Industry Dynamics, Investment and Business Cycles*

Julieta Caunedo†
Washington University in St Louis

February 14, 2013

Abstract

This paper investigates how features of the business cycle interact with technological restrictions at the firm level to generate dispersion in marginal products of ex ante identical firms. The model is able to deliver a non-monotonic relationship between dispersion in marginal products, aggregate productivity and the features of the business cycle. When aggregate uncertainty is low and dispersion in marginal products is low, aggregate productivity is high. But when aggregate uncertainty is high, aggregate productivity is low, and the allocation can be consistent with low dispersion in marginal products. These two alternative economies differ in their underlying industry dynamic. Hence, dispersion is an imperfect statistic of aggregate productivity in the model. Allocations are typically non efficient due to imperfect competition and non-convexities in production. I study the properties of the optimal industrial policy. In general, the efficient allocation does **not** dictate equalization of capital labor ratios across **all** firms. Allocations are dynamically optimal so there is no room for further reallocation. [JEL Codes: E32,L11,E23].

*I am greatly indebted to Rody Manuelli for his guidance and advice. I am very grateful to Yongs Shin for encouragement and invaluable suggestions. I thank Riccardo DiCecio, (...)
†Contact: jcaunedo@go.wustl.edu

1 Introduction

Dispersion in marginal products within narrowly defined industries is an stylized fact of modern economies. There are many reasons for which such disparity may arise: market incompleteness, subsidies/taxes to particular firms, only to name a few. Computed productivity at the firm level is proportional to marginal products of inputs in production. Recent research has identified computed productivity dispersion in and above USA levels, where market imperfections and industrial policy are relatively mild, as evidence of inefficiencies in production (Hsieh and Klenow, 2009). This paper contributes to this literature by exploring the implications of modeling technological restrictions at the firm level. I find that excess dispersion in marginal products can be consistent with efficient allocations (given a particular market structure) and is in general, not independent of other features of the economy, such as the business cycle. The paper highlights that dispersion in marginal products is an imperfect measure of the associated efficiency losses.

The study of the mechanism that originates cross sectional dispersion in productivity is critical to assess whether there are associated efficiency losses or not, and its impact on aggregate productivity. In this paper, I will focus on the role of business cycles in influencing dispersion in marginal products. Real options theory suggests that the underlying stochastic process generating the cycles, shapes investment decisions. In a partial equilibrium set up, they are mainly characterized by the persistency and volatility of the shocks, with an increase in the former inducing boosts of investment and an increase in the latter inducing inaction. This is turn, induces dispersion in marginal products. Such a set up, however, does not account for changes in factor prices in response to investment and disinvestment decisions, which may have a quantitative

[0](...)Adrian Peralta Alva, Juan M. Sanchez and Raul Santaeulalia-Llopis for many discussions,as well as seminar participants at LACEA(2011), Midwest Macroeconomic Meetings (2012), Vigo Workshop in Dynamic Macroeconomic (2012) and Santader Campus Nobel Event (2012), Washington University Graduate Student Conference (2012), in particular Isaac Bailey. I am grateful from support by the Center for Research in Economics and Strategy (CRES), in the Olin Business School, Washington University in St. Louis.

impact in those predictions. Hence, it is necessary to study the problem within a general equilibrium framework.

To understand the relationship between dispersion in marginal products, the business cycle and efficiency losses, consider the following example. Suppose we observe disparity in marginal products among two firms holding the same blueprint: an incumbent operating at high capacity and an entrant operating at lower scale. If we focus only on this snapshot, such disparity can be read as an inefficiency. Suppose, however, that the incumbent had entered the market at an earlier time during a boom, and current market conditions have worsened since. Therefore, the entrant finds optimal to wait to scale up capacity until market conditions improve, while the incumbent does not exit because its option value of remaining in the market is positive. Hence, a gap between the marginal products of capital for these two otherwise identical firms, is consistent with them holding optimal investment strategies.

This approach has implications for the relationship between business cycle features and cross sectional dispersion in marginal products, and is of particular interest for cross country analysis of aggregate productivity. If we compare two countries with the same aggregate productivity, but distinct features of the business cycle, dispersion in marginal products across them may differ. At a given period in time, the measure and type of firms that are holding capital away from the one they would have chosen if entering the market in the current period, depends on the degree of uncertainty in the economy. As mentioned before, this paper complements previous studies that draw conclusions as of aggregate productivity from observations in dispersion in productivity. Maps between dispersion and aggregate productivity, and dispersion and the features of the business cycle, are usually not one to one. Therefore, measures of efficiency gains stemming from elimination of dispersion in marginal products should be adjusted to account for these non-linear relationships.

The model I study in this paper generates a non-monotonic relationship between the volatility of the business cycle and cross sectional dispersion in productivity. At one extreme, when the volatility of the aggregate productivity process is low, there would be little churning in the industry. The economy approximates a stationary one, where the only determinant of investment decisions is the firm's idiosyncratic productivity. The mechanism explained before becomes irrelevant. At the other extreme, when the volatility of the process is very high, the value of waiting both for the incumbent and for the investor outside the market rises. Hence, incentives for entry and exit are lower, firm churning drops and the mechanism vanishes again. Although at the two extremes cross sectional dispersion in productivity will be relatively low, mean TFP could be substantially different, i.e. we expect the economy with higher volatility to be less productive on average.

Furthermore, suppose that the stochastic process determining aggregate productivity is an autoregressive one. Both, persistence and volatility of the shocks will determine the unconditional volatility of the process. When the process is approximated by a Markov chain with two states, the persistence parameter is fully determined by the diagonal elements of the transition matrix, while the volatility of the shock is determined by the range of values of the aggregate shock. The model implies that for a given level of unconditional volatility, the size of the shocks and the transition probabilities have different implications for the relationship between volatility of output, cross sectional dispersion and aggregate productivity. Hence, details about the distribution of shocks matter to explain the response of the economy to particular shocks. On the one hand, drops in persistence tend to shrink dispersion and increase volatility of output while lowering aggregate productivity. On the other hand, increases in the size of the shocks induce higher cross sectional dispersion, lower volatility of output and aggregate productivity.

This disparate behavior highlights the importance of addressing questions of efficiency in a fully fledged general equilibrium model. For a given market structure, cross sectional dispersion in productivity can be consistent with constrained efficiency, but market allocations need not be optimal due to other considerations, e.g. market power. In other words, a planner with the same technological restrictions at the firm level would not generate the decentralized allocation. Studying optimality in models with aggregate uncertainty, heterogenous firms and irreversibilities and indivisibilities is often not possible. I show that if the source of irreversibilities are sunk costs, there is a centralized problem whose allocation can be decentralized as a market allocation. To decentralize the optimal allocation, adjustment costs faced by firms need to vary across aggregate states, and will typically differ from the ones assumed in the centralized problem. Comparing the optimal and decentralized allocation proves revealing: for a given set of adjustment costs, the former implies subsidies to large and more productive firms and induces higher competition in the market than the latter. Still, wedges in marginal products across ex ante identical firms are not fully eliminated in the optimal allocation.

The rest of the paper is organized as follows. The following section reviews the literature, then we introduce the model. Finally we calibrate the model to the USA economy and analyze the implications of the model for cross sectional dispersion in productivity, aggregate productivity, industry dynamic and optimal industrial policy.

2 Literature Review

The relationship between business cycles and industry dynamics has received increased attention by the economics literature. Models of industry equilibrium with complete markets (for example Hopenhayn, 1992) typically display aggregation. Hence, the role of heterogeneity in equilibrium allocations is overshadowed. As marginal product are equalized and capital labor ratios are proportional to productivity differences, the model boils down to one of a representative firm with average productivity. When the relationship between productivity, size (employment or assets) and output is non-monotonous, heterogeneity matters in a non-trivial way. Marginal products may differ in excess of productivity differences as is typically the case in incomplete market models.

Some efforts departing from this one to one map are those by Lee and Mukoyama(2008), Clementi and Palazzo(2010), Khan and Thomas (2011) (in incomplete markets).Lee and Mukoyama(2008) provide evidence of disparate entry and exit behavior along the cycle and propose a model to quantitatively explain those facts. They analyze the effect of fluctuations in fixed production costs and labor adjustment costs on the industry dynamic in a model with no capital. Clementi and Palazzo (2010) analyze the propagation of aggregate shocks due to entry and exit of firms when firms are allowed to accumulate capital. Khan and Thomas (2011) study the effect of irreversibilities and collateral constraints in equilibrium allocations in an economy without industry dynamic. They find that both frictions reinforce each other in slowing down reallocation. In open economies, Alessandria (2008) and Bilbiie,Ghironi and Melitz (2007) studied the effect of exit, entry and export behavior on aggregate fluctuations too.

This paper contributes to the literature by studying the effect of interactions between features of the business cycle and non-convexities at the micro level. Irreversibilities and indivisibilities are a propagation mechanism of aggregate technological shocks, but its effect depends on the nature of aggregate fluctuations, i.e. persistence and (unconditional) volatility of the underlying process. More important, we study the model implications for the design of optimal industrial policy. As a subproduct, we provide an alternative algorithm to the standard approximation methods of Krusell and Smith (1998). The algorithm is useful for the study of allocations under a certain class of non-convexities in production[1].

One of the main drivers for the introduction of industry dynamic into standard RBC models seems to be the observed heterogeneity in firms behavior as of capital accumulation, employment (Moscarini & Postel-Vinay,2011, Davis, et.al., 2010), entry and exit (Barstelman, et.al.,2009) . However, as already pointed out by Melitz (2003), if reallocation is costless, economies with heterogeneous firms will be unidentifiable from an economy in which all firms hold the average productivity in the market. The result follows form the equalization of the shadow price of capital as in the Gorman(1953) tradition (in that case the shadow price of wealth).

In this model, the response of the economy to aggregate shocks depends on the measure of firms that hold capital away from the one they would have chosen if entering the market in the current period. Marginal products may differ for a given period of time (static inefficiency) but capacity choices are dynamically optimal, so there will be no room for reallocation. The result resembles earlier intuitions drawn in Caballero and Hammour (1998) when analyzing factor specificity.

[1] If irreversibilities and indivisibilities are binding for a non-negligible measure of firms (in the sense that, they matter for the actual dynamic of the endogenous distribution), then approximation methods may generate allocations arbitrarily far from the actual ones.

First generation models of irreversibility analyzed the impact of lumpy investment at the micro level on aggregate investment dynamic. Exploiting a partial equilibrium setup, they found substantial contribution of lumpiness to aggregate investment fluctuations (Caballero and Engel (1994),Caballero, Engel & Haltinwanger (1995), Abel & Eberly (1996), Doms and Dunne (1998)). In the second generation, the seminal paper of Veracierto(2002) and later Thomas(2002) argued that although in partial equilibrium their impact on investment dynamic is substantial, those effects fade away in general equilibrium. In both cases, a costless adjustment economy cannot be told apart from one with irreversibilities. Later work by Kahn and Thomas(2005 & 2007) found additional support for this hypothesis.

Such results suggest that irreversibilities can be dispensed with when analyzing aggregate fluctuations. Recent work by Bachman et.al. (2011) show that irreversibilities play an important role in explaining the conditional heteroscedasticity in aggregate investment behavior observed in the data. On a different line, Khan and Thomas (2011) show that quantitatively the interaction of irreversibilities with financial frictions may explain large drops in aggregate efficiency and slow recoveries. As exposit by Caballero (1999) irreversibilities might have important consequences for the aggregate behavior of the economy when interacted with market incompleteness or informational asymmetries.

This paper shows that even in complete market set up, there is a relevant mechanism through which irreversibility affect aggregate efficiency. When interacted with aggregate uncertainty in a fully fledged industry dynamic model, they generate a disparity between investment decisions of entrants and incumbents in the market. The vintage of the firm becomes relevant in explaining their investment behavior. The most salient different between this paper and previous work by Veracierto (2002), is that he abstracts from the industry dynamic problem. In that paper, entry and exit is exogenously given[2]. Endogenous industry dynamic in the context of aggregate uncertainty is the pillar of the mechanism that this paper analyzes. Also, the nature of noncovexities in production is different from the one exploited in that paper. While there partial irreversibility is

[2]As can be seen from table 4 in Veracierto (2002), when there is full irreversibility, the change in the exogenous death rate has considerable effect on investment dispersion across production units.

allowed, here there is full irreversibility that is induced by sunk costs associated to investment decisions.

This paper relates substantially to the literature on irreversible investment as started by Dixit (1988), Abel et.al. (1995), Mariotti et. al. (2006) and Caballero and Engel (1994). However, the set up of the model is closer to the work of Melitz and Ghironi (2004). Distinctively, I allow firms to endogenously determine efficiency through investments in distinct technologies. The latter implies that the size distribution of firms is endogenously determined. I take a radical stand as of the idiosyncratic productivities of the firms, they are constant. I assume though that the log productivity is drawn from an exponential distribution, so that the model can be interpreted as the limiting case of a model in which firms idiosyncratic productivity is stochastic and follow a Brownian Motion (See Luttmer (2010) for an example). The constancy of idiosyncratic shocks is an assumption made for tractability. The mechanism of the model does not vanish when they are allowed to change in time, and can even be reinforced (as negative idiosyncratic shocks may render previous investment decisions statically inefficient). In this version of the model we assume idiosyncratic shocks are fully insurable so that we can separate the technological inefficiencies from those stemming from market incompleteness.

On the theoretical side, we contribute to the work initiated by Lucas and Prescott (1971). They showed that a competitive equilibrium can be decentralized as an industry equilibrium in which the planner maximizes overall surplus in the economy by allocating labor across firms. I show that when firms compete monopolistically and there are irreversibilities in investment and indivisibilities in capital allocations, there is a pseudo planner problem whose equilibrium allocation coincides with the decentralized solution as long as state contingent subsidies and taxes are available.

I investigate the implications of the model for optimal industrial policy. This has been done for models of international trade under oligopolistic competition in prices and quantities (Eaton and Grossman,1986). For a model of industry dynamic without capital accumulation Lee and Mukoyama (2008) study the impact of alternative policies on labor regulations. However, their policies are ad hoc in the sense that there is no notion of efficiency associated to them. Rogerson and Restuccia (2008) study the impact of taxes to output and inputs in production over aggregate TFP, for both i.i.d. policies and policies correlated with the productivity of the firms. One of the main drivers of cross countries differences in aggregate efficiency is the skewness in the size distribution. Guner, Ventura, and Xu (2006) study policies that target the size of the establishment, which in turn is correlated with their idiosyncratic productivity, and find substantial role in shaping aggregate productivity.

3 Model

This is an infinite horizon economy with time indexed by $t = 0, 1, ...$

There are different goods in the economy. The first one is a final good which agents use for consumption and capital accumulation. It is produced by means of a continuum of intermediate goods. Intermediate goods are produced by combining capital and labor. Each intermediate good is perfectly differentiated and each firm producing it faces a constant elasticity demand.

Final goods are traded competitively while there is monopolistic competition in intermediate goods.

The technology for production of intermediate goods is endogenously chosen, and each one is associated to a minimum running capacity in terms of capital goods.

There is aggregate uncertainty in the economy. The exogenous shock is denoted s_t For simplicity, we assume it can only take two values, i.e. $\{g, b\}$, $g > b$ associated to the "good" and "bad" state, respectively. The transition probabilities are given by

$$P(s_{t+1} = g/s_t = g) = \gamma_g$$

$$P(s_{t+1} = b/s_t = b) = \gamma_b$$

3.1 Households

The representative household derives utility from consumption of the final good C_t.

The household is endowed with a unit of labor that for simplicity is supplied inelastically to the firms. She receives a wage w_t for the services. She can also accumulate capital K_t (priced in terms of the final good which is taken as a numeraire) and rent it at price r_t to the firms. The aggregate stock depreciates at rate $\widehat{\delta}$.The household can buy shares of two different mutual funds that entitled it to the dividends generated by the firms operating in the economy. The first mutual fund consist of all the firms running with low minimum capacity technology, and the second is build with all firms using a technology with higher minimum capacity. Shares to the mutual fund entitle the consumer with the dividends of the firms and the possibility to resell the assets in forthcoming periods.

Her problem reads

$$Max_{\ C_t, n_t^L, n_t^H, K_{t+1}} E_0 \left[\sum_{t=0}^{\infty} \beta^t U(C_t) \right] \tag{1}$$

subject to

$$C_t + K_{t+1} - (1-\widehat{\delta})K_t + \sum_{k=L,H} P_t^k n_t^k = w_t + r_t K_t + \sum_{k=L,H} \left(d_t^k + P_t^k \right) n_{t-1}^k$$

$$X_{t+1} = \Gamma_c(X_t)$$

where P_t^k is the price of shares n_t^k of a mutual fund of firms of technology k at period t+1, that pays dividends d_{t+1}^k and can be sold tomorrow at price P_{t+1}^k. In computing the return to the share holdings, the agent needs to forecast the law of motion of the distribution of firms in the market for each possible realization for the exogenous aggregate shock, s_t. The aggregate state $X_t = \left(s_t, v_t^L, v_t^H, K_t \right)$ is summarized by the exogenous shock and the distribution of firms per technology, v_t^k for $k = L, H$ and the available aggregate stock of capital. To save on notation we denote $v_t^k(z) \equiv v_t^k([0, z))$ the measure of firms with productivity at most z. The subjective law of motion for the representative consumer is denoted by Γ_c.U fulfills the standard assumptions of concavity, monotonicity and differentiability. $\beta \in (0, 1)$ is the discount factor.

The FOCs read

$$U'(C_t) = \varsigma_t \tag{C_t}$$

$$\varsigma_t P_t^k = \beta E_t \left[\varsigma_{t+1} \left(d_{t+1}^k + P_{t+1}^k \right) \right] \tag{X_t}$$

For a standard CES specification $U(C_t) = \frac{C_t^{1-\theta}}{1-\theta}$ we can rewrite

$$P_t^k = E_t \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} \frac{C_\tau^{-\theta}}{C_t^{-\theta}} d_\tau^k$$

which is the present discounted value of all future dividends for firms of vintage $t+1$with technology k, adjusted by the corresponding pricing kernel.

3.2 Final Goods Sector

There is a single competitive firm with a CES technology that produces final goods Y_t out of intermediate inputs y_{it}. The firm maximizes profits as

$$Max_{y_{it}}\ Y_t - \int p_{it} y_{it}\ di$$

subject to

$$Y_t \leq \left(\int y_{it}^{\rho} di \right)^{\frac{1}{\rho}}$$

where p_{it} is the cost of good y_{it}. The elasticity of substitution across intermediate goods will be given by

$$\sigma = \frac{1}{1-\rho}$$

We assume $\rho \in (0,1)$ so that goods are substitutes in production.

