Explaining Productivity Puzzles with Frictional Firm Entry: Endogenous Markups Versus Dynamic Reallocation

Anthony Savagar*

Latest draft click here†

November 12, 2015

Abstract

I present a theory of firm entry and exit in the business cycle that links short-run productivity overshooting to long-run persistence, a dynamic observed in contemporary *'productivity puzzles'*. The theory emphasizes two mechanisms: (1) slow firm entry/exit and (2) firm pricing that reflects the number of competitors in the market. Given these mechanisms, economic contraction causes a short-run exacerbated fall in productivity (overshooting) because the negative shock is absorbed by incumbents due to slow exit responses. This weakens incumbents' returns to scale, thus worsening productivity. However, the productivity overshooting recedes over time as firms exit which *dynamically reallocates* resources among incumbents, reviving the remainders returns to scale and thus productivity. This process of exit consolidating the market is not purely beneficial for productivity because the remaining firms face fewer competitors and thus charge higher markups which damages productivity. Therefore despite some reversion from the initial fall, there is a long-run persistent negative

*Thanks to the following people for helpful comments: Huw Dixon, Patrick Minford, Vivien Lewis, Tommaso Monacelli, Tom Holden, Leo Kaas, Frédéric Dufourt, Rudolph Dos Santos Ferreira, Paulo Brito, Luís Costa, Harald Uhlig, Akos Valentinyi, Stephen Millard, Martin Kaae Jensen, David Collie. Thanks to funders: ESRC, Julian Hodge Institute of Macroeconomics and Royal Economic Society Fellowship. I appreciate the hospitality of KU Leuven.

†asavagar@gmail.com. Working draft, please do not cite without permission. Project files are available on https://github.com/asavagar and Ipython notebooks replicating symbolic and numerical results are available on nbviewer.

effect on productivity due to higher markups responding to the fall in number of firms. To analyze the trade-off between productivity improving dynamic reallocation and productivity degrading endogenous markups, I develop a continuous time, analytically tractable DGE model. The main mechanisms are dynamic entry so firms are slow to respond causing initial overshooting, and endogenous markups so pricing behaviour depends on the number of competitors firms face.

JEL: E32, D21, D43, L13, C62, Endogenous markups, Entry, Endogenous Productivity, Imperfect product markets, dynamical systems

The paper proposes a business cycle theory in which firm entry and exit cause endogenous short-run and long-run productivity movements. Interest in endogenous productivity over the business cycle is high in light of Great Recession *productivity puzzles*[1]. The puzzles describe exacerbated productivity falls with weak recovery, and are prominent in several European countries shown in figure 1a for labor productivity[2]. The problem is especially pronounced in the UK, and empirical studies (Barnett et al. 2014) find that up to half of the shortfall in UK labor productivity relative to pre-crisis trend arose because of impaired resource allocation and unusually high firm survival rates[3]. This evidence emphasizes the importance of firm dynamics in explaining macroeconomic productivity, but traditional macroeconomic theory nullifies entry by assuming that the number of firms in an economy adjusts instantaneously to arbitrage profits. If entry is instantaneous, it can only affect productivity through an immediate change in the number of competitors which affects pricing markups, but it ignores the short-run effect of sluggish entry reallocating resources as firms adjust to arbitrage profit. In this paper I analyze the new trade-off that emerges when noninstantaneous entry is combined with competitive endogenous markups.

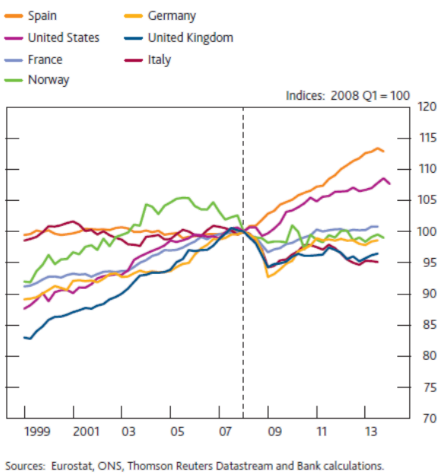
The main result is a theory to explain that shocks initially exacerbate productivity movements but the exacerbation relinquishes as firm entry/exit adjusts. Crucially productivity never regains long-run underlying productivity because of structural changes in competition due to long-run changes in the number of incumbents. To be clear, 'entry' is net entry, so when negative it is exit. Therefore entry and exit the same process–they cannot arise

[1] The term has been used extensively in the media and academia e.g. The Productivity Puzzle Under the Bonnet, The Economist, May 30, 2015; Budget 2015: How do you solve the 'productivity puzzle'?, BBC News, July 8, 2015

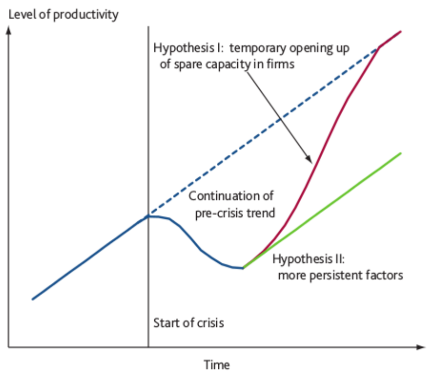
[2] The source of both figures is Barnett et al. 2014 and they have been reproduced in the press e.g. Emily Cadman's UK Productivity Puzzle: The Bank of England's Answers, Financial Times, August 14, 2014.

[3] Goodridge, Haskel, and Wallis 2014 show that accounting for labor and capital still leaves a TFP puzzle. I focus on TFP.

together. A contractionary shock will solely cause exit (negative net entry); an expansionary shock will solely cause entry (positive net entry). Hence the theory is a general explanation of endogenous productivity over the business cycle, with initially exacerbated and persistent positive productivity effects associated with entry in expansion and exacerbated and persistent negative productivity effects associated with exit in contraction. Although general, Great Recession productivity puzzles provide a contemporary view of the theory since they depict a short-run exacerbated fall in productivity followed by some persistence due to structural factors. Therefore my results provide a theory that combines the two hypotheses posed by the Bank of England in figure 1b. I demonstrate that a negative shock to the economy, modeled as a supply-side TFP shock, is first absorbed by incumbent firms because exit cannot arise initially. Therefore productivity falls drastically as the incumbents output falls and they suffer worse returns to scale (hypothesis I in figure 1b). Lower output per firms causes negative profits which leads to exit. As exit occurs productivity improves because resources are reallocated among incumbents and better returns to scale improves productivty, as shown by the hypothesis I reversion. However, this consolidation of resources among fewer firms reduces the competitive pressure on those who remain allowing them to charge higher markups. Higher markups mean each unit sold generates more revenue so that in a long-run zero profit equilibrium firms can produce less to cover their fixed cost of production. By choosing to produce less their scale suffers which creates an offsetting negative productivity effect that persists in equilibrium and hence links to hypothesis II in figure 1b.



(a) Cross-country Labor Productivity



(b) Productivity Puzzle Theory

I develop a tractable model of dynamic (endogenous) firm entry in the macroeconomy with imperfectly competitive product markets that cause endogenous markups. Dynamic entry means that firms slowly adjust to arbitrage profits, so short-run profits are nonzero. This entry friction arises because a *congestion effect* raises sunk entry costs as entry increases[4]. Imperfect competition creates a markup of factor prices above their marginal products, and the markup is endogenous because it depends on the number of firms. The relationship is negative and occurs because firms are large in their industry so they strategically interact under Cournot competition. With this model setup, I analyze the trade-off between *endogenous markups* and *dynamic reallocation*. Endogenous markups cause entry to increase productivity and exit to decrease productivity. Dynamic reallocation causes entry to decrease productivity and exit to increase productivity. For example, with endogenous markups exit (entry) weakens (strengthens) competition which raises (lowers) markups, thus decreases (increases) productivity. In opposition, dynamic reallocation means exit (entry) concentrates (dissipates) resources thus increasing (decreasing) incumbents' scale and therefore productivity. Dynamic reallocation emphasises not the amount of resources, but their division among firms. And entry determines this division. I analyze measured productivity, which is an adjusted measure of total factor productivity (TFP).

The model demonstrates procyclical profits, entry, employment and productivity, whereas markups are countercyclical. For example, a positive shock to technology is initially borne by incumbents who raise their output whilst entry is inert in the short-run. Through greater scale incumbents' productivity increases. However, by raising output incumbents accrue monopoly profits, these non-zero profits incentivise potential firms to begin entering. Entry reallocates resources and reduces output per firm which diminishes scale and therefore productivity. The influx in entry diminishes profits and through congestion raises the sunk entry cost which slows the rate of entry. Eventually the profits from incumbency arbitrage to zero so entry ceases and zero incumbency profits are balanced with zero sunk entry costs because there is no congestion. The long-run effect of a rise in the number of incumbents is that competition in the market is fiercer, so firms charge a lower markup. In order to cover fixed costs, firms with lower markups must raise revenue by increasing output, therefore in zero-profit (free-entry) equilibrium firm scale is increased which means there is a long-run permanent effect on productivity. In summary, the positive shock increases output, profit, employment (an input) and entry, whereas markups decrease because entry increases competition.

[4]A familiar notion in industrial organization (IO) theory e.g. Ericson and Pakes 1995

Formally the model follows a Ramsey-Cass-Koopmans setup. There is endogenous labor and capital, and the novel additions are firm entry and endogenous markups. A representative household chooses its consumption and labor exertion, but the household is limited by a budget which consists of labor income and investment income. Investment income consists of returns on capital and returns on firm ownership (firm profit). The return on capital is the economy's risk-free rate, which consequently determines the opportunity cost of investing in a firm. This balance between paying a cost to setup a firm and investing that cost at the market rate binds firm entry. It is a dynamic condition because sunk costs depend on the number of entering firms (congestion effect). Hence in free entry equilibrium[5] profits are zero, so a household is indifferent between creating a firm or investing that sunk cost at the risk free rate. For example, if the value of incumbency exceeds the risk free alternative, then there will be entry. Consequently congestion will stifle start-ups and entry will slow. On the firm side of the economy there is imperfect competition and generalized returns to scale (U-shaped cost curves). All firms produce with the same production function (firms are symmetric) which has a fixed overhead cost and nondecreasing marginal cost. The fixed overhead allows for imperfect competition which causes pricing markups. Firms are aggregated across two levels. The lowest level of aggregation is the firm level, and aggregating firms gives the industry level. The macroeconomy is the aggregate across all industries. I focus on symmetric equilibria so an industry is representative of the whole economy. Firms have price setting power within their industry, but are small in the aggregate economy. The influence of a firm on industry price causes endogenous markups[6]. Within an industry firms strategically interact with Cournot competition, so they maximise profits by choosing output to produce. This output choice is influenced by the number of competitors in the industry. When there are more competitors demand functions reflect a higher elasticity of market demand and therefore weaker markup setting power.

The model economy includes three core assumptions 1) endogenous entry 2) returns to scale 3) endogenous markups. The counterfactual of each assumption emphasizes its importance. First, in the absence of endogenous entry there is instantaneous free entry[7]. This counterfactual implies that there is no short-run productivity effect as incumbent firms bear shocks.Second, in the absence of increasing returns to scale, returns to scale are constant. This counterfactual makes entry impotent because firms produce at the same pro-

[5] The long-run equilibrium when firms have freely entered to arbitrage positive profits.

[6] If a firm were small in its industry, markups would be fixed as in the status-quo Dixit and Stiglitz 1977 case.

[7] This is a limiting case of my model, as is the other extreme a fixed number of firms.

ductivity regardless of size. Third, in the absence of endogenous markups, markups are fixed. This counterfactual implies there is no persistent effect on productivity because firms do not alter their markups.

A technically precise description of the model is that the economy is defined by a four dimensional dynamical system consisting of state variables capital, number of firms, consumption and entry. The fixed point (steady state) of the system is hyperbolic, and a saddle equilibria, so that there are two unstable eigenvalues and two stable eigenvalues. Restricting the unstable eigenvalues reduces the model to a two-dimensional stable manifold in capital and number of firms. So, consumption and entry are jump variables that may be immediately determined as functions of these two variables. Hence at any point in time, the economy is defined by its capital and number of firms. Whence the division of capital among firms is the crucial relationship. By solving the model within a neighbourhood of the steady state the general trajectories of the states are known and other variables behaviour follows. This allows derivations of the main endogenous productivity argument by showing that number of firms are slow to respond. An important feature of the model is that the stable manifold has dimension greater than one which is sufficient to open the complex plane thus allowing nonmonotone dynamics as the economy transition to steady state [8].

Related Literature This paper links Etro and Colciago 2010[9] to Jaimovich and Floetotto 2008. The first paper includes sluggish firm entry and endogenous markups, but does not discuss productivity. Their contribution is to improve business cycle moment-matching using Cournot and Bertrand strategic interactions; I use Cournot which the authors advocate. The second paper has endogenous markups and analyzes productivity, but firm entry is instantaneous. Their contribution is to explain the productivity effect of instantaneous entry on markups. This is equivalent to the long-run effect that causes productivity persistence in my paper. My link combines endogenous entry with endogenous markups to explain productivity over the business cycle. The result is that endogenous entry distinguishes short-run productivity dynamics from long-run productivity dynamics.

The endogenous entry setup of this paper follows Datta and Dixon 2002 which is close to industrial organization literature by Das and Das 1997. Importantly this differs from most recent endogenous entry literature that

[8] Only sink equilibria (models with indeterminacy) allow for nonmonotonicity in the single variable case (Jaimovich 2007 or Dufourt, Nishimura, and Venditti 2013). This paper has a saddle not a sink.

[9] Etro and Colciago 2010 also note that Cournot competition causes inefficiency through excess entry. Etro 2009 provides an excellent survey of macroeconomic models with endogenous entry and endogenous market structures.

uses Bilbiie, Ghironi, and Melitz 2012 (BGM)[10]. However, the interpretation of the two approaches is analogous. Both endogenous entry formulation reduce to an arbitrage condition that equates sunk entry costs to incumbency profits. A strength of the Datta and Dixon 2002 formulation is that dynamics stem from endogenous sunk costs, rather than fixed sunk costs in BGM. These endogenous sunk costs are called congestion effects since they increase as number of entrants (congestion) increases. It is then a lemma that sunk entry costs equate to profit from incumbency, rather than in BGM where this is assumed. The BGM setup is influential in discrete time, simulation exercises[11], whereas the model in this paper is continuous time and analytically tractable. BGM distinguish entry from exit (exit is exogenous), whereas this paper treats them as symmetric. Entry measures the change in the number of incumbent firms, so negative entry is exit. Lewis 2009, Lewis and Poilly 2012, Lewis and Stevens 2015 and Berentsen and Waller 2009 all recognise the importance of congestion effects in macroeconomic models with entry.

