Cardiff Economics Working Papers

Working Paper No. E2017/8

Firm Entry, Excess Capacity and Aggregate Productivity

Anthony Savagar and Huw Dixon

July 2017

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Firm Entry, Excess Capacity and Aggregate Productivity

Anthony Savagar* Huw Dixon†

July 7, 2017

Abstract

Slow firm entry over the business cycle causes measured TFP to vary endogenously because incumbent firms bear shocks. Our main theorem states that imperfect competition and dynamic firm entry are necessary and sufficient conditions for these endogenous productivity fluctuations. The result focuses on the short-run absence of entry and incumbents' output response given this quasi-fixity. Quantitatively we show the endogenous productivity effect is as large as a traditional capital utilization effect.

JEL: E32, D21, D43, L13, C62, dynamic entry, endogenous productivity, endogenous sunk costs, business stealing, business cycle, continuous time

---

*Corresponding author a.savagar@kent.ac.uk. Step-by-step guides to computational results are available in Python notebooks accessible here https://github.com/asavagar/SavagarDixon_EntryCU_public. The paper was completed as part of my ESRC and RES funded PhD. It was previously titled *"The Effect of Firm Entry on Capacity Utilization and Macroeconomic Productivity"*. Notable thanks to Leo Kaas, Frédéric Dufourt, Akos Valentinyi, David Baqaee, Swati Dhingra, Mathan Satchi, Ben Heijdra, Vivien Lewis, Jo Van Biesebroeck, Ricardo Reis and conference participants at SED Edinburgh 2017.

†dixonh@cardiff.ac.uk

How does firm entry affect aggregate productivity? Many answers to this question explore the productivity-enhancing effects of firm entry decreasing market power (Jaimovich and Floetotto 2008; Edmond, Midrigan, and Xu 2015) or reallocating resources among heterogeneous firms (Clementi and Palazzo 2016; Hsieh and Klenow 2014; Atkeson and Burstein 2010). Additionally, in imperfectly competitive economies, an entrant incurring an overhead cost, stealing incumbents' business and charging a markup to cover the overhead degrades productivity too as incumbents' scale falls. If firm entry is dynamic then firms respond slowly to arbitrage profits after a shock, which creates short-run reprieve from this *business stealing* effect, and consequently a short-run period in which incumbents may utilize their excess capacity, earn monopoly profits, improve returns to scale and strengthen productivity. In this paper we develop this insight to provide a tractable theory of firm entry, capacity utilization and endogenous productivity over the business cycle without requiring firm heterogeneity, endogenous markups or increasing returns aggregation.

Our main result (Theorem 1) gives a necessary and sufficient condition for dynamic firm entry to cause endogenous variations in measured TFP following a technology shock. Imperfect competition is necessary because it creates increasing returns to scale as firms underutilize their overhead costs. Dynamic firm entry is necessary because it creates a short-run period for incumbents to exploit these increasing returns, free from business stealing. Quantitatively we show the size of the endogenous productivity effect is similar to a traditional capital utilization (endogenous depreciation) effect.

Consider a positive technology shock. With dynamic firm entry and standard capital accumulation, both capital and number of firms are fixed stocks in the short run (quasi-fixed).[1] Therefore the technology improvement is initially borne by incumbents with quasi-fixed capital. In order to maximize profits, these incumbents will instantaneously change their production (intensive margin) through labor, and in turn productivity varies through

[1] This is the Marshallian definition of the short-run: at least one factor of production is fixed, and firm entry is yet to adjust. Usually it is not present in macroeconomic models as firms are fixed or instantaneously adjust, despite capital usually being a quasi-fixed input.

returns to scale that arise under monopolistic competition. However, after the short-run period firms begin to enter to arbitrage profits, and they steal business which reverses the incumbents' alteration in intensive margin until it returns to its initial level in the long run. Thus entry reverse the short-run productivity fluctuation. Crucially, it is the short-run absence of entry that causes endogenous procyclical productivity, and entry then reverses this fluctuation.