The corresponding input demand for each variety i emerges from the FOC of the problem. In particular,

$$Y_t^{1-\rho} y_{it}^{\rho-1} = p_{it}$$

3.3 Intermediate Goods

3.3.1 Problem of the firms

Suppose there is a continuum of heterogenous firms, each of which faces a constant elasticity demand function for a differentiated intermediate good i as described before.

Firms learn their idiosyncratic productivity $z \in [0, \overline{z}]$ before they decide whether to enter into the market or not[3]. The idiosyncratic productivity, z, is drawn from an exogenous distribution $G(z)$.

To produce the differentiated good, firms use capital and labor. Labor is available at cost w_t in units of the composite good. Capital is available at cost r_t in terms of the composite good also. For any given intermediate good there are two possible technologies for production. Output is described by a Cobb-Douglas production function as

$$y_t \leq s_t z \psi^j l_t^{1-\alpha} k_t^{\alpha}$$

Each technology is associated to a minimum capacity $\{k_L, k_H\}$ (low and high) and a productivity shifter, ψ^j. For simplicity we assume, $\psi^L = 1$ and $\psi^H > 1$. The capital choice sets are $[k_L, \infty)$ and $[k_H, \infty)$ for each technology, respectively. We interpret this indivisibility as the construction of a plant, or the set up of machinery in production which entails a particular capacity. The adoption of a particular technology (blueprint) is costly. Entrepreneurs pay I_L units of the final good to buy low minimum capacity technology. The usage of a high minimum capacity technology implies an additional sunk cost I_H. Both costs are paid at the time of adoption. While the purchase of the first technology allows for upgrades in technology selection, the second one does not allow further adjustments (to choose the low minimum capacity technology).

Firms are liquidated with **probability** δ, getting a scrap value of Π_e^f.They can select out voluntarily, getting a scrap value of Π_e (net of exit costs). Without loss of generality, we assume $\Pi_e^f = 0$, $\Pi_e > 0$, so that the option to exit is meaningful. Π_e can be interpreted as the scrap value of the installed capacity, brand name, etc. For expositional purposes we assume the latter is constant along the cycle and across sizes, but the model may accommodate richer structures.

The endogenous distribution of productivities of firms operating in the market is denoted by v_t^j for firms of technology $j = L, H$. In general, $v_t^j(z) \equiv v_t^j((0, z]$ is the

[3]As will become clear later, this assumption implies that there are no resources allocated to firms that do not get to produce in the current period. This is important because entry is costly in this model.

measure of firms with productivity at most z running technology j. Before describing the optimal decisions of the firms, define x_t as the vector of idiosyncratic state variables to the firm, i.e. $x_t = \left(z, \psi^j\right)$. Let X_t defined as before and define Γ_f as the law of motion for the aggregate state as perceived by any arbitrary firm; i.e. $X_{t+1} = \Gamma_f(X_t)$.

3.3.2 Pricing Strategy

First, we abstract from the options to exit, entry and technology selection to study the static problem of the firms. In other words, their pricing strategy and labor and capital decisions. Firms in this model are entirely equity owned.

The problem of a firm producing an intermediate good i in any period t is

$$\pi(x_t, X_t) = Max_{p_t,l_t,k_t,y_t}\ (p_t y_t - w_t l_t - r_t k_t)$$

subject to

$$\begin{aligned} y_t &\leq s_t z \psi^j l_t^{1-\alpha} k_t^{\alpha} \\ \left(\frac{Y(X_t)}{y_t}\right)^{1-\rho} &= p_t \\ k_t &= [k_j, \infty) \end{aligned}$$

where l_t is the labor allocation, k_t is the capital allocation and k_j is the minimum capacity associated to the current technology of the firm, j. We will analyze the problem of technology selection further ahead. Note that in choosing prices and labor, the dynamic of the economy is irrelevant. Both are intratemporal decisions.

Because the elasticity of the demand is constant, the optimal price set by a firm is a constant markup over marginal cost. In particular,

$$p_t = \frac{(r_t - \lambda_t)^{\alpha}\, w_t^{1-\alpha}}{\rho \alpha^{\alpha}(1-\alpha)^{1-\alpha} s_t z}$$

where $\lambda_t \geq 0$ is the lagrange multiplier associated to the feasible set for capital. If the minimum capacity requirement is not binding then, $\lambda_t = 0$, otherwise $\lambda_t > 0$ and the markup for this firm is lower than otherwise.

From the FOC of the firms we can compute the labor and capital demand as

$$l_t = \left[s_t z \psi^j \left(\frac{1-\alpha}{w_t} \right)^{\frac{1}{\rho}-\alpha} \left(\frac{\alpha}{r_t - \lambda_t} \right)^{\alpha} \right]^{\frac{\rho}{1-\rho}} \rho^{\frac{1}{1-\rho}} Y(X_t) \tag{2}$$

$$k_t = \max \left\{ k_j, \left[s_t z \psi^j \left(\frac{1-\alpha}{w_t} \right)^{(1-\alpha)} \left(\frac{\alpha}{r_t} \right)^{\frac{1}{\rho}-(1-\alpha)} \right]^{\frac{\rho}{1-\rho}} \rho^{\frac{1}{1-\rho}} Y(X_t) \right\} \tag{3}$$

The more the relative efficiency in production the higher the demand of labor for $\rho \in (0,1)$.

Capital labor ratios for unconstrained firms will be given by

$$\frac{k_t}{l_t} = \frac{w_t}{r_t} \frac{\alpha}{1-\alpha}$$

Constrained firms will set

$$\frac{k_t^c}{l_t^c} = \frac{w_t}{r_t - \lambda_t} \frac{\alpha}{1-\alpha}$$

where $l_t^c(\lambda_t)$ is the labor demand of a constrained firm. If the minimum capacity requirement is binding, the firm adjusts its resource allocation through the flexible factor, in this case labor. From the labor demands, we know that labor is decreasing in the marginal product of capital. What the last condition indicates is that constrained firms have both higher capital and labor demands, but their labor demand does not increase enough to equalize the firms capital labor ratio to that of unconstrained firms. In other words, their labor share drops relative to unconstrained firms.

Given the dynamic features of our model it will be possible to generate states in which firms with identical idiosyncratic productivity are running different technologies at the same point in time. If one of the firm is constrained by the minimum capacity requirement, i.e. is running high minimum capacity requirement technology, and the entrant in the market decides to use the low capacity requirement technology, then the ratio in capital labor ratios will be

$$\frac{k_t^c / l_t^c}{k_t / l_t} = \frac{r_t}{r_t - \lambda_t^i} > 1$$

This disparity is at the heart of the dynamic studied in this paper. In a static model of complete markets, disparate capital labor ratios are typically a sign of inefficiencies in the allocation. However, in a dynamic model with entry and exit this disparities might be dynamically efficient. Studying the allocation generated by a pseudo-planner with the same technological restrictions than those present in the decentralized problem, would shed light on how much of the disparity is or is not dynamically efficient.

The observed capital labor ratio for a constrained firm will depend on its own shadow value for capital. The latter is determined by the idiosyncratic productivity of the firm, the technology adopted and the aggregate state of the economy, solving

$$\lambda_t = r_t - MPK(k_j, \lambda_t)$$

$$\lambda_t = r_t - \alpha\rho Y(X_t)^{1-\rho} \left(s_t z \psi^j\right)^{\rho} \left[\left(\frac{l_t^c(\lambda_t)}{k_j} \right)^{1-\alpha} k_j \right]^{\rho-1} \left(\frac{l_t^c(\lambda_t)}{k_j} \right)^{1-\alpha}$$

The shadow value of capital moves continuously with the idiosyncratic productivity of the firm.

Labor and capital demand are proportional to the relative efficiency in production of the firm

$$l(x_t, X_t) = \frac{\frac{\psi^j z}{(r_t(X_t)-\lambda_t)^{\alpha}}^{\frac{\rho}{1-\rho}}}{\int \left(\frac{\psi_i^j z_i}{\left(r_t(X_t)-\lambda_t^i\right)^{\alpha}}\right)^{\frac{\rho}{1-\rho}} di} \tag{4}$$

$$k(x_t, X_t) = K_t \frac{\frac{\psi^j z}{(r_t(X_t)-\lambda_t)^{\frac{1-(1-\alpha)\rho}{\rho}}}^{\frac{\rho}{1-\rho}}}{\int \left(\frac{\psi_i^j z_i}{\left(r_t(X_t)-\lambda_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}}\right)^{\frac{\rho}{1-\rho}} di} \tag{5}$$

Because marginal product of capital may differ across production units, the efficiency in production consists of the idiosyncratic productivity of the firm discounted by the firm's MPK. Firms that are constrained (have lower marginal products) will demand relatively more of both capital and labor inputs, conditional on operating in the market. The increase in relative capital allocation will be higher than the increase in relative labor allocation, as can be seen from computing $\partial l(x_t, X_t)/\partial \left(r_t(X_t) - \lambda_t\right)$ and $\partial k(x_t, X_t)/\partial \left(r_t(X_t) - \lambda_t\right)$.

Note also that if there are no constrained firms, then labor and capital allocations are just proportional to the relative productivity of firm versus the average the economy.

$$l(x_t, X_t) = \frac{\psi^j z^{\frac{\rho}{1-\rho}}}{\int (\psi_i^j z_i)^{\frac{\rho}{1-\rho}} di}$$

$$k(x_t, X_t) = K_t \frac{\psi^j z^{\frac{\rho}{1-\rho}}}{\int \psi_i^j z_i^{\frac{\rho}{1-\rho}} di}$$

Finally, suppose that there is a positive shift in the productivity of the firms in the market[4], so that $\int (\frac{\psi_i^j z_i}{\left(r_t(X_t) - \lambda_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}})^{\frac{\rho}{1-\rho}} di$ increases. The set of firms that are capacity constrained gets larger and the relative capital allocation shifts from productive unconstrained firms, to less productive constrained ones. When this set expands, the initial effect on the allocation of capital across production units gets reinforced. A larger mass of constrained firms implies that the denominator in (5) raises even more. Hence, equilibrium capital demands of unconstrained firms (relatively more productive) are pushed further down. If there is exit in the economy, some of the reallocation of capital is slowed down by the exit of less productive firms. The overall effect on the allocation of resources and aggregate efficiency in production will depend on whether firm exit is enough to undo the reallocation induced by minimum capacity constraints kicking in.

3.4 Firm flows

So far, we have described the differences in the static profits of each of the firms. We now move to describing their value.

[4]Such shift can be induced by a positive aggregate shock that increases the measure of firms running the high minimum capacity technology versus the low minimum capacity.

3.4.1 Exit and Upgrade

A firm using a high minimum capacity technology may choose to operate or exit in the current period. If it operates it will get profits according to the realization of the aggregate shock and its own idiosyncratic state,

$$\pi(x_t, X_t) = (1-\rho)\, Y(X_t)^{1-\rho} \left(s_t z \psi^j l(x_t, X_t)^{1-\alpha} k(x_t, X_t)^{\alpha}\right)^{\rho}$$

Next period it may die with probability δ or continue operating. If it does, then it can exercise the option to exit irrespective of which state of the world s_t is realized. The value of the firm W_t follows

$$W_t(z, \psi^H, X_t) = Max \left\{ \Pi_e, \pi(z, \psi^H, X_t) + E_t \left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\, W_{t+1}(z, \psi^H, X_{t+1}) \right) \right\} \tag{6}$$

subject to

$$X_{t+1} = \Gamma_f(X_t)$$

$\widetilde{\beta}_{t+1}(X_t, X_{t+1})$ is the stochastic discount factor of the household, i.e. $\widetilde{\beta}_{t+1}(X_t, X_{t+1}) = \beta \frac{U'(C(X_{t+1}))}{U'(C(X_t))}$ (we will use $\widetilde{\beta}_{t+1}$ to save notation when necessary).

On the continuation region, when the option to exit is not exercised, the value of the firms is the present discounted value of all future expected profits. We call it $\widetilde{W}_t$.

If we compute the continuation value for a high capacity firm it reads

$$\widetilde{W}_t(z, \psi_H, X_t) = \pi(z, \psi_H, X_t) + E_t \left(\widetilde{\beta}_{t+1}\,(1-\delta)\, W_{t+1}(z, \psi_H, X_{t+1}) \right)$$

Proposition 1 $\widetilde{W}_t(z, \psi^H, X_t)$ *is monotonic increasing in idiosyncratic productivity,* z[5]*.*

Let the function $z^e(\psi^H, X_t)$ determines the threshold for exit of high capacity firms when the aggregate state of the economy is X_t.

Corollary 2 *The optimal strategy will be a trigger strategy in which only firms with productivity* $z < z^e(\psi_H, X_t)$ *exit the market.*

[5]The proof to propositions and theorems can be found in Appendix (B)

The corollary follows from the monotonicity of the continuation value and the definition of the value of the firm (6).

The value of the high capacity firms is

$$\begin{aligned} W_t(z, \psi^H, X_t) &= \widetilde{W}_t(z, \psi^H, X_t) \qquad \text{if } z \geq z^e(\psi^H, X_t) \\ W_t(z, \psi^H, X_t) &= \Pi_e \qquad \text{o/w} \end{aligned} \tag{7}$$

Next, we move to the problem of firms currently holding a low minimum capacity requirement technology. After observing the aggregate state, they may decide to exit the market, to upgrade capacity or to operate at the current one.

$$W_t(z, \psi^L, X_t) = Max\{\Pi_e, W_t(z, \psi^H, X_t) - I^H, \widetilde{W}_t(z, \psi^L, X_t)\} \tag{8}$$

subject to

$$X_{t+1} = \Gamma_f(X_t)$$

Their continuation value is

$$\widetilde{W}_t(z, \psi^L, X_t) \ = \pi(z, \psi^L, X_t) + E_t\left(\widetilde{\beta}_{t+1}\left(1-\delta\right)W_{t+1}(z, \psi^L, X_{t+1})\right)$$

Proposition 3 $\widetilde{W}_t(z, \psi^L, X_t)$ *is monotonic increasing in idiosyncratic productivity,* z.

Corollary 4 *The thresholds for exit* $z^e(\psi^L, X_t)$*and upgrade* $z^u(\psi^H, X_t)$ *are such that firms with* $z < z^e(\psi^L, X_t)$ *will exit the market; firms with* $z > z^u(\psi^H, X_t)$ *will upgrade capacity; and firms with intermediate values of productivity will remain with the low minimum capacity requirement technology.*

The corollary follows from the monotonicity of the continuation values and the description of the value function (8).

The value of a small capacity firm is

$$\begin{aligned} W_t(z,\psi^L,X^s) &= W_t(z,\psi^H,X_t) - I_H \qquad \text{if } z > z^u(\psi^H,X_t) \\ W_t(z,\psi^L,X_t) &= \widetilde{W}_t(z,\psi^L,X_t) \qquad \text{if } z \in \left[z^e(\psi^L,X_t), z^u(\psi^H,X_t)\right] \\ W_t(z,\psi^L,X_t) &= \Pi_e \qquad \text{o/w} \end{aligned} \tag{9}$$

3.4.2 Entry

Of the total mass of firms operating in the market M_t, a fraction δM_t are forced out of the market each period. There is a continuum of firms ready to enter the market at any period t. They observe their productivity before investing I_L units. If they decide to enter, its capacity is low, but they may choose to upgrade capacity immediately at cost I_H.For the problem to be well defined we need to assume $I_L \geq \Pi_e$. Otherwise, entrepreneurs could create resources just by entering and exiting immediately from the market. To simplify the analysis assume $I_L = \Pi_e$[6].

The mass of entrants M_t^{ent} is pinned down by the free entry condition,

$$I_L \geq W(z^e(\psi^L,X_t),\psi^L,X_t) \tag{10}$$

with equality if $M_t^{ent} > 0$.

It is worth noting that average effective productivity, which is used in the computation of the expected value of the firms, is indeed endogenously determined by the choice of exit and upgrade thresholds of firms in the market. As will be clear in the definition of equilibrium, entrants correctly anticipate their future expected profits, so that pre-entry expected profit equalize the post entry value. The entrant with the marginal productivity (the one that will induce exit) is indifferent between entering and not.

[6] The relationship between the cost of entry and the scrapt value of firms will determine the industry dynamic. The higher the entry cost with respect to scrap value, the lower the measure of entrants. If scrap values are allowed to differ across technologies, then this is analogous to assuming equalization with the scrap value of low capacity firms $I_L = \Pi_{e,L}$.

3.5 Features of the solution

This is because profits are monotonic in the firm efficiency in production and continuation values too. Hence, in doing comparative statics across idiosyncratic characteristics of the firms, it is without loss of generality to look at short run payoffs. The profits of the firms can be described as

$$\pi(x_t, X_t) = \Lambda(X_t)\left(\frac{z\psi^j}{MPK_t^{\frac{\alpha}{\rho}}}\right)^{\frac{\rho}{1-\rho}}$$

where $\Lambda(X_t) = \left(\frac{\sum_{j=L,H}\int(\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\alpha}})^{\frac{\rho}{1-\rho}}dv_t^j}{\sum_{j=L,H}\int(\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}})^{\frac{\rho}{1-\rho}}dv_t^j}\right)^{\alpha\rho}\frac{1}{\left(\sum_{j=L,H}\int(\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\alpha}})^{\frac{\rho}{1-\rho}}dv_t^j\right)^{\rho}}(1-\rho)Y_t^{1-\rho}s_t^{\rho}K_t^{\alpha\rho}.$

Proposition 5 *The optimal allocation satisfies*

1. $z^e(\psi_L, X_t) > z^e(\psi_H, X_t)$
2. $\frac{\partial z^e(\psi^j, X_t)}{\partial r_t} \geq 0$
3. $z^u(\psi_H, X_t) > z^e(\psi_H, X_t)$
4. M^e *is procyclical*

The first result indicates that firms running the low minimum capacity technology will find optimal to exit before firms of the same idiosyncratic productivity running the high capacity technology . The second, that increases in the cost of capital, increase the likelihood of voluntary exit. Upgrade thresholds are higher than exit ones for high capacity firms. Finally, the levels of entry are procyclical.