An important distinction of this paper is its focus on qualitative dynamical systems, rather than quantitative simulations in the afforementioned works. This follows Brito and Dixon 2013. Rather than productivity, their focus is on theorems to show that firm entry is sufficient for nonmonotone responses to fiscal shocks. Excluding imperfect competition removes the vital mechanism for generating increasing returns to scale that are necessary for productivity dynamics. This mechanism is present in Aloi and Dixon 2003 who use firm entry to explain productivity in an open economy without capital or endogenous markups. This mechanism between imperfect competition, increasing returns to scale and productivity is an established explanation for procyclical productivity over the business cycle (Hall 1989, Hall 1987, Caballero and Lyons 1992).

There are two competing formulations of endogenous markups. There is a supply-side approach, used in this paper, Etro and Colciago 2010, Jaimovich 2007, and Jaimovich and Floetotto 2008. There is a demand-side approach used by Bilbiie, Ghironi, and Melitz 2012. The supply-side approach relies on firms strategically interacting which affects market demand. The demand-side approach relies on consumers' elasticity of substitution varying as product variety changes with entry. Lewis and Poilly 2012 compare the methods. Empirical business cycle literature shows many examples of countercylical markups. A cornerstone work is Bils 1987, and there are many contributions by Rotemberg and Woodford surveyed in Rotemberg and Woodford 1999.

[10]For example, Lewis and Poilly 2012 and Etro and Colciago 2010.

[11]And not limited to macroeconomics. Examples include Loualiche 2014 in finance, Peters 2013 in growth and Hamano and Zanetti 2014 in macroeconomics.

These traditional explanations of countercyclical markups rely on price stickiness. Whereas, the study of entry provides a new factor to enrichen markup countercylicality. The idea stems from the ubiquity of the relationship in empirical IO. For example Campbell and Hopenhayn 2005 find a negative correlation between markups and entry in many sectors of the US economy. In macroeconomics, Portier 1995 shows that entry is procylical and markups countercyclical over the French business cycle. Other empirical features that relate to this paper are procyclical productivity Rotemberg and Summers 1990, and procyclical net business formation Bergin and Corsetti 2008.

Roadmap – Section 1 explains the intuition behind the model. Section 2 outlines a model of firm entry in the macroeconomy where firms compete with strategic interactions. Section 3 begins analysis by explaining how the competition effect of entrants reducing markups affects factor prices and profits. Section 4 investigates static outcomes showing that long-run output and productivity are endogenous since they depend on the number of operating firms. Section 5 is the main result which presents a theorem to explain productivity puzzles, where productivity overshoots on the impact of a shock, then relinquishes but leaving some persistence.

1 Intuition of Excess Capacity with Short-run and Long-run Capacity Utilization

Before developing a complex dynamic model with endogenous entry, imperfect competition and endogenous markups, a simple diagram can explain the intuition of how entry causes endogenous and persistent productivity dynamics. Figure 2 shows the cost curves and equilibria of a firm with increasing marginal costs and a U-shaped average cost due to a fixed overhead cost ϕ. Under imperfect competition a firm produces y_{IC}[12] which is less than the perfect competition outcome y_{PC}, which is also the efficient outcome as it minimizes costs. The difference between y_{IC} and y_{PC} is excess capacity[13] (labelled), and utilizing excess capacity lowers costs which improves firm productivity and in turn aggregate productivity. With an entry mechanism the underproduction of each firm in imperfect competition corresponds to excess entry. This means ther are 'too many' firms each underproducing, so a more

[12] This is the long-run Chamberlin-Robinson equilibrium in which marginal revenue equals marginal cost and profits are zero. I omit the MR and MC curves for clarity, and I assume the curves have fully shifted following any shock.

[13] Macroeconomists should note this definition of excess capacity which follows Vives 1999. It is distinct from *capital utilization* or any form of input intensity.

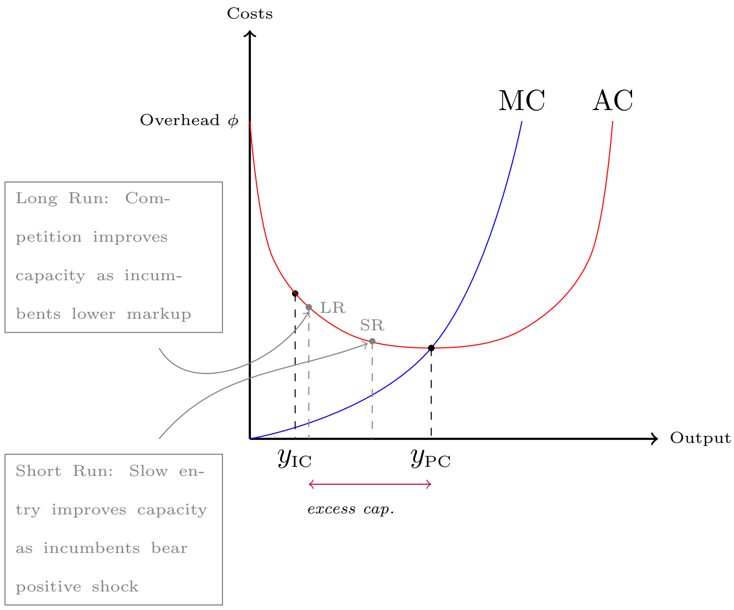


Figure 2: Excess Capacity, Short-run and Long-run Utilization

efficient outcome is fewer firms but each producing more, hence with y_{IC} closer to minimium cost y_{PC}.

If there is a positive shock to the economy, and entry is slow (endogenous), then that shock is initially borne by the incumbents so they utilize capacity and costs lower. A move from y_{IC} to point SR in the diagram. This indicates an immediate increase in productivity in aggregate. But in this new position, incumbents earn monpoly profits which attracts entrants. Over time entrants move into the market, gradually reducing the capacity of incumbents so excess capacity rises back towards the initial level, finally halting at a position like LR where profits are zero again. This mechanism corresponds to the gradual amelioration of the initial boost in productivity as firms adjust. The final part of the story is most important, because it explains why although the initial drastic effect subsides there is still some long-run effect on productivity, so output per firm returns to LR rather than initial y_{IC}. As firms enter to arbitrage profit to zero they must now each produce more $y_{LR} > y_{IC}$ because the positive shock encouraged entry (raised the number of competitors) which put downward pressure on markups. Therefore in long-run zero profit equilibrium firms charge lower markups than in the initial pre-shock position. Consequently each incumbent must raise revenue by increasing output to cover the costs of production ϕ and attain zero profit. So overall there is a small fall in costs per firm due to capacity utilization, and thus a small but persistent improvement in productivity after the initial positive shock.

2 Endogenous Entry Model with Imperfect Competition

The model follows a Ramsey-Cass-Koopmans setup. Additions are imperfect competition, firm entry, endogenous markups and capital accumulation. The model is deterministic, and labour is endogenous. There are two state variables: capital and number of firms $(K, n) \in \mathbb{M} \subseteq \mathbb{R}^2$, where $\mathbb{M}$ is the state space of the control problem that later forms a subset of the general dynamical system state (or phase) space. I solve the model as a decentralised equilibrium because imperfect competition distorts the optimising behaviour of the firm.

Notation and Terminology Y_x denotes the derivative of Y with respect to X, except when $X = t$ which denotes time dependence. For clarity I usually omit the (t) notation that denotes time dependence in ordinary differential equations (ODEs). To be clear, the primitive endogeneous model variables are $C(t), e(t), K(t), n(t)$, defined later. They are the state variables of the four dimensional dynamical system which forms the model economy. The four states depend on time and therefore so do functions of them $L(t), r(t), w(t), \pi(t), Y(t), y(t), \mu(t), \mathcal{P}(t), \Pi(t), Z(t)$. Time dependence is irrelevant in steady state, which I denote with an asterix Y^*. Also for clarity, I often suppress function domains. For example, after first introduction $F : K \times L \to \mathbb{R}$ is written F rather than $F(K, L)$.

2.1 Firm

In the economy there is a continuum of sectors of measure one. In each industry, there is a finite number of intermediate firms that each produce a homogenenous good. Since the goods are homogeneous, they are perfectly substitutable in the production of an industry good. However, at the next level of aggregation, industry goods are imperfect substitutes for eachother when aggregated into a final good. Entry and exit of firms into existing sectors occurs until profits are zero. This does not happen immediately but occurs in the long run. This is known as the free entry equilibrium. In the short-run profits will diverge from zero as they are arbitraged by entrants. Perfect factor markets mean that each firm faces the same price w for labor and r for capital, and the result is that aggregate capital and labour are divided equally among firms $k = \frac{K}{n}$ and $l = \frac{L}{n}$. A lowercase letter denotes *per firm*.

A fixed cost in production allows firms to compete under imperfect competition in the product market. Strategic interactions occur under imperfect competiton because firms are large in their industry so can influence industry price. This is why markups are endogenous (depend on number of incumbent firms), rather than fixed in the traditional Dixit-Stiglitz case where firms are small therefore do not affect industry price level. I focus on Cournot competition so firms strategically interact through their choice of output to maximise profits given the behaviour of others. The form of strategic interaction determines the markup of factor price above marginal cost. Specifically I focus on the level of factor price markup above the factor marginal product. That is the markup of wage above marginal product of labor and interest rate above marginal product of capital.

Final output Y is produced by a competitive firm using the output of a continuum of industries (*aka* intermediate goods or sectors) Q_j for $j \in [0, 1]$ as inputs in a CES production function with constant elasticity of substitution $\theta_I \in (0, \infty)$.

$$Y(t) = \left(\int_0^1 Q_j(t)^{\frac{\theta_I - 1}{\theta_I}} dj \right)^{\frac{\theta_I}{\theta_I - 1}}, \quad \theta_I \in (0, 1) \tag{1}$$

Cost minimization leads to conditional demand for industry j

$$Q_j(t) = \left(\frac{P_j}{P} \right)^{-\theta_I} Y \tag{2}$$

Thus the inverse demand function is $P_j = \left(\frac{Q_j}{Y} \right)^{-\frac{1}{\theta_I}} P$. Substituting the conditional industry demand (2) into the aggregate production function (1) gives the aggregate price index

$$P = \left(\int_0^1 P_j^{1-\theta_I} dj \right)^{\frac{1}{1-\theta_I}} \tag{3}$$

Notice that perfect competition in the final goods market requires equality of price and marginal cost P.

Assumption 1 (Firm Production with U-shaped Average Cost Curve)**.** *Firms are symmetric, so each has the same production technology. The ith firm in the jth industry produces output:*

$$y_{j,i}(t) := \max\{AF(k_{j,i}(t), l_{j,i}(t)) - \phi, 0\} \tag{4}$$

where $F : \mathbb{R}_+^2 \supseteq (k, l) \to \mathbb{R}_+$ is a firm production function with continuous partial derivatives which is homogenous of degree $\nu \in (0, 1)$ (hod-ν) on the *open cone $\mathbb{R}^2_+$, and $\phi \in \mathbb{R}_+$ a fixed cost denominated in output. The Hessian matrix of F has a symmetric main diagonal (Young's theorem), negative mixed derivatives (off-diagonal), and its determinant is positive so the concavity properties are*

$$F_{kl} = F_{lk} > 0, \qquad F_{kk}, F_{ll} < 0, \qquad F_{kk}F_{ll} - F_{kl}^2 > 0$$

Inada's conditions hold so that marginal products of capital and labor are strictly positive which rules out corner solutions.

$$F_k, F_l > 0$$

Although we shall focus on the case of U-shaped average costs many of our calculations hold without loss of generality for several cases:

- $\phi > 0, \nu \in (0,1)$ U-Shaped average cost and increasing marginal cost curve compatible with imperfect and perfect competition.
- $\phi = 0, \nu = 1$ Constant returns and no fixed cost so globally constant returns to scale. Average cost and marginal cost are equivalent.
- $\phi > 0, \nu = 1$ A fixed cost with constant marginal cost leads to globally decreasing average cost.
- $\phi \geqslant 0$ and $\nu \in (1, \infty)$ Both average and marginal costs are increasing, so there are globally increasing returns to scale. The extent to which ν exceeds 1 is bounded. Proof below. Equilibrium will only exist with imperfect competition.

Notice that we view number of firms as a factor of production $F(k,l) = F\left(\frac{K}{n}, \frac{L}{n}\right) = n^{-\nu}F(K,L)$. It is essentially a measure of organization which captures how resources are divided. The production function with a fixed cost and decreasing returns to scale cause a U-shaped average cost curve. Decreasing returns to scale arise because the variable production function $F : \mathbb{R}^2_+ \supseteq (k,l) \to \mathbb{R}_+$ is convex, $\nu \in (0,1)$, in capital and labour which causes increasing marginal cost. The fixed cost ϕ creates a nonconvexity which prevents some firms producing because an active firm must sell at least enough to cover the fixed cost. The fixed cost occurs each period, and is different to the sunk entry cost which is paid once to enter (see Entry Section 2.1.3)[14]. $A \in [1, \infty)$ is a scale parameter reflecting the productivity

[14] As in Jaimovich 2007 and Rotemberg and Woodford 1996 the role of this parameter is to reproduce the apparent absence of pure profits despite market power. It allows zero profits in the presence of market power.

level. It may be interpreted as total factor productivity (TFP). In Section 4.1.1 we derive *measured productivity* which is a function of TFP that captures the fixed cost and returns to scale effect. Since the average cost curve is U-shaped, there is an efficient level of production at minimum average cost, where average cost and marginal cost intersect. This is the Walrasian outcome that would arise under perfect competition.

2.1.1 Strategic Interactions and Endogenous Markups

Within each industry j there is Cournot monopolisitc competition among a set $\mathcal{I}(j)$ of $n(j) \in (1, \infty)$ firms. So the representative ith firm in industry j chooses output to maximise profits subject to the inverse demand function implicit in (2) and the quantities $y_{j,i'}$ supplied by other firms $i' \in \mathcal{I}(j)$ß. It takes as given the quantitiy of final output Y produced by the competitive sector, the aggregate price level P of the intermediate sector (it cannot influence this price level) and the factor market prices w and r. Therefore it solves

$$\max_{(y(j,i),l_i,k_i)} P\left(\frac{(y_i + \sum_{i'} y_{i'})}{Y}\right)^{-\frac{1}{\theta_I}} y_i - rk_i - wl_i \qquad (5)$$

$$\text{s.t.} \quad y_{j,i}(t) \leqslant AF(k_{j,i}(t), l_{j,i}(t)) - \phi \qquad (6)$$

Each firm's technology is symmetric with respect to intermediate inputs that are shared equally due to perfect factor markets $k = \frac{K}{n}$ and $l = \frac{L}{n}$. The result is a symmetric equilibrium outcome, so we can drop i, j indexes and focus on a single representative industry as the whole economy. Under symmetric equilibrium the first order conditions of the firms profit maximising problem lead to a markup $\mu(n(t)) \in (1, \infty)$ of price above marginal cost.

$$\mu(n(t)) = \frac{\theta_I n(t)}{\theta_I n(t) - 1} \qquad (7)$$

The markup is endogenous and decreasing in the number of firms $\mu_n = -\frac{\theta_I}{(\theta_I n - 1)^2} < 0$. The negativity of the derivative of the markup with respect to number of firms captures the competition effect of entry lowering markup. When there are many firms in the industry $n \to \infty$, the markup disappears $\mu \to 1$ so price equals marginal cost which is the perfect competition outcome. If $n(t) = 1$ then the economy is populated by a continuum of one firm industries each producing a differentiated product and the resulting fixed markup is the well-know monopolistic competition case (Dixit and Stiglitz 1977).