In the literature on microproduction theory and efficiency analysis, capacity utilization is the ratio of actual output to some measure of potential output (full capacity) given a firm's short-run stock of capital and other quasi-fixed factors of production (Nelson 1989).[2] In our work the potential output benchmark will be the efficient firm size which minimizes long-run average cost (and arises under perfect competition). Introductory treatments of monopolistic competition refer to this as capacity output, and excess capacity is underproduction relative to this level (Hall and Lieberman 2009, Ch. 11). Crucially it is distinct from the same term often used in RBC research to mean the more specific concept of *capital utilization* and endogenous depreciation, which also creates endogenous productivity fluctuations (King and Rebelo 1999).

The implications of our model are that less competitive economies have greater excess capacity, greater returns to scale and greater productivity fluctuations. Under perfect competition these fluctuations do not arise. In figure 1 we show the close relationshop between output, capacity utilization and net entry for quarterly US data 1994-2013.[3] The correlation between GDP and net entry is 0.64 and between GDP and capacity utilization is 0.84. Morrison 1992; Berndt and Morrison 1981 and Berndt and Fuss 1986 emphasize the positive relationship between productivity and capacity utilization. Emerging evidence by Rossi and Chini 2016 finds that entry is procyclical and

[2]Morrison 2012 and Nadiri and Prucha 2001 give good overviews, also see footnote 5.

[3]Data are logged and HP-filtered at a quarterly frequency. Raw data is taken from Federal Reserve FRED database, and the unique MNEMONICS are GDP and TCU. The definition of TCU is *"Capacity utilization is the percentage of resources used by corporations and factories to produce goods in manufacturing, mining, and electric and gas utilities for all facilities"*. Net entry is calculated from quarterly firm birth and death figures taken from the Bureau of Labor Statistics BED program.

Figure 1: US Procylical Capacity Utilization and Net Entry

sluggish in US data 1977-2013, which reaffirms Campbell 1998.

We develop a Ramsey-Cass-Koopmans model with endogenous labor, capital accumulation, monopolistic competition and dynamic entry. Firm entry is dynamic because the value of incumbency and sunk costs of entry do not instantaneously equate, so profits are not instantaneously zero. Our setup achieves this with an endogenous sunk cost akin to a capital adjustment cost, in the sense that entry is more expensive the more entry is taking place. Since the sunk cost is increasing in entry, a prospective entrant has an incentive to delay entry if the net present value of incumbency is less than the sunk cost, and in turn the sunk cost will diminish in the future if less entry takes place today. In long-run steady state there is no role for firm dynamics. Entry is zero, profits are zero and the sunk cost is zero so the static outcomes are the same as a model without entry or sunk costs. Our interest is the short-run transition to this zero-profit, zero-entry steady state. Our model is deterministic and in continuous time and we study transition under an unexpected once-and-for-all technology shock.

The model has two state variables: capital and number of firms. Consequently capital per firm is quasi-fixed because state (predetermined) variables cannot adjust instantaneously. Whereas, labor (through consumption) and entry can jump on impact to put the economy on its stable manifold which is defined by capital and number of firms. Subsequently the economy evolves along this stable manifold as capital per firm adjusts. Our theoretical contribution is to show that our model which is defined by four endogenous variables (consumption, entry, capital, number of firms) always has two positive and two negative eigenvalues. This implies that it is always determinate, and has a two dimensional stable manifold (a saddle path exists), which formalizes the Marshallian definition of the short run in a macro DGE context.[4]