Points 2. and 4. are particularly important as when a positive productivity shock hits the economy there are two opposing forces describing the industry dynamic. The first one is the improvement in productivity efficiency, which should increase profits

$\frac{\partial \Lambda(X_t)}{\partial s} > 0$. The free entry condition implies that new firms would want to operate in the market. Further competition for resources (labor is inelastically supplied and the stock of capital is fixed for the current period) implies that factor prices raise, generating exit. The raise in overall efficiency may generate lower equilibrium markups and profits.

4 Aggregation?

Replacing capital and labor demands in the aggregate production function, we have

$$Y(X_t) = TFP_t K_t^{\alpha}$$

. Let the measure of firms operating in the market is $M_t = v_t^L(z^u(.)) + v_t^H(\overline{z})$ and define a scaled measure $\widehat{v}_t^j = \frac{v_t^j}{M_t}$. Hence, aggregate efficiency is described by

$$TFP_t = s_t M_t^{\frac{\rho}{1-\rho}} \frac{\left[\sum_{j=L,H} \int \left(\frac{\psi_i z_i}{\left(MPK_t^i\right)^{\alpha}}\right)^{\frac{\rho}{1-\rho}} d\widehat{v}_t^j\right]^{\frac{1-(1-\alpha)\rho}{\rho}}}{\left[\sum_{j=L,H} \int \left(\frac{\psi_i z_i}{\left(MPK_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}}\right)^{\frac{\rho}{1-\rho}} d\widehat{v}_t^j\right]^{\alpha}}$$

Hence, aggregate efficiency is determined by the realization of the exogenous shock, the measure of firms operating in the market (as is usual in models of monopolistic competition), the distribution of firms across technologies and the measure of firms that are capacity constrained, as summarized by a moment of the marginal products of capital in the market. As alpha goes to zero, disparity in marginal products become irrelevant for aggregate productivity, because the share of the factor for which the minimum constraint may bind becomes negligible. Because capital labor ratios may differ across production units, there are multiple allocations (distributions across technologies) that yield the same TFP$_t$ but potentially different responses to aggregate shocks in the economy.

Differences in the allocation between a centralized and decentralized problem may stem from two sources: the aggregate measure of firms operating, or the allocation of firms across technologies. Suppose first that the distribution of firms across technologies is exogenously given and there are no minimum capacity constraints. Then if the decentralized and centralized allocation may differ, they can only do it through the number of operating firms in the market. This is the case in models with monopolistic competition and no aggregate uncertainty because there is a 1-1 map between that economy and one with aggregate decreasing returns to scale[7]. When there is aggregate uncertainty however, this is no longer true[8]. Second, suppose that the number of firms in the market is constant but the allocation of firms is determined endogenously. Because when there is aggregate uncertainty the social and private value of a firm type may differ, the optimal allocation decisions may differ two.

Before moving to the definition of the equilibrium, let's recap on the pricing equation for the shares of the mutual fund. From the household problem we obtained

$$P_t^k = E_t \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} \left(\frac{C(X_\tau)}{C(X_t)} \right)^{-\theta} d_\tau^j(X_\tau)$$

From the optimal policies of the firms we know that in equilibrium the dividends of the firms in the market in each period t correspond to

$$\begin{aligned} d_t^L(X_t) &= \int \pi(z, \psi^L, X_t)\ dv_t^L(z) + \Pi_e v_{t-1}^L(z^e(\psi^L, X_t))\,(1-\lambda) - I_L M_t^{ent}(X_t) \\ d_t^H(X_t) &= \int \pi(z, \psi^H, X_t)\ dv_t^H(z) + \Pi_e v_{t-1}^H(z^e(\psi^H, X_t))\,(1-\lambda) - I_H M_t^u\ (X_t) \end{aligned}$$

where $M_t^u\ (X_t)$ is the measure of upgrades in state X_t and $M_t^{ent}(X_t)$ the corresponding measure of entrants. In other words, they equalize total profits of the firms operating

[7]The private value of a firm in the monopolistic competition case is given by the PDV of its profits, i.e. $(1-\rho)\, Y(X_t)^{1-\rho} \left(s_t z \psi^j l^*(x_t, X_t)^{1-\alpha} k^*(x_t, X_t)^\alpha\right)^\rho$. If there is no aggregate uncertainty, one can normalize $Y(X_t)^{1-\rho} = 1$. Now, if we study the problem of a central planner with technology, $Y = \int \left(s_t z \psi^j l_t^{1-\alpha} k_t^\alpha\right)^\rho dz$, it is possible to show that the "social" value of firm type is given by $(1-\rho)\left(s_t z \psi^j l^*(x_t, X_t)^{1-\alpha} k^*(x_t, X_t)^\alpha\right)^\rho$ where labor and capital demands coincide with those in the decentralized solution before.

[8]If there is uncertainty, we cannot normalize aggregate output as we did before and the map breaks.

in the market this period, plus the scrap value of those that exit the market, less the cost of setting up new firms, and upgrades.

5 Equilibrium

5.1 Competitive Equilibrium

The measure of firms per technology v_t^j,belongs to the space of bounded and continuous measures on $[0, \overline{z}.]$, $i.e.$ C^v. The mapping that describes the law of motion for the distribution of firms across technologies is $\Gamma : C^v \times C^v \times \{s_g, s_b\} \rightarrow C^v \times C^v$, characterized by

$$
\begin{aligned}
v_t^L(z) &= (1-\delta)\left[v_{t-1}^L(z) - v_{t-1}^L(z_t^e(\psi^L))\chi_{\left(z_t^e(\psi^L) > z_{t-1}^e(\psi^L))\right)}\right] \\
+M_t^{ent}\frac{G(z)-G(z_t^{e*}(\psi^L))}{1-G(z_t^{e*}(\psi^L))} \quad z_t^u &> z > z_t^e(\psi^L) \\
&= 0
\end{aligned}
$$

If $z_{t-1}^u \leq z_t^u$

$$
\begin{aligned}
v_t^H(z) &= (1-\delta)\left[v_{t-1}^H(z) - v_{t-1}^H(z_t^{e*}(\psi^H))\chi_{\left(z_t^e(\psi^H) > z_{t-1}^e(\psi^H)\right)}\right] & z_t^u > z > z_t^e(\psi^H) \\
&= (1-\delta)\ \ v_{t-1}^H(z) + M_t^{ent}\frac{G(z)-G(z_t^{u*})}{1-G(z_t^e(\psi^L))} & 1 > z > z_t^u
\end{aligned}
$$

If $z_{t-1}^u > z_t^u$

$$
\begin{aligned}
v_t^H(z) &= (1-\delta)\ \ v_{t-1}^H(z) - v_{t-1}^H(z_t^e(\psi^H))\chi_{\left(z_t^e(\psi^H) > z_{t-1}^e(\psi^H)\right)} & z_t^u > z > \ z_t^e(\psi^H) \\
&= (1-\delta)\left[v_{t-1}^H(z) + v_{t-1}^L(z) - v_{t-1}^L(\ z_t^u)\right] + M_t^{ent}\frac{G(z)-G(z_t^u)}{1-G(z_t^e(\psi^L))} & z_{t-1}^u > z > \ z_t^u \\
&= (1-\delta)\, v_{t-1}^H(z) + v_{t-1}^L(z_{t-1}^u) - v_{t-1}^L(\ z_t^u) + M_t^{ent}\frac{G(z)-G(z_t^u)}{1-G(z(\psi^L))} & 1 > z > z_{t-1}^u
\end{aligned}
$$

Definition 6 *A competitive equilibrium is a sequence of thresholds* $\left\{z^e(\psi^j, X_t), z^u(X_t)\right\}_{t=0}^{\infty}$, *distribution of firms* $\left\{v_t^L(z), v_t^H(z)\right\}_{t=0}^{\infty}$ *a law of motion for the dynamic of the distributions of firms,* Γ, *entrants* $\{M_t^{ent}\}_{t=0}^{\infty}$ *with productivities drawn from* $G(z)$, *and a sequence of consumption, aggregate capital and share holdings,*$\left\{C_t, K_{t+1}, n_t^H, n_t^L\right\}_{t=0}^{\infty}$ *such that given the prices* $\left\{r_t, w_t, P_t^k\right\}_{t=0}^{\infty}$, *the cost structure* $\Upsilon_c = \left[\Pi^e, I^H, I^L\right]$, *the initial stock of capital in the economy* K_0 *and share holdings,* $n_0^H = n_0^L = 1$ *and the exogenous dynamic of the aggregate shocks* $\{s_t\}_{t=0}^{\infty}$

i) The representative consumer maximizes utility (as in (1)*)*

ii)Firms in the intermediate goods sector maximize their value as described by (7) and (9) given their residual demand and productivity z.

iii)Firms in the final good sector maximize profits

iv) $W_t(z^e(1, X_t), 1, X_t) \leq I_L$ *with equality if* $M_t^{ent} > 0$

v) $M_t = M_t^{ent} + (1-\delta) M_{t-1} - M_t^{ext}$ *where* $M_t = v_t^L(z_t^u(X_t)) + v_t^H(\overline{z}), M_t^{ext} = (1-\delta)\left[v_{t-1}^L(z_t^e(\psi_L, X_t)) + v_{t-1}^H(z_t^e(\psi_H, X_t))\right]$

vi) markets clear:1) $\int l(z, \psi_L, X_t) dv_t^L(z) + \int l(z, \psi_H, X_t) dv_t^H(z) = 1$

2) $\int k(z, \psi_L, X_t) dv_t^L(z) + \int k(z, \psi_H, X_t) dv_t^H(z) = K_t$

2) $n_t^k = 1, k = L, H$

3) $C_t + K_{t+1} - (1-\widehat{\delta})K_t + I_L M_t^{ent} + I_H M_t^{upt} = Y_t + \Pi_e M_t^{ext}$

vii) consistency $\Gamma = \Gamma_f = \Gamma_c$, *where* Γ *is described as before.*

To prove existence of the equilibrium we will define a centralized problem, whose solution coincides with the decentralized allocation under certain conditions. Because there is monopolistic competition in the intermediate good sector the allocation will be in general not surplus maximizing. We will define a pseudo-planner problem in which the latter is compensated through taxes or subsidies to entry, exit and upgrade, so that the allocation of firms coincides with the decentralized solution.

5.2 Industry Equilibrium (PP)

Before going into the core of this section, let's define some notation.

Let $\widetilde{X}_t = \left(s_t, v_{t-1}^L, v_{t-1}^H, K_t\right)$ the relevant aggregate state space from the point of view of the planner.

The cost structure will given by tuples $\Upsilon(\widetilde{X}_t) = \left[I^{L*}, I^{H*}, \Pi^{e*}(\psi_L), \Pi^{e*}(\psi_H), T\right]$. The first element in the vector is the cost associated to entry of new firms, the second is that associated to upgrades, Π^{e*} is the scrap value of firms exiting the market for each type of technology. The last element, T, is a lump sum transfer to guarantee budget balance. Each of the four elements described before equalizes the cost/scrap value from the decentralized allocation plus a wedge. To assure that the goods associated to those wedges are not lost, we transfer back the total value of the wedges as a lump sum transfer T.

The corresponding productivity thresholds are $z^{e*}(\psi_L, \widetilde{X}_t), z^{e*}(\psi_H, \widetilde{X}_t), z^{u*}(\widetilde{X}_t)$for exit of low and high capacity firm and upgrades. The level of entry is $M^{ent*}(\widetilde{X}_t)$. For notational convenience we will drop their dependence on the state of the economy. We will index them only by time but dependence on the full state should be understood.

Proposition 7 *There exist sequences* $\{\Upsilon_t\}_{t=0}^{\infty}$ *such that the allocation of firms that solves the planner problem (PP) coincides with the competitive allocation*

$$V\left(\widetilde{X}_t\right) = \underset{C_t,\ K_{t+1},\ Y_t,\ z_t^{e*}(\psi_L), z_t^{e*}(\psi_H),\ z_t^{u*}, M_t^{ent}, l_t^i, k_t^i}{Max} U\left(C_t\right) + \beta EV\left(\widetilde{X}_{t+1}\right)$$

subject to

$$\begin{aligned} & C_t + I_t^{L*} M_t^{ent} + I_t^{H*} M_t^u + K_{t+1} - K_t(1-\widehat{\delta}) \\ \leq \ & Y_t + T_t + \Pi_t^{e*}(\psi_L) M_t^{ext}\left(\psi_L\right) + \Pi_t^{e*}(\psi_H) M_t^{ext}\left(\psi_H\right)] \end{aligned}$$

$$s_t \left(\int \left(\psi_L z_i l_i^{1-\alpha} k_i^{\alpha}\right)^{\rho} dv_t^L(z_i) + \int \left(\psi_H z_i l_i^{1-\alpha} k_i^{\alpha}\right)^{\rho} dv_t^H(z_i) \right)^{\frac{1}{\rho}} = Y_t$$

$$\int l_i di = 1, \textit{ and } \int k_i di = K_t$$

$$k_i \geq k_L \textit{ if } \psi_i = \psi_L, \textit{ and } k_i \geq k_H \textit{ if } \psi_i = \psi_H$$

$$\widetilde{X}_{t+1} = \Gamma\left(\widetilde{X}_t\right)$$

where $M_t^u = (1-\delta)\left[v_{t-1}^L(z_{t-1}^{u*}) - v_{t-1}^L(z_t^{u*})\right]\chi_{\left(z_t^u < z_{t-1}^u\right)} + M_t^{ent}\frac{1-G(z_t^{u*})}{1-G(z_t^{ent*})}$, $M_t^{ext}(\psi_H) = (1-\delta)\left[v_{t-1}^H(z_t^e(\psi_H, X_t))\right]$ *and* $M_t^{ext}(\psi_L) = (1-\delta)\left[v_{t-1}^L(z_t^e(\psi_L, X_t))\right]$

For expositional purposes the full proof can be found in the Appendix. Heuristically it goes as follows. First we compute the optimality conditions for the allocation of firms across sizes, and the actual measure of active firms. Then we show that for any set of thresholds (including those that solve the decentralized problem) there exist a unique set of subsidies/taxes such that the centralized solution coincides with those thresholds. To show this the linearity of adjustment costs is critical.

Proposition 8 *For a given transfer scheme, the solution to the centralized problem exists and it is unique.*

The problem above would be a standard concave problem if there were no sunk costs to technology adoption and no minimum capacity constraint that may bind in equilibrium. The presence of a continuum of heterogenous firms mitigates potential non-convexities as in Mas-Collel (1977). The continuum of firms works as the divisible commodity necessary to convexify the aggregate feasible set.

Corollary 9 *The solution to the competitive equilibrium exists*

We cannot say much about the determinacy of the competitive equilibrium. Suppose that the household decides to consume more in the current period and less in the following one. If there was a unique equilibrium the latter will decrease the marginal utility of consumption today and increase it tomorrow. However, higher demand today for final goods implies higher demand for all intermediate goods, which triggers entry, raising labor demand and wages. The productivity cutoff for exit needs to raise, as profits are now lower than before on average. The shift in the cutoff implies that the average productivity in the market goes up, (and the average markup drops) justifying the initial additional consumption of the households.

6 Numerical Results

For the numerical results presented in this section we assume a particular case of the general version of the model presented before. Technologies are associated to a particular capacity, and there is no productivity shifter (because there is no intensive margin selection of capital, once capacities are selected, productivity shifter are redundant). Production under each alternative technology is given by

$$y_t = s_t z k_j^{\alpha} l(x_t, X_t)^{1-\alpha} \text{ for } j = L, H$$

where, $k_L < k_H$ is assumed[9].

This version also abstract away from aggregate capital dynamic. This is not a trivial abstraction as the endogenous allocation of firms across capacities determines the return to capital, and therefore frictions that induce inefficient allocation of firms may also affect the path of capital accumulation[10]. In particular, we assume that there is a stock of capital ready to be used in any particular company (the stock is large enough so that any firm that decides to invest in capacity or enter the market can be supplied with the corresponding stock).

[9] I am currently working on an arbitrary finite amount of different technologies.

[10] For a different kind of irreversibilities in production, Veracierto (2002) and Khan and Thomas (2005) show that lumpiness at the micro level do not translate in aggregate investment booms that are observed in the aggregate. Their work asks much more to the model in terms of micro features that we can ever do with our stylized model. However, they typically do not study the industry dynamic, i.e. entry and exit are not incorporated in the model.

6.1 Algorithm

Given a cost structure, Υ,the solution to the pseudo-planner problem is a set of functions $z^{e*}(k_L, \widetilde{X}_t; \Upsilon), z^{e*}(k_H, \widetilde{X}_t; \Upsilon)$, $z^{u*}(\widetilde{X}_t; \Upsilon)$ that solves the corresponding optimality conditions.

We start the algorithm by assuming an arbitrary cost structure $\Upsilon^c = \left[\Pi^e, \Pi^e, I^H, I^L, 0\right]$ (with no transfers, T).

We implement standard value function iteration over the centralized problem. We use Smolyak algorithm to generate the interpolation grid for the value function. Then we use tensor products for interpolations.The grid of firms operating in the market has at most 50 productivities, but the search algorithm for the optimal policy functions is set over a continuous multidimensional space.

Given the main theorem of the paper, we know that the optimal allocation of firms that solves the pseudo-planner's problem can be decentralized under a state contingent transfer scheme. Under the particular set up of the model, i.e. only two aggregate states and persistent idiosyncratic shocks, we can describe the optimality conditions for the upgrade, exit thresholds and the entry level in the decentralized allocation [11] (See Appendix for detailed description).

The optimality conditions for the firms, as well as those of the centralized problem are linear in the adjustment costs. Hence, if we replace the allocation that solves the pseudo planner problem into the system of equations that solves the decentralized allocation, we can infer the cost structure that decentralizes the allocation. At the centralized allocation, the optimality conditions from the decentralized problem would

[11] An arbitrary finite set of aggregate states can be accomodated, but the number of past thresholds that need to be included in the state space of the planner grows with the number of grid points of the aggregate state. For example, with two states we only need to keep track of last period thresholds (4) and aggregate state, with three states we may potentially need to keep track of the set of thresholds of the last two periods (8) and corresponding aggregate states. And so on and so forth.