An optimizing firm's choice of labor and capital correspond to an imperfectly competitive factor market equilibrium such that the price of a factor does not reflect its marginal product.

Proposition 1 (Factor Market Equilibirum)**.** *Under symmetric inter and intra-industrial equilibrium the optimal price setting rules are a markup of firms' marginal products.*

$$AF_k(k(t), l(t)) = \mu(n)r(t) \tag{8}$$
$$AF_l(k(t), l(t)) = \mu(n)w(t) \tag{9}$$

The marginal revenue product of capital (MRPK) $\frac{AF_k}{\mu(n)}$ equates to the price of capital and the MRPL $\frac{AF_l}{\mu(n)}$ equals to the price of labor. As markups increase the marginal revenue from an additonal unit of production is less. Because the MRPs are nonmonotone functions of n_t there is the possibility of multiple equilibria. Different numbers of firm cause the factor market relationship to hold. I do not investigate these implications, instead I assume a unique solution.

2.1.2 Profit

Operating profit $\pi(t) : (K, L, n) \to \mathbb{R}$ is the profit of an incumbent firm in a given period. Operating profits exclude the one-time sunk entry cost that is included in aggregate profits, discussed after we cover the entry process.. Therefore operating profit of a firm is $\pi(t) := y(t) - r(t)k(t) + w(t)l(t)$ and by substituting in factor prices and using Euler's homogeneous function theorem we get profit under imperfect competition[15]

$$\pi(L, K, n; A, \phi) = \left(1 - \frac{\nu}{\mu(n(t))}\right) AF(k, l) - \phi \tag{10}$$
$$\tag{11}$$

Profit is increasing in the markup and is greater than the perfect competition case of $\mu \to 1$. Profits are nonzero in the short run, but in long-run steady state we shall see they are zero (Section 4).

[15]Rearranging the profit function gives the income identity which makes it clearer how markups enter output per firm and is equivalent to the production approach as follows $y(t) := r(t)k(t) + w(t)l(t) + \pi(t) = \frac{AF_l}{\mu(n(t))}l + \frac{AF_k}{\mu(n(t))}k + \left(1 - \frac{\nu}{\mu(n(t))}\right) AF - \phi = \frac{\nu AF}{\mu(n(t))} + \left(1 - \frac{\nu}{\mu(n(t))}\right) AF - \phi = AF - \phi = y(t)$.

2.1.3 Firm Entry

I use the entry setup developed in Datta and Dixon 2002. The process of entry determines the number of firms $n(t)$ and the amount of entry $e(t)$ in a period. It is important to emphasize that 'entry' is 'net entry', so it measures the change in the stock of firms. If the stock of firms increases then net entry is positive so there has been entry, whereas if the stock of firms decreases then net entry is negative so there has been exit. This emphasizes that entry is a single symmetric process incorporating both entry and exit, and they cannot occur together. This is unlike papers that treat entry and exit as different processes. For example recent macroeconomics literature models a process of firm creation (entry), but treats exit as a fixed exogenous process (analogous to depreciation of capital). The importance of this point is that a positive shock to the economy will always cause solely entry and a negative shock solely exit. I shall focus on negative exit-inducing shocks, but the inverse argument would hold for positive shocks.

An endogenous sunk entry cost and an entry arbitrage condition determine the number of firms operating at time t. The sunk entry cost increases with the the number of entrants, and the arbitrage condition equates sunk cost with incumbency profits. Das and Das 1997 term the endogenous sunk cost an entry adjustment cost; in macroeconomics, Lewis 2009 and Berentsen and Waller 2009 use the term *congestion effect*, since more entrants cause congestion in entry that increases the sunk cost. The justification for congestion effects is that resources used to setup a firm are in inelastic supply, so that more entrants raises competition for the resources and therefore increases sunk cost. For example, when introducing a new product, if more firms are entering there is a negative entry externality because it is more costly to differentiate a product. Additionally to evidence for entry externalities and their prevalence in industrial organization literature, the assumption provides an analytical framework to study short-run dynamics away from steady state. It is the sunk entry cost that prevents instantaneous adjustment of firms to steady state[16].

Assumption 2 (Sunk Entry Cost (congestion effect)). *Sunk entry cost $q \in \mathbb{R}$ increases with the number of entrants $\dot{n}$ in t.*

$$q(t) = \gamma \dot{n}, \qquad \gamma \in (0, \infty) \tag{12}$$

[16]The entry adjustment costs theory is analogous to capital adjustment cost models which recognise that investment (deinvestment) in capital is more costly for larger investment (deinvestment). The cost of investment depends on level of innvestment which is the flow of capital; analogously, the cost of entry depends of the level of entry which is the flow of no. firms. See Stokey 2008 for a modern account of capital adjustment costs.

Entry and exit are symmetic for simplicity. A prospective firm pays sunk cost q to enter, and an incumbent firm pays $-q$ to exit. When firms are exiting $\dot{n} < 0 \implies q < 0$, hence $-q > 0$ so the cost of exit is positive. The congestion parameter γ is the marginal cost of entry, and its bounds are the two well-known cases: less sensitivity to congestion $\lim_{\gamma \to 0} q(t)$ implies instantaneous free entry, and more congestion sensitivity $\lim_{\gamma \to \infty} q(t)$ implies fixed number of firms. An extension of the sunk cost assumption to have a fixed cost and the congestion effect, where the fixed cost is paid regardless of the number of entering firms. This setup is closer to Das and Das 1997, and captures the classic case of fixed sunk costs as in Hopenhayn 1992 and Jovanovic 1982, but leads to multiple equilibria in our setup.

Assumption 3 (Entry Arbitrage)**.** *Gain from entry equals return from investing the cost of entry at the market rate $r(t)$.*

$$\dot{q}(t) + \pi(t) = r(t)q(t) \tag{13}$$

The two assumptions are well supported. The congestion effect assumption is common in industrial organization literature. Mata and Portugal 1994 show empirically that firm failure and industry entry rates are positively correlated, and theoretically Das and Das 1997 and Ericson and Pakes 1995 both assume sunk entry costs that rise with number of entrants[17]. The intuition for congestion is that there is more competition for a fixed resource needed to setup. For example, many firms entrants raise initial advertising costs to make consumers aware of the product. If many firms are entering there will be many startup advertising campaign vying for attention. There several synonyms for congestion effects in the literature "negative network effects" or "entry adjustment costs". In terms of the macroeconomy, Lewis 2009 statistically models congestion externalities in entry and concludes that the mechanism improves model fit because it dampens impact responses of entry. Both Lewis and Poilly 2012 and Berentsen and Waller 2009 model congestion effects in a DSGE model. They differ slightly to our setup because entry reduces the probability of survival for example by reducing the likelihood of a sale.

The arbitrage assumption implies that the return to investing in a firm is equal to the return of that investment at the risk free rate r. And an implication is that the value of a firm is equal to present discounted value of future profits as in Bilbiie, Ghironi, and Melitz 2012. Therefore there is an *intertermporal zero profit condition* that implies expected profits of an

[17]Ericson and Pakes 1995 assume the sunk cost is non-decreasing in number of entrants. The assumption includes the simple case of fixed cost not responding to entrants, which they assume in the numerical exercise.

entrant are always zero; if they were ever non-zero, a firm would revise entry to a more profitable time. The zero-profit condition is dynamic rather than static. In the static case current profits, rather than expected future profits, are instantaneously zero (e.g. Jaimovich and Floetotto 2008), so the value of the firm equates to current profits.

The two assumptions form a dynamical system in number of firms and cost of entry $\{n, q\}$ which reduces to a second-order ODE in number of firms

$$\gamma \ddot{n}(t) - r(t)\gamma \dot{n}(t) + \pi(t) = 0 \tag{14}$$

By defining entry, this second-order ODE is separable into two first-order ODEs

Definition 1 (Entry and Exit)**.** Entry (or exit) is measured by the change in the number of firms, *aka* net entry. Negative entry is exit.

$$e(t) = \dot{n} \tag{15}$$

Therefore the model of industry dynamics which determines the number of firms is two ODEs

$$\dot{n}(t) = e(t) \tag{16}$$

$$\dot{e}(t) = -\frac{\pi(t)}{\gamma} + r(t)e(t), \quad \gamma \in (0, \infty) \tag{17}$$

With entry defined as the change in number of firms (15), the arbitrage condition's (17) interpretation depends on the rate of change of entry $\dot{e}(t)$, which is acceleration in number of firms $\ddot{n}(t)$. For example the rate of entry is increasing $\dot{e} > 0$ if the outside option $r(t)e(t)$ exceeds the profit from entering $\frac{\pi(t)}{\gamma}$. This is because when households invest in the more attractive outside option, as opposed to setting up firms, the cost of setting up a firm falls because there is less congestion. The result is an increase in the amount of entry. Initially, it is counterintuitive that the rate of entry decreases with profits, but this captures that when profits are high entry is high, so via congestion the cost of entry is high, and thus the rate of entry slows. The dynamic sunk cost causes firms to respond overtime rather than immediately. Intuitively a firm cannot instantaneously know its cost of entry. A propspective entrant must wait an instance in order to observe the amount of entry and therefore its sunk cost. Consider the contradiction that entry cost is fixed so observable in an instance, $q(t) = \gamma \ \forall t$. In which case the second-order ODE that dictate industry dynamics becomes static $\pi = r\gamma$, so there is no dynamic entry. Rather than the intertemporal zero-profit condition, there is an instataneous alignment of current profit and oppurtunity cost. As shown in Datta and Dixon 2002 an implication of the model is that net present value of the firm (stock market value) equates to the sunk entry costs. In this sense the model is equivalent to BGM's approach, except the advantage here is that it is a lemma whereas in BGM is is assumed and then firms dynamics follow.

Aggregate profits $\Pi(t) \in \mathbb{R}_+$ are distinct from operating profit $\pi(t)$ because they include the fixed cost paid by each firm to enter the market. The first firm to enter in a period pays 0, whereas the second firm will pay γ and the third firm to enter 2γ and so on. Therefore the economy wide cost of entry, $Z(t)$, is

$$Z(t) = \gamma \int_0^{e(t)} i \, di = \gamma \frac{e(t)^2}{2} \tag{18}$$

Hence aggregate profits are aggregate operating profits less entry costs

$$\Pi(t) = n(t)\pi(t) - Z(t) \tag{19}$$

$$\Pi = n \left[AF(k, l) \left(1 - \frac{\nu}{\mu(n)} \right) - \phi \right] - \gamma \frac{e(t)^2}{2} \tag{20}$$

Aggregate profits are an important factor driving capital investment $\dot{K}$. Ceteris paribus entry reduces aggregate profits. It increases the aggregate sunk costs of entry, and diminishes supernormal operating profits through the competition effect on markups.

2.2 Household

The economy is inhabited by a continuum of identical households whose mass is normalized to one. A household seeks policy functions of consumption $\{C(t)\}_0^\infty \in \mathbb{R}$ and labor supply $\{L(t)\}_0^\infty \in [0, 1]$ that maximise lifetime utility $\mathrm{U} : \mathbb{R}^2 \to \mathbb{R}$. We assume $u : \mathbb{R} \times [0, 1] \to \mathbb{R}$ is jointly concave and differentiable in both of its arguments. It is strictly increasing in C and strictly decreasing in L. A representative household solves the utility maximisation problem:

$$\mathrm{U} := \int_0^\infty u(C(t), 1 - L(t)) e^{-\rho t} dt \tag{21}$$

$$\text{s.t.} \quad \dot{K}(t) = rK(t) + wL(t) + \Pi(t) - C(t) \tag{22}$$

The household owns capital $K \in \mathbb{R}$ and takes equilibrium rental rate and wage $r, w \in \mathbb{R}$ as given. Similarly they own firms and take profits as given $\Pi \in \mathbb{R}_+$. The household maximizes utility subject to a budget constraint reaaranged to the law of motion of capital (22). The budget constraint shows that income is earned from capital income, labor income and profit from owning firms and it is spent on consumption or investment in more capital. Maximising utility subject to the capital equation of motion gives six Pontryagin conditions for optimal consumption and labor[18]. The six simplify to three: an intertemporal consumption Euler equation (23), an intratemporal labor-consumption trade-off (24) and the resource constraint (22).

$$\dot{C}(t) = \frac{C(t)}{\sigma(C(t))}(r(t) - \rho), \quad \text{where } \sigma(C(t)) = -C(t)\frac{u_{CC}(C(t))}{u_C(C(t))} \tag{23}$$

$$w(t) = -\frac{u_L(L(t))}{u_C(C(t))} \tag{24}$$

The solution of the dynamic optimization problem for household consumption will be one of the solutions of this system of two differential equations that satisfy our arbitrarily chosen initial condition for capital $K(0) = K_0 \in \mathbb{R}_+$. To complete the boundary value problem we add two transversality conditions on the upper boundary. This completes the unique solution for the boundary value problem that characterizes the optimal path of consumption and labor: three variables (C, K, n), three equations (22)-(24), three boundary conditions (25).

$$\lim_{t\to\infty} K(t)e^{-\rho t} \geqslant 0 \quad \lim_{t\to\infty} K(t)\lambda(t)e^{-\rho t} = 0 \quad K_0 = K(0) \tag{25}$$

In general equilibrium these equations hold and boundary conditions hold, with factor prices and profit determined endogenously from factor market equilibrium $r, w, \Pi : C \times K \times n \to \mathbb{R}_+$.

2.3 Canonical Model in General Equilibrium

Combining the equilibrium conditions from the household and the firm side of the economy defines the model economy as a four dimensional dynamical system that determines consumption, entry, capital and number of firms (C, e, K, n). Importantly labor supply L does not enter the system as an independent variable because it can be defined in terms of C, K, n by combining household intratemporal equilibrium condition with the factor market equilibrium from the firm problem. By understanding the trajectories of labor, we can trace how the *competition effect* of entry reducing markups affects the model through factor price equilibrium and consequently profits.

[18] Appendix A constructs and solves the Hamiltonian.

