_{i,t}$  for  $i = 1, \dots, 4$  are decreasing functions of  $z_t^*$ . From product demand, equation (9), we can now confirm that firm entry and the productivity cut-off have a positive relationship.

**Lemma 2** Make the change of variables,  $u = z/z_t^*$ . Let  $g(u_t)$  be a density function with elasticity,

$$\epsilon(u_t) = -\frac{u_t g'(u_t)}{g(u_t)} \tag{17}$$

which is weakly increasing. Let  $j_1(u_t)$  and  $j_2(u_t)$  be positive functions such that  $\frac{j_1(u_t)}{j_2(u_t)}$  is strictly increasing. The ratio  $\frac{J_1(z_t^*)}{J_2(z_t^*)}$  is a decreasing function of  $z_t^*$ , where  $J_i(z_t^*) = \int_{z_t^*}^{z_{\max}} j_i(u) g(z) dz$ , for  $i = 1, 2$ .

**Proof** See Appendix. ■

Equation (17) holds for all well-known distributions, with one important exception, the Pareto distribution, which features a constant elasticity, equal to one plus the shape parameter. Nevertheless, the result of Lemma 2 is correct for the bounded Pareto distribution ( $z_{\max} < +\infty$ ), and in what follows, we will assume the distribution of productivity is such that it satisfies Lemma 2 or it is Pareto. In fact, Lemma 2 also shows us why the unbounded Pareto distribution is a special case. Whereas the majority of aggregators move with the productivity cut-off,  $z_t^*$ , in the unbounded Pareto case, they are fixed. Feenstra (2018) contains a discussion on bounded and unbounded Pareto distributions, and our paper offers a natural extension of his analysis using the elasticity of the density function,  $\epsilon(u_t)$ .

We can now investigate firm entry and assess the factors affecting the productivity cut-off. In Proposition 3, we establish the existence of a solution for  $z_t^*$ .

**Proposition 3** Define  $z_{\min}$  as the infimum of the domain for the distribution function  $G(z)$ .

An internal solution,  $z_t^* > z_{\min}$  to,

$$\frac{\pi_{3,t}}{\pi_{2,t}} \frac{1}{\pi_{1,t}} = \left(\frac{\zeta}{f}\right) a_t L_t \quad (18)$$

exists if and only if  $\left[\frac{\pi_{2,t}(z_{\min})}{\pi_{3,t}(z_{\min})}\right] \pi_{1,t}(z_{\min}) > \left(\frac{f}{\zeta}\right) / a_t L_t$  and if condition (17) is satisfied.

**Proof** Combining free entry with the factor price equation for wages and product demand generates equation (18). Lemma 1 implies that  $\pi_{1,t}$  is a decreasing function and

$\lim_{z_t^* \rightarrow z_{\max}} \pi_{1,t} = 0$ . Lemma 2 implies that the ratio  $\frac{\pi_{2,t}}{\pi_{3,t}}$  decreases with  $z_t^*$ . Therefore, the solution to (18) exists and it is unique. ■

Proposition 3 shows that the productivity cut-off exists for any set of parameter values when  $z_{\min} = 0$ . This rests on the fact that  $\pi_{2,t} > \pi_{3,t}$ , for any  $z$ , and that if  $z_{\min} = 0$ , then  $\lim_{z_t^* \rightarrow \infty} \pi_{1,t} = +\infty$ , a proof of which is presented in the Appendix. In the case of the

Pareto distribution - commonly used in the analysis of models with firm-level heterogeneity - a solution exists as long as entry costs are not too large, and if not, all firms that enter produce, with  $z_t^* \rightarrow z_{\min}$ . An important implication of Proposition 3 is that, if  $z_t^*$  exists, it is positively related to the total supply of labor,  $L_t$ .<sup>17</sup> By Lemma 1, firm entry and labor supply are also positively related.

### 3.2. Productivity, the Herfindahl Index, and Markup

In this section, we discuss the properties of average productivity, the Herfindahl index, the aggregate markup, and measured TFP.

**Proposition 4** Average productivity, defined as  $\bar{z}_t = \frac{1}{1-G(z_t^*)} \int_{z_t^*}^{z_{\max}} z dG(z)$ , increases with total labor supply. The Herfindahl index,  $H_t$ , defined in equation (11), decreases with total labor supply. The aggregate markup, defined as  $\mathbf{m}_t \equiv \frac{\pi_{1,t}}{1-G(z_t^*)}$ , decreases with total labor supply.

**Proof** Recall,  $1 - G(z_t^*) = \int_{z_t^*}^{z_{\max}} dG(z)$ . Lemma 2 implies that the Herfindahl index and average markup decline with the cut-off and Proposition 3 implies that  $z_t^*$  increases with total labor supply. ■

It is also interesting to observe that the share of capital and labor depend on the productivity cut-off. Recall that  $\frac{w_t L_t}{y_t} = \frac{\pi_{3,t}}{\pi_{1,t}}$  and  $\frac{r_t K_t}{y_t} = \alpha \frac{\pi_{4,t}}{\pi_{1,t}}$ . The labour share is larger than  $(1 - \alpha)$  because labor is used for entry, whilst the share of capital is also larger than  $\alpha$ .<sup>18</sup> Lemma 2 allows us to establish that when  $z_t^*$  increases, the labor share,  $\pi_{3,t}/\pi_{1,t}$ , declines. This is because the amount of labor used for entry grows more slowly than employment in the production sector. Contrarily, the share of capital,  $\pi_{4,t}/\pi_{1,t}$ , increases with  $z_t^*$ . Thus, as

<sup>17</sup>Applying implicit function theorem, we have  $\frac{dz_t^*}{d(\zeta L/F)} = \pi_1 / \frac{\partial}{\partial z_t^*} \left( \frac{\pi_{3,t}}{\pi_{2,t}} - \frac{\zeta L_t}{f} \pi_{1,t} \right) > 0$ , for  $a_t = 1$ .

<sup>18</sup>For the former,  $\pi_{3,t} - (1 - \alpha)\pi_{1,t} = \alpha\pi_{2,t} > 0$ , and for the latter,  $\pi_{4,t} - \pi_{1,t} = \pi_{2,t} > 0$ .

$z_t^*$  rises, the utilization of labor increases, whilst the utilization of capital declines. These ratios are constant when we assume productivity has an unbounded Pareto distribution.

We can also use Lemma 2 to characterize measured TFP, defined as,  $TFP_t \equiv y_t/K_t^\alpha L_t^{1-\alpha}$ .

**Proposition 5** Measured TFP is,

$$TFP_t = (a_t mc_t) Z_t \tag{19}$$

where  $Z_t \equiv \frac{(1-\alpha)^{1-\alpha}}{\pi_{3,t}^{1-\alpha} \pi_{4,t}^\alpha} \pi_{1,t}$  is a decreasing function of  $z_t^*$  and  $mc_t$  is defined in equation (11).

**Proof** See Appendix. ■

Measured TFP consists of three terms. The first two terms depends on marginal cost,  $mc_t = z_t^* \exp(\mathbf{H}_t/2\zeta)$ . The first part is the productivity cut-off,  $z_t^*$ , which increases with labor supply. The second part, the Herfindahl index, is related to crowding in the product space, or diminishing returns to new varieties, as discussed in Feenstra and Weinstein (2017). The final term,  $Z_t$ , is defined through the demand for capital and labor (see equation (12) and (13)) and we refer to this as misallocation. Since the second and third terms (diminishing returns and misallocation), in general, are decreasing functions of  $z_t^*$ , we already know that these mute any change of productivity onto measured TFP.

## 4. Mechanism in a Special Case

In this section, we present analytical results for a simplified version of our model. At this point it is worth stressing the focus of our analysis: we are interested in characterizing aggregate outcomes conditional on the selection effect. The strength of selection depends on the distribution of firm-level productivity.

### 4.1. The Model with a Fixed Stock of Capital

We assume the aggregate capital stock is fixed and  $\alpha \rightarrow 0$ . This implies,  $\pi_{1,t} = \pi_{3,t}$ , such that the free entry condition, as presented in Table 1, can be written as,  $\pi_{2,t} = \frac{f/a_t}{\zeta L_t}$ . In this case, the wage rate for the economy is simply,  $w_t = \frac{y_t}{L_t}$ , which is consistent with  $TFP_t = a_t m c_t$  and  $Z_t \equiv 1$ . We also assume utility is logarithmic in consumption. Given the economy-wide resource equation, this implies,  $-L_t u_L(t) = 1$ , such that the total supply of labor is unaffected by aggregate technology, and so we set  $L_t = 1$ . If we now re-consider the free entry condition, a rise in aggregate technology,  $a_t$ , lowers  $\pi_{2,t}$ , and since  $\pi_{2,t}$  is declining in  $z_t^*$ , we can immediately conclude that aggregate technology and the productivity cut-off have a positive relationship. Thus, a positive shock to technology raises average productivity,  $\bar{z}_t = \frac{1}{1-G(z_t^*)} \int_{z_t^*}^{z_{\max}} z dG(z)$ .