Also, if idyosincratic shocks are i.i.d. (no persistence) the optimality conditions of the decentralized allocation can be computed.

typically not hold. To bring the equilibrium about, we redefine the adjustment costs faced by firms as $\Pi_e^c(1,\widetilde{X}_t) = \Pi_e(1+\tau^e(1,\widetilde{X}_t))$, $\Pi_e^c(k,\widetilde{X}_t) = \Pi_e(1+\tau^e(k,\widetilde{X}_t))$, $I_H^c = I_H(1+\tau^u(\widetilde{X}_t))$, $I_L^c = I_L(1+\tau^{ent}(\widetilde{X}_t))$ and solve a system of nonlinear equations for the tax/subsidy scheme. $\Upsilon^c = \left[\Pi_e^c(1,\widetilde{X}_t), \Pi_e^c(k,\widetilde{X}_t), I_H^c, I_L^c, 0\right]$

Now, as long at the cost structures do not coincide, aggregate transfers in the pseudo planner will be non zero. However, we know from the characterization of the solution, that transfers do not affect the optimal allocation of firms across capacities or the level of operating firms. Hence, output is the same irrespective of the size of the transfer and only aggregate consumption changes. As mentioned in the analysis of the model, a difficulty in solving the decentralized allocation directly is that firms need to forecast the set of prices for forthcoming periods. Among others, they need to forecast the correct discount rate (the pricing kernel) that allows them to asses their value at each realization of the aggregate state. This problem is solved in our framework because given the equivalence of allocations we can use the equilibrium marginal utility of consumption (and aggregate consumption) that we obtain from the solution to the centralized problem. Aggregate consumption in the decentralized allocation equalizes the aggregate consumption of the pseudo planner problem plus the value of transfers per state.

Finally, we address the question of optimal industrial policy. We recompute the optimal allocation under Υ^c, i.e. the cost structure that delivers the calibrated allocation.

In a nutshell the algorithm goes as follows

1. Assume, $\Upsilon = \left[\Pi^e, \Pi^e, I^H, I^L, 0\right]$ compute the optimal allocation of the PP.
2. Calibrate the model.
3. Replace the allocation into the optimality condition of the DP, compute $\Upsilon^c = \left[\Pi_e^c(1,\widetilde{X}_t), \Pi_e^c(k,\widetilde{X}_t), I_H^c, I_L^c\right]$.
4. Let Υ_T^c be the cost structure for the calibrated model.Set $\Upsilon = [\Upsilon^c, 0]$ and compute the optimal allocation.

6.2 Calibration

The simplified version of the model is calibrated to the USA economy. We will take this calibrated economy as the benchmark for our analysis. In particular, we assume that the features of the business cycle and cost structure in the USA generate the minimal heterogeneity in marginal products that we could observe in an economy with non-convexities in production of the type described in the model and competitive complete markets. This allocation however, may not be efficient. We know from the previous analysis that unless a particular set of transfers are implemented, the centralized and decentralized allocation will differ. The object of study for the calibration exercise is the decentralized allocation.

Although business cycle statistics are typically presented at quarterly frequency, industry dynamics statistics are only available on a yearly basis. Hence, the time unit of the model is a year.

Some of the calibrated parameters are standard in the RBC literature. The persistence of expansions and recession periods were set to match the average duration of the phase of the business cycle (in years) from the USA. In particular, $\gamma_s = 1 - 1/T_s$ where T_s is the average length of a particular phase of the business cycle $s = g, b$. The average duration of an expansion was set to 3.175 years (or 12.7 quarters), and that of a recession to 1.425 years (or 5.7 quarters). The discount factor was set to match a steady state interest rate of 4 %, $1 + r = \beta^{-1}$. For simplicity we assumed log utility[12].

The next parameter to calibrate is the substitutability across intermediate goods in the final good aggregator. Atkenson and Kehoe (2005) set a value of 15% to the returns to entrepreneurship, whose analogous in the model is $1-\rho$. The substitutability parameter is set to 0.85. Finally, the share of capital in production is set to 1/3 as is standard in the literature.

The size of aggregate shocks was calibrated to match a measure of volatility of

[12] In a model with non trivial capital accumulation dynamic, theta should be carefully calibrated, hopefully to match the degree of smoothing in the aggregate consumption profile along time.

output. In particular, we construct hp-filtered series of log GDP from annual data 1930 to 2011. We target the volatility of the cyclical component of the series.

Parameter	**Matching**	**Value**
γ_g	Persistence of Expansions	.685
γ_b	Persistence of Recessions	.298
β	Average Annual Interest Rate	.98
s_g, s_b	Volatility Output (Cyclical Component)	$exp(.025), exp(-.025)$
α	Share of Capital	33%
$\sigma(\rho)$	Returns to entrepreneurship	6.66 (0.85)
θ	Intertemporal Elasticity of Substitution	1 (log utility)

In replicating features of the size distribution of firms it is important to generate levels of employment comparable to those observed in the data. The model predicts that relative labor demands are described by

$$\frac{l_i}{l_j} = \frac{\left(z_i k_i^{\alpha}\right)^{\frac{\rho}{1-\rho(1-\alpha)}}}{\left(z_i k_j^{\alpha}\right)^{\frac{\rho}{1-\rho(1-\alpha)}}}$$

where k_i represents the capacity level chosen by the firm i and z_i tis corresponding idiosyncractic productivity. The log differential of employment levels will be proportional to the log differential in capacity levels and idiosyncratic productivities. For a given set of capacities (k^L, k^H) the grid of productivities is set so that the upper bound in log-employment levels corresponds to the upper bound in the reported size distribution by employment by the Business Dynamic Statistics of the Census Bureau. As a matter of fact the grid range as well as the set of capacities are jointly calibrated with the rest of the cost structure in the economy. Capacities and costs (in particular the upgrade cost) are jointly determining the size distribution of firms as well as the dynamic of firm investment, and exit rates for high capacity firms.

In general, there are no closed form solutions for many of those statistics. To calibrate them we simulate the model economy via Montecarlo. In particular we run the optimal policies for 1000 periods and run 100 paths of those for a variety of parameter specifications. The distribution of sizes in the economy inherits some of the properties of the exogenous distribution of idiosyncratic productivity, $G(z)$. From previous studies (Axtell, 2001 & Cabral & Mata, 2003) we know that for the USA, the distribution follows a Pareto distribution with parameter close to 1. The model is calibrated so that log(z) is exponential with parameter,$\lambda_G = 1.1$. In other words, $G(z)$ is Pareto with parameter λ_G. The Pareto like distributions imply that there are a lot small firms with low levels of employment and a few that account for a large share of total employment. In view of this feature, both the grid to approximate the distribution, as well as the grids to search for the optimal allocation across technologies, were constructed equally spaced in logarithms.

Parameter	**Value**
$k^H; k^L$	2;1
$[0, \overline{z}]$	$[0.01, 3.4]$
λ_G	1.1
Π^{e*}	2
I_L^*	2
I_H^*	1
λ	.05

The hazard rate for exogenously determined exit was set to 5% which corresponds to a lower bound in the numbers reported in Lee and Mukoyama (2007) based on statistics from the Annual Census of Manufactures.The remaining set of parameters that need to be calibrated are those that characterize the cost structure Υ. To do so, we target statistics of the industry dynamic. Average exit rates and entry rates are as reported in Lee & Mukoyama (2007). We also target the percentage of firms with positive investment spikes as reported by Cooper and Haltinwanger, (2006). A spike is defined as firm that reports an investment rate of 20% or higher.

Targets	**Data**	**Model**
GDP Volatility (Cyclical Component)	.209	.211
Entry Rate	6.2%	6.5%
Exit Rate	5.5%	4.4%
Firms with a Positive Investment Spike	18%	5.4%+entrants=12%

The values for the industry dynamic reported under the "Model" column ,correspond to the the average obtained through montecarlo simulations .Additional statistics are reported in the table below,

Targets	**Model**	**s.d.**	**max**	**min**
Entry Rate	6.5%	6%	13%	0%
Exit Rate	4.4%	1.4%	5.9%	0%
Firms with a Positive Investment Spike	5.4%[13]	6.4%	13%	0%

Exit rates are tight around the average, but standard errors for entry and investment rates are large. Within one standard deviation of the average, both investment an entry rates can double.

The costs displayed in the table below are the primitives that we used to solve the pseudo planner problem. Given the equilibrium allocations, we can compute the implied set of costs in the decentralized allocation

			good times	**bad times**
Π^{e*}	2	Π^e_L	.99	1.58
		Π^e_H	1.02	1.6
I^*_L	2	I_L	.99	1.58
I^*_H	1	I_H	-.75	.55

The identified scrap values are slightly higher for high capacity firms than there are for low capacity firms.Entry costs are equal to the scrap value of low capacity firms

[13] This number does not include entrants.

by construction. Finally, upgrade costs are higher during recession than in expansions. Note that the cost of upgrade during an expansion is negative. In other words, the model generated industry dynamic is generated by subsidizing upgrades in expansions. This is consistent with countercyclical investment good prices as reported by Christiano and Fisher (1998).

The following pictures depict the distribution of employment and establishments by employment size as predicted by the model and in the data. The model is consistent with the features of the USA firm distribution.

Figure 1: Establishment Distribution

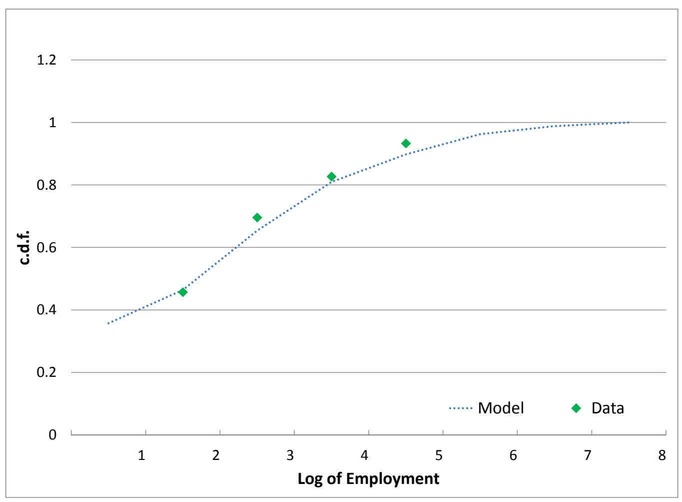
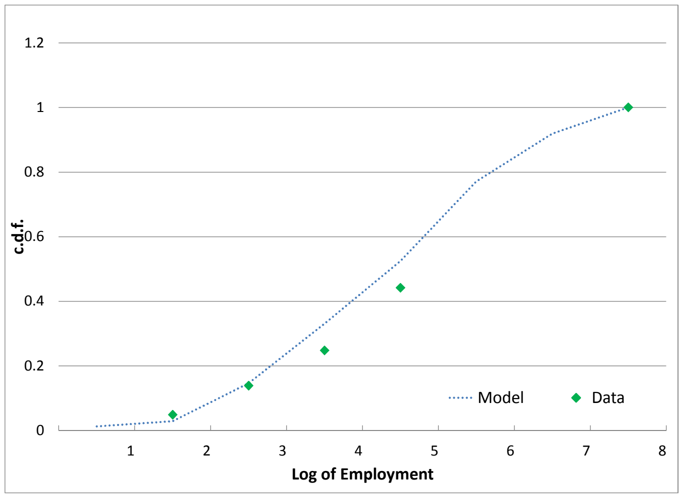


Figure 2: Employment Distribution



6.3 Results

6.3.1 The propagation role of non-convexities at the micro level

We first describe the predictions of the model for aggregate TFP.

Given that technology in our model is constant returns to scale, total factor productivity is proportional to a geometric average of capital and labor productivity. A measure of total factor physical productivity ($TFPQ$) for a production unit is proportional to

$$TFPQ_z \triangleq MPK_z^{\alpha} MPL_z^{1-\alpha}$$

where the marginal product of capital and marginal product of labor are defined as

$$MPK = \rho\alpha\frac{y_i}{k_i}$$

$$MPL = \rho(1-\alpha)\frac{y_i}{l_i}$$

,respectively.

Aggregate TFP can be characterized as

$$TFP_t = \left(\sum_{j=L,H}\int \left(TFPQ_{tz_i}\right)^{\frac{1}{\rho}} dv_t^j(z_i)\right)^{\rho}$$

a weighted moment of the distribution of factor productivity across production units. Alternatively, it can be written in terms of total factor revenue productivity, which includes the optimal pricing of the firms.

$$TFP_t == \left[\sum_{j=L,H}\int \left(z_i\frac{TFPR}{TFPR_{z_i}}\right)^{\frac{\rho}{1-\rho}} dv_t^j(z_i)\right]^{\frac{1-\rho}{\rho}} \tag{TFP}$$

Hence, aggregate productivity is determined by the joint distribution of idiosyncratic productivities and total revenue factor productivity.

Note that one of the frictions in our model might be irrelevant in determining the total factor productivity. Monopolistic competition does not per se generate dispersion in efficiency in production. It might affect productivity through the equilibrium number of operating firms in the market. If the only friction in the model is imperfect competition, aggregate total factor productivity can be computed as

$$TFP_t = \left[\sum_{j=L,H} \int (z_i)^{\frac{\rho}{1-\rho}} \, dv_t^j(z_i) \right]^{\frac{1-\rho}{\rho}} \tag{MCTFP}$$

We will use this measure as benchmark to compare our efficiency measures against.

A usual assumption in the literature that studies misallocation as inferred from the data, is that policies (that affect marginal products at the firm level) are uncorrelated with idiosyncratic productivity, i.e. Hsieh and Klenow (2009). In our model, such assumption does not hold because disparities in marginal products are only observed for capacity constrained firms. Firms that are capacity constrained are typically at the bottom of the distribution of idiosyncratic productivity for a given capacity. Furthermore, their shadow value of capital is proportional to their productivity.

The table below shows the propagation effect of non-convexities in production on computed aggregate TFP. All values are reported in log points. The first column reports the statistic described in ($MCTFP$) with the measure of firms normalized to 1. The second column reports (TFP), and the third one reports the same statistic for a normalized measure of firms equal to one. The second row displays dispersion statistics for the time series of log TFP.

	Baseline	**Model TFP**	**Model TFP**[14]
mean TFP	0.855	0.601	0.52
s.d. TFP	3.3%	1.3%	1.2%

The model with non-convexities generates TFP measures 29.8% below the ones generated by the model with monopolistic competition only, which is equivalent to 0.25

[14]corresponds to TFP measure with normalized measure of firms operating in the market

log point losses. For this particular exercise, non-convexities are forced to have a substantial role. There are only two levels of capital, which induces substantial dispersion in marginal products and through it loses in efficiency. Typically, the level of firms operating in the market need not be the same in both alternative scenarios. If we force the number of firms of the full model to coincide with those in the monopolistic competition case (which was normalized to one), then the aggregate computed TFP reaches only 0.52 log points. In other words, there are two mechanisms through which non-convexities at the micro level shape aggregate productivity. On the one hand, dispersion in capital labor ratios worsens aggregate productivity .If the market is allowed to correct the distortion by inducing entry which lowers markups and pushes the allocation of production towards the most productive firms, then some of the losses can be overcame. In an economy with decreasing returns to scale instead of monopolistic competition the intuition will go along the same line. When entry is allowed to adjust in the market, the equilibrium allocation induces more firms to be operating in equilibrium which under decreasing returns to scale improves the allocation of resources (cost minimization).

The volatility of fluctuations in TFP is three fold lower than in the baseline case. This is what in the lexicon of Bachmann et. al. (1996) would be called the PE-smoothing. Although they refer to the smoothing effect that nonlinear lumpy adjustment cost have on aggregate investment rates, the intuition of the result goes along the same lines. Because reallocation is slowed down by the micro frictions in our model, aggregate TFP does not react as much to aggregate shocks as it does in a model with no such frictions (the monopolistic competition baseline).

6.3.2 Features of the Business Cycle: Persistence, vs the Size of the Shocks, vs Unconditional Volatility

We have shown that the technological friction in our model can propagate aggregate shocks. Now, we investigate how the propagation effect is shaped by different features of the business cycle. We are particularly interested in changes in the volatility and persistence of the aggregate shock. Although we have used a reduced form characterization of the aggregate shock (a two state Markov chain), to gain intuition as of the sources of fluctuations in aggregate volatility, let's characterize the underlying process. Suppose the aggregate shock s_t follows an AR(1)

$$s_t = \rho s_{t-1} + e_s$$

where ρ is the persistence of the shock and e_s an i.i.d. shock with mean zero and standard deviation σ_e. The unconditional volatility of the aggregate shock is

$$\sigma_s^2 = \frac{\sigma_e^2}{1-\rho^2}$$

Hence, both changes in the persistence and variance of the e_s shock affect aggregate volatility. If the AR(1) process is approximated by a two state Markov chain, a la Rouwenhorst (1995), then

$$\left(\frac{s_g - s_b}{2}\right)^2 = \sigma_e^2$$

and

$$\gamma_g + \gamma_b - 1 = \rho$$

Our exercise is as follows. We first compute the statistics under analysis for shorter expansion periods $(\gamma_g = .5)$, about 2 years on average. Let the unconditional variance in this scenario be $\widehat{\sigma}_s^2$. The shock increases the underlying unconditional volatility from 0.025^2 to $\widehat{\sigma}_s^2 .= 0255^2$.

Then we set γ_g back to its calibration value, and we increase $s_g - s_b$ to .051 so that we generate the same unconditional variance $\widehat{\sigma}_s^2$. The following table reports TFP measures under alternative definitions, as well as measures of volatility of output, and the relative dispersion of low capacity versus high capacity firms.

	Baseline	**Case I**	**Case II**
	γ_g= **.685**; $\mathbf{s}_g - \mathbf{s}_b$= **.05**	γ_g= **.5**; $\mathbf{s}_g - \mathbf{s}_b$= **.05**	γ_g= **.685**; $\mathbf{s}_g - \mathbf{s}_b$= **.051**
TFP	100	100.9	100.3
TFP^{MC}/**TFP**	1.423	1.418	1.421
$\boldsymbol{\sigma}_{TFP}$	1.3%	1.19%	1.36%
$\boldsymbol{\sigma}_{Y}$	2.11%	2.36%	1.97%
$\boldsymbol{\sigma}_H^{tfp}/\boldsymbol{\sigma}_L^{tfp}$	54.27%	54.18%	54.21%

When the unconditional volatility of the aggregate shock increases the propagation role of the micro frictions drops. The intuition for this result is that as volatility increases the adjustment costs become less of a constraint for the investment decisions of the firms. Losses are always bounded by the option to exit the market and the associated scrap value of the firm. Hence as volatility goes up, the average value of the firms goes up, making the adjustment cost less stringent.