Proposition 2 (General Equilibrium Labor Supply)**.** *Consumption reduces labor supply, whereas capital and number of firms increase labor supply* $L(\underset{-}{C},\underset{+}{K},\underset{+}{n})$

$$L_C(C,K,n) < 0, \quad L_K(C,K,n) > 0, \quad L_n(C,K,n) > 0$$

Proof. The effects arise through combining factor market equilibrium (9), which determines wage, with the intratemporal condition (24), which determines consumption-labor choice. Then by the implicit function theorem differentiate the intratemporal condition with labor defined implicitly by $L(C,K,n)$. Derivations in appendix B. □

The important determinant of the sign of labor response to a variable change is the response of labor marginal product which influences wage. For example, entry (a rise in n) raises labor supply because it raises wage. Wage rises because labor per firm falls which rises its marginal product due to decreasing returns$\nu < 0$. With constant returns there is no effect $L_n = 0$ as firms employ labor at equal productivity regardless of size. An interesting result from the lemma is that entry and labor supply are positively related. Vice-versa, exit leads to a fall in labor supply. In this paper, we investigate the influence of entry on measured productivity, whereas much empirical discussion is based on labor productivity. The result shows that our model encapsulates labor productivity arguments as a specific case. A lack of exit will buoy labor supply and worsen labor productivity.

Definition 2 (General Equilibrium)**.** Competitive equilibrium is the equilibrium paths of aggregate quantities and prices $\{C(t), L(t), K(t), n(t), e(t), w(t), r(t)\}_{t=0}^{\infty}$, with prices strictly positive, such that $\{C(t), L(t)\}_{t=0}^{\infty}$ solve the household problem. $\{K(t)\}_{t=0}^{\infty}$ satisfies the law of motion for capital. Labour and capital $\{L(t), K(t)\}_{t=0}^{\infty}$ maximise firm profits given factor prices. The flow of entry causes the arbitrage condition on entry to hold (price of entry equals net present value of incumbency). State variables $\{K(t), n(t)\}_{t=0}^{\infty}$ satisfy transversality. Factor prices are set according to factor market equilibrium (9) and ensure goods and factor markets clear.

The dynamic equilibrium conditions from the previous section are the capital accumulation equation, the number of firms definition, the consumption Euler, and the entry arbitrage condition.

Definition 3 (Nonlinear system)**.** The dynamical system is a pair $(\mathbb{X}, g)$, where $\mathbb{X} = (\mathbb{R}^4, \psi)$ is Euclidean space and metric. It defines at a point in time $t \in \mathbb{R}$ the state of the system $(C(t), e(t), K(t), n(t)) = x(t) \in \mathbb{X} \subseteq \mathbb{R}^4$ described by a C^1 vector valued transition map $g : \mathbb{R}^{5+\mathbb{P}} \supseteq \mathbb{X} \times \mathbb{R} \times \Omega \longrightarrow \mathbb{R}^4$. The parameterization $(\phi, \nu, \gamma, \rho)$ defined on an open set $\Omega \in \mathbb{R}^{\mathbb{P}}$

Drop time dependence for clarity.

$$\dot{K} = Y - \frac{\gamma}{2}e^2 - C, \qquad Y = n(F(k,l) - \phi) \tag{26}$$

$$\dot{n} = e \tag{27}$$

$$\dot{C} = \frac{C}{\sigma(C)}(r - \rho), \qquad \sigma(C) = -\frac{Cu_{CC}}{u_C} \tag{28}$$

$$\dot{e} = re - \frac{\pi}{\gamma}, \qquad \pi = AF(k,l)(1 - (1-\mu)\nu) - \phi \tag{29}$$

where factor prices

$$r = \frac{AF_k(k,l)}{\mu}, \qquad w = \frac{AF_l(k,l)}{\mu} \tag{30}$$

Substituting in factor prices, profits and output which are all in terms of (C, K, n) and noting that by the intratemporal condition and wage equilibirum L is implicitly defined as $L(C, K, n)$ the model economy is a system of four ODEs in consumption, entry, capital, number of firms (C, e, K, n). Also by Euler's homogeneous function theorem note $F(k,l) = n^{-\nu}F(K,L)$ and $F_k(k,l) = n^{1-\nu}F_K(K,L)$.

$$\dot{K} = n(An^{-\nu}F(K, L(C,K,n)) - \phi) - \frac{\gamma}{2}e^2 - C \tag{31}$$

$$\dot{n} = e \tag{32}$$

$$\dot{C} = -\frac{u_C}{u_{CC}}\left(\frac{An^{1-\nu}F_K(K, L(C,K,n))}{\mu} - \rho\right) \tag{33}$$

$$\dot{e} = \frac{An^{1-\nu}F_K(K, L(C,K,n))e}{\mu} - \frac{An^{-\nu}F(K, L(C,K,n))(1 - \frac{\nu}{\mu}) - \phi}{\gamma} \tag{34}$$

3 Competition Effect on Factor Prices and Profit

The competition effect enters the model through factor market equilibrium affecting factor prices r and w and in turn affecting profit π. After outlining these mechanisms in this section, the following section on steady state analysis shows that the mechanism propagtes to long-run outcomes, where is raises output per firm and productivity.

Definition 4 (Competition Effect)**.** The competition effect is that markups $\mu(n(t))$ decrease in the number of firms competiting $\mu_n < 0$

The effect on wage $w_\mu\mu_n$ alters the intratemporal condition that determines the static trade-off between labor and consumption. By raising wage, ceteris paribus, it increases labor supply $L_n = L_n^{\bar{\mu}} + L_\mu\mu_n$. The effect on interest rates $r_\mu\mu_n$ alters the households intertemporal decision on consumption now versus investment for future consumption. Therefore it alters the dynamic capital accumulation equation. Since both wage and interest rate determine profit, the competition effect will change profit outcomes $\pi_\mu\mu_n$, which determine the amount of entry through the dynamic entry condition.

There are three effects of an entering firm on factor prices, and analogous three effects on operating profits which are a function of factor prices and form a key result (proposition 3).

$$w_n = \frac{A}{\mu(n)}\left[(1-\nu)n^{-\nu}F_L + n^{1-\nu}F_{LL}L_n - \frac{n^{1-\nu}F_L}{\mu}\mu_n\right] \tag{35}$$

$$r_n = \frac{A}{\mu(n)}\left[(1-\nu)n^{-\nu}F_K + n^{1-\nu}F_{KL}L_n - \frac{n^{1-\nu}F_K}{\mu}\mu_n\right] \tag{36}$$

The three effects are a positive scale effect, an ambiguous labor effect and a positive markup effect. The scale effect captures that an extra firm reduces per firm inputs, but raises aggregate number of firms. The reduction in labor or capital per firm raises their marginal product and therefore price due to decreasing returns to scale. Hence with constant returns $\nu = 0$ the effect is not present. The labor effect captures that entry increases labor supply and therefore lowers wage or raises interest rate. Lastly the competition effect that arises from endogenous markups $\mu_n < 0$ captures that an extra firm increases competition and lowers markups which raises the marginal revenue product of labor (capital) and so wage (interest rate) must increase to maintain equilibrium.

Profits are increasing in markups, which adds an extra effect of a firm entering the market. The result is that profits diminish faster, than if markups were fixed.

Proposition 3. *Entry has two negative effects on operating profit.*

$$\pi_n = \frac{A}{\mu}\left[(\mu-\nu)\left(-\nu n^{-\nu-1}AF + n^{-\nu}F_L L_n\right) + \frac{\nu n^{-\nu}F}{\mu}\mu_n\right] < 0 \tag{37}$$

Proof. By substituting in L_n it can be shown that the negative scale effect dominates the positive labor effect in the second component of π_n. See appendix C □

The three effects of an entering firm on factor prices feed through to profits. The scale effect decreases profit, the labor effect increases profit, the markup effect decreases profit. The markup (competition effect) and scale (business stealing) effect reinforce eachother. Making negative effect of entry on profits larger than the case with fixed markups. However, there is a positive effect on profits from labor, since an entrant raises the market wage leading to a higher supply of labor which raises per firm output and revenues. The effect can be shown to be dominated hence the inequality.

In sum the competition effect of entry depresses markups which raises wage and interest rate in factor market equilibrium. This effect of higher factor prices causes profits to fall more from each entrant. The result is that zero profit arises when fewer firms have entered and so each firm has a larger market share. Conversely a negative shock that leads to negative profits and exit means each exiter raises incumbents profits less regaining zero profits requires more exit and then remaining firms can produce less. So the mechanisms in this section are responsible for the results in the next section that show output per firm and productivity are increasing in the number of firms.

4 Steady State Outcomes

This section first derives the efficient outcomes that correspond to minimum average cost, or the number of firms that maximises output. It then analyzes the fixed point of the dynamical system, which corresponds to the zero profit outcome, often assumed instantaneously in other papers.

4.1 Efficient Output and Productivity

In symmetric equilibrium aggregate output is the number of firms in a representative sector multiplied by the amount a firm produces. It is homogeneous of degree 1 in K, L, n. This captures that capital and labor per firm do not change if all factors are changed equally, so output per firm is homogeneous of degree 0, but aggregation across all firms causes a proportional increase because of a proportional change in the number of firms that are being aggregated.

$$Y(t) = n(t)y(t) = n(t)^{(1-\nu)}AF(K(t), L(t)) - n(t)\phi \tag{38}$$

The effect of an entrant on aggregate output is

$$Y_n = (1-\nu)n^{-\nu}AF(K, L) + n^{1-\nu}AF_L(K, L)L_n(C, K, n) - \phi \tag{39}$$

$$Y_{nn} = (1-\nu)n^{-1}[-\nu n^{-\nu}AF(K, L) + n^{1-\nu}AF_L(K, L)L_n(C, K, n)] \tag{40}$$

There are three effects of firm entry on aggregate output: an ambiguous returns to scale effect, a positive labor effect and a negative fixed cost effect (resource duplication). The scale effect is positive with decreasing returns $\nu \in (0,1)$ and zero with constant returns $\nu = 0$. The effect captures that business reallocation among more firms improves aggregate output when there are decreasing returns.

At the point where the positive returns to scale and labor effect equate the fixed cost effect, there is an optimal efficient number of firms $Y_n = 0|_{\nu<1}$. These outcomes are those that would arise under Walrasian perfect competition, and if there were no markups in our model ($\mu = 1$). In the AC-MC diagram this is where they intersect at minimum average cost. To ensure the outcomes are defined we assume rising marginal cost $\nu \in (0,1)$ and to ensure production is nonnegative we assume fixed cost effect exceeds positive labor effect $\nu\phi \geqslant An^{1-\nu}F_L(K,L)L_n$ so there are initially decreasing returns as costs decrease toward the minimum.

Proposition 4. *When output is maximised with respect to number of firms in the economy the efficient levels of output are*

$$F(k^e, l^e) = \frac{1}{(1-\nu)}\left(\frac{\phi}{A} - F_l(k,l)L_n\right) \tag{41}$$

$$y^e = AF(k^e, l^e) - \phi = \frac{1}{1-\nu}\left(\phi\nu - AF_l(k,l)L_n\right) \tag{42}$$

It is notable that most papers dicussing entry focus on constant returns to scale. As this section has shown this implies there is no optimal firm size; analogously there is no perfect competition equilibrium because the market tends to a natural monopoly due to the fixed cost teamed with constant returns to scale. Firm size, in terms of factors it employs, is unimportant because all firms produce at the same efficiency. This limits the role of entry, so that productivity results arise solely from the competition effect of more firms reducing markups. How output is divided among firms does not matter.

4.1.1 Homogeneous Degree Zero Productivity

I call productivity at a point in time *measured productivity*. The name distinguishes it from *underlying productivity* that prevails in long-run equilibrium. The measure is equivalent to TFPR (R for revenue) in Peters 2013 or Foster, Haltiwanger, and Syverson 2008.

Corresponding to the efficient levels of output is a definition of productivity that is also maximised at these efficient output levels.

Definition 5 (Measured Productivity)**.** Measured productivity $\mathcal{P} : K, L, n \to \mathbb{R}_+$ is the amount of output an economy produces for a given technology, with technology normalized to be homogeneous of degree 1 to remove scale effects

$$\mathcal{P}(t) = \frac{Y(t)}{F(K(t), L(t))^{\frac{1}{\nu}}} \tag{43}$$

This aggregate measure is the same as the per firm measure $\mathcal{P} = \frac{n(t)y(t)}{F(K(t),L(t))^{\frac{1}{\nu}}} = \frac{y(t)}{F(k(t),l(t))^{\frac{1}{\nu}}}$.

A more productive economy has larger measured productivity because it combines inputs more efficiently and produces more output with the same technology as another economy. An outcome of this definition of measured productivity is that when it is maximised with respect to number of firms $\mathcal{P}_n = 0$ the corresponding levels of output are the efficient outcomes y^e. Therefore the maximum attainable productivity $\mathcal{P}^e$ that arises at the efficient level of production is

$$\mathcal{P}^e = \frac{y^e}{F(k^e, l^e)^{\frac{1}{\nu}}} = \frac{(\phi\nu - AF_l L_n) A^{\frac{1}{\nu}} (1-\nu)^{\frac{1}{\nu}-1}}{(\phi\nu - AF_l L_n)^{\frac{1}{\nu}}}, \quad \phi\nu > AF_l L_n, \ \nu \in (0,1) \tag{44}$$

And in the constant returns limit the maximum attainable measured productivity is equivalent to TFP $\lim_{\nu \to 1} \mathcal{P}^e = A$.

Since production technology in the denominator is $hod - \nu$ we need to normalize it to be $hod - 1$. Then productivity will be $hod - 0$ in inputs. That means that the scale of inputs K, L, n does not affect productivity. Whereas with a typical non-normalized measure an economy with more inputs would always appear less productive. Hence we capture changes in efficiency of technology use, how effectively the inputs are combined with a given technology, rather than how many inputs there are. Consider an example of two economies $\mathcal{A}$ and $\mathcal{B}$. They are identical in every sense, except economy $\mathcal{B}$ is endowed with $\lambda \in (1, \infty)$ times more factors K, L, n. Since the economies are identical, except for scale of factors, then a good productivity measure should reflect that both economies have the same productivity: they combine factors with the same efficiency to produce output. Now assume the contradiction that we do not normalize technology and use a standard, non-normalized, TFP measure $\hat{\mathcal{P}}$. Then $\hat{\mathcal{P}}^{\mathcal{A}} = \frac{Y(t)}{F(K(t),L(t))}$ and since Y is $hod - 1$ and F is $hod - \nu$ then $\hat{\mathcal{P}}^{\mathcal{B}} = \frac{\lambda Y(t)}{\lambda^{\nu} F(K(t),L(t))} = \lambda^{1-\nu} \hat{\mathcal{P}}^{\mathcal{A}}$. So $\hat{\mathcal{P}}^{\mathcal{A}} < \hat{\mathcal{P}}^{\mathcal{B}}$ we conclude erroneously that economy $\mathcal{B}$ is more productive simply because it has more factors, not because it combines those factors more efficiently. Under our normalized measure $\mathcal{P}^{\mathcal{B}} = \frac{\lambda Y(t)}{(\lambda^{\nu} F(K(t),L(t)))^{\frac{1}{\nu}}} = \mathcal{P}^{\mathcal{A}}$. With constant returns $\nu = 1$ there are no scale effects, so our measure collapses to the common definition.

4.2 Steady State

Now I shall show that the steady state of our economy corresponds to zero profits. And leads to levels of output and productivity that depend endogenously on the number of firms, and these levels are strictly less than the efficient levels that would arise under perfect competition defined in section 4.1.