Combining the pricing equation with resources we solve for output as a function of the productivity cut-off in the following way.

$$y_t = TFP_t \quad \text{and} \quad \pi_{2,t} = \frac{f/a_t}{\zeta} \quad (20)$$

A change in  $a_t$  has a direct and indirect effect on output. The indirect effect is reflected in the change in  $z_t^*$ , the strength of which is controlled by the free entry condition, which is the second equation in (20). The explanation for the impact of a change in the productivity cut-off on output (measured TFP) is a *selection effect* - which relates to  $z_t^*$  - and *crowding in the product space* - which relates to the Herfindahl index,  $H_t$ . Feenstra and Weinstein (2017) discuss how changes the Herfindahl index relate to crowding in the product space when there are translog preference. The term crowding refers the to the idea that, with a lower Herfindahl index, new varieties are more substitutable and, therefore, less desirable. All else equal, when the Herfindahl index falls so does the demand for goods, and therefore, total output.

Proposition 4 also allows us to link the results for output to the aggregate markup. A positive technology shock encourages firm entry and leads to a lower aggregate markup.

Despite the similarities, our results differ from, for example, Bilbiie *et al.* (2012) and Etro and Colciago (2010), where an aggregate technology shock also leads to increased firm entry (and, by definition, greater product variety) and a lower markup. In our model, whilst there is more entry, the fraction of successful entrants falls, and this affects product variety. As the product space becomes more crowded the Herfindahl index falls. We relate these factors to the productivity cut-off and the strength of firm selection.

#### 4.2. Bounded Pareto Distribution

In this section we assume firm productivity is Pareto distributed.<sup>19</sup> In general, there are two effects from our distributional assumptions: we refer to these as bounding and elasticity effects. The elasticity effect depends on the shape of the distribution, characterized through the elasticity of the density,  $\epsilon(u_t)$ , which, in general, is a weakly decreasing function. The Pareto assumption implies a constant elasticity. The bounding effect, is controlled by  $z_{\max}$ , and a lower value for  $z_{\max}$  (i.e. lowering the upper bound of the distribution) curtails the selection effect because it alters the reallocation of labor to the most productive firms.

When firm-level productivity is Pareto,

$$G(z) = \frac{1 - z^{-\kappa}}{1 - z_{\max}^{-\kappa}} \quad \text{and} \quad \epsilon(u) = k + 1 \quad (21)$$

where  $z_{\min} = 1$  and  $k > 0$  is the shape parameter. Although Lemma 2 does not apply to the Pareto case, there is a unique solution for  $z_t^* > 1$ , and we have following result.

**Corollary** When productivity is Pareto distributed with shape parameter  $\kappa$ ,

$$\widehat{\pi}_{i,t} = -(\kappa + \zeta_i) \widehat{z}_t^* \quad (22)$$

for  $i = 1, \dots, 3$  and where a caret denotes the log deviation of a variable from its steady-state value. The parameters  $\zeta_3 > \zeta_2 > \zeta_1 > 0$  and for  $z_{\max} \rightarrow \infty$ ,  $\zeta_i \rightarrow 0$ , for all  $i$ .

---

<sup>19</sup>We use a Pareto distribution in this section for analytical convenience and comparability with other studies. The Pareto assumption is by far the most popular assumption over firm-level productivity.

When firm productivity is distributed Pareto, the multipliers of the auxiliary variables depend on the shape parameter and ranked composite parameters, all of which decline with  $z_{\max}$ .

We summarize the response of productivity to an aggregate technology shock as,

$$\widehat{z}_t^* = \left( \frac{1}{\kappa + \zeta_2} \right) \widehat{a}_t \quad \text{and} \quad \widehat{\bar{z}}_t = \phi(z^*) \left( 1 - \frac{z^*}{\bar{z}} \right) \widehat{z}_t^* \quad (23)$$

where  $\phi(z^*) \equiv \frac{g(z^*)z^*}{1-G(z^*)} = \frac{\kappa}{1-(z^*/z_{\max})^\kappa} > 0$  is the log hazard ratio. A rise in technology leads to a higher productivity cut-off and a higher average level of productivity. As the upper bound of the distribution,  $z_{\max}$ , rises, the cut-off productivity is more sensitive to the shock because labor is reallocated to increasingly more productive firms.

The output response to a change in technology is,

$$\begin{aligned} \widehat{y}_t &= \widehat{a}_t + \widehat{m}c_t \\ &= \left\{ 1 + \frac{1}{\kappa + \zeta_2} \left[ 1 + \frac{\zeta_1 - \zeta_3}{2} \left( \frac{H}{\bar{\zeta}} \right) \right] \right\} \widehat{a}_t \end{aligned} \quad (24)$$

There are competing forces that explain the response of output in equation (24). Mechanically, there are two reasons the sensitivity of output lowers as the upper bound of the productivity distribution falls. First, the selection effect is weaker, as in equation (23), and thus the response of  $z_t^*$  is smaller. Second, the Herfindahl index falls with  $z_t^*$  because there is an increase in product variety. These are the *selection* and *crowding* effects described above.

Finally, we very briefly consider the case,  $z_{\max} \rightarrow \infty$ , where  $\widehat{z}_t^* = \widehat{\bar{z}}_t = \frac{1}{\kappa} \widehat{a}_t$  and  $\widehat{y}_t = \left( 1 + \frac{1}{\kappa} \right) \widehat{a}_t$ . When the productivity distribution is unbounded, a positive change in aggregate technology encourages firm entry (i.e.,  $\widehat{N}_t > 0$ ), the strength of selection leads to no change in product variety,  $\widehat{n}_t = -[\phi(z^*)] \widehat{z}_t^* + \widehat{N}_t = 0$ , and no change in the markup. It is instructive to consider what happens as  $\kappa$  falls and variance of the productivity distribution rises. As  $\kappa$

falls, the average markup rises, as does the Herfindahl index.<sup>20</sup> In this case, there are fewer, very productive firms, and output is more sensitive to aggregate technology shocks.

## 5. Quantitative Exercise

In this section, we present a quantitative analysis of the model. We assume a standard form for utility,  $u(c_t, L_t) = \ln c_t - \frac{\chi}{1+\nu} L_t^{1+\nu}$ , where  $\frac{1}{\nu}$  is the Frisch elasticity, and we consider two commonly used distributions for firm-level productivity: a bounded Pareto distribution,  $G(z; z_{\max}, \kappa)$ , introduced and discussed above, and an unbounded log normal distribution,  $G(z; \mu, \sigma)$ , where  $\mu$  ( $\sigma$ ) is the location (scale) parameter.

### 5.1. Calibration

Our calibration strategy is the following. We start by picking some standard parameter values. In particular, we assume that a period in the model is a year and  $\beta = 0.96$  and  $\delta = 0.1$ . These parameters determine the risk-free rate and rental rate of capital. We normalize aggregate technology to  $a = 1$ , set the parameter that determines the capital used in production of goods at one third, so  $\alpha = 1/3$ , we assume the Frisch elasticity is 0.72, a value suggested by Heathcote *et al.* (2010), and that individuals spend a third of their time in work, so  $L = 1/3$ . The remaining parameters of the model are calibrated to steady-state targets.

Table 2 presents all parameters used to characterize the steady-state and their respective targets.

===== **Table 2 Here** =====

---

<sup>20</sup>The markup is given by,  $m = \kappa \left\{ \int_1^\infty (\Omega - 1) u^{-\kappa-1} du \right\}$ . If firms are (approximately) homogeneous in productivity ( $\kappa \rightarrow \infty$ ), the average markup approaches zero. This particular case is also studied in Rodriguez-Lopez's (2011) monetary model of an open economy.