Given on the same unconditional volatility, the effect of persistence and the size of the shocks is disparate upon the volatility of the productivity measure. While persistence generates smoothing in TFP measures, changes in the size of the shocks increases their volatility. If the size of the shock is held constant and fluctuations occur more often, then the value of the firms get discounted accordingly. For a given set of cost, the regions of inaction get enlarged. This is not to say the aggregate TFP drops, because the less favorable conditions induce selection in the market. To survive, firms need to be more productive which shifts the relative productivity of those operating in the market. Note that the relative dispersion of high capacity and low capacity firms shrinks the most in the first scenario.

These two types of shocks affect disparately the volatility of the cyclical component of output also. While drops in persistence increase its volatility, increases in the size of the shocks, smooth fluctuations of output. The size of the impact is asymmetric: volatility raises about 0.25 percentage points in the first case, and drops 0.14 percentage points in the second case. To understand the origins of this disparate effect we need to describe the industry dynamic.

	Baseline	**Case I**	**Case II**
Entry Rate	6.5%	4.93%	6.55%
Exit Rate	4.4%	13%	4.43%
Upgrade Rate	5.4%	0.8%	5.2%

In case 1, the drop in persistence of positive shocks induce drops in the value of the firms and hence drops in entry levels. Exit rates increase accordingly and upgrades in technology get discouraged. There are less firms operating in equilibrium at higher average markups. The volatility of computed aggregate productivity drops as reshuffling of firms is larger in this economy than in the baseline case. The volatility of output increases in line with the drop in persistence of the shock.

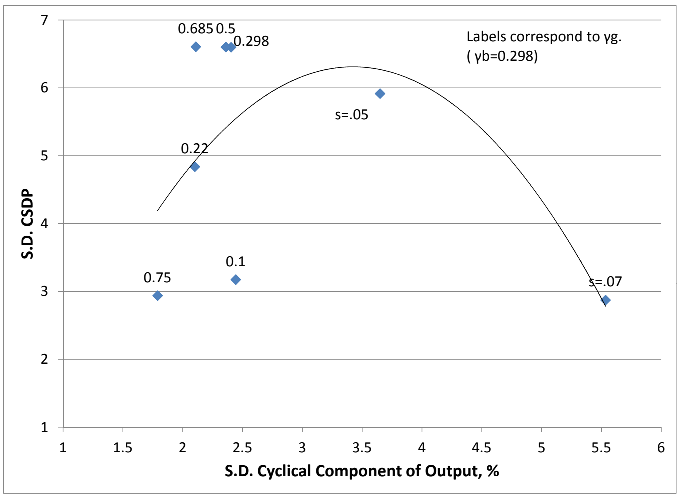
In case II instead, the persistence of the shock is maintained by the unconditional volatility of the shock increases slightly. The industry dynamic barely changes so that the volatility of TFP increases slightly with respect to the baseline. The volatility of the cyclical component of output drops because there is reallocation of resources towards more productive firms as the aggregate productivity shows.

6.3.3 Inverted U relationship between Output volatility and Cross Sectional Dispersion

Now we test whether the model can qualitatively describe the U shape relationship between the volatility of the cyclical component of output and observed cross sectional dispersion in computed TFP across firms. To do so, we compute the solution to the model for a range of persistence parameters keeping constant the size of the shocks. From the previous section we know that there is a negative correlation between the persistence parameter and the volatility of output for a given size of the shocks. And there is in turn a positive correlation between the volatility of time series aggregate log TFP and the persistence parameter.

The relationship between the volatility of output and the cross sectional dispersion in productivity is non-monotonic (See Appendix A below). Hence, the persistence of the underlying aggregate shock has a non-monotonic relation with the observed cross sectional dispersion. The figure below display a scatter plot comparable to the figures in the introduction of the paper.

Figure 3: Dispersion in TFP and Cyclical Component of GDP, Model



Qualitatively the model can replicate the shape of the relationship. By changing the size of the shocks as well as the persistence parameter it might be possible to replicate the full range of volatility of output and productivity dispersion observed in the data.

The observed dispersion in productivities generated by the model is bounded by the range of the exogenous productivity z and the capacity levels, which we used to generate the distribution of log employment. By including more levels of capital we could drop the underlying dispersion in idiosyncratic productivity and still generate enough variability in log of employment to match the size distribution. This is a relevant issue to address when generating comparable measures of cross sectional dispersion between the model and the data.

6.3.4 Optimal Policy

Given the cost structure that was backed out in the calibration exercise we can study recommendations for optimal policy. The taxing scheme described in the body of the paper is obtained numerically by averaging over the observed allocations along the simulated paths in the Montecarlo experiment. We solve the planner problem with the cost structure implied by the calibrated decentralized allocation, i.e.

	good times	**bad times**
Π^e_L	.99	1.58
Π^e_H	1.02	1.6
I_L	.99	1.58
I_H	-.75	.55

The allocations obtained from the centralized problem differ from those studied in the baseline case before. The optimal policy induces a larger mass of establishments holding higher levels of employment, and a higher share of total employment.

Figure 4: Establishment Distribution

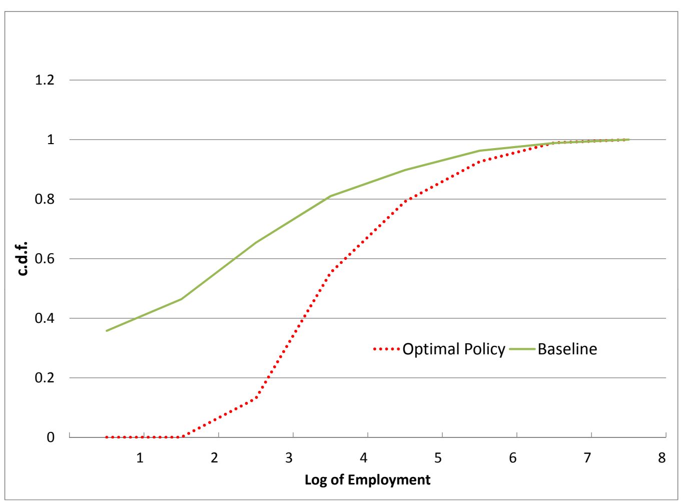
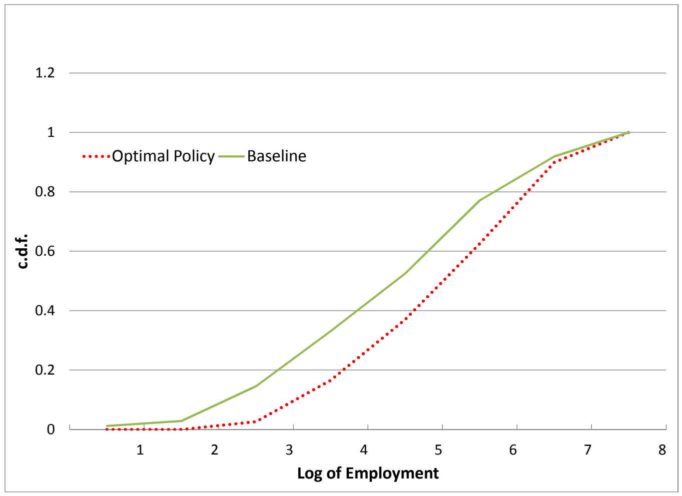


Figure 5: Employment Distribution



Average productivity increases about 13% while volatility drops and cross sectional dispersion drops 25%, mostly within low capacity firms. The table below summarizes those results and includes a measure of the cross sectional dispersion in computed log productivities. Mean TFP and the cross sectional dispersion were normalized to 1 in the baseline.

	Baseline	**Optimal Policy**
mean TFP	1	1.13
s.d. TFP	1.3%	1.4%
CSDP	1	0.75

To gain intuition, we report the cost structure that should prevail in the market for the decentralized allocation to replicate the optimal policy.

	good times	bad times
Π_L^e	.53	.90
Π_H^e	.56	.85
I_L	.53	.90
I_H	-.95	-.43

In lines with the prediction that we described for the distribution of establishments in the economy, investment in capacity should be highly subsidized, and the size of the subsidy should be procyclical. In other others, the price of investment goods should be highly countercyclical.

Entry costs should be lower than under the baseline, as higher entry rates induce further competition for resources and selection towards the most productive firms. Higher competition, implies lower markups and larger and more productive establishments on average. The scrap value for low capacity firms parallels the behavior of the entry cost by construction.

7 Conclusion

This paper explores the implications of technological restrictions at the firm level for aggregate efficiency in production when there is aggregate uncertainty. I find that dispersion in marginal products among firms operating analogous technologies can be consistent with (constrained) efficient allocations. Furthermore, observed dispersion is not independent of other features of the economy, such as the business cycle or more broadly the degree of demand uncertainty that firms face. The paper highlights that dispersion in marginal products is an imperfect measure of the associated efficiency losses.

When the industry dynamic is incorporated in a general equilibrium framework, low aggregate productivity allocations can be associated with relatively low dispersion in marginal products and very little firm churning. On the other hand, high aggregate productivity allocations are associated to low dispersion and can be consistent with extensive firm churning, i.e. USA. Hence the relationship between productivity dispersion and aggregate efficiency is non-monotonous.

Partial irreversibility and higher divisibility in capital allocations will undo some of the quantitative implications of the model. However, as long as the movements in investment thresholds between expansion and recession periods is such that the measure of incumbents firms holding capital away from the one they would have chosen if entering the market in the current period is non-trivial, non-convexities at the micro level will induce dispersion in marginal products and computed productivity.

As usual in models with non-convexities, the volatility of the underlying shock is relevant in explaining equilibrium allocations. In this model also, the size and persistence of aggregate shocks generate disparate aggregate responses and firm dynamics for a given unconditional volatility. The study of firm dynamic as an identifying mechanism for features of the aggregate productivity process is a promising path for further research.

Finally, the paper contributes to the study of non-convex economies with heterogeneous agents by providing an equivalence result. The allocations of the problem here presented can be generated by a constrained optima problem. The equivalence result eases the way to study the features of the optimal policy. Although our equivalence result relies heavily in the existence of a continuum of firms and the presence of sunk costs, potentially other schemes of adjustment costs can be accomodated.

References

[1] Abel, Andrew & Eberly, Janice, 1997 "Optimal Investment with Costly Reversibility" Review of Economic Studies, Vol 63(4) pg. 581-93, October .

[2] Alessandria, George & Choi, Horag, 2007 "Do Sunk Costs of Exporting Matter for Net Export Dynamics?" Quarterly Journal of Economics, pp. 289-336, February.

[3] Asker, John, Collard-Wexler, Allan & De Loecker, Jan, 2011. " Productivity Volatility and the Misallocation of Resources in Developing Countries", NBER Working Paper No. 17175.

[4] Barstelman, Eric, Haltiwanger, John & Scarpetta, Stefano, 2009. "Measuring and Analyzing Cross Country Differences in Firm Dynamics", in Price Dynamics: New Evidence from Micro Data, edited by Dunne, T., Jensen, B. Roberts, M.. University of Chicago Press, January.

[5] Bertola, Giuseppe & Caballero, Ricardo. 1991 "Irreversibility and Aggregate Investment" Review of Economic Studies, Vol 61(2) pg. 223-43, April.

[6] Bilbiie, Florin & Ghironi, Fabio & Melitz, Marc, 2007. "Endogenous Entry, Product Variety, and Business Cycles," NBER Working Papers 13646, National Bureau of Economic Research, Inc.

[7] Bloom, N., Floetotto, M. & Jaimovich, N., 2010. "Really Uncertain Cycles" Working Paper, Stanford University.

[8] Busso, Matias, Madrigal, Lucia &Pages, Carmen, 2012 "Productivity and Resource Misallocation in Latin America" IDB Working Paper Series No. IDB-WP-306

[9] Caballero, Ricardo, 1999."Aggregate Investment" Chapter 12 in Woodford and Taylor. Handbook of Macroeconomics, vol. IB. Amsterdam.

[10] Caballero, Ricardo & Engel, Eduardo, 1994. "Explaining investment dynamics in U.S manufacturing: A Generalized S,s approach" NBER, Working Paper 4887.

[11] Campbell, Jeffrey, 1998. "Entry, Exit, Embodied Technology, and Business Cycles," Review of Economic Dynamics, Elsevier for the Society for Economic Dynamics, vol. 1(2), pages 371-408, April.

[12] Clementi, Gian Luca & Palazzo, Dino, 2010. "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," Working Paper Series 27_10, Rimini Centre for Economic Analysis.

[13] Dixit, Avinash & Pindick, Robert, 1994 "Investment under Uncertainty" Princeton

[14] Eslava, Marcela & Haltiwanger, John & Kugler, Adriana &Kugler, Maurice, 2009. "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia" NBER Working Papers 14935, National Bureau of Economic Research, Inc.

[15] Gertler, Mark & Gilchrist, Simon, 1993. "The cyclical behavior of short-term business lending: Implications for financial propagation mechanisms," European Economic Review, Elsevier, vol. 37(2-3), pp 623-63.

[16] Ghironi, F. & Melitz, M.,2004. "International Trade and Macroeconomic Dynamics with Heterogenous Firms", NBER working papers 10540.

[17] Hopenhayn, Hugo A, 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," Econometrica, Econometric Society, vol. 60(5), pp 1127-50, September.

[18] Hsieh, C. & Klenow, P., 2009. "Misallocation and Manufacturing TFP in China and India", Quarterly Journal of Economics, Vol. CXXIV, November, Vol.4.

[19] Jovanovic, Boyan, 1982. "Selection and the Evolution of Industry," Econometrica, vol. 50(3), pages 649-70

[20] Jovanovic, Boyan, 1997. "Vintage Capital and Inequality", NBER Working Paper 6416.

[21] Kehrig, Mattias, 2011. "The cyclicality of Productivity Dispersion", Working paper, Northwestern University.

[22] Kiyotaki, Nobuhiro & Moore, Nobuhiro,1997. "Credit Cycles" Journal of Political Economy, vol.105(2), pp.211-248.

[23] Kydland, Fynn & Prescott, Edward, 1982. "Time to Build and Aggregate Fluctuations", Econometrica, vol. 50, pp. 1345-1370.

[24] Lee,Yoonsoo & Mukoyama, Toshihiko, 2008. "Entry, Exit, and Plant-Level Dynamics over the Business Cycle," Working Papers 08-17, Center for Economic Studies, U.S. Census Bureau

[25] Lucas, Robert & Prescott, Edward, 1971. "Investment under Uncertainty". Econometrica, vol. 39, pp. 659-681.

[26] Luttmer, Erzo,2010. "Models of Growth and Firm Heterogeneity", Annual Review of Economics, vol.2, pp.547-576.

[27] Luttmer, Erzo, 2005. "Growth and the Size Distribution of Firms" University of Minnesota and Federal Reserve Bank of Minneapolis.

[28] Mas-Colell, Andreu, 1977. "Indivisible commodities and general equilibrium theory," Journal of Economic Theory, Elsevier, vol. 16(2), pages 443-456, December.

[29] McDonald, R. and Siegel, (1986), "The Value of Waiting to Invest," The Quarterly Journal of Economics, vol. 101, No. 4. (Nov., 1986), pp. 707-728

[30] Melitz, Marc, 2003."The impact of Trade on Intre-Industry reallocations and Aggregate Industry Productivity", Econometrica, vol.71,No. 6.

[31] Midrigan, Virgiliu & Xu, Daniel, 2010. "Finance and Misallocation: Evidence from Plant-level Data," NBER Working Papers 15647, National Bureau of Economic Research, Inc.

[32] Moscarini, Giuseppe & Postel-Vinay, Fabien, 2009 "Large Employers Are More Cyclically Sensitive" NBER Working Paper 14740.

[33] Mariotti, Thomas; Decamps, J. & Villeneuve, S., 2006 "Irreversible investment in alternative projects", Economic Theory, 28 , pp.425-448

[34] Veracierto, Marcelo, 2002. "Plant-Level Irreversible Investment and Equilibrium Business Cycles," American Economic Review, American Economic Association, vol. 92(1), pages 181-197, March.

[35] Rouwenhorst, K.G., 1995. "Asset Pricing Implications of Equilibrium Business Cycle Models," In: Cooley, T.F. (Ed.), Frontiers of Business Cycle Research. Princeton University Press, Princeton, NJ, 294-330.

[36] Stokey, Nancy, 2009 " The Economics of Inaction" Princeton University Press.

8 Appendix (A)

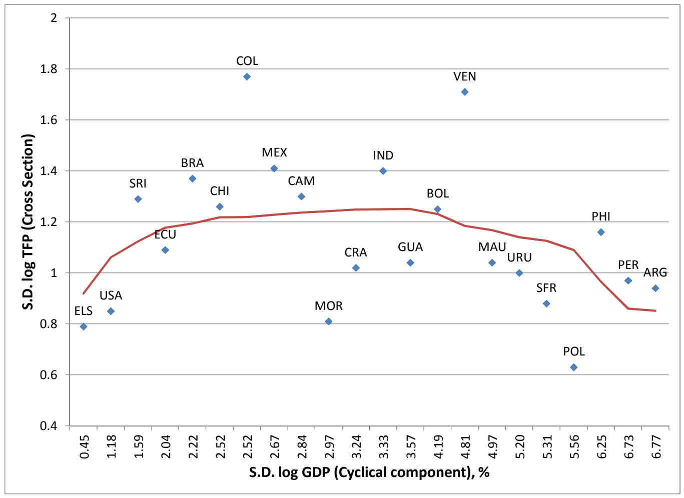
The following figure illustrates a U-shaped relationship between the volatility of output and the observed CSDP. On the vertical axis there are measures of productivity dispersion from micro level data for a sample of Latin American countries (See Busso, et.al., 2012) combined with Asker, et.al. (2011) measures of productivity based on the World Bank Enterprise Survey data. The methodology used to compute the dispersion measure follows Hsieh and Klenow(2009) as is explained in detail below. In the horizontal axis, I report the standard deviation of the cyclical component of log GDP (in percentage points). Let country i CSDP be σ_i^{cs}, and its correspondent volatility of output be σ_i^y. If we run a regression of the form

$$\sigma_i^{cs} = a_0 + a_1\sigma_i^y + a_2\left(\sigma_i^y\right)^2 + a_3\left(\sigma_i^y\right)^3 + \epsilon_i$$

we find a hump-shaped relationship between them[15]. This relationship is statistically significant in the sense that a standard ANOVA F-test, rejects a linear trend versus a quadratic one (p-value 0.024) and the quadratic trend cannot be rejected against an alaternative of a cubic trend (p-value 0.54). The predicted values of the regression are plotted as the trend line in the scatter below.