Assume that a solution of the system converges to steady state $(K^*, n^*, C^*, e^*) \to (K^s, n^s, C^s, e^s)$ as $t \to +\infty$[19]. In steady state $\dot{K} = \dot{n} = \dot{C} = \dot{e} = 0$, which immediately implies entry is zero via (46), which in turn, via 48, implies profits are zero.

$$\dot{K} = 0 \Leftrightarrow Y^*(C^*, K^*, n^*) = C^* \tag{45}$$
$$\dot{n} = 0 \Leftrightarrow e^* = 0 \tag{46}$$
$$\dot{C} = 0 \Leftrightarrow r^*(C^*, K^*, n^*) = \rho \tag{47}$$
$$\dot{e} = 0 \Leftrightarrow \pi^*(C^*, K^*, n^*) = 0 \tag{48}$$

In steady state aggregate output equates to consumption; entry is zero; the interest rate equals the discount factor and profits are zero. Intuitively when profits are zero entry ceases as there is no entry incentive, and when the discount factor and interest rate are equated there is indifference between consumption and saving so all output is consumed. Rewriting the system in terms of underlying variables (C, e, K, n), again with labor defined implicitly $L(C, K, n)$, shows that ouput per firm and therefore measured productivity depend endogenously on the number of firms.

$$n^* \left[A n^{-\nu} F(K^*, L^*) - \phi \right] = C^* \tag{49}$$
$$e^* = 0 \tag{50}$$
$$\frac{A {n^*}^{1-\nu} F_K(K^*, L^*)}{\mu(n^*)} = \rho \tag{51}$$
$${n^*}^{-\nu} F(K^*, L^*) = \frac{\phi}{A\left(1 - \frac{\nu}{\mu(n^*)}\right)} \tag{52}$$

[19]Ignore the trivial steady state that arises when the state vector is the zero vector.

Theorem 1 (Endogenous Steady State Output and Productivity)**.** *Steady state output per firm y^* and measured productivity $\mathcal{P}^*$ are endogenous because they depend on the markup which depends on the endogenous variable $n(t)$, the number of active firms*[20].

$$y^*(\underset{+}{n^*}, \underset{-}{\mu(n^*)}) = \frac{\phi\nu}{\mu(n^*) - \nu} \tag{53}$$

$$\mathcal{P}^*(\underset{+}{n^*}, \underset{-}{\mu(n^*)}) := \frac{y^*}{F(k^*, l^*)^{\frac{1}{\nu}}} = \nu \left[\frac{A}{\mu(n^*)} \left(\frac{\mu(n^*) - \nu}{\phi} \right)^{1-\nu} \right]^{\frac{1}{\nu}} \tag{54}$$

Proof. From the zero-profit condition variable production $AF(k^*, l^*) = \frac{\phi\mu(n^*)}{(\mu(n^*)-\nu)}$, and then $y^* = AF(k^*, l^*) - \phi$. Then substitute into the expression for productivity. □

Corollary 1. *Steady state output per firm and measured productivity are increasing in the number of firms in the economy.*

$$y^*_n > 0 \qquad \mathcal{P}^*_n > 0$$

Proof.

$$\frac{\partial \mathcal{P}}{\partial n} = -\frac{A\nu}{\mu^2} \frac{(\mu - 1)}{(\mu - \nu)^\nu \phi^{1-\nu}} \cdot \frac{\partial \mu(n^*)}{\partial n} > 0 \tag{55}$$

□

Since markups are decreasing in number of firms, output is increasing in number of firms, and similarly productivity is increasing in number of firms. The simpler case of constant marginal cost gives a similar outcome $\mathcal{P}^*|_{\nu=1} = \frac{A}{\mu}$ as in Jaimovich and Floetotto 2008 and Peters 2013. Using the variable production expression we can infer the behavior of number of firms in steady state.

Corollary 2. *The number of active firms is procyclical.*

$$n^* = \left[AF(K^*, L^*(C^*, K^*, n^*)) \left(\frac{\mu(n^*) - \nu}{\mu(n^*)\phi} \right) \right]^{\frac{1}{\nu}} = \left(\frac{\mu(n^*) - \nu}{\phi\nu} \right) Y^* \tag{56}$$

[20] Parameter changes have second-order effects since they change n^*. The signs for ϕ and ν exclude second order effects.

When there is a rise in aggregate ouput the number of firms also increases since $\left(\frac{\mu(n^*)-\nu}{\phi\nu}\right) > 0$. Constant returns to scale weakens the effect, conversely firms are more procylical with greater decreasing returns.

With fixed markups output per firm always returns to a constant level that is the amount of sales required to cover the fixed cost. Now μ is endogenous if n rises then more sales are needed to cover the fixed cost and therefore output per firm in equilibrium depends on the number of firms in the market.

$$Y^* = n^* y^* = (AF(K^*, L^*(C^*, K^*, n^*)))^{\frac{1}{\nu}} \left(\frac{\mu(n^*)\phi}{\mu(n^*) - \nu}\right)^{1-\frac{1}{\nu}} \frac{\nu}{\mu(n^*)} \tag{57}$$

Aggregate output is much simpler with constant returns $Y^*|_{\nu=1} = \frac{AF(K^*,L^*)}{\mu(n^*)}$. There is productive inefficiency from the markup, which reduces Y^*, but the fixed cost component $\left(\frac{\mu(n^*)\phi}{\mu(n^*)-\nu}\right)^{1-\frac{1}{\nu}} \nu$ is unimportant as all firms use the fixed cost with the equal efficieny. It is useful to compare the endogenous markup case, to the better-known case of fixed markups in steady state. With a fixed markup number of firms does not affect aggregate output through y^* which is always fixed exogenously as a function of given parameters, so an extra firm simply contributes this fixed extra amount to output. With endogenous markup an extra firm alters per firm output y^* since a firm needs to produce more to cover fixed costs due to fiercer markup competition. $\frac{\partial Y^*}{\partial n} = y^* + n^* y^*_n$. With fixed markups only the first effect is present; an extra firm raises output by the same fixed y^*, but with endogenous markup there is also the comeptition effect which raises output per firm.

Since markups are now negatively related to the number of firms, a single firm needs to sell more output to cover its fixed cost and break even in a free entry equilibrium. Hence with more firms, output per firm rises. It is important to notice this always arises in the long-run, but in the short run we could move from a position of higher output per firm to an equilibrium with lower output per firm, that is still above the initial level. Or we could move to higher output per firm and then with more entry, the extra competition raises output per firm more.

The fixed cost and decreasing returns to scale cause a dampening effect that captures that productivity is less sensitive to markups when the fixed cost is high. This component falls out when there are constant returns to scale $\phi^{1-\nu}|_{\nu\to 1} \to 1$.

The markup causes extra profit that helps us to understand the mechanism through which entry is affected. Profits offer entry incentives, and incentives rise when markups are higher which encourages more entry than would arise under perfect competition. However since markups up decreasing in number of firms the excessive amount of profit will diminish faster as firms enter, and zero profit will arise when fewer firms have entered so the business stealing effect is dampened and larger firms remain. Underlying productivity is a function of number of firms whereas with fixed markups firms entry always returned the economy to a position with the same productivity. The extra mechanism is important, since a prospective firm now considers how fierce competition in the market is, whereas with fixed markup it took for granted that it could enter and charge a given markup.

The conclusions from the static analysis are that output per firm increases with number of firms, and productivity increases with number of firms. These results arise because number of firms degrade monopoly power, and this effect will always prevail over the dynamic business reallocation effect that causes output per firm to decrease as each firm enters because any shock rises productivity and output per firm too much on impact. From the long-run perspective more firms are always better.

5 Main Result: Productivity Dynamics

The main result shows that the impact effect of a TFP shock causes an exacerbated response in short-run productivity that relinquishes over time but leaves some long-run persistence due to the competition effect. This means that the difference between measured productivity on impact and measured productivity in the long-run is dampened because of a persistent change in productivity draws it closer to the initial level.

Theorem 2 (Permanent Change in TFP)**.** *On impact of a shock productivity overshoots the long-run effect, but there is no reversion to underlying productivity due to a persistent change in degree of competition.*

$$\mathcal{P}(0)_A - \mathcal{P}^*_A = \mathcal{P}(0)_A - \mathcal{P}^{*\bar{\mu}}_A - \mathcal{P}^*_\mu \mu(n^*)_A \tag{58}$$

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}^*_A = (\mu^* - 1)\frac{\mathcal{P}^*}{\nu}\left[\underbrace{\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right)}_{\textit{Alloc. Effect (+)}} + \underbrace{\frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n^*_A}_{\textit{Comp. Effect (-)}}\right] \tag{59}$$

where $\mathcal{P}^* = \nu\left[\frac{A}{\mu(n^*)}\left(\frac{\mu(n^*)-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}}$

Proof. Details in appendix D. The proof follows from analytically solving the system within a neighborhood of the steady state. The analytical solution shows that at $t = 0$ neither capital nor firms respond to a shock. From corrollary 2 equation 56 $n^*_A > 0$ improved TFP raises the number of firms. □

The positive allocation effect captures that on impact there is a direct effect on incumbents' productivity from having a different TFP and there is a reinforcing labor effect that also responds immediately. The negative competition effect captures that the long-run level of productivity moves in the same direction as the initial effect which closes the gap between initial and long-run productivity.

Corollary 3. *In more competitive economies ($n \to \infty$) the volatility of productivity is reduced.*

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} = \mathcal{P}^*_A \tag{60}$$

Proof. Since the markup disappears as firms increase $\lim_{n\to\infty} \mu(n_t) = 1$ then

$$\lim_{n\to\infty} \left(\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}^*_A \right) = 0$$

□

The result implies that productivity puzzles are weakened when there is more competition. Conversly overshooting is greater when there are few firms per sector. This strengthens imperfect competition and therefore markups are higher. It also means that the long-run structural change to competition will be greater. Consequently, the impact of a technology shock causes a large change in measured productivity initially but it then reverts to a similar but weaker level of productivity in the long-run. Contrarily, if the sector is very competitive, there are many firms in the sector and the initial effect on productivity is small, likewise there is little structural change to competition from more firms entering because there are still many firms competing. The implication is that more competition, which is synonymous to more firms, implies less volatile productivity and less persistence in productivity shocks[21]. As number of firms in the economy gets large then the long-run effect arises immediately $\lim n \to \infty \mu(n) = 1$ so $\mathcal{P}(0)_A = \mathcal{P}^*_A$

Figure 3 shows that the long-run competition effect tightens the gap between impact and long-run effect. A negative shock to technology from A_1 to A_2 causes an initially big fall in measured productivity (dashed arrow), but it recovers as firms begin to exit. However, the red curve shows that with fixed markups $\bar{\mu}$ productivity recovers to regain the underlying level that incorporates the new worse technology A_2, whereas the blue time path shows that despite some recovery there is always persistently worse productivity in the long run (shown by the gray box), and this is because the markup μn^* is allowed to fall thus reducing long-run competition.

[21] This is an interesting testable implication to expand upon empirically

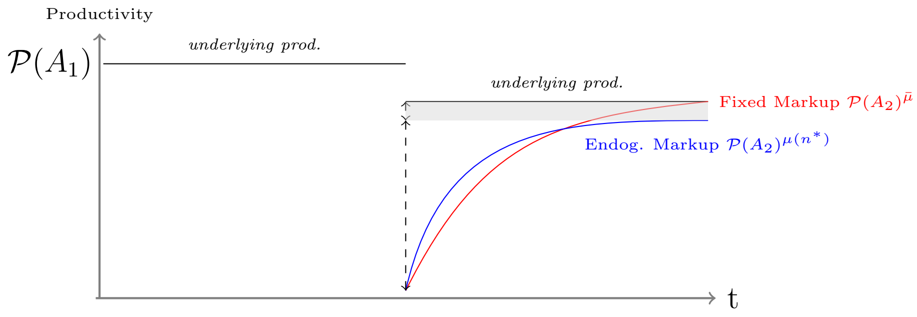


Figure 3: Exacerbated Productivity Followed by Long-run Persistence

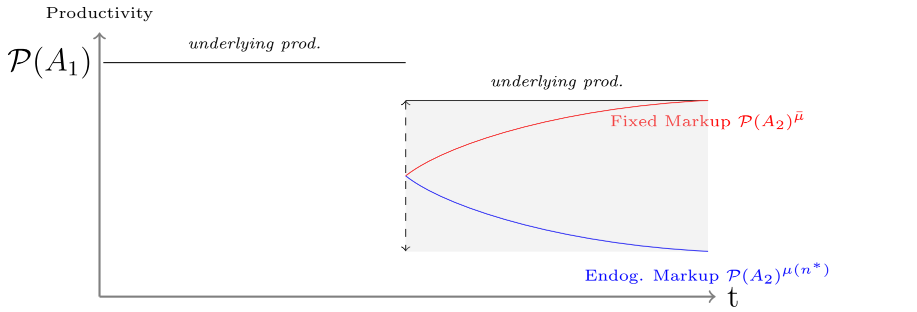


Figure 4: Short-run Productivity Undershooting

5.1 Competition Effect Strengthens Misallocation Effect

A special case that may arise is if the negative competition effect is larger than the positive scale effect. The previous discussion assumed that $\mathcal{P}(0)_A - \mathcal{P}_A^* > 0$. However, if the competition effect is large then $\mathcal{P}(0)_A - \mathcal{P}_A^* < 0$, so the initial movement in measured productivity is less than the long-run change in productivity. In terms of a positive shock to TFP this would mean an increase in measured productivity on impact as incumbent firms benefit from the improved technology, but then as firms begin to enter their negative effect of reallocating business is less than their positive effect reducing markups, so as they enter productivity continues to improve. If there is a negative TFP shock as in figure 3 the result is an initiall fall in productivity, followed by further worsening of productivity to a long-run level below the initial movement. After the initial fall in productivity the further worsening occurs as firms exit and weaker competition reduces productivity more than the reallocation of resources among incumbents.

5.2 Numerical Exercise

A numerical exercise is useful to compare the two effects to see whether undershooting or overshooting arises. The baseline RBC model assumes isoelastic (constant elasticity) separable subutilties and a Cobb-Douglas production function.

5.2.1 Utility

$$U(C,L) = \frac{C^{1-\sigma}-1}{1-\sigma} - \xi\frac{L^{1+\eta}}{1+\eta} \tag{61}$$

The derivatives are

$$U_C = C^{-\sigma}, \quad U_{CC} = -\sigma C^{-\sigma-1}, \quad U_L = -\xi L^{\eta} \tag{62}$$

The degree of relative risk aversion is constant $\sigma(C) = -C\frac{U_{CC}}{U_C} = \sigma$. Isoelastic utility implies there is constant elasticity of utility with respect to each good. For example, a percent change in consumption always causes a $(1-\sigma)\%$ change in utility. $\sigma \neq 1$ is the constant coefficient of relative risk aversion. $\sigma \to \infty$ implies infinite risk aversion, so consumption has little effect on utility. η is Frisch elasticity of labor supply.