Based on Davis *et al.* (2006) we assume that the failure rate of firms is 14 percent. We then set the the shape parameter of the Pareto distribution at  $\kappa = 1.653$  following the analysis of Nigai (2017; Table 9; ‘variable markups’). Since the translog parameter,  $\zeta > 0$ , acts to scales the mass of firms, we normalize it to unity, and we use  $z_{\max}$  - the truncation in the distribution - to achieve an aggregate markup of 23 percent.<sup>21</sup> This calibration implies a share of physical investment in GDP of 17.4 percent and a Herfindahl index of 3789. The former is close to values for the US, whereas the latter is somewhat higher than typical concentration ratios. When we assume firm productivity is log normally distributed, we leave the distribution unbounded, but ensure our economies have the same average markup and average productivity. To do this, we set  $\sigma = 0.321$  to hit the markup of 23 percent. We then set  $\mu = 0.445$  so that we achieve  $\bar{z} = 1.76$  (which is the outcome when firm productivity is Pareto distributed). With a log normal distribution of firm productivity the share of physical investment in GDP is 18 percent and the Herfindahl index is 3214.

## 5.2. Impulse Responses and Second Moments

Technology shocks are the only source of aggregate uncertainty in our model. We assume aggregate technology follows an AR(1) process such that,

$$\hat{a}_t = \rho \hat{a}_{t-1} + \hat{e}_t$$

We set  $\rho = 0.979^4$  and  $\sigma_e = 0.0072 \times \sqrt{4}$  consistent with King and Rebelo (1999).

Figure 1 plots Impulse Response Functions of selected variables for a 1 percent positive innovation to technology. The two sets of dashed lines are the model with firm heterogeneity, for the different productivity distributions, and the solid (black) lines are generated from a benchmark RBC model, using the same calibration as outlined above.<sup>22</sup> The vertical axis

<sup>21</sup>When we allow  $z_{\max} \rightarrow \infty$  the markup rises to 34.1 percent.

<sup>22</sup>Since this version of the model is entirely standard we omitted the details.

is the percentage deviation from the steady state and the number of years after the shock is reported on the horizontal axis.

===== **Figure 1 Here** =====

We first consider standard variables - depicted in the two upper rows. It is clear that, in both cases, i.e., with a bounded Pareto or a log normal distribution, there is a significant amplification from firm selection: the rise in aggregate output generated from the shock to productivity is around 50 percent higher than the benchmark model. This is the result of both consumption and investment in physical capital rising considerably more when firms are heterogeneous. Consistent with this pattern of magnification, factor prices are considerably more sensitive to the shock.

The third row of figure 1 depicts those variables specific to the model with selection. Consider measured TFP ( $TFP_t$ ), which Proposition 5 decomposes into three separate terms. When the productivity distribution is Pareto (log normal), the initial change in the productivity cut-off is 0.84 (0.4) percent, whereas the response of the Herfindahl index (not reported) is  $-0.30$  ( $-0.46$ ) percent. The remaining change in measured TFP derives from changes in the utilization of factors. In general, these results suggest that around 20 percent (22, in the case of Pareto and 16, in the case of log normal) of the variation in measured TFP can be attributed to firm selection. Recall, firm selection magnifies the effect of technology shocks over the business cycle because factor inputs become sensitive. Moreover, despite bounding the Pareto distribution from above, the impact of a change in technology onto measured TFP is larger than when firm-level productivity is distributed log normal.

Our model also feature pro-cyclical firm entry and a counter-cyclical markup. As we discuss above, in this sense, our model is similar to, for example, Jaimovich and Floetotto (2008)

and Etro and Colciago (2010). In the former, a 1 percent shock to technology generates a fall in the markup of around 0.30 percent, and in the latter, the movement in the markup is twice as large. With a Pareto distribution, the initial response of the markup is somewhere in between these values at 0.15 percent, whereas the markup response with log normally distributed productivity is much stronger. Empirically, whilst Rotemberg and Woodford (1991) estimate the elasticity of the markup with respect to output to be around 0.2, more recently, Hong (2017) finds that firm-level markups have an average elasticity of around 0.9. Our model implies an elasticity (upon impact) of between 0.19 and 0.44.

To further evaluate the properties of our model, we compute the second moments of aggregate variables. Table 3 presents the moments for the three cases presented above.

===== **Table 3 Here** =====

The results in Table 3 show that the entry and exit of heterogeneous firms, in our model, generates improvements relative to the benchmark RBC setting, consistent with the impulse responses reported in Figure 1. In all cases, despite the Pareto distribution being bounded, and the selection effect mitigated (relative to  $z_{\max} \rightarrow +\infty$ ), it performs marginally better in terms producing movements in output, consumption, and physical investment closer to the data. The model necessarily fails to produce enough volatility in hours worked, but given the empirically relevant Frisch elasticity we assume, this is not surprising. Finally, although the volatility of output is considerably higher than that reported for the RBC model this is not only attributable to changes in TFP but also to variation in inputs.

## 6. Conclusion

This paper studies the macroeconomic implications of firm selection. We provide analytical results that show how firm selection is related to the underlying distribution of firm-level

productivity. Selection, diminishing returns to new varieties (crowding in the product space), and misallocation all relate to measured TFP. We use this result to connect the magnification of aggregate technology shocks, via the extensive margin, with the elasticity of the density function of firm productivity. Although diminishing returns and misallocation act to dampen the magnification of aggregate technology shocks, the reallocation of resources to relatively more productive firms generates quantitatively important variations in output.

## Appendix A

In this Appendix we present proofs for Lemmas and Propositions not reported in the text.

### A.1. Proof of Proposition 2

The cost minimization problem for firm  $i$  is,

$$\min_{k_t(i), l_t(i)} [w_t l_t(i) + r_t k_t(i)] + \lambda_t \{y_t(i) - a_t z_t(i) [k_t(i)]^\alpha [l_t(i)]^{1-\alpha}\} \quad (25)$$

where  $\lambda_t$  is a Lagrange multiplier. This implies,  $l_t(i) = [(1 - \alpha) / \alpha] (r_t / w_t) k_t(i)$ , and so,  $w_t l_t(i) + r_t k_t(i) = (r_t / \alpha) k_t(i) = [y_t(i) / a_t z_t(i)] w_t^{1-\alpha} r_t^\alpha / \psi$ , where  $\psi \equiv (1 - \alpha)^{1-\alpha} \alpha^\alpha$ .

Marginal cost is,

$$mc_t(i) = \frac{\partial}{\partial y_t(i)} [w_t l_t(i) + r_t k_t(i)] = \frac{mc_t}{z_t(i)} ; \quad mc_t \equiv \frac{1}{a_t} \frac{w_t^{1-\alpha} r_t^\alpha}{\psi} \quad (26)$$

The profit maximization problem for firm  $i$  is to maximize,

$$\max_{\rho_t(i)} \left[ \rho_t(i) - \frac{mc_t}{z_t(i)} \right] y_t(i) - \mu_t [\rho_t(i) y_t(i) - s_t(i) y_t] ; \quad s_t(i) = -\zeta \ln \frac{\rho_t(i)}{\rho(z_t^*)} \quad (27)$$

where  $\rho_t(i) y_t(i) = s_t(i) y_t$  is the demand for good  $i$  and  $\mu_t$  is a Lagrange multiplier. This implies,

$$\rho_t(i) y_t(i) = \zeta \left[ \frac{z_t(i) \rho_t(i)}{mc_t} - 1 \right] y_t \quad (28)$$

We use  $\rho_t(i) y_t(i) = s_t(i) y_t$  and (27) to rewrite (28) as,  $-\ln \frac{\rho_t(i)}{\rho(z_t^*)} = \left[ \frac{z_t(i) \rho_t(i)}{mc_t} - 1 \right]$ , or,

$$\frac{z \rho_t(z)}{mc_t} = \ln \left[ \frac{\rho(z_t^*)}{\rho_t(z)} \exp \right] \quad \text{and} \quad \rho(z_t^*) = \frac{mc_t}{z_t^*}$$

which implies,  $z \rho_t(z) / mc_t = -\ln [z_t \rho_t(z) / mc_t] + \ln \left( \frac{z_t}{z_t^*} \exp \right)$ . We then use the Lambert-W function - which implies  $\ln \Omega \left( \frac{z_t}{z_t^*} \exp \right) + \Omega \left( \frac{z_t}{z_t^*} \exp \right) = \ln \left( \frac{z_t}{z_t^*} \exp \right)$  - to rewrite this equation as,

$$\frac{z \rho_t(z)}{mc_t} = \Omega \left( \frac{z_t}{z_t^*} \exp \right) \quad (29)$$

which produces equation (5) in Proposition 2. To derive equation (6) in Proposition 2, note that, by definition,  $s(z_t^*) = 0$ . Applying this to the condition for the expenditure share implies,

$$0 = \frac{1}{n_t} + \zeta \left[ \frac{1}{n_t} \int \ln p_t(j) dj - \ln p(z_t^*) \right] \quad \rightarrow \quad \zeta \ln p(z_t^*) = \frac{1}{n_t} + \zeta \left[ \frac{1}{n_t} \int \ln p_t(j) dj \right]$$