[15] I run the same exercise for an alternative sample of countries using labor productivity dispersion. A similar pattern is found (Data from Bartelsman, et.al. (2009).

Figure 6: Dispersion in TFP and Volatility of Output



8.1 Asker, Collard-Wexler and De Loecker (2011)

The notion of productivity is that of revenue productivity, as in Foster, Haltinwanger and Syverson (2008). This is the standard notion as typically prices and quantities are not separatedly observed at the producer level. As the authors clearly describe, deviations across producers in our measure of productivity, or its covariance with firm size, are directly informative of any type of distortion, be it adjustment costs, markups or policy distortions.

The main source of data is the World Bank Enterprise Research Data, which supplies data collected in a coordinated way across 33 countries. These data were collected by the World Bank across 41 countries and many different industries between 2002 and 2006. Standard output and input measures are reported in a harmonized fashion. Out of the 41 countries in the data, 33 have usable rm-level observations. This primarily because, for many years and countries, the World Bank did not collect multi-year data on capital stock.They use 5,558 observations that have full information, which is needed to compute productivity volatility.

Out of the total 33 countries we select out 16 for which GDP measures are available for relatively long periods of time. Countries and number of observations are listed below

Country	**Firms**
Morocco	376
Malawi	125
Mauritious	52
South Africa	199
Sri Lanka	114
Philipines	278
Indonesia	426
Costa Rica	273
Ecuador	109
El Salvador	190
Guatemala	162
Brazil	85
Chile	745
Peru	31
Poland	63
Guyana	29

The data is available from http://www.enterprisesurveys.org and was accessed by the authors on December 15th, 2010. The population of industries to be included in the Enterprise Surveys and Indicator Surveys, includes (according to ISIC,revision 3.1): all manufacturing sectors (group D), construction (group F), services (groups G and H), transport, storage, and communications (group I), and subsector 72 (from Group K). The survey only includes firms with 5 employees or more, and fully owned government establishments are excluded.

The survey used a stratified sampling procedure, in which firms were sampled randomly within groups based on the firm's sector of activity, firm size, and geographicallocation. The structure of the sampling leads to an oversampling of larger firms (relative to random sampling of all firms in the economy). The exact structure of the stratification varies by the size of the economy in question. The authors did not do any sampling correction.

Firms were surveyed between 2002 and 2006. The majority of firms within a country were surveyed in the same year. The survey asked questions about activity in the current year and the previous two years. Thus, the panel-data aspect of these data, relating to activity in year t to t-2, comes from the recollections and records of managers in year t.

If markets are complete and there are no adjustment costs, profit maximization implies that the coefficient in the revenue production function associated to a particular input equals the share of the input in sales,

$$\beta_x = \frac{P_{it}^x X_{it}}{S_{it}}$$

for any input X. Productivity is measured as a residual

$$w_{it} = s_{it} - \beta_k k_{it} - \beta_l l_{it} - \beta_m M_{it}$$

To recover the share of capital in revenue production, they assume constants returns in physical production, so that

$$\beta_k = (1 - \frac{1}{e}) - \beta_l - \beta_m$$

where e is the value of the elasticity parameter, which they set equal to 4 (As in Bloom, 2009).

8.2 Busso, Madrigal and Pages (2011)

They use micro-data from manufacturing firms in 10 Latin American countries. The results are obtained from the analysis of establishment-level data produced by countries' statistical offices.Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, El Salvador, Mexico, Uruguay and Venezuela.

Data coverage varies across countries but, in general, data are drawn from censuses of the largest, formally established firms, with a random sample of smaller firms. The datasets are representative at the national level and span a period of nearly 10 years. In most cases, coverage is restricted to the manufacturing sector, although in Mexico and Uruguay data coverage extends beyond manufacturing to other economic sectors. Most of the analysis is restricted to firms of 10 or more employees.

The methodology follows Hsieh and Klenow (2009) very closely. That is, wedges in final good production and capital rental cost are assumed. The analysis is conducted evaluating revenue product productivity and constant returns to scale are assumed for the underlying production function.

9 Appendix (B)

9.1 Equivalence with the decentralized solution

Proof.

Claim 10 *The allocation of firms across capacities solves the following systems of equations*

$$\frac{\partial Y_t}{\partial z_t^{e*}(\psi_L)} + E\left[\widetilde{\beta}_{t+1}\Pi_{t+1}^{e*}(\psi_L)\eta_t^L(\psi_L)\left(1-\delta\right)\right] + I_t^{H*}M_t^{ent}\frac{1-G(z_t^{u*})}{\left(1-G(z_t^{e*}(\psi_L))\right)^2}g(z_t^{e*}(\psi_L))$$
$$\geq \Pi_t^{e*}(\psi_L)\left(1-\delta\right)\eta_{t-1}^L(z_t^{e*}(\psi_L))$$

$$\frac{\partial Y_t}{\partial z_t^{e*}(\psi_H)} + E\left[\widetilde{\beta}_{t+1}\Pi_{t+1}^{e*}(\psi_H)\eta_t^H(z_t^{e*}(\psi_H))\left(1-\delta\right)\right] \geq \Pi_t^{e*}(\psi_H)\left(1-\delta\right)\eta_{t-1}^H(z_t^{e*}(\psi_H))$$

$$\frac{\partial Y_t}{\partial z_t^{u*}} + E\left\{\widetilde{\beta}_{t+1}\left(I_{t+1}^{H*}\left(1-\delta\right)\eta_t^L(z_t^{u*})\right)\right\}$$
$$= I_t^{H*}\left[\left(1-\delta\right)\eta_{t-1}^L(z_t^{u*}) + \frac{M_t^{ent}}{1-G(z_t^{e*}(\psi_L))}g(z_t^{u*})\right]$$

jointly with the equilibrium measure of entrants.

$$I_t^{L*} + I_t^{H*}\frac{1-G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))}$$
$$= \frac{1-\rho}{\rho}\frac{Y_t}{M_t} + E\left\{\widetilde{\beta}_{t+1}\left(1-\delta\right)\left(\Pi_{t+1}^{e*}(\psi_L)\frac{G(z_t^{u*})-G(z_t^{e*}(\psi_L))}{1-G(z_t^{e*}(\psi_L))} + \Pi_{t+1}^{e*}(\psi_H)\frac{1-G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))}\right)\right\}$$

where $M_t = v_t^L + v_t^H$, $z_t^{ent*} = z_t^{e*}(1)$

Aggregate output and consumption are as defined in the problem; Υ *given;* $\eta_t^j(z)$ *is the Radon-Nykodin derivative of* v^j *so that*

$$v_t^j(z') \equiv v_t^j((0,z']) = \int_{z_t^{e*}(.)}^{z'} \eta_t^j(z)dz$$

[16]

[16] v^j is absolutely continuous with respect to the standard euclidean measure on the real line.

The allocation problem is

$$V\left(\widetilde{X}_t\right) = \underset{C_t,\ K_{t+1},\ Y_t,\ z_t^{e*}(\psi_L), z_t^{e*}(\psi_H),\ z_t^{u*}, M_t^{ent}, l_t^i, k_t^i}{Max} U\left(C_t\right) + \beta E V\left(\widetilde{X}_{t+1}\right)$$

subject to

$$\begin{aligned}
& C_t + K_{t+1} - K_t(1-\widehat{\delta}) + \\
& + I_t^{L*} M_t^{ent} + I_t^{H*}\left\{(1-\lambda)\left[v_{t-1}^L(z_{t-1}^{u*}) - v_{t-1}^L(z_t^{u*})\right]\chi_{\left(z_t^u < z_{t-1}^u\right)} + M_t^{ent}\frac{1-G(z_t^{u*})}{1-G(z_t^{e*}(1))}\right\} \\
\leq \;& Y_t + T_t + \Pi_t^{e*}(1)\,(1-\delta)\,v_{t-1}^L(z_t^e(\psi_L)) + \Pi_t^{e*}(k)\,(1-\delta)\,v_{t-1}^H(z_t^e(\psi_H))]
\end{aligned}$$

$$s_t\left(\int\left(\psi_L z_i l_i^{1-\alpha} k_i^{\alpha}\right)^{\rho} dv_t^L(z_i) + \int\left(\psi_H z_i l_i^{1-\alpha} k_i^{\alpha}\right)^{\rho} dv_t^H(z_i)\right)^{\frac{1}{\rho}} = Y_t$$

$$\int l_i di = 1, \text{ and } \int k_i di = K_t$$

$$k_i \geq k_L \text{ if } \psi_i = \psi_L, \text{ and } k_i \geq k_H \text{ if } \psi_i = \psi_H$$

and the Law of motion for the distributions

$$\begin{aligned}
v_t^L(z) &= (1-\delta)\left[v_{t-1}^L(z) - v_{t-1}^L(z_t^{e*}(\psi_L))\chi_{\left(z_t^e(\psi_L) > z_{t-1}^e(\psi_L)\right)}\right] + M_t^{ent}\frac{G(z) - G(z_t^{e*}(\psi_L))}{1-G(z_t^{e*}(\psi_L))} \qquad z_t^{u*} > z > z_t^{e*} \\
&= 0 \qquad \text{o/w}
\end{aligned}$$

If $z_{t-1}^{u*} \leq z_t^{u*}$

$$\begin{aligned}
v_t^H(z) &= (1-\delta)\ \left[v_{t-1}^H(z) - v_{t-1}^H(z_t^{e*}(\psi_H))\chi_{\left(z_t^e(\psi_H) > z_{t-1}^e(\psi_H)\right)}\right] \qquad z_t^{u*} > z > z_t^{e*}(\psi_H) \\
&= (1-\delta)\ v_{t-1}^H(z) + M_t^{ent}\frac{G(z) - G(z_t^{u*})}{1-G(z_t^e(\psi_L))} \qquad 1 > z > z_t^{u*}
\end{aligned}$$

If $z_{t-1}^{u*} > z_t^{u*}$

$$\begin{aligned}
v_t^H(z) &= (1-\delta)\ v_{t-1}^H(z) - v_{t-1}^H(z_t^{e*}(\psi_H))\chi_{\left(z_t^e(\psi_H) > z_{t-1}^e(\psi_H)\right)} \qquad z_t^{u*} > z > \ z \\
&= (1-\delta)\left[v_{t-1}^H(z) + v_{t-1}^L(z) - v_{t-1}^L(\ z_t^{u*})\right] + M_t^{ent}\frac{G(z) - G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))} \qquad z_{t-1}^{u*} > z > \ z_t^{u*} \\
&= (1-\delta)\,v_{t-1}^H(z) + v_{t-1}^L(z_{t-1}^{u*}) - v_{t-1}^L(\ z_t^{u*}) + M_t^{ent}\frac{G(z) - G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))} \qquad 1 > z > z_{t-1}^{u*}
\end{aligned}$$

Keep in mind that $v_{t-1}^{j}(z_t^{e*}(\psi_L)) = \int_{z_{t-1}^{e*}(\psi_L)}^{z_t^{e*}(\psi_L)} \eta_{t-1}^{j}(z)dz$, so the whole problem can be rewritten in terms of the measures η^j.

Let's first compute the optimality conditions that have an obvious counterpart in the decentralized allocation. The optimality condition for capital accumulation reads

$$U'(C_t) = E_t\left[U'(C_{t+1})\,\beta\left[\lambda_{t+1}^{k} + 1 - \delta\right]\right] \qquad (K_{t+1})$$

If allocations are the same in the decentralized and centralized allocations, then aggregate TFP (as shown in the aggregation section of the paper) will be the same, and the marginal product of capital too. Typically, $\lambda_{t+1}^{k} \equiv \frac{\partial Y_{t+1}}{\partial K_{t+1}} = \alpha s_{t+1} TFP_{t+1} K_{t+1}^{\alpha-1}$.

The optimality conditions for capital and labor allocations read,

$$(1-\alpha)\, Y_t^{1-\rho} y_t^{\rho-1} \psi_j z_i l_i^{-\alpha} k_i^{\alpha} = \lambda_t^{l}$$

$$\alpha Y_t^{1-\rho} y_t^{\rho-1} \psi^j z_i l_i^{1-\alpha} k_i^{\alpha-1} = \lambda_t^{k} - \varrho_t^{j}$$

where λ_t^l and λ_t^k are the lagrange multipliers associated to the feasibility constraint on labor and capital, respectively. $\varrho_t^j \geq 0$ is the lagrange multiplier associated to the minimum capacity constraint on capital. Combining both expressions we obtain

$$\frac{k}{l} = \frac{\lambda_t^{l}}{\lambda_t^{k} - \varrho_t^{j}} \frac{\alpha}{1-\alpha}$$

If prices in the decentralized allocation equalize the shadow values of the feasibility constraints, then capital labor allocation (conditional on a technology distribution across firms) will be the same.

Let's now analyze the optimality conditions with respect to the dynamic decisions of the firms, where the wedges between the centralized and decentralized allocation show up.

The optimality condition with respect to the exit threshold for low capacity firms is

$$\begin{aligned} 0 \;\geq\; & U'(C_t)\left[\frac{\partial Y_t}{\partial z_t^{e*}(\psi_L)} + \Pi_t^{e*}(\psi_L)\,(1-\delta)\,\eta_{t-1}^{L}(z_t^{e*}(\psi_L))\right] \\ & + U'(C_t)\; I_t^{H*} M_t^{ent} \frac{1 - G(z_t^{u*})}{(1 - G(z_t^{e*}(\psi_L)))^2} g(z_t^{e*}(\psi_L)) + \beta E \frac{\partial V(\widetilde{X}_{t+1})}{\partial z_t^{e*}(\psi_L)} \quad (z_t^{e*}(1)) \end{aligned}$$

$$\frac{\partial V(\widetilde{X}_t)}{\partial z^{e*}_{t-1}(\psi_L)} = -U'(C_t)\Pi^{e*}_t(1)\eta^L_{t-1}(z^{e*}_{t-1}(\psi_L))\,(1-\delta)$$

Replacing,

$$\begin{aligned}
&\frac{\partial Y_t}{\partial z^{e*}_t(\psi_L)} + E\left[\widetilde{\beta}_{t+1}\Pi^{e*}_{t+1}(\psi_L)\eta^L_t(\psi_L)\,(1-\delta)\right] \\
&+I^{H*}_t M^{ent}_t \frac{1-G(z^{u*}_t)}{(1-G(z^{e*}_t(\psi_L)))^2} g(z^{e*}_t(\psi_L)) \\
\geq\;& \Pi^{e*}_t(\psi_L)\,(1-\lambda)\,\eta^L_{t-1}(z^{e*}_t(\psi_L))
\end{aligned}$$

where $\widetilde{\beta}_{t+1}$ is the endogenous discount factor as defined before. The equation holds with equality if $z^{e*}_t(\psi_L) > z^{e*}_{t-1}(\psi_L)$ i.e. there is exit of firms in the market. Note that η^L_{t-1} is not defined for $z^{e*}_t(\psi_L) < z^{e*}_{t-1}(\psi_L)$, arbitrarily set equal to zero. At the optimal exit threshold, the marginal cost of exiting firms of type $z^{e*}_t(\psi_L)$: marginal output loss plus the expected future return to exiting tomorrow, plus the increase in the relative number of upgrades given an entry level; equalizes the marginal benefit of exit: the value in consumption units of the stock of capital that is being liquidated ($\Pi^{e*}_t(\psi_L)$).

Analogously the exit threshold for high capacity solves

$$\frac{\partial Y_t}{\partial z^{e*}_t(\psi_H)} + E\left[\widetilde{\beta}_{t+1}\Pi^{e*}_{t+1}(\psi_H)\eta^H_t(z^{e*}_t(\psi_H))\,(1-\delta)\right] \geq \Pi^{e*}_t(\psi_H)\,(1-\delta)\,\eta^H_{t-1}(z^{e*}_t(\psi_H))$$

with equality if $z^{e*}_t(\psi_H) > z^{e*}_{t-1}(\psi_H)$. If there is exit, the productivity threshold is such that the value of exiting today has to equalize the value of operating in the current period and exiting one period later.

The optimality condition with respect to the upgrade threshold is

$$\begin{aligned}
0 \;\leq\;& U'(C_t)\left[\frac{\partial Y_t}{\partial z^{u*}_t} + I^{H*}_t\left[(1-\delta)\,\eta^L_{t-1}(z^{u*}_t)\chi_{\left(z^{u*}_{t-1} > z^{u*}_t\right)} + \frac{M^{ent}_t}{1-G(z^{e*}_t(\psi_L))} g(z^{u*}_t)\right]\right] \\
&+\beta E\frac{\partial V(\widetilde{X}_{t+1})}{\partial z^{u*}_t}
\end{aligned}$$

where

$$\frac{\partial V(\widetilde{X}_t)}{\partial z^{u*}_{t-1}} = U'(C_t)\left[-I^{H*}_t\,(1-\delta)\,\eta^L_{t-1}(z^{u*}_{t-1})\right]$$

Hence,

$$\begin{aligned} & \frac{\partial Y_t}{\partial z_t^{u*}} + E\left\{ \widetilde{\beta}_{t+1}\left(I_{t+1}^{H*}\left(1-\delta\right)\eta_t^L(z_t^{u*})\right)\right\} \\ = \; & I_t^{H*}\left[\left(1-\delta\right)\eta_{t-1}^L(z_t^{u*}) + \frac{M_t^{ent}}{1-G(z_t^{e*}(\psi_L))} g(z_t^{u*})\right] \end{aligned}$$

At the upgrade threshold, the cost of increasing the upgrade threshold, i.e. the differential in output for an additional type running at high capacity plus the cost of potential future upgrades given by the delayed upgrade today, should equalize the cost savings of upgrading current low capacity firms, plus the marginal entrants that may be running at high capacity.

Finally, the optimality condition for the entry measure M^{ent}. To compute it, it will be useful to recall that aggregate output in the economy can be expressed as $Y_t(X_t) = s_t M_t^{\frac{\rho}{1-\rho}} \frac{\left[\sum_{j=L,H} \int (\frac{\psi_i z_i}{\left(MPK_t^i\right)^{\alpha}})^{\frac{\rho}{1-\rho}} d\widehat{v}_t^j \right]^{\frac{1-(1-\alpha)\rho}{\rho}}}{\left[\sum_{j=L,H} \int (\frac{\psi_i z_i}{\left(MPK_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}})^{\frac{\rho}{1-\rho}} d\widehat{v}_t^j \right]^{\alpha}}$. Also, $M_t = M_{t-1}\left(1-\delta\right) + M_t^{ent}$.