5.2.2 Production

$$F(k,l) = k^{\alpha}l^{\beta} = K^{\alpha}L^{\beta}n^{-(\alpha+\beta)} = F(K,L)n^{-(\alpha+\beta)} \tag{63}$$

$$F_k = \alpha k^{\alpha-1}l^{\beta} = \alpha K^{\alpha-1}L^{\beta}n^{1-(\alpha+\beta)}, \quad F_l = k^{\alpha}\beta l^{\beta-1} = K^{\alpha}\beta L^{\beta-1}n^{1-(\alpha+\beta)} \tag{64}$$

Cobb-Douglas production conforms to our assumptions on the production function derivatives, and it is homogeneous of degree $\alpha+\beta$, so $\nu = \alpha+\beta$ in our general notation. α and β are capital and labor shares respectively. This implies increasing marginal costs if $\alpha+\beta < 1$.

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}_A^* = \frac{\mathcal{P}^*}{\nu(n^*-1)}\left[\frac{2}{A} + \beta\frac{n_A^*}{n^*} - \frac{\nu}{\mu(n^*)\left(\mu(n^*)-\nu\right)}\frac{n_A^*}{n^{*2}}\right] \tag{65}$$

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}_A^* = \frac{\mathcal{P}^*}{(\alpha+\beta)(n^*-1)}\left[\frac{2}{A} + \left(\beta - \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1}-(\alpha+\beta)\right)}\right)\frac{n_A^*}{n^*}\right] \tag{66}$$

A sufficient condition for undershooting $\mathcal{P}(0)_A - \mathcal{P}^*_A < 0$ is

$$\frac{2}{A} + \left(\beta - \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1} - (\alpha+\beta)\right)}\right)\frac{n^*_A}{n^*} < 0$$

therefore a necessary condition is $\beta < \frac{(\alpha+\beta)}{\frac{n^{*2}}{n^*-1}\left(\frac{n^*}{n^*-1}-(\alpha+\beta)\right)}$, where the right-hand side is the competition effect that arises from endogenous markups. The competition effect is zero with exogenous (fixed) markups, therefore the necessary condition is always violated and undershooting cannot arise. This formalizes the logic that if there is no persistent effect on productivity then undershooting of the long-run level cannot arise. Similarly, any effect that weakens the long-run persistent effect of competition will reduce the likelihood of undershooting. Formally that is any effect that reduces $\mathcal{P}^*_\mu$. For example, the competition effect is increasing in returns to scale ν and decreasing in the markup μ, and therefore increasing in number of firms [22]. Therefore a higher number of firms, and closer to constant returns to scale $\nu := \alpha + \beta \to 1$ will strengthen the competition effect, leading to undershooting. Figure 5 shows the competition effect is largest with close to constant returns and increases with number of firms for low values of firms, but once firms become large so the markup is negligible the competition effect weakens since there is near constant returns to scale and the dominant effect of an additional firm is fewer input among incumbents. Figure 6 shows a calibration that strenghtens the necessary condition by choosing close to CRTS with a low β. The negative region is quantitiatively small, so is unlikely to offset the initial effect required to meet the sufficient condition, unless n^*_A is unrealistically large. For example $n^* = 5$ and $\nu = \alpha + \beta = 0.6 + 0.2 = 0.8$ meets the necessary condition for negativity, so the sufficient condition, where assuming $A = 1$, is $2 - 2\frac{n^*_A}{5} < 0$. Therefore the condition is $n^*_A > 5 = n^*$, so there must be a 100% change in market size if the competition effect from entry is to exceed the initial misallocation effect..

6 Conclusion

The paper investigates the effect of firm entry on measured productivity over the business cycle. I consider that entry is noninstantaneous and entry affects the price markups that incumbents charge. Together these mechanisms

[22] Although this only holds for small values of n, when n is large and the markup dissipates to 1 so the effect of competition is unimportant in which case the competition effect starts decreasing in n

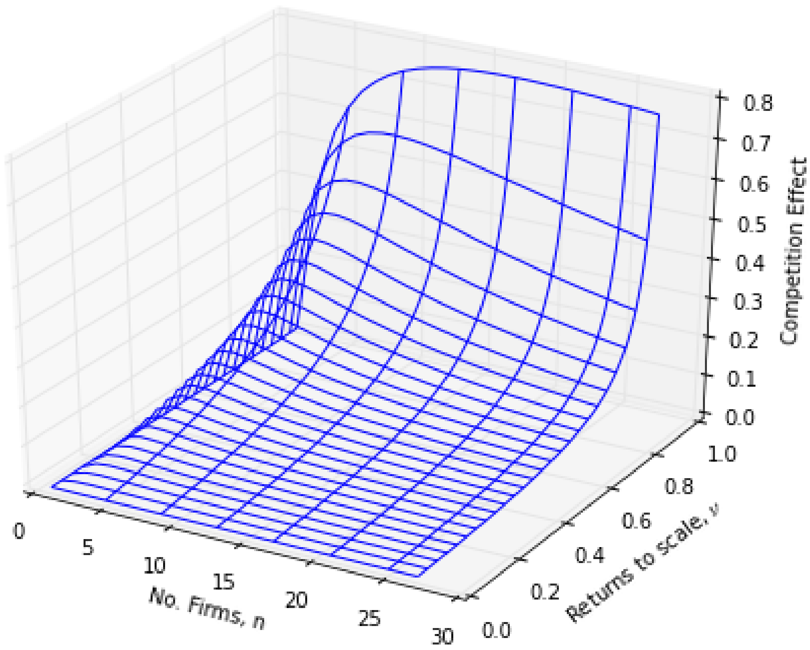


Figure 5: Competition Effect

can explain short-run procylical productivity and weaker long-run persistence. Contemporary productivity puzzles provide a lens to view the theory through. In relation to productivity puzzles, the theory explains that productivity is exacerbated on impact, since firms cannot adjust immediately so incumbents bear shocks, and in the long run underlying productivity is not be regained because subsequent adjustment of firms causes structural changes in competition. The structural changes in competition reflect that entry strengthens competition which improves productivity in the long run (inversely, exit weakens competition, decreases productivity). Furthermore I show that in highly competitive industriesthe distinction between short-run and long-run productivity is small, so measured productivity quickly and accurately reflects underlying productivity.

A growing number of DSGE papers show promising quantitative simulation results from adding a firm entry process. Despite these appealing data matching properties, little research has reduced models to minimal state variables to understand the analytical effect of entry. This paper allows economists to understand how the entry variables interact with the model in a general setup before specifying functional forms or numerical calibrations. We learn that firm entry dynamics can explain short-run dynamic changes in

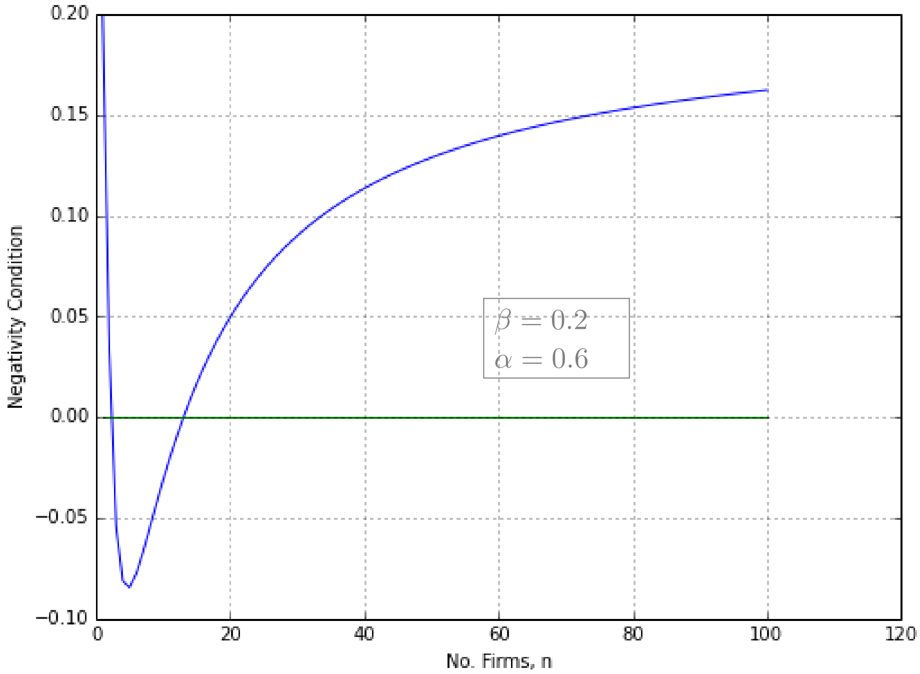


Figure 6: Initial Overshooting Versus Undershooting

productivity over the business cycle and long-run static changes that persist. The two explanations arise from two different effects of entry, a dynamic reallocation effect that redistributes resources as firms adjust and a static competition effect that alters firms' pricing markup decisions in response to competition from entry.

A Household Optimization Problem

Use the Maxmimum Principle to obtain the necessary conditions for a solution to the household's utility maximisation problem. The current value Hamiltonian is

$$\hat{\mathcal{H}}(t) = u(C(t), L(t)) + \lambda(t)(w(t)L(t) + r(t)K(t) + \Pi(t) - C(t)) \tag{67}$$

The costate variable λ_t is the shadow price of wealth in utility units. The Pontryagin necessary conditions are

$$\hat{\mathcal{H}}_C(K, L, C, \lambda) = 0 \implies u_C - \lambda = 0 \tag{68}$$

$$\hat{\mathcal{H}}_L(K, L, C, \lambda) = 0 \implies u_L + \lambda w = 0 \tag{69}$$

$$\hat{\mathcal{H}}_K(K, L, C, \lambda) = \rho\lambda - \dot{\lambda} \implies \lambda r = \rho\lambda - \dot{\lambda} \implies \frac{\dot{\lambda}}{\lambda} = -(r - \rho) \tag{70}$$

$$\hat{\mathcal{H}}_\lambda := \dot{K}_t \implies \dot{K} = rK + wL + \Pi - C \tag{71}$$

The four Pontryagin conditions (68)-(70) reduce to two equations: a differential equation in consumption (*consumption Euler equation* or *intertemporal condition*), and a static injective mapping between labor and consumption (*intratemporal condition*).

B Optimal Labour Derivatives

Partially differentiate the intratemporal Euler with respect to each variable treating labor as an implicit function, and with wage set at the imperfect competition market rate $w(K, L, n) = \frac{An^{1-\nu}F_L(K,L)}{\mu(n)}$.

$$u_L(L) + u_C(C)w(K, L, n) = 0 \tag{72}$$

$$u_L(L) + u_C(C)\frac{An^{1-\nu}F_L(K, L)}{\mu(n)} = 0 \tag{73}$$

Recall the utility and production function assumptions: $F_{LL}(K, L),\ u_{CC}(C),\ u_{LL}(L) < 0$ and $u_C(C),\ F_L(K, L),\ F_{LK}(K, L) > 0$

$$u_{LL}L_C + u_C\frac{An^{1-\nu}F_{LL}L_C}{\mu(n)} + \frac{u_{CC}An^{1-\nu}F_L}{\mu(n)} = 0 \tag{74}$$

$$L_C = \frac{-u_{CC}An^{1-\nu}F_L\mu(n)^{-1}}{u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1}} < 0 \tag{75}$$

$$u_{LL}L_K + u_C\frac{An^{1-\nu}F_{LL}L_K}{\mu(n)} + \frac{u_C An^{1-\nu}F_{LK}}{\mu(n)} = 0 \tag{76}$$

$$L_K = \frac{-u_C An^{1-\nu}F_{LK}\mu(n)^{-1}}{u_{LL} + u_C An^{1-\nu}F_{LL}\mu(n)^{-1}} > 0 \tag{77}$$

$$u_{LL}L_n + \frac{u_C A(1-\nu)n^{-\nu}F_L}{\mu(n)} + \frac{u_C A n^{1-\nu}F_{LL}L_n}{\mu(n)} + u_C A n^{1-\nu}F_L\frac{-\mu(n)_n}{\mu(n)^2} \quad (78)$$

$$L_n = \frac{u_C A n^{1-\nu}F_L\mu_n\mu^{-2} - u_C A(1-\nu)n^{-\nu}F_L\mu^{-1}}{u_{LL} + u_C A n^{1-\nu}F_{LL}\mu(n)^{-1}} > 0, \quad \nu \in (0,1) \quad (79)$$

In all cases the denominator is the intratemporal condition differentiated with respect to labor, and it is negative $u_{LL} + u_C A n^{1-\nu}F_{LL}\mu(n)^{-1} < 0$. Therefore the numerator distinguishes signs. Concavity of the production and utility functions, assumptions above, are sufficient to determine the signs of the numerator except for L_n which depends on returns to scale of the technology ν. With decreasing returns $\nu < 1$ labor increases; with increasing returns labor decreases and with constant returns $\nu = 1$ labor would be irresponsive to entry if there were fixed markups $\mu_n = 0$, but the endogenous markup $\mu_n < 0$ means labor increases with entry even with constant returns. This is because although the marginal product of labor does not change because of constant returns, the fall in markups reduces the wedge between marginal product of labor and wage, so wage increases.

The economic intuition is easier to understand in terms of wages, where $w^{\bar{L}}$ is wage with labor fixed.

$$L_C = \frac{-u_{CC}w}{u_{LL} + u_C w_L}, \quad L_K = \frac{-u_C w_K^{\bar{L}}}{u_{LL} + u_C w_L}, \quad L_n = \frac{w\frac{\mu_n}{\mu} - u_C w_n^{\bar{L}}}{u_{LL} + u_C w_L} \quad (80)$$

C Optimal interest rate, profit, output

Given optimal labour choice we can evaluate how interest rate, profit and output respond.

C.1 Output

$$Y(L(C,K,n),K,n) = n^{1-\nu}[AF(K,L(C,K,n)) - \phi] \quad (81)$$

$$Y_C = An^{1-\nu}F_L(K,L)L_C(C,K,n) < 0 \quad (82)$$

$$Y_K = An^{1-\nu}[F_K(K,L) + F_L(K,L)L_K(C,K,n)] > 0 \quad (83)$$

$$Y_n = (1-\nu)AF - \phi + AF_l L_n \lesseqgtr 0 \quad (84)$$

Furthermore in steady state when $F(\frac{K}{n},\frac{L}{n})^* = \frac{\phi}{A(1-(1-\mu)\nu)}$ then $\hat{Y}_n^* = \frac{-\phi\mu\nu}{1-(1-\mu)\nu} + AF_l L_n$ which is positive or negative depending whether the negative component outweighs the positive labour effect.