Related Literature The most relevant papers for this research are Datta and Dixon 2002; Jaimovich and Floetotto 2008 and Bilbiie, Ghironi, and Melitz 2012 (BGM). Datta and Dixon 2002 provide a continuous time dynamic entry model in partial equilibrium that we adapt to a business cycle DGE environment, as in Brito and Dixon 2013 under perfect competition. Jaimovich and Floetotto 2008 investigate endogenous productivity with firm entry through the channel of endogenous markups, but entry is static (instantaneous). BGM has popularized dynamic entry in business cycle modeling by providing a quantitative model that improves moment matching. This has been successfully adopted in several studies (Lewis and Poilly 2012; Etro and Colciago 2010). We use a different dynamic entry setup that allows tractability. It is conceptually and mathematically similar to BGM in the sense that there are two state variables in capital and number of firms. We exclude endogenous markups, instead using a monopolistic competition setup with fixed markups and firms with U-shaped average cost curves. Our firm setup is similar to Devereux, Head, and Lapham 1996 but we include dynamic entry and exclude increasing returns aggregation (aka returns to specialization or thick markets). They argue that static entry and returns to specialization cause endogenous productivity effects.

Hall 1986 emphasizes that measured TFP has important endogenous components. Hall 1987 explains productivity procylicality arises from variations in output per firm that lead to movements down the average cost curve. Our contribution is to microfound sluggish entry as an explanation for why this movement down the average cost curve happens temporarily, and link it to

[4] See footnote 1.

microproduction theory on capacity utilization and quasi-fixity.[5] We focus on the delay in the demand curve shifting, as firms are quasi-fixed, which causes short-run monopolistic profits and intertemporal capacity utilization effects. Whereas the papers by Devereux, Head, and Lapham 1996; Chatterjee and Cooper 1993 and Jaimovich and Floetotto 2008 explain how entry causes endogenous productivity movements by changing the slope of demand curves in the long run (which is equivalent to the short run since they have instantaneous firm entry) either through endogenous markups or increasing returns to aggregation. Our work generalizes those papers that have instantaneous entry with papers that have a fixed number of firms (Blanchard and Kiyotaki 1987; Hall 1990; Rotemberg and Woodford 1992; Hornstein 1993), so that rather than an immediate extensive margin adjustment, or a permanent intensive margin adjustment, in the short-run the intensive margin adjusts, but is unchanged in the long-run as the extensive margin compensates.[6] The business stealing effect of entry decreasing incumbents' output was introduced by Mankiw and Whinston 1986 in industrial organization literature, and international trade uses it to explain falling measured TFP of domestic producers following foreign entry Harrison and Aitken 1999.

1 Simple Model with Constant Marginal Cost

This section introduces the main idea of the paper in a partial equilibrium model with constant marginal costs and no capital. A firm produces with the following production function

$$y = A\frac{L}{n} - \phi \tag{1}$$

It pays an overhead cost ϕ and employs labor L which is divided equally among all n firms. Assuming a constant returns to scale aggregator, aggregate output is the number of producers multiplied by their production $Y = AL -$

[5] Morrison 2012; Nadiri and Prucha 2001; Berndt and Morrison 1981; Hulten 1986.

[6] These two cases arise in our model as limiting cases of the endogenous sunk cost.

$n\phi$, and defining productivity as output per unit of labor gives

$$\mathcal{P} = \frac{Ay}{y+\phi} \tag{2}$$

which implies productivity is increasing in output due to the overhead cost $\mathcal{P}_y = \frac{A\phi}{(y(t)+\phi)^2} > 0$.[7] The effect of an additional firm on aggregate output is $Y_n = y + ny_n$. The first term is the entrants own output contribution (1) (the extensive margin contribution), and the second term $ny_n = -A\frac{L}{n}$ represents aggregate business stealing (the intensive margin contribution). Therefore the overall effect of an entrant on aggregate output is $Y_n = -\phi$ which is negative and equal to the fixed cost an entrant incurs.