Optimality yields

$$\begin{aligned} & I_t^{L*} + I_t^{H*}\frac{1-G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))} \\ = \; & \frac{1-\rho}{\rho}\frac{Y_t}{M_t} + E\left\{ \widetilde{\beta}_{t+1}\left(1-\delta\right)\left(\Pi_{t+1}^{e*}(\psi_L)\frac{G(z_t^{u*})-G(z_t^{e*}(\psi_L))}{1-G(z_t^{e*}(\psi_L))} + \Pi_{t+1}^{e*}(\psi_H)\frac{1-G(z_t^{u*})}{1-G(z_t^{e*}(\psi_L))}\right)\right\} \end{aligned}$$

In other words, the average cost of entry has to equalize the average return on the additional firms in the market which includes the value of producing today plus the additional scrap value if liquidated in the following period. ■

Claim 11 *There exist an industrial policy* $\left\{ \tau^e(\psi_L, \widetilde{X}), \tau^e(\psi_H, \widetilde{X}), \tau^{ent}(\widetilde{X}), \tau^u(\widetilde{X})\right\}$ *such that given a cost structure for the planner* Υ *with* $T_t = \left(\tau^e(\psi_L, \widetilde{X}) + \tau^e(\psi_H, \widetilde{X})\right)\Pi^e + I^H\tau^u(\widetilde{X}) + I^L\tau^{ent}(\widetilde{X})$*; the associated set of thresholds and entry level* $Z^*(\widetilde{X}) = \left\{ z_t^{e*}(\psi_L), z_t^{e*}(\psi_H), z^{u*}, M^{e}\right.$ *satisfy*

$$\Pi^e = F_1(Z^*(\widetilde{X}))$$

$$\Pi^e = F_2(Z^*(\widetilde{X}))$$
$$I^H = F_3(Z^*(\widetilde{X}))$$
$$I^L = F_4(Z^*(\widetilde{X}))$$

where F_i $i = 1, 2, 3, 4$ correspond to the optimality conditions for the exit and upgrade thresholds, and entry level in the decentralized economy. In other words, $Z^*(\widetilde{X})$ characterizes the decentralized allocation.

The key to understand how the policy is designed is to observe that the optimality conditions of the planner with respect to allocation of firms across technologies are linear in the costs.

Proof. Let the tax/subsidy structure satisfy

$$\begin{aligned}
1+\tau^e(\psi_L, X_t) &= \frac{\frac{\partial Y_t}{\partial z_t^e(\psi_L)} + I^H(1+\tau^u(X^s))M_t^{ent}\frac{1-G(z^u)}{(1-G(z_t^e(\psi_L))^2}g(z^e(1,X^s))}{\eta_{t-1}^L(z_t^e(\psi_L))\,(1-\delta)\,\Pi^e} \\
&\quad + E_t\left[\widetilde{\beta}_{t+1}\frac{\eta_t^L(z_t^e(\psi_L))}{\eta_{t-1}^L(z_t^e(\psi_L))}\,(1+\tau^e(\psi_L, X_{t+1}))\right]
\end{aligned}$$

$$\begin{aligned}
1+\tau^e(\psi_H, X_t) &= \frac{\frac{\partial Y_t}{\partial z_t^e(\psi_H)}}{\eta_{t-1}^L(z_t^e(\psi_H))\,(1-\delta)\,\Pi^e} \\
&\quad + E_t\left[\widetilde{\beta}_{t+1}\frac{\eta_t^L(z_t^e(\psi_H))}{\eta_{t-1}^L(z_t^e(\psi_H))}\,(1+\tau^e(\psi_H, X_{t+1}))\right]
\end{aligned}$$

$$1+\tau^u(X^s) = \varphi(z^u)E_t\left[\widetilde{\beta}_{t+1}(1+\tau^u(X_{t+1}))\,(1-\delta)\,\eta_t^L(z^u(k,X^s))\right] + \frac{\partial Y_t}{\partial z_t^e(\psi_H)}$$

where $\varphi(z^u) = \left[(1-\lambda)\,\eta_{t-1}^L(z^u) + g(z^u)\frac{M_t^{ent}}{1-G(z_t^e(\psi_L))}\right]^{-1}$

$$\begin{aligned}
(1+\tau^{ent}(X_t))I_L &= \frac{1-\rho}{\rho}\frac{Y_t(X_t)}{M_t(X_t)} \\
&\quad + \Pi^e E\left\{\widetilde{\beta}_{t+1}\,(1-\delta)\,\{(1+\tau^e(\psi_L, X_{t+1}))\,\frac{G(z_t^u)-G(z_t^e(\psi_L)))}{1-G(z_t^e(\psi_L))} + (1+\tau^e(\psi_H, X_{t+1}))\right. \\
&\quad - I^H(1+\tau^u(X_t))\frac{1-G(z_t^u)}{1-G(z_t^e(\psi_L))}
\end{aligned}$$

where $M_t = M_{t-1}(1-\lambda) + M_t^{ent}$.

Note that the LHS are always evaluated at the equilibrium thresholds and the corresponding measure of firms from the decentralized problem. $\widetilde{\beta}_{t+1}$ is computed at the consumption levels indicated by the level of output and investment dictated by the thresholds of the decentralized problem.

The optimal policy today depends on the expected policy tomorrow. In computing the expected value of the transfer the planner uses the transition functions for the distribution described at the beginning of the problem. All in all, the vector of optimal transfers will be a solution to a system of linear equations defined by the equations we just described in both states.

Because the system is linear in tax/subsidy rates, an no equations are colinear,it has a solution. ■

Intuitively, there is a tax structure that equalizes the thresholds of the decentralized and centralized problem.

9.2 Existence and Uniqueness of the centralized allocation

Lemma 12 *The measure associated to the distribution of types is absolutely continuous(AC) with respect to the lebesque measure on the real line*

Proof. The claim follows from the absolute continuity of the exogenous distribution of types. We prove by induction.

By definition

$$v_1^L(z) = \left[\frac{G(z) - G(z^{e*}(1, X))}{1 - G(z^{e*}(1, X))}\right]$$

$$v_1^H(z) = \left[\frac{G(z) - G(z^{u*}(k, X))}{1 - G(z^{e*}(1, X))}\right]$$

Take a sequence of of intervals $(a_k, b_k)_{k=1}^K$ and let

$$\sum_{k=1}^{K} \left| v_1^j(b_k) - v_1^j(a_k) \right| \leq \varepsilon$$

Replacing by the definition

$$\sum_{k=1}^{K} \left| \frac{1}{1 - G(z_1^{eL})} \left(G(b_k) - G(a_k) \right) \right| \leq \varepsilon$$

Let $\widehat{\varepsilon} = \varepsilon \left[1 - G(z^{e*}(1, X)) \right]$ .By AC of G, there exist $\widehat{\delta}$ such that

$$\sum_{k=1}^{K} |b_k - a_k| \leq \widehat{\delta}$$

Because ε was arbitrary, and $(a_k, b_k)_{k=1}^K$ too, v_1^j is absolutely continuous.

Suppose v_N^j is absolutely continuous. By definition,

$$v_{N+1}^L(z) = (1-\delta) \left[v_N^L(z) - v_N^L(\psi_L) \right]_{\chi_{z_t^e(\psi_L) > z_{t-1}^e(\psi_L)}} + \left[\frac{G(z) - G(z^{e*}(\psi_L))}{1 - G(z^{e*}(\psi_L))} \right] \quad z_t^{u*} > z > z^{e*}(\psi_L)$$

If $z_{t-1}^{u*} \leq z_t^{u*}$

$$\begin{aligned} v_{N+1}^H(z) &= (1-\delta)\ v_N^H(z) && z_t^{u*} > z > \ z_t^{e*}(\psi_H) \\ &= (1-\delta) v_N^H(z) + M_t^{ent} \frac{G(z) - G(z_t^{u*})}{1 - G(z_t^{e*}(\psi_L))} && 1 > z > z_t^{u*} \end{aligned}$$

If $z_{t-1}^{u*} > z_t^{u*}$

$$\begin{aligned} v_{N+1}^H(z) &= (1-\delta)\ v_N^H(z) && z_t^{u*} > z > \ z_t^{e*}(\psi_H) \\ &= (1-\delta) \left[v_N^H(z) + v_N^L(z) - v_N^L(\ z_t^{u*}) \right] \ + M_t^{ent} \frac{G(z) - G(z_t^{u*})}{1 - G(z_t^{e*}(\psi_L))} && z_{t-1}^{u*} > \\ &= (1-\delta) v_N^H(z) + v_N^L(\ z_{t-1}^{u*}) - v_N^L(\ z_t^{u*}) + M_t^{ent} \frac{G(z) - G(z_t^{u*})}{1 - G(z_t^{e*}(\psi_L))} && 1 > z > z_{t-1}^{u*} \end{aligned}$$

therefore, the sum of absolutely continuous functions. Hence, v_{N+1}^j is absolutely continuous. ■

Proposition 13 *The solution to the Planner problem exist and it is unique*

Proof. Let $\Xi_{t-1} = v_{t-1}^L, v_{t-1}^H$, $s_t = \{s_g, s_b\}$. $E_\gamma$ia the expected value under the transition probabilities for the exogenous shock described in the body of the paper, γ_g, γ_b. We can write the planner's problem in terms of the operator $\digamma$ as

$$\digamma V(s_t, \Xi_{t-1}) = \underset{\Xi_t \in \Gamma(s_t, \Xi_{t-1})}{Max} U(\Xi_{t-1}, \Xi_t) + \beta E_\gamma \left[V(s_{t+1}, \Xi_t)\right]$$

where $\Xi_{t-1} \in \Theta$, the set of bounded absolutely continuous functions on R^+.The utility function is defined as $U : \Theta x \Theta \to R$

Let $H(\Theta, \theta)$ be the set of functions (functional) $f : \Theta \to R$ that are homogenous of degree $(1-\theta)$, continuous except potentially at the origin if $\theta > 1$ and bounded in the norm

$$\|f\| = \sup_{\|\Xi_{t-1}\|=1, \Xi_{t-1} \in \Theta} \|f(\Xi_{t-1})\|$$

$\digamma : H(\Theta, \theta) \to H(\Theta, \theta)$.

First we prove the proposition for $\theta < 1$.

Claim 14 *$U : \Theta x \Theta \to R$ is bounded and continuous if $\theta < 1$.*

Proof. *As defined in the body of the paper $U(.)$ is CES with parameter θ. If $\theta < 1$ then $U(C(\Xi_t, \Xi_{t+1}))$ is bounded below as $U(0) = 0$. Now, potentially U is unbounded above. However, we will show that the feasible measure of firms in the market is always bounded above, so that consumption is bounded and $U(.)$ too, along the relevant state space.*

By definition, the total measure of firms in the market is $M_t = v_t^L(1) + v_t^H(1)$. Using the aggregation results, one could right the feasibility constraint of the planner as

$$\begin{aligned} M_t^{\frac{1-\rho}{\rho}} y_t^* + T_t + \Pi_t^{e*}(\psi_L) M_L^e + \Pi_t^{e*}(\psi_H) M_H^e - I_t^{L*} M_t^{ent} - I_t^{H*} M_t^{upg} &= C_t \\ M_t - (1-\delta) M_{t-1} - M_t^{ent} + (M_L^e + M_H^e) &= 0 \end{aligned}$$

where M_L^e is the measure of low capacity firms that exit and M_H^e the analogous for high capacity firms, M_t^{upg} is the measure of firms that upgrade capacity.M_0 given.

A strategy to make the measure of firms grow without bound would be to never exit firms and enter as much as possible. Now, because entry is costly, optimality dictates that the marginal cost of an entrant equalizes the marginal return,

$$\frac{1-\rho}{\rho}\left(\left(1-\delta\right)M_{t-1}+M_t^{ent}\right)^{\frac{1-\rho}{\rho}-1}y_t^*=I_t^{L*}$$

which pins down a finite level of entry at each t. Replacing the entry level into the dynamic equation for the measure of firms we obtain

$$M_t=\left(\frac{\rho}{1-\rho}\frac{I_t^{L*}}{y_t^*}\right)^{\frac{\rho}{1-\rho}}$$

which is bounded.

$U(.)$ is continuous by assumption so the claim is proved. ■

■

Proof. Given the structure of the stochastic process for s_t, it has the Feller property.

Also, Θ is a convex set as each convex combination of two AC functions is AC. Furthermore the space Θ is a complete normed space (the proof is analogous to the proof for the set of continuous bounded functions).

Then $\digamma$ is a contraction, with a unique fixed point in Θ.

If instead $\theta \geq 1$, the return U is discontinuous at zero and potentially unbounded below. In the same line of the previous proof, we will show that the set of firms in the market is always bounded below at a positive measure. The set of productivities in the market is not growing, so the latter rules out unboundness.

Claim 15 *$U:\Theta x\Theta\rightarrow R$ is bounded if $\theta\geq 1$.*

Proof. *Write the feasibility constraint of the planner as*

$$\begin{aligned} M_t^{\frac{1-\rho}{\rho}} y_t^* + T_t + \Pi_t^{e*}(\psi_L) M_L^e + \Pi_t^{e*}(\psi_H) M_H^e - I_t^{L*} M_t^{ent} - I_t^{H*} M_t^{upg} &= C_t \\ M_t - (1-\delta) M_{t-1} - M_t^{ent} + (M_L^e + M_H^e) &= 0 \end{aligned}$$

A strategy to make the measure of firms shrinks without bound would be to never enter firms and exit as many as possible. Now, because exit is costly (in terms of output), optimality dictates

$$\frac{1-\rho}{\rho} \left((1-\delta) M_{t-1} + M^e\right)^{\frac{1-\rho}{\rho}-1} y_t^* = \Pi_t^{e*}$$

which pins down a finite level of entry at each t. Replacing the entry level into the dynamic equation for the measure of firms we obtain

$$M_t = \left(\frac{\rho}{1-\rho} \frac{\Pi_t^{e*}}{y_t^*}\right)^{\frac{\rho}{1-\rho}-1}$$

which is bounded at a positive number. Hence, U is bounded below. For $\theta = 0$ (log utility) boundness above is also an issue. The argument made in the previous claim applies in this case. ■

■

Proof. Hence U is continuous for the relevant state space of the economy (the point of discontinuity is not feasible, except if the economy starts with measure 0 of firms which we rule out). The remaining of the proof is as explained before. ■

9.3 Firm Flows

9.3.1 Exit and Upgrade

The value for a high capacity firm reads

$$W_t(z,k,X_t) = Max\left\{\Pi_e, \pi(z,k,X_t) + E_t\left(\widetilde{\beta}_{t+1}(1-\lambda) W_{t+1}(z,k,X_{t+1})\right)\right\}$$

Under the guess $E_t\left[v^H(z^u(X_{t+1})) - v^H(z^u(X_t))\right] < \lambda$, firm distributions are stationary across aggregate states as indexed by s.

To compute the exit threshold , rewrite the continuation value $\widetilde{W}(.)$ as

$$\widetilde{W}(z,k,X) = \pi(z,k,X) + E\left[\widetilde{\beta}\,(1-\delta)\,\Pi_e\right] + E_t\left(\widetilde{\beta}\,(1-\lambda)\max\left[W(z,k,X) - \Pi_e, 0\right]\right) \tag{11}$$

Let $X^j = (s_j, v_j^L, v_j^H)$ for $j = g, b$. Then one can write

$$\begin{aligned}
& \widetilde{W}(z^e(k,X^g),k,X^b) - \widetilde{W}(z^e(k,X^g),k,X^g) \\
= \; & \pi(z^e(k,X^g),k,X^b) - \pi(z^e(k,X^g),k,X^g) \\
& + E\left\{\left[\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right](1-\lambda)\,\Pi_e\right\} \\
& + \left[\widetilde{\beta}(X^b,X^b)\gamma_b - (1-\gamma_g)\widetilde{\beta}(X^g,X^b)\right](1-\lambda)\,Max\left[W(z^e(k,X^g),k,X^b) - \Pi_e, 0\right]
\end{aligned}$$

If $\widetilde{W}(z^e(k,X^g),k,X^b) - \widetilde{W}(z^e(k,X^g),k,X^g) > 0$,

$$\begin{aligned}
\widetilde{W}(z^e(k,X^g),k,X^b) - \widetilde{W}(z^e(k,X^g),k,X^g) \; = \; & (1-\kappa_{bb}+\kappa_{gb})^{-1}\left[\pi(z^e(k,X^g),k,X^b) - \pi(z^e(k,X^g),k,X\right. \\
& \left. -E\left\{\left[\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right](1-\lambda)\,\Pi_e\right\}\right]
\end{aligned}$$

where $\kappa_{bb} = (1-\delta)\,\widetilde{\beta}(X^b,X^b)\gamma_b$ and $\kappa_{gb} = (1-\delta)\,(1-\gamma_g)\widetilde{\beta}(X^g,X^b)$. Note also that if shocks are i.i.d., $E\left[\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right] = 0$ and the second term drops.

Replacing in (11) we obtain the equation describing the exit threshold $z^e(k,X^g)$

$$\begin{aligned}
& \widetilde{W}(z^e(k,X^g),k,X^g)\left\{1 - \kappa_{gg} - \kappa_{gb}\left(\frac{1-\kappa_{gg}+\kappa_{bg}}{1-\kappa_{bb}+\kappa_{gb}}\right)\right\} = \pi(z^e(k,X^g),k,X^g) + \\
& + \frac{\kappa_{gb}}{1-\kappa_{bb}+\kappa_{gb}}\left[\pi(z^e(k,X^g),k,X^b) - \pi(z^e(k,X^g),k,X^g)\right]\chi_{\left(\pi_{bg}>\pi_g\right)}
\end{aligned}$$

where $\pi_{bg} \equiv \pi(z^e(k,X^g),k,X^b)$ and $\pi_g \equiv \pi(z^e(k,X^g),k,X^g)$ and κ_{ig} is defined analogously to κ_{ib}.