C.2 Rents

$$r = \frac{1}{\mu} A F_k \tag{85}$$

$$r_C = \frac{1}{\mu}\frac{A}{n} F_{kl} L_C < 0 \tag{86}$$

$$r_K = \frac{1}{\mu}\frac{A}{n}[F_{kk} + F_{kl} L_K] < 0 \tag{87}$$

$$r_n = \frac{1}{\mu}\frac{A}{n}[(1-\nu)F_k + F_{kl} L_n] - \frac{1}{\mu^2}\mu_n A F_k > 0 \tag{88}$$

Both r_K and π_n require extra work to derive the signs. They are found by substituting in L_K and L_n respectively, then they can be rearranged into a form including $F_{KK}F_{LL} - F_{KL}^2$ which by the second partial derivative test for concavity assumption is greater than zero.

C.3 Profit

$$\pi = AF(k,l)(1 - \frac{\nu}{\mu}) - \phi \tag{89}$$

$$\pi_C = AF_l \frac{L_C}{n}(1 - \frac{\nu}{\mu}) < 0 \tag{90}$$

$$\pi_K = \frac{A}{n}(F_k + F_l L_K)(1 - \frac{\nu}{\mu}) > 0 \tag{91}$$

$$\pi_n = \frac{A}{n}(-\nu F + F_l L_n)(1 - \frac{\nu}{\mu}) + AF(k,l)\frac{\nu}{\mu^2}\mu_n < 0 \tag{92}$$

For any K, L, n profit is higher when imperfect competition μ increases, but not necessarily higher for any given K, n. This explains that even if imperfect competition increases and therefore higher profits are available, the number of firms can (counterintuitively) decrease. The offsetting factor is the the indirect labor effect: labor supply is discouraged by the rise in imperfect competition andas there is a bigger wedge between wage and marginal product of labor. Therefore labor supply falls such that profits are lower for a given K, n. I typically assume these secondary labor effects to be too small to offset the primary mechanisms. So operating profit increases with imperfect competition, even as L is allowed to adjust.

D Productivity Dynamics

Throughout the derivations remember that the markup is a function of number of firms $\mu(n)$, but for simplicity I write μ.

$$\mathcal{P}(t)_A = \frac{n^{1-\nu}F + yn_A + (n^{1-\nu}A - ny\frac{1}{\nu}F^{-1})(F_K K_A + F_L L_A)}{F^{\frac{1}{\nu}}} \tag{93}$$

The crucial step with dynamic firms and capital is that state variables do not move on impact $K_A = 0$ and $n_A = 0$. This is what causes the distinction between short-run and long-run productivity that is not present with instantaneous free entry. Therefore at $t = 0$ the change in productivity depends on the direct effect of better technology, and its indirect effect on labor, which increases labor supply.

$$\mathcal{P}(0)_A = \frac{n^{1-\nu}F + (n^{1-\nu}A - ny\frac{1}{\nu}F^{-1})F_L L_A}{F^{\frac{1}{\nu}}} \tag{94}$$

$$\mathcal{P}(0)_A = n^{1-\nu}F^{1-\frac{1}{\nu}} + \left(F^{-\frac{1}{\nu}}n^{1-\nu}A - \mathcal{P}\frac{1}{\nu}F^{-1}\right)F_L L_A \tag{95}$$

Assuming that the economy is initially in steady state when the shock occurs, evaluate the expression with all variables x at steady state $x(0) = x^*$. From $\pi = y - rK - wL$ then $y^* = rK^* + wL^*$ so $y^* = \frac{A\nu n^{*-\nu}F}{\mu^*}$ and thus $\mathcal{P}^* = \frac{n^* y^*}{F^{*\frac{1}{\nu}}} = \frac{A\nu n^{*1-\nu}F^{1-\frac{1}{\nu}}}{\mu}$. This expression for productivity makes it easier to represent the impact effect of a TFP shock in terms of steady state productivity $\mathcal{P}^*$ as follows

$$\mathcal{P}(0)_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + \left(\frac{\mu^*\mathcal{P}^*}{\nu F^*} - \frac{\mathcal{P}^*}{\nu F^*}\right)F_L^* L_A^* \tag{96}$$

$$\mathcal{P}(0)_A|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^* - 1)\frac{\mathcal{P}^*}{\nu F^*}F_L^* L_A^* \tag{97}$$

Comparing the short-run impact effect to the long-run steady-state effect $\mathcal{P}_A^* = \mathcal{P}_A^{*\bar{\mu}} + \mathcal{P}_\mu^* \mu_A = \frac{\mathcal{P}^*}{A\nu} + \mathcal{P}_\mu^* \mu_n n_A^*$ shows that the endogenous productivity effect dampens the difference between short-run and long-run effects

$$\mathcal{P}(0)_A - \mathcal{P}_A^*|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^* - 1)\frac{\mathcal{P}^*}{\nu F^*}F_L^* L_A^* - (\mathcal{P}_A^{*\bar{\mu}} + \mathcal{P}_\mu^* \mu_A) \tag{98}$$

$$\mathcal{P}(0)_A - \mathcal{P}_A^*|_{x(0)=x^*} = \frac{\mu^*\mathcal{P}^*}{A\nu} + (\mu^* - 1)\frac{\mathcal{P}^*}{\nu F^*}F_L^* L_A^* - \left(\frac{\mathcal{P}^*}{A\nu} + \mathcal{P}_\mu^* \mu_n n_A^*\right) \tag{99}$$

$$\mathcal{P}(0)_A - \mathcal{P}_A^*|_{x(0)=x^*} = (\mu^* - 1)\frac{\mathcal{P}^*}{A\nu} + (\mu^* - 1)\frac{\mathcal{P}^*}{\nu F^*}F_L^* L_A^* - \mathcal{P}_\mu^* \mu_n n_A^* \tag{100}$$

$$\mathcal{P}(0)_A - \mathcal{P}_A^*|_{x(0)=x^*} = (\mu^* - 1)\frac{\mathcal{P}^*}{\nu}\left(\frac{1}{A} + \frac{F_L^* L_A^*}{F^*}\right) - \mathcal{P}_\mu^* \mu_n n_A^* \tag{101}$$

The expression for $\mathcal{P}_\mu^*$ can simplyify the expression further

Lemma 1.

$$\mathcal{P}^*_\mu = -\frac{\mathcal{P}^*(\mu-1)}{\mu(\mu-\nu)} < 0 \tag{102}$$

Proof.

$$\mathcal{P}^*_\mu = \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}-1}\left(\frac{-A}{\mu^2}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu} + \frac{A}{\mu}(1-\nu)\left(\frac{\mu-\nu}{\phi}\right)^{-\nu}\frac{1}{\phi}\right) \tag{103}$$

$$= \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}}\left(-\frac{1}{\mu} + \frac{1-\nu}{\mu-\nu}\right) \tag{104}$$

$$= \left[\frac{A}{\mu}\left(\frac{\mu-\nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}}\left(\frac{\nu(1-\mu)}{\mu(\mu-\nu)}\right) = -\frac{\mathcal{P}^*(\mu-1)}{\mu(\mu-\nu)} < 0 \tag{105}$$

□

Notice that with constant returns the expression is simply

$$\mathcal{P}^*_\mu\Big|_{\nu\to 1} = -\frac{A}{\mu^2}$$

and with many firms the markup tends to unity, therefore long-run underlying productivity reflects true TFP.

$$\mathcal{P}^*_\mu = -\frac{\mathcal{P}^*(\mu^*-1)}{\mu^*(\mu^*-\nu)} \tag{106}$$

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = (\mu^*-1)\frac{\mathcal{P}^*}{\nu}\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) + \frac{\mathcal{P}^*(\mu^*-1)}{\mu^*(\mu^*-\nu)}\mu_n n^*_A \tag{107}$$

$$\mathcal{P}(0)_A - \mathcal{P}^*_A|_{x(0)=x^*} = (\mu^*-1)\mathcal{P}^*\left[\frac{1}{\nu}\left(\frac{1}{A} + \frac{F^*_L L^*_A}{F^*}\right) + \frac{1}{\mu^*(\mu^*-\nu)}\mu_n n^*_A\right] \tag{108}$$

E Productivity Dynamics with Functional Forms

First let us restate the short-run versus long-run productivity effect, and focus attention on the square bracketed component Γ that represents allocation versus competition effect.

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}_A^* = (\mu^* - 1)\frac{\mathcal{P}^*}{\nu}\left[\left(\frac{1}{A} + \frac{F_L^* L_A^*}{F^*}\right) + \frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n_A^*\right] \tag{109}$$

$$\text{Define } \Gamma = \frac{1}{A} + \frac{F_L^* L_A^*}{F^*} + \frac{\nu}{\mu^*(\mu^* - \nu)}\mu_n n_A^* \tag{110}$$

From the state-state labor per firm l^* we have

$$L^* = n^* \left[\frac{1}{A}\left(\frac{\rho}{\alpha}\right)^{\alpha}\left(\frac{1 - \frac{\nu}{\mu}}{\frac{\phi}{\mu}}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} \text{ thus } L_A^* = L^*\left(\frac{n_A^*}{n^*} + \frac{1}{\beta A}\right) \tag{111}$$

We can also substitute out the following simplifications

$$\frac{F_L^*}{F^*} = \frac{\beta K^{\alpha} L^{\beta-1} n^{-(\alpha+\beta)}}{K^{\alpha} L^{\beta} n^{-(\alpha+\beta)}} = \beta L^{*-1} \tag{112}$$

$$\mu = \frac{n}{n-1} \text{ therefore } \mu_n = -\frac{1}{n^2} \tag{113}$$

Therefore the allocation versus competition effect component becomes

$$\Gamma = \frac{2}{A} + \left(\beta - \frac{\nu}{\left(\frac{n^*}{n^*-1}\right)^2 (n^* - (n^* - 1)\nu)}\right)\frac{n_A^*}{n^*} \tag{114}$$

And short-run versus long-run productivity dynamics simplify to

$$\mathcal{P}(0)_A\big|_{x(0)=x^*} - \mathcal{P}_A^* = \left(\frac{1}{n^* - 1}\right)\frac{\mathcal{P}^*}{\nu}\left[\frac{2}{A} + \left(\beta - \frac{\nu}{\left(\frac{n^*}{n^*-1}\right)^2 (n^* - (n^* - 1)\nu)}\right)\frac{n_A^*}{n^*}\right] \tag{115}$$

$$\mathcal{P}^* = \nu\left[\frac{A(n^* - 1)^{\nu}}{n^*}\left(\frac{n^*(1-\nu) + \nu}{\phi}\right)^{1-\nu}\right]^{\frac{1}{\nu}} \tag{116}$$

F Jacobian

$$\mathbf{J} = \begin{bmatrix} \dot{C}_C^* & \dot{C}_e^* & \dot{C}_K^* & \dot{C}_n^* \\ \dot{e}_C^* & \dot{e}_e^* & \dot{e}_K^* & \dot{e}_n^* \\ \dot{K}_C^* & \dot{K}_e^* & \dot{K}_K^* & \dot{K}_n^* \\ \dot{n}_C^* & \dot{n}_e^* & \dot{n}_K^* & \dot{n}_n^* \end{bmatrix} = \begin{bmatrix} \frac{C^*}{\sigma} r_C^* & 0 & \frac{C^*}{\sigma} r_K^* & \frac{C^*}{\sigma} r_n^* \\ -\frac{\pi_C^*}{\gamma} & \rho & -\frac{\pi_K^*}{\gamma} & -\frac{\pi_n^*}{\gamma} \\ Y_C^* - 1 & 0 & Y_K^* & Y_n^* \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{117}$$

We have a system of four ODEs in four variables $\{C, e, K, n\}$ and labour is implicitly defined by these $L(K, n, e)$. Rather than per firm arguments for the production function, we take n outside the function using *Euler's Homogeneous Function Theorem.* [23] Since

- F is hod-ν then $F(\frac{K}{n}, \frac{L}{n}) = n^{-\nu} F(K, L)$
- and F is hod-$\nu - 1$ its first derivative $F_k(\frac{K}{n}, \frac{L}{n}) = n^{1-\nu} F_K(K, L)$
- and F is hod-$\nu - 2$ its second derivative $F_{kk}(\frac{K}{n}, \frac{L}{n}) = n^{2-\nu} F_{KK}(K, L)$
- $F_k k + F_l l = \nu F$
- $x F_{xx} + y F_{xy} = (\nu - 1) F_x$

[23]The subscript signifies which argument of the function the derivative is with respect to. So F_k is the derivative with respect to the first argument $\frac{K}{n}$ in $F(\frac{K}{n}, \frac{L}{n})$ and F_K is derivative with respect to first argument K in $F(K, L)$

$$\dot{C}_C = -\left(1 - \frac{u_C u_{CCC}}{u_{CC}^2}\right)((1-\mu)An^{1-\nu}F_K - \rho) + \frac{-u_C}{u_{CC}}(1-\mu)n^{1-\nu}AF_{KL}L_C \tag{118}$$

$$\dot{C}_e = 0 \tag{119}$$

$$\dot{C}_K = \frac{-u_C}{u_{CC}}(1-\mu)An^{1-\nu}(F_{KK} + F_{KL}L_K) \tag{120}$$

$$\dot{C}_n = \frac{-u_C}{u_{CC}}(1-\mu)A\left[(1-\nu)n^{-\nu}F_K + n^{1-\nu}F_{KL}L_n\right] \tag{121}$$

$$\dot{e}_C = (1-\mu)An^{1-\nu}F_{KL}L_C e - \frac{(1-(1-\mu)\nu)}{\gamma}An^{-\nu}F_L L_C \tag{122}$$

$$\dot{e}_e = (1-\mu)An^{1-\nu}F_K \tag{123}$$

$$\dot{e}_K = (1-\mu)An^{1-\nu}(F_{KK} + F_{KL}L_K)e - \frac{(1-(1-\mu)\nu)}{\gamma}An^{-\nu}(F_K + F_L L_K) \tag{124}$$

$$\dot{e}_n = (1-\mu)A[(1-\nu)n^{-\nu}F_K + n^{1-\nu}F_{KL}L_n]e - \frac{1-(1-\mu)\nu}{\gamma}A[(-\nu)n^{-\nu-1}F + n^{-\nu}F_L L_n] \tag{125}$$

$$\dot{K}_C = n^{1-\nu}AF_L L_C - 1 \tag{126}$$

$$\dot{K}_e = -\gamma e \tag{127}$$

$$\dot{K}_K = n^{1-\nu}A(F_K + F_L L_K) \tag{128}$$

$$\dot{K}_n = A[(1-\nu)n^{-\nu}F + n^{1-\nu}F_L L_n] - \phi \tag{129}$$

$$\dot{n}_C = 0 \tag{130}$$

$$\dot{n}_e = 1 \tag{131}$$

$$\dot{n}_K = 0 \tag{132}$$

$$\dot{n}_n = 0 \tag{133}$$

$$\begin{bmatrix} \frac{C^*}{\sigma}(1-\mu)\frac{A}{n}F_{kl}L_C & 0 & \frac{C^*}{\sigma}(1-\mu)\frac{A}{n}[F_{kk} + F_{kl}L_K] & \frac{C^*}{\sigma}(1-\mu)\frac{A}{n}[(1-\nu)F_k + F_{kl}L_n] \\ (1-\mu)\frac{A}{n}F_{kl}L_C e - \frac{AF_l L_C(1-(1-\mu)\nu)}{\gamma n} & (1-\mu)AF_k & (1-\mu)\frac{A}{n}[F_{kk} + F_{kl}L_K]e - \Xi & (1-\mu)\frac{A}{n}[(1-\nu)F_k + F_{kl}L_n]e - \Psi \\ AF_l L_C - 1 & -\gamma e & A(F_k + F_l L_K) & (1-\nu)AF - \phi + AF_l L_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{134}$$

where $\Xi = \frac{A(F_k + F_l L_K)(1-(1-\mu)\nu)}{n\gamma}$ and $\Psi = \frac{A(-\nu F + F_l L_n)(1-(1-\mu)\nu)}{n\gamma}$. Then in steady state we have $e = 0$, $F_k = \frac{\rho}{A(1-\mu)}$, and $F = \frac{\phi}{A(1-(1-\mu)\nu)}$

$$\begin{bmatrix} \frac{-u_C}{u_{CC}}(1-\mu)\frac{A}{n}F_{kl}L_C & 0 & \frac{-u_C}{u_{CC}}(1-\mu)\frac{A}{n}[F_{kk}+F_{kl}L_K] & \frac{-u_C}{u_{CC}}(1-\mu)\frac{A}{n}[(1-\nu)F_k+F_{kl}L_n] \\ -\frac{AF_lL_C(1-(1-\mu)\nu)}{\gamma n} & (1-\mu)AF_k & -\frac{A(F_k+F_lL_K)(1-(1-\mu)\nu)}{n\gamma} & -\frac{A(-\nu F+F_lL_n)(1-(1-\mu)\nu)}{n\gamma} \\ AF_lL_C-1 & 0 & A(F_k+F_lL_K) & (1-\nu)AF-\phi+AF_lL_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{135}$$

There is a direct effect of μ which appears as $1-\mu$ in several elements of the Jacobian matrix. There is also an indirect effect which changes value of the functions F and u and their derivatives because they are valuated at a different steady state.