Re-inserting this expression,

$$\begin{aligned} s_t(z) &= \zeta [\ln p(z_t^*) - \ln p_t(z)] = \zeta \left[ \ln \frac{p_t(z_t^*)}{p_t(z)} \right] = \zeta \left[ \ln \frac{z}{z_t^*} - \ln \Omega \left( \frac{z_t}{z_t^*} \exp \right) \right] \\ &= \zeta \left[ \ln \left( \frac{z}{z_t^*} \exp \right) - \ln \Omega \left( \frac{z_t}{z_t^*} \exp \right) - 1 \right] = \zeta \left[ \Omega \left( \frac{z_t}{z_t^*} \exp \right) - 1 \right] \end{aligned} \quad (30)$$

which implies,  $s(z_t) = \zeta \left[ \Omega \left( \frac{z_t}{z_t^*} \exp \right) - 1 \right]$ . Finally, for equation (7), we use,  $\vartheta_t(i) = [\rho_t(i) - mc_t(i)] y_t(i)$ . Combining with the price equation,  $\rho_t(i) = \Omega_t \frac{mc_t}{z_t(i)}$ , and the definition for marginal costs,  $mc_t(i) = \frac{mc_t}{z_t(i)}$ , we have,  $\pi_t(z) = \left(1 - \frac{1}{\Omega_t}\right) s_t(z) y_t$ , where  $\Omega_t \equiv \Omega \left( \frac{z_t}{z_t^*} \exp \right)$ . The demand curve for a good produced by a firm with productivity  $z$  is,  $s_t(z) y_t = \rho_t(z) y_t(z)$ , which implies,  $\vartheta_t(z) = (\Omega_t - 1) \left(1 - \frac{1}{\Omega_t}\right) \zeta y_t$ .

In the text, we also claim that the elasticity of price with respect to marginal costs decline with firms' productivity. Here we prove that claim. First, define  $u_t \equiv z/z_t^*$  and write the price as  $\rho_t(u) = \Omega_t(u \exp) mc_t/uz_t^*$ .

$$\begin{aligned} \epsilon_{\rho_t, mc_t}(u) &= \frac{d \ln \rho_t(u)}{d \ln mc} = \left\{ \left[ \frac{\partial \Omega_t(u \exp)}{\partial \ln u} - 1 \right] \frac{\partial \ln(u)}{\partial \ln z_t^*} - 1 \right\} \frac{\partial \ln z_t^*}{\partial \ln mc} + 1 \\ &= 1 - \Upsilon(u_t) \times \frac{\partial \ln z_t^*}{\partial \ln mc} < 1 \end{aligned}$$

where  $\Upsilon(u) \equiv \frac{\partial \ln \Omega(u \exp)}{\partial \ln u} = \frac{\partial}{\partial u} \Omega(u \exp) \frac{u}{\Omega(u \exp)} = [1 + \Omega(u \exp)]^{-1} < 1$  and  $\frac{\partial \ln(u)}{\partial \ln z_t^*} = -1$  and  $\frac{\partial \ln z_t^*}{\partial \ln mc} > 0$ . By direct differentiation,

$$\frac{d}{du} \Upsilon(u) = -\frac{\Omega'(u \exp)}{[1 + \Omega(u \exp)]^{-2}} < 0 \quad (31)$$

Therefore  $\frac{d}{du} \epsilon_{\rho_t, mc_t}(u) > 0$ , which implies larger, more productive firms (i.e., higher  $u_t$ ) have greater price pass-through. Rodriguez-Lopez (2011) refers to this case as the exact translog case.

## A.2. Proof of Lemma 1 and 2

For Lemma 1, make the change of variables,  $u = z/z_t^*$ . Totally differentiating,  $J'(z_t^*) = -j(1)g(z_t^*) - \frac{1}{(z_t^*)^2} \int_{z_t^*}^{z_{\max}} z j' \left( \frac{z}{z_t^*} \right) dG(z) < 0$ . Consider the integral  $J(a) = \int_a^{z_{\max}} j \left( \frac{z}{a} \right) dG(z)$ , where  $a > 0$  is any number. Since the integral exists, for any  $\varepsilon > 0$ , there should exist  $k(\varepsilon, a) > a$ , such that  $\int_{k(\varepsilon, a)}^{z_{\max}} j \left( \frac{z}{a} \right) dG(z) < \varepsilon$ . Since  $j' \geq 0$ , and  $k(\varepsilon, a) > a$ , it is true that  $j \left( \frac{z}{a} \right) > j \left( \frac{z}{k(\varepsilon, a)} \right)$  and  $\int_{k(\varepsilon, a)}^{z_{\max}} j \left( \frac{z}{k(\varepsilon, a)} \right) dG(z) < \int_{k(\varepsilon, a)}^{z_{\max}} j \left( \frac{z}{a} \right) dG(z) < \varepsilon$ . This completes the proof.

Now let  $g(u)$  be any PDF and  $j_1(u)$  and  $j_2(u)$  two positive increasing functions defined at the same domain. Let  $g(u)$  be a density function such that it's elasticity,

$$\epsilon(u) = -\frac{ug'(u)}{g(u)} \quad (32)$$

is weakly increasing. Let  $j_1(u)$  and  $j_2(u)$  be positive functions such that  $\frac{j_1(u)}{j_2(u)}$  is strictly increasing. The ratio  $\frac{J_1(z_t^*)}{J_2(z_t^*)}$  is a decreasing function of  $z_t^*$ , where  $J_i(z^*) = \int_{z^*}^{z_{\max}} j_i \left( \frac{z}{z^*} \right) g(z) dz$ .

To prove this statement, we start by using the following substitution  $u = \frac{z}{z^*}$ , which implies  $J_i(z^*) = z^* \int_1^{z_{\max}/z_t^*} j_i(u) g(z_t^*u) du$ . We differentiate this as,

$$J_i'(z) = \frac{J_i(z_t^*)}{z_t^*} - \frac{z_{\max}}{z_t^*} j_i(z_{\max}/z_t^*) g(z_{\max}) + \int_1^{z_{\max}/z_t^*} j_i(u) (uz_t^*) g'(z_t^*u) du \quad (33)$$

Now consider the ratio,

$$\begin{aligned} \mathcal{J}(z_t^*) &= (J_2(z_t^*))^2 \frac{d}{dz_t^*} \left[ \frac{J_1(z_t^*)}{J_2(z_t^*)} \right] = J_1'(z_t^*) J_2(z_t^*) - J_1(z_t^*) J_2'(z_t^*) \\ &= \left( \frac{J_1(z_t^*)}{z_t^*} - \frac{z_{\max}}{z_t^*} j_1 \left( \frac{z_{\max}}{z_t^*} \right) g(z_{\max}) + \int_1^{z_{\max}/z_t^*} j_1(u) (uz_t^*) g'(z_t^*u) du \right) J_2(z_t^*) \\ &\quad - \left( \frac{J_2(z_t^*)}{z_t^*} - \frac{z_{\max}}{z_t^*} j_2 \left( \frac{z_{\max}}{z_t^*} \right) g(z_{\max}) + \int_1^{z_{\max}/z_t^*} j_2(u) (uz_t^*) g'(z_t^*u) du \right) J_1(z_t^*) \end{aligned}$$

therefore,

$$\begin{aligned}
\mathcal{J}(z_t^*) &= \int_1^{z_{\max}/z_t^*} j_2(u) g(uz_t^*) du \int_1^{z_{\max}/z_t^*} j_1(u) g(uz_t^*) \left[ \frac{uz_t^* g'(uz_t^*)}{g(uz_t^*)} \right] du \\
&\quad - \int_1^{z_{\max}/z_t^*} j_1(u) g(uz_t^*) du \int_1^{z_{\max}/z_t^*} j_2(u) g(uz_t^*) \left[ \frac{uz_t^* g'(uz_t^*)}{g(uz_t^*)} \right] du \\
&\quad + z_{\max} g(z_{\max}) \left[ \int_1^{z_{\max}/z_t^*} j_1(u) g(uz_t^*) j_2\left(\frac{z_{\max}}{z_t^*}\right) du \right] \\
&\quad - z_{\max} g(z_{\max}) \left[ \int_1^{z_{\max}/z_t^*} j_2(u) g(uz_t^*) j_1\left(\frac{z_{\max}}{z_t^*}\right) du \right]
\end{aligned} \tag{34}$$

We consider the terms in the first two lines and the final two lines of equation (34) separately.