Per firm profits are revenue less costs, where output price is the numeraire, $\pi = y - w\frac{L}{n}$. Under imperfect competition wages are a fraction of marginal products $w = (1-\zeta)A$ where $\zeta \in (0,1)$ is the Lerner Index of market power, and A is the marginal product of labor.[8] At this imperfectly competitive wage, profits are increasing in market power as costs diminish:

$$\pi = y - (1-\zeta)(y+\phi). \tag{3}$$

The free-entry condition on firm dynamics implies that in the long-run firms enter the market to arbitrage profits to zero which gives steady state output and labor per firm

$$y^* = \frac{1-\zeta}{\zeta}\phi, \qquad \frac{L^*}{n^*} = \frac{\phi}{A\zeta}, \qquad \zeta \in (0,1),\ A > 0 \tag{4}$$

so steady-state productivity is $\mathcal{P}^* = (1-\zeta)A$. Firms' output is decreasing in market power because with an increasing markup between price and marginal cost firms can produce less and still cover their fixed cost of production in zero profit equilibrium. Per firm production (firm size) is independent of technology A in the long run. An improvement in technology raises profits

[7]Subscripts denote total derivatives e.g. $\mathcal{P}_y = \frac{d\mathcal{P}}{dy}$.

[8]The Lerner index measures the difference between price and marginal cost as a proportion of price $\zeta = \frac{p-mc}{p} \in (0,1)$. In related literature, macroeconomists have often preferred the markup notation $\mu = \frac{p}{mc} \in (1,\infty)$, implying $w = \frac{1}{\mu}A$.

and therefore entry until all firms return to producing the same output. The intensive margin y^* is fixed, but the extensive margin $Y^* = n^* y^*$ adjusts by the change in number of firms. From (2) the effect of a technology shock on productivity is

$$\mathcal{P}_A(t) = \frac{\mathcal{P}(t)}{A} + \mathcal{P}_y(t) y_A(t) \tag{5}$$

so it depends on the direct effect of a shift in the level of technology $\frac{\mathcal{P}(t)}{A} = \frac{y(t)}{y(t)+\phi}$ and it depends on the indirect response of output per firm y_A interacted with returns to scale $\mathcal{P}_y$. Therefore if the economy begins in steady state $y(0) = y^*$, before a new technological advancement, the short-run effect of the shock evaluated at steady state is

$$\mathcal{P}_A(0)|^* = \frac{y^*}{y^* + \phi} + \frac{A\phi}{(y^* + \phi)^2} y_A(0)|^* \tag{6}$$

And since it will return to its initial level in the long run $y(t \to \infty) = y^*$ and $y_A(\infty) = y_A^* = 0$ and $\mathcal{P}_A(\infty) = \mathcal{P}^*$, the effect of the shock in the long run is

$$\mathcal{P}_A^* = \frac{y^*}{y^* + \phi} \tag{7}$$

Therefore comparing (6) and (7), shows there is a short-run endogenous productivity effect.

$$\mathcal{P}_A(0)|^* - \mathcal{P}_A^* = \frac{A\zeta^2}{\phi} y_A(0)|^*, \qquad \phi > 0 \tag{8}$$

Equation (8) is analogous to the main result of the paper (Theorem 1). It states that, given increasing returns ($\phi > 0$), imperfect competition ($\zeta \in (0, 1)$) combined with short-run output variation of incumbents ($y_A(0)|^* \neq 0$) causes endogenous productivity effects in response to a technology shock. The sign of the effect (overshooting or undershooting) depends on output per firm's short-run response. Therefore imperfect competition and short-run output variation are necessary and sufficient conditions. We argue that the short-run intensive margin response ($y_A(0)|^* \neq 0$) arises because of slow firm entry.[9] That is, in the short-run absence of entry incumbents' output varies. But in the long-run, entry will adjust, which causes business stealing, until profits are zero so output per firm returns to its original level.