The μ parameter has a direct effect in two rows. In the first row the direct effect which is visible as $(1-\mu)$ decreases the absolute value of $\dot{C}$ responses to changes in the four equilibrium variables, so consumption is less prone to change i.e. consumption volatility dampened. In the second row, $(1-\mu)$ raises the absolute value response of $\dot{e}$ to C, K, n changes, and lowers the response to an e change. The change in entry rates are exacerbated, so a period of lots of entry activity, will be followed by a period of less entry relative to the perfect competition case. The direct effect of entry deviating from equilibrium causes a greater change in the rate of entry than if there were perfect competition.

The indirec effect of μ is to change the steady state, and therefore alter where the functions $F_{kk}, F_k, F_l, u_{CC}, u_C$ are evaluated. The labour derivatives, are defined in terms of the other variables earlier in the appendix. But although these marginal effects may increase or decrease their signs and thus the general dynamics will not change.

G Eigenvectors

To calculate the four eigenvectors solve $(\mathbf{J}-\lambda^i_j\mathbf{I})\mathbf{P}^i_j = 0$ for $\mathbf{P}^i_j$, where there are four separate cases to solve for and hence four eigenvectors, since there are four eigenvalues two stable $\mathbf{P}^S_1, \mathbf{P}^S_2$, two unstable $\mathbf{P}^U_1, \mathbf{P}^U_2$. Since the eigenvalues are chosen such that $|\mathbf{J}-\lambda\mathbf{I}| = 0$ and a zero determinant means the matrix $\mathbf{J}-\lambda\mathbf{I}$ is completely linearly dependent (perfectly coupled). Then the eigenvectors are unique only up to a scalar multiple. Hence choose $P_{1,4} = 1$ as the normalization. Then from row four it follows $P_{1,2} = \lambda^i_j$. With $P_{1,2} = \lambda^i_j, P_{1,4} = 1$, we get from row 1 and 3

$$P_{1,1} = \frac{1}{\frac{Cr_C}{\sigma} - \lambda^i_j}\left[\frac{-Cr_n}{\sigma} - \frac{Cr_K}{\sigma}P_{1,3}\right] \tag{136}$$

$$P_{1,1} = \frac{1}{Y_C - 1}\left[-Y_n - (Y_K - \lambda^i_j)v_{1,3}\right] \tag{137}$$

Equating and solving

$$P_{1,3} = \frac{\frac{C}{\sigma} r_n (Y_C - 1) - Y_n(\frac{C}{\sigma} r_C - \lambda_j^i)}{(\frac{C r_C}{\sigma} - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma} r_K (Y_C - 1)} \tag{138}$$

Plug back in

$$P_{1,1} = \frac{\frac{C}{\sigma}(r_K Y_n - r_n(Y_K - \lambda_j^i))}{(\frac{C}{\sigma} r_C - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma} r_K (Y_C - 1)} \tag{139}$$

So our eigenvector

$$\begin{bmatrix} P_{i,1} \\ P_{i,2} \\ P_{i,3} \\ P_{i,4} \end{bmatrix} = \begin{bmatrix} \frac{\frac{C}{\sigma}(r_K Y_n - r_n(Y_K - \lambda_j^i))}{(\frac{C}{\sigma} r_C - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma} r_K (Y_C - 1)} \\ \lambda_j^i \\ \frac{\frac{C}{\sigma} r_n (Y_C - 1) - Y_n(\frac{C}{\sigma} r_C - \lambda_j^i)}{(\frac{C r_C}{\sigma} - \lambda_j^i)(Y_K - \lambda_j^i) - \frac{C}{\sigma} r_K (Y_C - 1)} \\ 1 \end{bmatrix} \tag{140}$$

G.1 Cost function

Static optimization problem so drop time subscripts, as in Jaimovich 2007 p309

$$C(r, w, y) = \min_{l,k} wl + rk + \phi \quad \text{s.t.} y \leqslant A k^\alpha l^\beta - \phi \tag{141}$$

With Cobb-Douglas production the total cost function from substituting Lagrangean obtained conditional input demands $k(r, w, y) = \left[\left(\frac{w\alpha}{r\beta}\right)^\beta \left(\frac{y+\phi}{A}\right)\right]^{\frac{1}{\alpha+\beta}}$ and $l(r, w, y) = \left[\left(\frac{r\beta}{w\alpha}\right)^\alpha \left(\frac{y+\phi}{A}\right)\right]^{\frac{1}{\alpha+\beta}}$ into the cost function is

$$C(r, w, y) = (\alpha + \beta) \left(\frac{y + \phi}{A}\right)^{\frac{1}{\alpha+\beta}} \left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + \phi \tag{142}$$

Where the firm takes factor prices as given. The average cost $AC := \frac{C}{y}$ is U-shaped and the marginal cost $MC := \frac{dC}{dy}$ is increasing in output with $\alpha + \beta < 1$.

$$MC = \frac{\partial C(r, w, y)}{\partial y} = \frac{(y + \phi)^{\frac{1}{\alpha+\beta} - 1}}{A^{\frac{1}{\alpha+\beta}}} \left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \tag{143}$$

$$\frac{\partial MC}{y} = \left(\frac{1}{\alpha + \beta} - 1\right) \frac{(y + \phi)^{\frac{1}{\alpha+\beta} - 2}}{A^{\frac{1}{\alpha+\beta}}} \left(\frac{r}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \left(\frac{w}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} \tag{144}$$

The leading multiplier $\frac{1}{\alpha+\beta} - 1$ determines how marginal cost responds to changing output. This shows that it is increasing when $\alpha + \beta < 1$ but is zero with constrant returns to scale $\alpha + \beta = 1$ which reflects a flat marginal cost curve.

H Markup Properties

If $\theta_I = 1$ industry goods are imperfectly substitutable, and therefore the aggregate good is a Cobb-Douglas composite of industry goods. The result is that the markup is a common asympototic function with eloquent properties see Wolfram Alpha for details.

Remark 1 (Endogenous markup)**.** *With many firms per industry the markup is 1*

Proof.

$$\lim_{n\to+\infty} \mu(n(t)) = \lim_{n\to+\infty} \frac{n(t)}{n(t)-1} \tag{145}$$

$$= \lim_{n\to+\infty} \left(\frac{1}{n(t)-1}\right) + 1 = 1, \quad n(t) \in (0, \infty] \tag{146}$$

□

References

Aloi, Marta and Huw Dixon (2003). "Entry Dynamics, Demand Shocks and Induced Productivity Fluctuations". In: *Ekonomia* 6.2, pp. 115–146 (cit. on p. 7).

Barnett, Alina, Sandra Batten, Adrian Chiu, Jeremy Franklin, and Maria Sebastia-Barriel (2014). "The UK productivity puzzle". In: *Bank of England Quarterly Bulletin*, Q2 (cit. on p. 2).

Berentsen, Aleksander and Christopher J. Waller (2009). *Optimal stabilization policy with endogenous firm entry*. Working Papers 2009-032. Federal Reserve Bank of St. Louis (cit. on pp. 7, 15, 16).

Bergin, Paul R. and Giancarlo Corsetti (2008). "The extensive margin and monetary policy". In: *Journal of Monetary Economics* 55.7, pp. 1222–1237 (cit. on p. 8).

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz (2012). "Endogenous Entry, Product Variety, and Business Cycles". In: *Journal of Political Economy* 120.2, pp. 304 –345 (cit. on pp. 7, 16).

Bils, Mark (1987). "The cyclical behavior of marginal cost and price". In: *The American Economic Review*, pp. 838–855 (cit. on p. 7).

Brito, Paulo and Huw Dixon (2013). "Fiscal policy, entry and capital accumulation: Hump-shaped responses". In: *Journal of Economic Dynamics and Control* 37.10, pp. 2123–2155 (cit. on p. 7).

Caballero, Ricardo J and Richard K Lyons (1992). "External effects in US procyclical productivity". In: *Journal of Monetary Economics* 29.2, pp. 209–225 (cit. on p. 7).

Campbell, Jeffrey R. and Hugo A. Hopenhayn (2005). "Market Size Matters". In: *Journal of Industrial Economics* 53.1, pp. 1–25 (cit. on p. 8).

Das, Sanghamitra and Satya P Das (1997). "Dynamics of entry and exit of firms in the presence of entry adjustment costs". In: *International Journal of Industrial Organization* 15.2, pp. 217–241 (cit. on pp. 6, 15, 16).

Datta, Bipasa and Huw Dixon (2002). "Technological Change, Entry, and Stock-Market Dynamics: An Analysis of Transition in a Monopolistic Industry". In: *American Economic Review* 92.2, pp. 231–235 (cit. on pp. 6, 7, 15, 18).

Dixit, Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity". English. In: *The American Economic Review* 67.3, pp. 297–308 (cit. on pp. 5, 13).

Dufourt, Frédéric, Kazuo Nishimura, and Alain Venditti (2013). *Indeterminacy and Sunspot Fluctuations in Two-Sector RBC models: Theory and Calibration*. Working Papers halshs-00796703. HAL (cit. on p. 6).

Ericson, Richard and Ariel Pakes (1995). "Markov-Perfect Industry Dynamics: A Framework for Empirical Work". English. In: *The Review of Economic Studies* 62.1, pp. 53–82 (cit. on pp. 4, 16).

Etro, Federico (2009). *Endogenous market structures and the macroeconomy*. Springer Science & Business Media (cit. on p. 6).

Etro, Federico and Andrea Colciago (2010). "Endogenous Market Structures and the Business Cycle*". In: *The Economic Journal* 120.549, pp. 1201–1233 (cit. on pp. 6, 7).

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" In: *American Economic Review* 98.1, pp. 394–425 (cit. on p. 24).

Goodridge, PR, Jonathan Haskel, and Gavin Wallis (2014). "The UK productivity puzzle is a TFP puzzle: current data and future predictions". In: (cit. on p. 2).

Hall, Robert E (1987). "Productivity and the business cycle". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 27. Elsevier, pp. 421–444 (cit. on p. 7).

Hall, Robert E (1989). *Invariance properties of Solow's productivity residual*. Tech. rep. National Bureau of Economic Research (cit. on p. 7).

Hamano, Masashige and Francesco Zanetti (2014). "Endogenous Establishment Destruction and Macroeconomic Dynamics". In: (cit. on p. 7).

Hopenhayn, Hugo A. (1992). "Entry, Exit, and firm Dynamics in Long Run Equilibrium". English. In: *Econometrica* 60.5, pp. 1127–1150 (cit. on p. 16).

Jaimovich, Nir (2007). "Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations". In: *Journal of Economic Theory* 137.1, pp. 300–325 (cit. on pp. 6, 7, 12, 45).

Jaimovich, Nir and Max Floetotto (2008). "Firm dynamics, markup variations, and the business cycle". In: *Journal of Monetary Economics* 55.7, pp. 1238 –1252 (cit. on pp. 6, 7, 17, 27).

Jovanovic, Boyan (1982). "Selection and the Evolution of Industry". In: *Econometrica* 50.3, pp. 649–70 (cit. on p. 16).

Lewis, Vivien (2009). "Business Cycle Evidence On Firm Entry". In: *Macroeconomic Dynamics* 13 (05), pp. 605–624 (cit. on pp. 7, 15, 16).

Lewis, Vivien and Céline Poilly (2012). "Firm entry, markups and the monetary transmission mechanism". In: *Journal of Monetary Economics* 59.7, pp. 670–685 (cit. on pp. 7, 16).

Lewis, Vivien and Arnaud Stevens (2015). "Entry and markup dynamics in an estimated business cycle model". In: *European Economic Review* (cit. on p. 7).

Loualiche, Erik (2014). *Asset pricing with entry and imperfect competition*. Tech. rep. Working paper (cit. on p. 7).

Mata, José and Pedro Portugal (1994). "Life Duration of New Firms". English. In: *The Journal of Industrial Economics* 42.3, pp. 227–245 (cit. on p. 16).

Peters, Michael (2013). "Heterogeneous mark-ups, growth and endogenous misallocation". In: (cit. on pp. 7, 24, 27).

Portier, Franck (1995). "Business Formation and Cyclical Markups in the French Business Cycle". English. In: *Annales d'Économie et de Statistique* 37/38, pp. 411–440 (cit. on p. 8).

Rotemberg, Julio J and Lawrence H Summers (1990). "Inflexible Prices and Procyclical Productivity". In: *The Quarterly Journal of Economics* 105.4, pp. 851–74 (cit. on p. 8).

Rotemberg, Julio J and Michael Woodford (1996). "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption". In: *American Economic Review* 86.1, pp. 71–89 (cit. on p. 12).

Rotemberg, Julio J. and Michael Woodford (1999). "The cyclical behavior of prices and costs". In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and M. Woodford. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 16, pp. 1051–1135 (cit. on p. 7).

Stokey, Nancy L (2008). *The Economics of Inaction: Stochastic Control models with fixed costs*. Princeton University Press (cit. on p. 15).

Vives, X. (1999). *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press (cit. on p. 8).