The sign of the first two lines can be written as,

$$\frac{\int_1^{z_{\max}/z_t^*} \epsilon(uz_t^*) j_2(u) g(uz_t^*) du}{\int_1^{z_{\max}/z_t^*} j_2(u) g(uz_t^*) du} - \frac{\int_1^{z_{\max}/z_t^*} \epsilon(uz_t^*) j_1(u) g(uz_t^*) du}{\int_1^{z_{\max}/z_t^*} j_1(u) g(uz_t^*) du} \tag{35}$$

where  $\epsilon(z) = -\frac{zg'(z)}{g(z)}$ . Now consider CDFs defined as,

$$G_i(u) = \int_{-\infty}^u g_i(y) dy = \frac{\int_1^y j_i(u) g(uz_t^*) du}{\int_1^{z_{\max}/z_t^*} j_i(u) g(uz_t^*) du} \quad \text{for } i = 1, 2; \quad 1 < y < z_{\max}/z_t^*$$

Damjanovic (2005) shows (formula 3) that if  $\frac{j_1(u)}{j_2(u)}$  is an increasing function of  $u$  then,  $G_1(u) < G_2(u)$ . This further implies that for any weakly increasing function  $\epsilon(u)$ ,

$$\int_{-\infty}^{+\infty} \epsilon(u) g_1(u) du > \int_{-\infty}^{+\infty} \epsilon(u) g_2(u) du \tag{36}$$

and (35) is negative if  $\epsilon(u)$  is increasing.

The sign of the second two lines in (34) can be written as,

$$\int_1^{z_{\max}/z_t^*} \left[ j_1(u) j_2\left(\frac{z_{\max}}{z_t^*}\right) - j_2(u) j_1\left(\frac{z_{\max}}{z_t^*}\right) \right] g(uz_t^*) du \tag{37}$$

However, as  $j_1/j_2$  is an increasing function, for any  $u < \frac{z_{\max}}{z_t^*}$ , then  $\frac{j_1(u)}{j_2(u)} < \frac{j_1\left(\frac{z_{\max}}{z_t^*}\right)}{j_2\left(\frac{z_{\max}}{z_t^*}\right)}$ , and the integral in equation (37) is negative. In this case, so is the final term in equation (34).

Here we note that the first two lines of equation (34) and the term in equation (35) equals zero for Pareto distribution. For a Pareto distribution,  $\epsilon(u) = k + 1$ , where  $k > 0$  is the shape parameter, and the ratio  $\frac{J_1(z_t^*)}{J_2(z_t^*)}$  is constant. The final line of equation (34) is negative for any truncated distribution including truncated Pareto. For a truncated Pareto distribution, or any other truncated distribution which satisfies (32),  $\frac{J_1(z_t^*)}{J_2(z_t^*)}$  is a decreasing function of  $z_t^*$ .

#### A.4. Proof of Proposition 5

We need to prove that the following function is increasing,

$$Z(z_t^*) = \left[ \frac{\pi_{3,t}}{\pi_{1,t}} (1 - \alpha) \right]^{1-\alpha} \left( \frac{\pi_{4,t}}{\pi_{1,t}} \right)^\alpha$$

First, note that,  $\alpha\pi_{4,t} = \pi_{1,t} - \pi_{3,t}$ , such that the factor prices can be written as,

$$\ln Z(z_t^*) = (1 - \alpha) \ln \pi + \alpha \ln(1 - \pi) + \Sigma$$

where we define  $\pi = \frac{\pi_{3,t}}{\pi_{1,t}}$  and  $\Sigma = \ln [(1 - \alpha)^{1-\alpha} (1/\alpha)^\alpha]$ . Differentiating,

$$\frac{d}{dz_t^*} \ln Z(z_t^*) = \left( \frac{1 - \alpha}{\pi} - \frac{\alpha}{1 - \pi} \right) \pi' = \frac{\pi'}{\pi(1 - \pi)} [(1 - \alpha) - \pi]$$

Lemma 2 implies that  $\pi'$  and so,

$$\begin{aligned} [(1 - \alpha) - \pi] &= \frac{1}{\pi_{1,t}} [(1 - \alpha) \pi_{1,t} - \pi_{3,t}] \\ &= \frac{1}{\pi_{1,t}} \left\{ \int_{z_t^*}^{z_{\max}} \frac{\Omega_t - 1}{\Omega_t} [(1 - \alpha) \Omega_t - (\Omega_t - \alpha)] dG(z) \right\} \\ &= -\frac{\alpha}{\pi_{1,t}} \left[ \int_{z_t^*}^{z_{\max}} \frac{(\Omega_t - 1)^2}{\Omega_t} dG(z) \right] < 0 \end{aligned}$$

This completes the proof.

#### A.5. Corollary

The proof follows immediately from Lemma 2 which implies that the ratio  $\frac{\pi_{1,t}}{\pi_{2,t}}$  is an increasing function of  $z_t^*$  for a truncated Pareto. This implies that  $\widehat{\pi}_{1,t} > \widehat{\pi}_{2,t}$ , and similarly,  $\widehat{\pi}_{2,t} > \widehat{\pi}_{3,t}$ , where a caret denotes the log deviation of a variable from its steady-state value. Thus,  $\widehat{\pi}_{1,t} > \widehat{\pi}_{2,t} > \widehat{\pi}_{3,t}$  also implies  $\zeta_3 > \zeta_2 > \zeta_1$ .

To complete the proof we show that  $\zeta_1 > 0$ . Recall,  $\pi_{1,t} = \int_{z_t^*}^{z_{\max}} \left[ \Omega \left( \frac{z}{z_t^*} \exp \right) - 1 \right] g(z) dz$ . Using  $u = z/z_t^*$ , we write this as,  $\pi_{1,t} = z_t^* \int_1^{z_{\max}/z_t^*} (\Omega_t(u \exp) - 1) g(uz_t^*) du$ , where  $g(uz_t^*) = \frac{\kappa}{1 - z_{\max}^{-\kappa}} (uz_t^*)^{-1-\kappa}$ , and so,

$$\pi_{1,t} = \frac{\kappa (z_t^*)^{-\kappa}}{1 - z_{\max}^{-\kappa}} \int_1^{z_{\max}/z_t^*} (\Omega_t - 1) u^{-1-\kappa} du.$$

Now consider the log deviation of  $\pi_{1,t}$ ,

$$\begin{aligned} \widehat{\pi}_{1,t} &= \frac{\partial \pi_{1,t}}{\partial z_t^*} \frac{z^*}{\pi_{1,t}} \widehat{z}_t^* = \left( -\kappa - \frac{\kappa (z_t^*)^{-\kappa}}{1 - z_{\max}^{-\kappa}} \frac{\Omega - 1}{(z_{\max}/z_t^*)^\kappa} \frac{1}{\pi_{1,t}} \right) \widehat{z}_t^* = -(\kappa + \zeta_1) \widehat{z}_t^* \\ \zeta_1 &\equiv \frac{\Omega(z_{\max}/z_t^* \times \exp) - 1}{\int_1^{z_{\max}/z_t^*} [\Omega(u \exp) - 1] u^{-1-\kappa} du} (z_{\max}/z_t^*)^{-\kappa} > 0 \end{aligned} \quad (38)$$

This completes the proof.

## Appendix B

In this Appendix we present additional details for the aggregated equations in Section 2.4 of the main text.

### B.1. Allocations and Factor Prices

The demand curve for good  $i$  is  $\rho_t(i) y_t(i) = s_t(i) y_t$ , where  $s_t(i)$  is defined in Proposition 2 and is the expenditure share. We can write the optimal price equation for firm with productivity  $z_t(i)$  in terms of the expenditure share or the Lambert-W functions, as,  $\rho_t(i) = \left[1 + \frac{s_t(i)}{\zeta}\right] \frac{mct}{z_t(i)}$  or  $\rho_t(i) = \Omega\left(\frac{z_t}{z_t^*} \exp\right) \frac{mct}{z_t(i)}$ , as described used in Proposition 1. The relationship between  $s_t(i)$  and  $\Omega\left(\frac{z_t}{z_t^*} \exp\right)$  is,

$$s_t(z) = \zeta (\Omega_t - 1) \quad \text{where} \quad \Omega_t \equiv \Omega\left(\frac{z_t}{z_t^*} \exp\right) \quad (39)$$

for all  $i$ . This expression is reported in Proposition 2 and we use it repeatedly in generating aggregate conditions.