1.1 Output per Firm Variation

To understand the intertemporal variation in firms' intensive margin which creates endogenous productivity fluctuations, recall that the production function of a firm has one input (labor per firm $\frac{L}{n}$), and thus varies through aggregate labor and total number of firms. For now, assume that labor is determined endogenously $L(A; C(A))$. It depends on technology directly, and indirectly through consumption $C(A)$. Therefore the effect of a technology shock on output per firm comes through three channels:

$$y_A(t) = \frac{L(t)}{n(t)} \left[1 + \varepsilon_{LA}(t) - \varepsilon_{nA}(t)\right] \tag{9}$$

where notation $\varepsilon_{XY} = X_Y \frac{Y}{X}$ is the elasticity of X with respect to Y.[10] In the short run, firms are slow to respond (quasi-fixed) so only labor adjusts

$$y_A(0)|^* = \frac{\phi}{\zeta A} \left[1 + \varepsilon(0)_{LA}|^*\right] \tag{10}$$

therefore the negative business stealing effect of entrants on incumbents' output is not present.[11] The short-run response of output per firm only depends on the short-run elasticity of labor. This will be positive with a dominant substitution effect, or negative with a dominant income effect (assuming leisure is a normal good).[12] Hence from substituting into (8) the size and sign of the endogenous productivity effect depend on market power, and the initial

[9] In general equilibrium we endogenously generate this through an endogenous sunk cost.

[10] The result follows from taking the derivative of $y = A\frac{L}{n} - \phi$ with respect to A: $y_A = \frac{L}{n} + \frac{A}{n^2}[nL_A - Ln_A]$.

[11] Assuming technology creates entry $n_A > 0$, not exit $n_A < 0$.

[12] A sufficiently strong income effect $\varepsilon(0)_{LA}|^* < -1$ reduces output per firm and causes productivity undershooting.

elasticity of labor to technology.

$$\mathcal{P}_A(0)|^* - \mathcal{P}_A^* = \zeta(1 + \varepsilon(0)_{LA}|^*) \tag{11}$$

With instantaneous entry this short-run effect is not present because number of firms respond instantaneously to ensure the change in labor is offset by a change in number of firms such that labor per firm is instantaneously at the scale that causes zero profits in (3).

The remainder of the paper formalizes this insight in a more realistic environment, with capital, increasing marginal costs, and microfoundations for slow firm entry and imperfect competition. Unlike the constant marginal costs of this simple example, increasing marginal costs teamed with an overhead cost create an efficient firm scale (full capacity benchmark). Thus we shall view short-run variations in incumbents' intensive margin as *excess capacity utilization*. This is a measurable concept which ties our work to a long-tradition in econometrics and microproduction theory Berndt and Morrison 1981.

2 Full Model

2.1 Household

The economy is inhabited by a continuum of infinitely-lived identical household who maximize utility subject to a resource constraint.

$$\max \mathrm{U} := \int_0^\infty u(C(t), 1 - L(t)) \exp(-\rho t) dt \tag{12}$$

$$\text{s.t.} \quad \dot{K}(t) = rK(t) + wL(t) + \Pi(t) - C(t) \tag{13}$$

Individual utility $u : \mathbb{R}_+ \times [0,1] \to \mathbb{R}_+$ is strictly increasing in consumption $u_C > 0$ and strictly decreasing in labor $u_L < 0$. Both goods are normal $u_{CC},\ u_{LL} < 0$, so marginal utility of consumption and disutility of labor are diminishing, and utility is additively separable $u_{CL} = 0$. exp is the exponential function and $\rho \in (0,1)$ is the discount factor over time $t \in \mathbb{R}_+$.[13] The household owns capital $K \in \mathbb{R}_+$, which does not depreciate, and it takes equilibrium rental rate r and wage rate w as given by the market rate (determined in section (2.2)).[14] Households own firms and receive firm profits $\Pi \in \mathbb{R}$. Solving the optimization problem simplifies to three conditions for optimal consumption and labor.[15] They are the intertemporal consumption Euler equation (14), intratemporal labor-consumption trade-off (15) and the resource constraint (13).

$$\dot{C} = \frac{C}{\sigma(C)}(r - \rho), \quad \text{where } \sigma(C) = -C\frac{u_{CC}(C)}{u_C(C)} \tag{14}$$

$$w = -\frac{u_L(L)}{u_C(C)} \tag{15}$$

The two boundary conditions for a unique solution are

$$K_0 = K(0) \tag{16}$$

$$\lim_{t\to\infty} K(t)u_C(t)\exp(-\rho t) = 0. \tag{17}$$

2.2 Firm Production and Strategic Interactions

There is monopolistic competition in the product market and perfect competition in the factor market, so firms are price setters for their output, and price takers for their inputs. Since each firm faces the same factor prices resources are divided equally among firms in symmetric equilibrium. Per firm variables are in lower case where $n(t) \in \mathbb{R}_+$ is the measure of firms.