Note that if $\pi_{bg} \leq \pi_g$

$$\widetilde{W}(z^e(k,X^g),k,X^g) = \frac{\pi(z^e(k,X^g),k,X^g)}{1-\kappa_{gg}-\kappa_{gb}\left(\frac{1-\kappa_{gg}+\kappa_{bg}}{1-\kappa_{bb}+\kappa_{gb}}\right)}$$

the threshold is increasing in the current state of the demand, as summarized by the marginal utility of consumption today. Higher demand today, implies lower marginal utility of consumption and higher discount factor $\widetilde{\beta}$, the RHS increases and hence the exit threshold drop.

Now if $\pi_{bg} > \pi_g$,the foregone value when exercising the option to exit includes potential future losses due to changes in the value of profits. The additional value increases with lower marginal utility of consumption in the current state ($\widetilde{\beta}(X^g, X^b)$ increases) so that the exit threshold drop.

The threshold in state b can be solved following the same strategy. Evaluate (11) at $z^e(k, X^b)$

$$\widetilde{W}(z^e(k,X^b),k,X^b)\left\{1 - E\left[\widetilde{\beta}\,(1-\lambda)\right]\right\} = \pi(z^e(k,X^b),k,X^b) +$$
$$+\frac{\kappa_{bg}}{1-\kappa_{gg}+\kappa_{bg}}\left[\pi(z^e(k,X^b),k,X^g) - \pi(z^e(k,X^b),k,X^b)\right]\chi_{\left(\pi_{gb}>\pi_b\right)} + \kappa_{bg}(1 - \frac{1-\kappa_{bb}+\kappa_{gb}}{1-\kappa_{gg}+\kappa_{bg}})\Pi_e$$

where $\pi_{gb} \equiv \pi(z^e(k, X^b), k, X^g)$ and $\pi_b \equiv \pi(z^e(k, X^g), k, X^b)$.

The value of small capacity firms can be described as

$$\begin{aligned} W(z,1,X) &= Max\{\Pi_e, W(z,k,X) - I^H, \pi(z,1,X) \\ &\quad +E\left(\widetilde{\beta}\,(1-\lambda)\,W(z,1,X)\right)\} \end{aligned}$$

Their continuation value is

$$W(z,1,X) = \pi(z,1,X) + E\left(\widetilde{\beta}\,(1-\lambda)\,W(z,1,X)\right)$$

Monotonicity implies that the optimal policy is a trigger strategy such that exit occurs whenever $z < z^e(1, X)$ and upgrade whenever $z > z^u(k, X)$. For convenience, rewrite the value function as

$$\begin{aligned} \widetilde{W}(z,1,X) &= \pi(z,1,X) + E\left[\widetilde{\beta}\,(1-\lambda)\,\Pi_e\right] + \\ &\quad +E_t\left(\widetilde{\beta}\,(1-\lambda)\max\left[0, \widetilde{W}(z,1,X) - \Pi_e, \widetilde{W}(z,k,X) - I_H - \Pi_e\right]\right) \end{aligned}$$

Let's first compute the exit threshold, $z^e(1,X)$. Using the fact that $\widetilde{W}(z^e(1,X),1,X) = \Pi_e$

$$\widetilde{W}(z,1,X)-\Pi_e = \pi(z,1,X)-\pi(z^e(1,X),1,X)+E_t\left(\widetilde{\beta}\,(1-\lambda)\max\left[0,\widetilde{W}(z,1,X)-\Pi_e\right]\right)$$

where we have used that for $z < z^u(k,X)$, optimality yields$\widetilde{W}(z,k,X) - I_H < \widetilde{W}(z,1,X)$. Throughout the analysis we are focusing in cost structures such that $z^e(1,X) < z^u(k,X)$.Otherwise all firms in the market will be running at high capacity and the capacity decision becomes trivial. Also note that optimality implies $z^e(k,X) < z^u(k,X)$ as it cannot be optimal to incur a cost I_h of building capacity to exit the market immediately.

Following a similar procedure as the one for high capacity firms, we obtain

All in all the exit thresholds,

$$\begin{aligned}&\widetilde{W}(z^e(1,X^g),1,X^g)\,(1-\kappa_{gg}-\kappa_{gb}) = \pi(z^e(1,X^g),1,X^g)+\\&+\frac{\kappa_{gb}}{1-\kappa_{bb}+\kappa_{gb}}\left[\pi(z^e(1,X^g),1,X^b)-\pi(z^e(1,X^g),1,X^g)\right]\chi_{\left(\pi_{bg}>\pi_g\right)}+\kappa_{gb}(1-\frac{1-\kappa_{gg}+\kappa_{bg}}{1-\kappa_{bb}+\kappa_{gb}})\Pi_e\end{aligned}$$

where $\pi_{bg} \equiv \pi(z^e(1,X^g),1,X^b)$ and $\pi_g \equiv \pi(z^e(1,X^g),1,X^g)$ and κ_s is as defined before.

$$\begin{aligned}&\widetilde{W}(z^e(1,X^b),1,X^b)\,(1-\kappa_{bb}-\kappa_{bg})\\=\;&\pi(z^e(1,X^b),1,X^b)\\&+\frac{\kappa_{bg}}{1-\kappa_{gg}+\kappa_{bg}}\left[\pi(z^e(1,X^b),1,X^g)-\pi(z^e(1,X^b),1,X^b)\right]\chi_{\left(\pi_{gb}>\pi_b\right)}\\&+\kappa_{bg}(1-\frac{1-\kappa_{bb}+\kappa_{gb}}{1-\kappa_{gg}+\kappa_{bg}})\Pi_e\end{aligned}$$

where $\pi_{gb} \equiv \pi(z^e(1,X^b),1,X^g)$ and $\pi_b \equiv \pi(z^e(1,X^g),1,X^b)$.

To compute the optimal upgrade threshold rewrite the value function as

$$\begin{aligned}\widetilde{W}(z,1,X) \;&=\; \pi(z,1,X)+E\left[\widetilde{\beta}\,(1-\delta)\,\widetilde{W}(z,1,X)\right]+\\&+E_t\left(\widetilde{\beta}\,(1-\delta)\max\left[0,\Pi_e-\widetilde{W}(z,1,X),\widetilde{W}(z,k,X)-I_H-\widetilde{W}(z,1,X)\right]\right)\end{aligned}$$

At the upgrade threshold $\widetilde{W}(z^u(k,X),k,X) - I_H - \widetilde{W}(z^u(k,X),1,X) = 0$. We are paying attention to equilibrium in which $z^e(1,X) < z^u(k,X)$ so that there are both types of capacities operating in the market. In other words, assume $\Pi_e - \widetilde{W}(z^u(k,X),1,X) < 0$.

$$\begin{aligned} \widetilde{W}(z^u(k,X^g),1,X^g) &= \pi(z^u(k,X^g),1,X^g) + E\left[\widetilde{\beta}(X^g,X)\,(1-\delta)\,\widetilde{W}(z^u(k,X^g),1,X^g)\right] \\ &\quad + \widetilde{\beta}(X^g,X^b)(1-\gamma_g)\,(1-\delta)\left[\widetilde{W}(z^u(k,X^g),1,X^b) - \widetilde{W}(z^u(k,X^g),1,X^g) - I_H\right] \end{aligned} \tag{12}$$

To evaluate the last term, we can write $\widetilde{W}(z^u(k,X^g),1,X^b)$ as

$$\begin{aligned} \widetilde{W}(z^u(k,X^g),1,X^b) &= \pi(z^u(k,X^g),1,X^b) + E\left[\widetilde{\beta}(X^b,X)\,(1-\delta)\,\widetilde{W}(z^u(k,X^g),1,X^g)\right] \\ &\quad + \widetilde{\beta}(X^b,X^b)\gamma_b(1-\lambda)\left[\widetilde{W}(z^u(k,X^g),1,X^b) - \widetilde{W}(z^u(k,X^g),1,X^g) - I_H\right] \end{aligned} \tag{13}$$

Substracting (12) from (13) one obtains

$$\begin{aligned} &\widetilde{W}(z^u(k,X^g),1,X^b) - \widetilde{W}(z^u(k,X^g),1,X^g) = \pi(z^u(k,X^g),1,X^b) - \pi(z^u(k,X^g),1,X^g) + \\ &+E\left[\left(\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right)(1-\delta)\,\widetilde{W}(z^u(k,X^g),1,X^g)\right] \\ &+\left[\widetilde{\beta}(X^b,X^b)\gamma_b - \widetilde{\beta}(X^g,X^b)(1-\gamma_g)\right](1-\delta)\left[\widetilde{W}(z^u(k,X^g),1,X^b) - \widetilde{W}(z^u(k,X^g),1,X^g) - I_H\right] \end{aligned}$$

Rewriting

$$\begin{aligned} &\left[\widetilde{W}(z^u(k,X^g),1,X^b) - \widetilde{W}(z^u(k,X^g),1,X^g)\right](1-\kappa_{bb}+\kappa_{gb}) = \pi(z^u(k,X^g),1,X^b) - \pi(z^u(k,X^g),1, \\ &+E\left[\left(\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right)(1-\lambda)\,\widetilde{W}(z^u(k,X^g),1,X^g)\right] \\ &+(\kappa_{gb}-\kappa_{bb})\,I_H \end{aligned}$$

Replacing in (12)

$$\begin{aligned} &\widetilde{W}(z^u(k,X^g),1,X^g)\left\{1 - E\left[\widetilde{\beta}(X^g,X)\,(1-\delta)\right]\right\} = \pi(z^u(k,X^g),1,X^g) + \\ &+\frac{\kappa_{gb}}{1-\kappa_{bb}+\kappa_{gb}}\left[\begin{array}{c} \left[\pi(z^u(k,X^g),1,X^b) - \pi(z^u(k,X^g),1,X^g)\right]\chi_{\left(\pi_{bg}>\pi_g\right)} + (\kappa_{gb}-\kappa_{bb})\,I_H \\ +E\left[\left(\widetilde{\beta}(X^b,X) - \widetilde{\beta}(X^g,X)\right)(1-\lambda)\,\widetilde{W}(z^u(k,X^g),1,X^g)\right] \end{array}\right] \end{aligned}$$

where $\pi_{bg} \equiv \pi(z^u(k, X^g), 1, X^b)$ and $\pi_g \equiv \pi(z^u(k, X^g), 1, X^g)$ and as before, if shocks are i.i.d so that $(1-\gamma_g) = \gamma_b$ (and viceversa), the expectation term in the RHS drops. Otherwise

$$\widetilde{W}(z^u(k, X^g), 1, X^g) \left\{1 - \kappa_{gg} - \kappa_{gb}(\frac{1 - \kappa_{gg} + \kappa_{bg}}{1 - \kappa_{bb} + \kappa_{gb}})\right\} = \pi(z^u(k, X^g), 1, X^g) +$$
$$+\frac{\kappa_{gb}}{1 - \kappa_{bb} + \kappa_{gb}} \left\{\left[\pi(z^u(k, X^g), 1, X^b) - \pi(z^u(k, X^g), 1, X^g)\right] \chi_{\left(\pi_{bg} > \pi_g\right)} + \left(\kappa_{gb} - \kappa_{bb}\right) I_H\right\}$$

and whenever the shocks are i.i.d $(\frac{1-\kappa_{gg}-\kappa_{gb}}{1-\kappa_{bb}+\kappa_{gb}}) = 1$

To determine the threshold, we equalize the return to upgrade with the value of waiting an additional period and having the option readily available again.

$$\widetilde{W}(z^u(k, X), k, X) - I_H = \widetilde{W}(z^u(k, X), 1, X)$$

To solve it, we need to compute $\widetilde{W}(z^u(k, X), k, X)$ for each possible exogenous aggregate state.

After some algebra one can show that $\widetilde{W}(z^u(k, X^g), k, X^g)$ reads

$$\widetilde{W}(z, k, X^g) = \vartheta(X^g) \left[\pi(z, k, X^g) + \pi(z, k, X^b) \frac{\kappa_{gb}}{1 - \kappa_{bb}}\right]$$

$\vartheta(X^g) = \frac{1-\kappa_{bb}}{1-\kappa_{gg}-\kappa_{bb}-\kappa_{gg}\kappa_{bb}-\kappa_{bg}\kappa_{gb}}$

We can follow the same procedure to compute the upgrade threshold in the bad state. The value of a small capacity firm at the threshold reads

$$\widetilde{W}(z^u(k, X^b), 1, X^b) \left\{1 - \kappa_{bb} - \kappa_{bg}(\frac{1 - \kappa_{bb} + \kappa_{gb}}{1 - \kappa_{gg} + \kappa_{bg}})\right\} = \pi(z^u(k, X^b), 1, X^b) +$$
$$+\frac{\kappa_{bg}}{1 - \kappa_{gg} + \kappa_{bg}} \left[\left[\pi(z^u(k, X^b), 1, X^g) - \pi(z^u(k, X^b), 1, X^b)\right] \chi_{\left(\pi_{gb} > \pi_b\right)} + \left(\kappa_{bg} - \kappa_{gg}\right) I_H\right]$$

where $\pi_{gb} \equiv \pi(z^u(k, X^b), 1, X^g)$ and $\pi_b \equiv \pi(z^u(k, X^b), 1, X^b)$.

And the value of the high capacity firm at the threshold reads

$$\widetilde{W}(z, k, X^b) = \vartheta(X^b) \left[\pi(z, k, X^b) + \pi(z, k, X^g) \frac{\kappa_{bg}}{1 - \kappa_{gg}}\right]$$

9.4 Features of the Solution

Proposition 16 *Continuation Values* $\widetilde{W}$ *are monotonic increasing in idiosyncratic productivity,* z*.*

Proof. First notice that $\pi(x_t, X_t)$ is bounded and continuous in z. (Replace the optimal factor demands in the profit function).

Second, let $W^*(\ x, X)$ be the unique fixed point to the operator T,

$$T(W(x, X_t)) = Max\left\{\Pi_e, \pi(x, X_t) + E_t\left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\,W(x, X_{t+1})\right)\right\}$$

We first show first that $W^*(\ x, X)$ is non-decreasing in z.

Let $C(Z)$ be the set of continuous bounded functions in z, and let $C'(Z)$ a closed subspace of non-decreasing functions. Take $W \in C(Z)$ and $z_1 < z_2$. then

$$\begin{aligned} T(W(z_1, \psi^j, X_t)) &= Max\left\{\Pi_e, \pi(z_1, \psi^j, X_t) + E_t\left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\,W(x, X_{t+1})\right)\right\} \\ &\leq Max\left\{\Pi_e, \pi(z_2, \psi^j, X_t) + E_t\left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\,W(x, X_{t+1})\right)\right\} \\ &= T(W(z_2, \psi^j, X_t)) \end{aligned}$$

so that $T(C'(Z)) \subseteq C'(Z)$. Hence by the Contraction Mapping Theorem $W^* \in C'(Z)$.

Now, we want to prove that for each ψ^j, X, the function $\widetilde{W}\ (z,\psi^j, X_t)$ is strictly increasing in z. Note that the expectation operator in the last term of the previous equation defined over the aggregate of the economy and independent of the productivity of the firm except through the function W. Let W be any continuous non decreasing bounded function in z, and take $z_1 < z_2$

$$\begin{aligned} \widetilde{W}(z_1, \psi^j, X_t) &= \pi(z_1, \psi^j, X_t) + E_t\left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\,W^*(z_1, \psi^j, X_{t+1})\right) \\ &< \pi(z_2, \psi^j, X_t) + E_t\left(\widetilde{\beta}_{t+1}(X_t, X_{t+1})\,(1-\delta)\,W^*(z_2, \psi^j, X_{t+1})\right) \\ &= \widetilde{W}(z_2, \psi^j, X_t) \end{aligned}$$

which proves the claim. ■

$$\pi(x_t, X_t) = \Lambda(X_t) \left(\frac{z\psi^j}{MPK_t^{\frac{\alpha}{\rho}}} \right)^{\frac{\rho}{1-\rho}}$$

where $\Lambda(X_t) = \left(\frac{\sum_{j=L,H} \int (\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\alpha}})^{\frac{\rho}{1-\rho}} dv_t^j}{\sum_{j=L,H} \int (\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\frac{1-(1-\alpha)\rho}{\rho}}})^{\frac{\rho}{1-\rho}} dv_t^j} \right)^{\alpha\rho} \frac{1}{\left(\sum_{j=L,H} \int (\frac{\psi_i^j z_i}{\left(MPK_t^i\right)^{\alpha}})^{\frac{\rho}{1-\rho}} dv_t^j\right)^{\rho}} (1-\rho) Y_t^{1-\rho} s_t^{\rho} K_t^{\alpha\rho}.$

Proposition 17 *The optimal allocation satisfies*

1. $z^e(\psi_L, X_t) > z^e(\psi_H, X_t)$
2. $\frac{\partial z^e(\psi^j, X_t)}{\partial r_t} \geq 0$
3. $z^u(\psi_H, X_t) > z^e(\psi_H, X_t)$
4. M^e *is procyclical*

1. **Proof.** Note first that the profit function $\pi(x_t, X_t) = \Lambda(X_t) \left(\frac{z\psi^j}{MPK_t^{\frac{\alpha}{\rho}}} \right)^{\frac{\rho}{1-\rho}}$ is monotonic in the firm idiosyncratic productivity and the technology shifter. We have proved that firms' continuation values are also monotonic. Hence $W(z, \psi_H, X_t) > W(z, \psi_L, X_t)$ for all z. The scrap value of the firms is constant and independent of the technology of the firm. The optimality condition for the exit thresholds equalizes the value of the firm to the scrap value and therefore, $z^e(\psi_L, X_t) > z^e(\psi_H, X_t)$. ■

   Proof. Note that the profit function is such that $\frac{\partial \pi(x_t, X_t)}{\partial r_t} < 0$. Following the same strategy than for the monotonicity in idiosyncratic productivity one can show that $W(z, \psi^j, X_t)$ is non increasing in the cost of capital and the continuation value $\widetilde{W}(z, \psi^j, X_t)$ is decreasing in r_t. As in the previous proof, the result follows from the optimality condition for the exit threshold. ■

Proof. Given that upgrades in technology are costly and the scrap value at exit is independent of the technology operated by the firm. It cannot be optimal to upgrade and exit immediately. For this strategy, I_H units of goods are paid , while exiting while running the low minimum capacity technology yields the same scrap value and no associated cost. ■

Proof. $\widetilde{W_t}(z, \psi_L, X_t)$ is increasing in the aggregate state of technology s. From the free entry condition the result follows. ■

The fact that the scrap value of the firm is independent of the cost capital and the idiosyncratic characteristics of the firm is critical to prove the previous results.