We start with the demand curve. Aggregating over  $z > z_t^*$ , we have,

$$\begin{aligned} \int_{z_t^*}^{z_{\max}} \rho_t(z) y_t(z) dG(z) &= \zeta y_t \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z) \\ \Leftrightarrow 1 &= \zeta \mathbf{N}_t \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z) \right] \end{aligned} \quad (40)$$

where  $\mathbf{N}_t$  is the mass of entrants. This is equation (9) in the main text. Now consider the profit function, which we write as,  $\pi_t(z) = (\Omega_t - 1) (1 - 1/\Omega_t) \zeta y_t$ . The free entry is such that,  $\mathbb{E} \vartheta_t(z) \equiv \int_{z_t^*}^{z_{\max}} \vartheta_t(z) dG(z) = w_t (f/a_t)$ , and in this case,

$$\zeta \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) (1 - 1/\Omega_t) dG(z) \right] y_t = w_t \frac{f}{a_t} \quad (41)$$

which is equation (10) in the main text.

Labor is used for entry and production. Total labor demand is  $L_t = \mathbf{N}_t \int_{z_t^*}^{z_{\max}} l_t(z) dG(z) + \mathbf{N}_t f$ . We want an expression for  $l_t(z)$ . Cost minimization implies,  $w_t l_t(z) = [(1 - \alpha) / \alpha] r_t k_t(z)$ , where  $r_t k_t(i) = \alpha (y_t / a_t z_t) w_t^{1-\alpha} r_t^\alpha / \psi$ . In combination with the demand for a product,  $w_t l_t(z) = \frac{w_t}{a_t z_t(i)} \frac{s_t(z) y_t}{\rho_t(z)} \left[ \left( \frac{1-\alpha}{\alpha} \right) \frac{r_t}{w_t} \right]^\alpha = (1 - \alpha) s_t(z) y_t / \Omega_t$ , where we used  $mc_t = \frac{1}{a_t} \frac{w_t^{1-\alpha} r_t^\alpha}{(1-\alpha)^{1-\alpha} \alpha^\alpha}$  and  $\rho_t(z) = \Omega_t \frac{mc_t}{z_t}$ . Finally,  $\int_{z_t^*}^{z_{\max}} w_t l_t(z) dG(z) = (1 - \alpha) \left( \frac{\Omega_t - 1}{\Omega_t} \right) \zeta y_t$ , using the expression for market share. Combining terms, total labor costs are,

$$w_t L_t = \zeta \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - \alpha) (1 - 1/\Omega_t) dG(z) \right] \mathbf{N}_t y_t \quad (42)$$

In combination with equation (40), this implies,

$$y_t = w_t L_t \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z) \right] / \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - \alpha) (1 - 1/\Omega_t) dG(z) \right]$$

which generates  $y_t = w_t L_t$  for  $\alpha \rightarrow 0$ . The demand for capital is,  $K_t = \mathbf{N}_t \int_{z_t^*}^{z_{\max}} k_t(z) dG(z)$ , where  $r_t k_t(z) = \left( \frac{\alpha}{1-\alpha} \right) w_t l_t(z)$ . This leads to,

$$r_t K_t = \mathbf{N}_t \int_{z_t^*}^{z_{\max}} \left( \frac{\alpha}{1-\alpha} \right) w_t l_t(z) dG(z) = \alpha \zeta \left[ \int_{z_t^*}^{z_{\max}} (1 - 1/\Omega_t) dG(z) \right] \mathbf{N}_t y_t \quad (43)$$

Using the definition of the markup provided in the text, equations (42) and (43) are used to generate equations (12) and (13) in the main text.

## B.2. Price and Herfindahl Index

In this section, we derive equation (11). Recall, the expenditure function is,

$$\ln(e_t) = \ln c_t + \nu_t + \frac{1}{n_t} \int_{i \in \Delta} \ln p_t(i) di + \frac{\zeta}{2n_t} \int_{i \in \Delta} \int_{j \in \Delta} \ln p_t(i) [\ln p_t(j) - \ln p_t(i)] dj di \quad (44)$$

where  $e_t$  is the minimum expenditure required to obtain  $c_t$  and  $\nu_t \equiv 1/2\zeta n_t$ . The price index, denoted  $P_t$ , is the minimum expenditure needed to buy a unit of  $c_t$ . Using equation (44) the price index in our model is:

$$\ln P_t = \nu_t + \frac{1}{n_t} \int \ln p(i) di + \frac{\zeta}{2n_t} \left[ \int \ln p(i) di \right]^2 - \frac{\zeta}{2} \int [\ln p(i)]^2 di \quad (45)$$

Inserting  $\frac{1}{n_t} \int \ln p_t(i) di = \frac{1}{1-G_t} \int_{z_t^*} \ln p_t(z) dG(z)$  and  $n_t = \mathbf{N}_t(1-G_t)$  into equation (45), we find:

$$(1-G_t)(\ln P_t - \nu_t) = \int_{z_t^*} \ln p_t(z) dG(z) + \frac{\zeta \mathbf{N}_t}{2} \left[ \int_{z_t^*} \ln p_t(z) dG(z) \right]^2 - \zeta \frac{\mathbf{N}_t(1-G_t)}{2} \int_{z_t^*} [\ln p_t(z)]^2 dG(z)$$

where  $G_t \equiv G(z_t^*)$ . We subtract  $\ln p(z_t^*)$  from both sides, so,

$$(1-G_t)[- \ln \rho(z_t^*) - \nu_i] = \int_{z_t^*} \ln \frac{\rho_t(z)}{\rho(z_t^*)} dG(z) - \zeta \frac{\mathbf{N}_t(1-G_t)}{2} \int_{z_t^*} \left[ \ln \frac{\rho_t(z)}{\rho(z_t^*)} + \ln p(z_t^*) \right]^2 dG(z) + \frac{\zeta \mathbf{N}_t}{2} \left[ \int_{z_t^*} \left( \ln \frac{\rho_t(z)}{\rho(z_t^*)} \right) dG(z) + (1-G_t) \ln p(z_t^*) \right]^2$$

where  $\rho_t(i) \equiv p_t(i)/P_t$ . Now recall,

$$- \ln \frac{\rho(z)}{\rho(z_t^*)} = z_t \frac{\rho(z)}{m c_t} - 1 = \Omega \left( \frac{z}{z_t^*} \exp \right) - 1 \quad \text{and} \quad \rho(z_t^*) = m c_t / z_t^*$$

where  $\Omega \left( \frac{z}{z_t^*} \exp \right)$  is the Lambert-W function and the latter expression is the reservation price. Substituting for demand - equation (40) - we find,

$$- \ln \rho(z_t^*) - \nu_t = - \frac{1}{\zeta \mathbf{N}(1-G_t)} + \frac{\zeta \mathbf{N}_t(1-G_t)}{2} \left[ \ln p(z_t^*) - \frac{1}{\zeta \mathbf{N}(1-G_t)} \right]^2 - \frac{\zeta \mathbf{N}_t}{2} \int_{z_t^*} [\ln p(z_t^*) - (\Omega_t - 1)]^2 dG(z)$$

which simplifies to,

$$\ln \frac{m c_t}{z_t^*} = \frac{\zeta}{2} \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1)^2 dG(z) \right] \mathbf{N}_t \tag{46}$$

where  $\Omega_t \equiv \Omega \left( \frac{z}{z_t^*} \exp \right)$  and  $\pi_{3,t} \equiv \int_{z_t^*}^{z_{\max}} (\Omega_t - 1)^2 dG(z)$  is reported in the main text.

Finally we can consider the Herfindahl index. Recall,  $s(z_t) = \zeta(\Omega_t - 1)$ , with the Herfindahl index defined as,  $\mathbf{H}(z_t^*) \equiv \mathbf{N}_t \int_{z_t^*}^{z_{\max}} s^2(z_t) dG(z)$ . Note that we expect all shares equal to

one, and so,  $\mathbf{N}_t \int_{z_t^*}^{z_{\max}} s(z_t) dG(z) = 1$ . In combination, we have,

$$\mathbf{H}(z_t^*) \equiv \zeta \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1)^2 dG(z) \right] / \left[ \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z) \right] \quad (47)$$

which is reported in the main text and  $\pi_{1,t} \equiv \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z)$ .