$$k(t) \equiv \frac{K(t)}{n(t)}, \tag{18}$$

$$l(t) \equiv \frac{L(t)}{n(t)} \tag{19}$$

[13]For clarity we follow the continuous time literature by suppressing time dependence $X(t)$ to X after initial introduction.

[14]Section 6 introduces depreciation.

[15]Supplementary appendix solves the Hamiltonian problem.

$Y : \mathbb{R}^2_+ \supseteq (n, y) \to \mathbb{R}_+$ is the final good and is a constant-returns CES aggregate of each $i \in n$ firms' output.

$$Y(t) = n^{1-\frac{\theta}{1-\theta}} \left[\int_0^n y(i)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{1-\theta}} \tag{20}$$

A firm is a 1-firm industry, so $\theta \in (1, \infty)$ is intersector substitutability.[16] The $n^{1-\frac{\theta}{1-\theta}}$ component removes love-of-variety. With the unit price of the aggregate good as the numeraire the sectoral demand $y(i)$ directed at each 1-firm industry takes constant elasticity form

$$y(i) = p(i)^{-\theta} Y, \quad \forall i \in (0, n) \tag{21}$$

with inverse demand for industry i given by $p(i)$. Firms have the same production technology

$$y(t) = \max\{AF(k, l) - \phi, 0\} \tag{22}$$

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 $\phi \in \mathbb{R}_{++}$ is an overhead cost denominated in output. F has concavity properties $F_k, F_l, F_{kl} = F_{lk} > 0,\ F_{kk}, F_{ll} < 0,\ F_{kk}F_{ll} - F^2_{kl} > 0$, and $A \in \mathbb{R}_{++}$ is a technology parameter. This production function gives a U-shaped average cost curve because there are initally increasing returns from the overhead, but these diminish due to increasing marginal costs in the production function.[17] The overhead cost is the nonconvexity which prevents some firms producing, and it occurs each period, which distinguishes it from the sunk entry cost that is paid once to enter (see section 2.3). The

[16] Sector, firm, industry and product are synonyms in this model.

[17] The increasing marginal costs assumption ($v < 1$) is necessary for existence of a perfectly competitive equilibrium when $\phi > 0$. As in Section 1, if we study constant returns ($\nu = 1$), results only exist under imperfect competition $\zeta \in (0, 1)$. This precludes the pefect competition benchmark $\zeta = 0$ we use to study capacity utilization. On some occasions we shall remark on the constant returns case, but assume $\nu \in (0, 1)$ unless otherwise stated.

increasing returns that $\phi > 0$ causes is a common outcome in the firm entry in macroeconomics literature (e.g. Jaimovich and Floetotto 2008; Devereux, Head, and Lapham 1996).[18] Under symmetry aggregate output is[19]

$$Y(t) = n(t)y(t) \tag{23}$$

From (22), firm production is homogeneous of degree 0 (hod-0) in aggregate inputs (K, L, n), whereas the aggregate production function (23) is hod-1 in (K, L, n). For example, double all inputs (capital, labor and number of firms): firm output is unaffected, but aggregate output doubles since there are twice as many firms all producing the same output. Therefore the intensive margin y is unchanged, but the extensive margin Y doubles.