## Appendix C

In this Appendix we present additional details for the simplified equations in Section 3.3 of the main text. Using the equation for factor prices when  $\alpha \rightarrow 0$  we have  $y_t = w_t L_t$ . Absent investment,  $y_t = c_t$ , such that, when  $c_t u_c(t) = 1$ , we find,  $-L_t u_L(t) = 1$ . When  $\alpha \rightarrow 0$ , we also know  $\pi_{1,t} = z_t^* \int_1^{z_{\max}/z_t^*} (\Omega_t - 1) g(uz_t^*) du = \pi_{3,t}$ . We also assume the distribution of productivity is,  $g(z) = \kappa z^{-(\kappa+1)} / [1 - z_{\max}^{-\kappa}]$  and  $G(z) = (1 - z^{-\kappa}) / [1 - z_{\max}^{-\kappa}]$ , with  $z_{\min} = 1$ . This implies the log hazard ratio is,  $\phi(z^*) \equiv [g(z^*) z^*] / [1 - G(z^*)] = \kappa / [1 - (z_{\max}/z^*)^{-\kappa}]$ . We express the solution to the model as,

$$1 = \zeta \pi_{1,t} \mathbf{N}_t \quad ; \quad z_t^* \exp\left(\frac{\mathbf{H}_t}{2\zeta}\right) = w_t/a_t \quad ; \quad w_t = y_t \quad ; \quad w_t \left(\frac{f/a_t}{\zeta}\right) = \pi_{2,t} y_t \quad (48)$$

where  $L_t = 1$  and  $\mathbf{H}_t \equiv \mathbf{H}(z_t^*)$  which is a system of four equations that determine the variables  $\{z_t^*, \mathbf{N}_t, w_t, y_t\}$ . We have already established that  $z_t^*$  falls with  $L_t$ , which is now assume to be the aggregate shock. This leads to,

$$(\kappa + \zeta_2) \widehat{z}_t^* = -\widehat{\pi}_{2,t} = \widehat{a}_t \quad \text{and} \quad \widehat{\mathbf{N}}_t = -\widehat{\pi}_{1,t} = (\kappa + \zeta_1) \widehat{z}_t^* \quad (49)$$

which implies that as labor supply rises so does the productivity cut-off falls and firm entry. Product variety is,  $\widehat{n}_t = \widehat{\mathbf{N}}_t - [\phi(z^*)] \widehat{z}_t^* = \left\{ \zeta_1 - \frac{\kappa(z_{\max}/z^*)^{-\kappa}}{1 - (z_{\max}/z^*)^{-\kappa}} \right\} \widehat{z}_t^* > 0$ , where  $\zeta_1 > 0$  is reported above. Finally, we map the shock into output and the Herfindahl index, in the following way,

$$\widehat{y}_t = \widehat{a}_t + \widehat{z}_t^* + \left(\frac{\mathbf{H}}{2\zeta}\right) \widehat{\mathbf{H}}_t \quad ; \quad \widehat{\mathbf{H}}_t = (\zeta_1 - \zeta_3) \widehat{z}_t^* < 0 \quad (50)$$

I also derive expressions for average productivity and the markup. For average productivity,

$$\bar{z}_t = \frac{1}{1 - G(z^*)} \int_{z_t^*}^{z_{\max}} z dG(z) \quad \Rightarrow \quad \widehat{\bar{z}}_t = \phi(z^*) \left(1 - \frac{z^*}{\bar{z}}\right) \widehat{z}_t^* \quad (51)$$

Given the definition of the markup in the text,  $\mathbf{m}_t \equiv \pi_{1,t} / [1 - G(z_t^*)]$ , we have,

$$\begin{aligned} \mathbf{m}_t &= \frac{\kappa}{1 - (z_{\max}/z_t^*)^{-\kappa}} \int_1^{z_{\max}/z_t^*} (\Omega_t - 1) u^{-1-\kappa} du \\ &\Rightarrow \quad \widehat{\mathbf{m}}_t = \kappa \frac{(z^*/z_{\max})^\kappa}{1 - (z^*/z_{\max})^\kappa} \left[1 - \left(\frac{\Omega - 1}{\mathbf{m}}\right)\right] \widehat{z}_t^* \end{aligned} \quad (52)$$

such that as  $z_{\max} \rightarrow \infty$  then  $\mathbf{m}_t \rightarrow \kappa \left\{ \int_1^\infty (\Omega - 1) u^{-\kappa-1} du \right\}$  and  $\widehat{\mu}_t \rightarrow 0$ .

## References

- ARKOLAKIS, C., COSTINOT, A., DONALDSON, D. and RODRIGUEZ-CLARE, A. (2017), “The Elusive Pro-Competitive Effects of Trade”, *Review of Economic Studies*, forthcoming.
- BARSEGHYAN, L. and DICECIO, R. (2011), “Entry Costs, Industry Structure, and Cross-Country Income and TFP Differences”, *Journal of Economic Theory* 106, 1828-1851.
- BARTELSMAN, E. and DOMS, M. (2000), “Understanding Productivity: Lessons from Longitudinal Microdata”, *Journal of Economic Literature* 38, 569-594.
- BERGIN, P. and FEENSTRA, R. (2000), “Staggered Price Setting, Translog Preferences, and Endogenous Persistence”, *Journal of Monetary Economics* 45, 657-680.
- BERNARD, A. EATON, J., JENSEN, B. and KORTUM, S. (2003), “Plants and Productivity in International Trade”, *American Economic Review* 93, 1268-1290.
- BILBIIE, F., GHIRONI, F. and MELITZ, M. (2012), “Endogenous Entry, Product Variety, and Business Cycles”, *Journal of Political Economy* 120, 304-345.
- de BLAS, B. and RUSS, K. (2015), “Understanding Markups in the Open Economy”, *American Economic Journal: Macroeconomics* 7, 157-180.
- CHATTERJEE, S. and COOPER, R. (1993), “Entry and Exit, Product Variety and the Business Cycle”, NBER Working Papers 4562.
- COMBES, P., DURANTON, G., GOBILLON, L., PUGA, D. and ROUX, S. (2012), “The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection”, *Econometrica* 80, 2543-2594.
- DAMJANOVIC, T. (2005), “Lorenz dominance for transformed income distributions: A simple proof”, *Mathematical Social Sciences* 50, 234-237.
- DAVIS, S., FABERMAN, R. and HALTIWANGER, J. (2006), “The Flow Approach to Labor

Markets: New Data Sources and Micro-Macro Links”, *Journal of Economic Perspectives* 20, 3-26.

DEVEREUX, M., HEAD, A. and LAPHAM, B. (1996), “Aggregate Fluctuations with Increasing Returns to Specialization and Scale”, *Journal of Economic Dynamics and Control* 20, 627-656.

ETRO, F. and COLCIAGO, A. (2010), “Endogenous Market Structures and the Business Cycle”, *Economic Journal* 120, 1201-1233.

FEENSTRA, R., and WEINSTEIN, D. (2017), “Globalization, Markups and US Welfare”, *Journal of Political Economy*, 125, 1041-1074.

FEENSTRA, R. (2018), “Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity”, *Journal of International Economics* 110, 16-27.

HONG, S. (2017), “Customer Capital, Markup Cyclicalities, and Amplification”, *St Louis Fed Working Paper* 2017-033A.

HOPENHYAN, H. and ROGERSON, R. (1993), “Job Turnover and Policy Evaluation: A General Equilibrium Analysis”, *Journal of Political Economy* 101, 915-938.

JAIMOVICH, N. and FLOETOTTO, M. (2008), “Firm Dynamics, Markup Variations and the Business Cycle”, *Journal of Monetary Economics* 55, 1238-1252.

GALI, J. (1994), “Monopolistic Competition, Business Cycles, and the Composition of Aggregate Demand”, *Journal of Economic Theory* 63, 73-96.

KING, R. and REBELO, S. (1999), “Resuscitating Real Business Cycles”, *Handbook of Macroeconomics*, in: J. B. Taylor & M. Woodford (ed.), *Handbook of Macroeconomics*, edition 1, volume 1, chapter 14, 927-1007.

- LEWIS, V. and POILLY, C., (2015), “Firm Entry, Markups and the Monetary Transmission Mechanism”, *Journal of Monetary Economics* 59, 670-685.
- MELITZ, M. (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, *Econometrica* 71, 1695-1725.
- MINNITI, A. and TURINO, F. (2013), “Multi-Product Firms and Business Cycle Dynamics”, *European Economic Review* 57, 75-97.
- MRAZOVA, M., NEARY, P. and PARENTI, M. (2016), “Sales and Markup Dispersion: Theory and Empirics”, mimeo, Oxford.
- OTTAVIANO, G. (2012), “Firm Heterogeneity, Endogenous Entry, and the Business Cycle”, NBER International Seminar on Macroeconomics, University of Chicago Press, 57-86.
- RESTUCCIA and ROGERSON, R. (2017), “The Causes and Costs of Misallocation”, *Journal of Economic Perspectives* 31, 151-174.
- RODRIGUEZ-LOPEZ, A. (2011), “Prices and Exchange Rates: A Theory of Disconnect”, *Review of Economic Studies* 78, 1135-1177.
- ROTEMBERG, J. and WOODFORD, M. (1991), “Markups and the Business Cycle”, *NBER Macroeconomics Annual*, 63-129.
- SYVERSON, C. (2011), “What Determines Productivity?”, *Journal of Economic Literature* 49, 326-365.

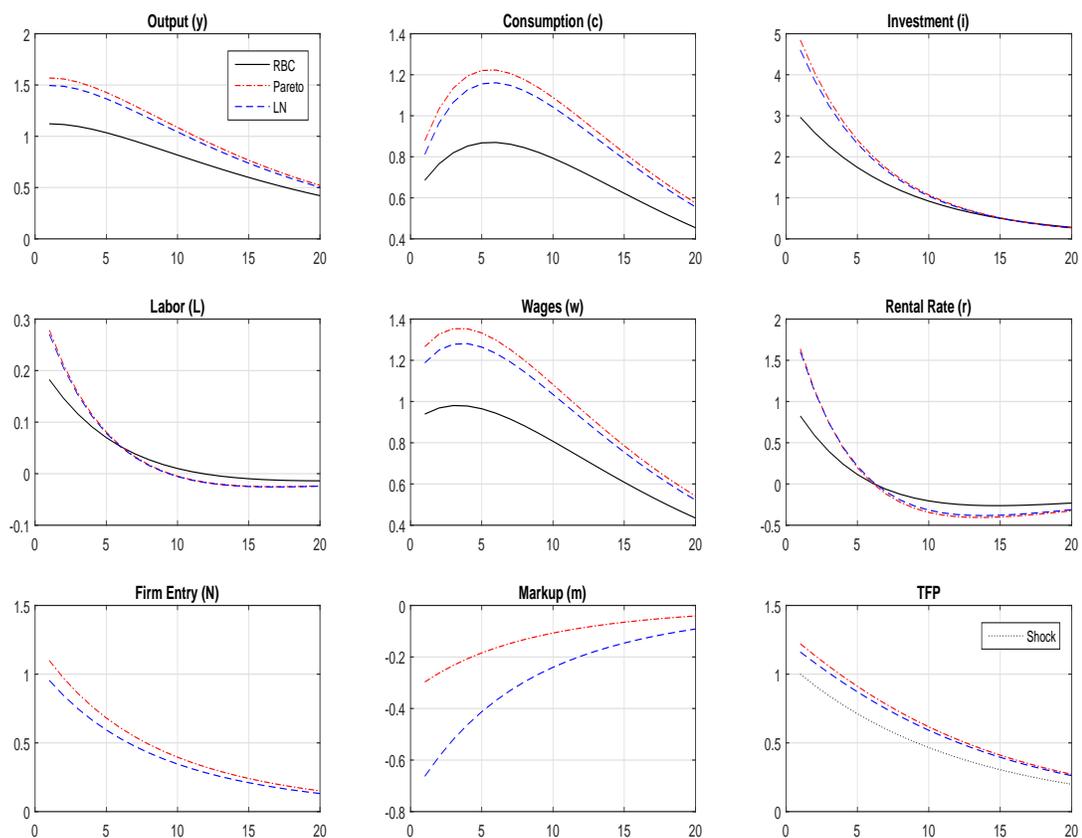
**Table 1: Model Equations**

| Description            | Model Equations   | Auxiliary Variables  |
|------------------------|---|--|
| Product Demand         | $\mathbf{N}_t = 1/\zeta \pi_{1,t}$  | $\pi_{1,t} = \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) dG(z)$                       |
| Free Entry             | $\pi_{2,t} y_t = (f/\zeta) (w_t/a_t)$   | $\pi_{2,t} = \int_{z_t^*}^{z_{\max}} (\Omega_t - 1) (1 - 1/\Omega_t) dG(z)$      |
| Price Index            | $a_t z_t^* \times \exp(\mathbf{H}_t/2\zeta) = w_t^{1-\alpha} r_t^\alpha / \psi$       | $\mathbf{H}_t = \int_{z_t^*}^{z_{\max}} (\Omega_t - 1)^2 dG(z) / \pi_{1,t}$      |
| Factor Price (labor)   | $w_t = \frac{y_t}{L_t} \pi_{3,t} / \mathbf{m}_t \int_{z_t^*}^{z_{\max}} dG(z)$        | $\pi_{3,t} = \int_{z_t^*}^{z_{\max}} (\Omega_t - \alpha) (1 - 1/\Omega_t) dG(z)$ |
| Factor Price (capital) | $r_t = \alpha \frac{y_t}{K_t} \pi_{4,t} / \mathbf{m}_t \int_{z_t^*}^{z_{\max}} dG(z)$ | $\pi_{4,t} = \int_{z_t^*}^{z_{\max}} (1 - 1/\Omega_t) dG(z)$                     |
| Labor Supply           | $w_t = -u_L(t) / u_c(t)$  | $\mathbf{m}_t = \pi_{1,t} / \int_{z_t^*}^{z_{\max}} dG(z)$                       |
| Capital Accumulation   | $i_t = K_{t+1} - (1 - \delta) K_t$  | —  |
| Euler Equation         | $u_c(t) = \beta \mathbb{E}_t u_c(t+1) [r_{t+1} + (1 - \delta)]$                       | —  |
| Resource Constraint    | $y_t = c_t + i_t$   | —  |

**Table 2: Exogenous Parameters and Calibration**

| <b>Parameters Set Exogenously</b>                |            |        |                                |                              |
|--|------------|--------|--------------------------------|------------------------------|
| Statistic  | Parameter  | Value  | Target/Source                  |                              |
| Discount factor                                  | $\beta$    | 0.96   | 4% risk-free rate              |                              |
| Depreciation rate                                | $\delta$   | 0.1    | 10%                            |                              |
| Capital share                                    | $\alpha$   | 1/3    | -                              |                              |
| Translog scaling                                 | $\zeta$    | 1      | Normalization                  |                              |
| Inverse Frisch                                   | $\nu$      | 1.39   | Heathcote <i>et al.</i> (2010) |                              |
| <b>Parameters for Productivity Distributions</b> |            |        |                                |                              |
| Distribution                                     | Parameters | Values | Target(s)                      | Source                       |
| Pareto   |            |        |                                |                              |
| shape  | $\kappa$   | 1.653  | –                              | Nigai (2017)                 |
| location   | $z_{\max}$ | 3.835  | $m = 0.23$                     | Martins <i>et al.</i> (1996) |
| Log normal                                       |            |        |                                |                              |
| shape  | $\sigma$   | 0.321  | $m$                            | -                            |
| location   | $\mu$      | 0.445  | $\bar{z}^{\text{Pareto}}$      |                              |

Figure 1: Impulse Responses to a Technology Shock



**Table 3: Business Cycle Moments<sup>23</sup>**

| Variable $x$ | Second Moments |            |        |      |
|--------------|----------------|------------|--------|------|
|              |                | $\sigma_x$ |        |      |
|              | Data           | RBC        | Pareto | LN   |
| $y_t$        | 1.81           | 1.01       | 1.41   | 1.35 |
| $c_t$        | 1.35           | .60        | .76    | .70  |
| $i_t$        | 5.30           | 2.79       | 4.63   | 4.40 |
| $L_t$        | 1.79           | .18        | .28    | .27  |
| $w_t$        | .68            | .83        | 1.12   | 1.05 |
| $r_t$        | .30            | .84        | 1.68   | 1.64 |
| $TFP_t$      | —              | .92        | 1.12   | 1.07 |
| $N_t$        | —              | —          | 1.03   | .89  |
| $m_t$        | .99            | —          | .28    | .62  |

<sup>23</sup>Source for data and RBC moments: King and Rebelo (1999) and Etro and Colciago (2010).