Under monopolistic competition a firm maximises profits subject to sectoral demand (21), taking real wage w and interest rate r as given. The result is the following factor market equilibrium, where the Lerner Index is the inverse of intersector substitutability $\zeta = \frac{1}{\theta} \in [0, 1)$.

$$AF_k(k, l)(1 - \zeta) = r \tag{24}$$

$$AF_l(k, l)(1 - \zeta) = w \tag{25}$$

This shows that the marginal revenue product of capital equates to the cost of capital and the marginal revenue product of labor equates to the wage. The Lerner Index of market power is the difference between price and marginal cost as a proportion of price $(\frac{P-MC}{P})$. The limits capture no market power $\zeta = 0$ when goods are highly substitutable (perfectly elastic demand) and total market power $\zeta \to 1$ when goods are completely differentiated.[20]

---

[18] As in Jaimovich 2007; Rotemberg and Woodford 1992; Devereux, Head, and Lapham 1996; Chatterjee and Cooper 1993 the role of overhead costs is to reproduce zero profits despite market power.

[19] From (20), re-parameterizing the multiplier to $n^{\kappa - \frac{\theta}{1-\theta}}$ introduces external increasing returns to scale at the aggregate level $Y = n^{\kappa} y^{*}$. Caballero and Lyons 1992 set $\kappa = 1.30$ reflecting thick market effects. Devereux, Head, and Lapham 1996 investigate endogenous productivity through this channel.

[20] In terms of a 'price-over-marginal-cost' markup $\mu = \frac{1}{1-\zeta} = \frac{\theta}{\theta-1}$.

2.2.1 Costs, Operating Profit and TFP

Under the imperfectly competitive factor market outcomes, total variable costs are decreasing in imperfect competition ζ.[21]

$$wl + rk = (1 - \zeta)\nu AF(k, l) \tag{26}$$

Conversely operating profits $\pi(t) = y - wl - rk$ are increasing in imperfect competition

$$\pi = (1 - (1 - \zeta)\nu)AF(k, l) - \phi \tag{27}$$

The extra profit $\zeta\nu AF(k, l)$ from imperfect competition relative to perfect competition causes a static inefficiency which can lead to excessive entry. There is a distortion between the benefit of an extra producer to the consumer, and the profit incentive of an entrant. Rearranging (27) shows that firm output varies positively with current operating profits

$$y(t) = \frac{\pi(t) + \nu(1 - \zeta)\phi}{1 - (1 - \zeta)\nu} \tag{28}$$

Proposition 1. *Aggregate output can be expressed as a function of inputs and measured TFP.*

$$Y(t) = \mathrm{TFP}(t)F(K, L)^{\frac{1}{\nu}} \tag{29}$$

where

$$\mathrm{TFP}(t) \equiv \left(\frac{A}{\pi(t) + \phi}\right)^{\frac{1}{\nu}} (1 - (1 - \zeta)\nu)^{\frac{1}{\nu}-1}[(1 - \zeta)\nu\phi + \pi(t)] \tag{30}$$

Proof. See Appendix C □

The inclusion of operating profits in measured TFP leads to endogenous measured TFP dynamics when profits are not instantaneously zero.[22] If prof-

[21] Using Euler's homogeneous function theorem that $F_l l + F_k k = \nu F(k, l)$ then the result follows from substitution of factor prices $AF_l(1 - \zeta)l + AF_k(1 - \zeta)k$.

[22] The result generalizes Jaimovich and Floetotto 2008, eq. 17 appendix. They acknowl-

its were instantaneously zero, then (31) would be fixed. As per firm output and operating profits are in a one-one mapping from (28), we shall interpret this endogenous TFP movement through changes in y (primal approach), so-called capacity utilization.[23] The relationship is convex which relates to the U-shaped AC curve.[24]

$$\mathrm{TFP}_{\pi} = A^{\frac{1}{\nu}}(1-(1-\zeta)\nu)^{\frac{1}{\nu}-1}(1-\nu)\left[\frac{\zeta\nu\phi}{(1-\nu)}-\pi\right] \tag{31}$$

As Jaimovich and Floetotto 2008 have argued, and Etro and Colciago 2010 acknowledge, the standard Solow residual is an upward biased measure of technology in the presence of endogenous markups that respond to entry and exit. In this paper we have a fixed markup so that bias is not present, but we explore the bias that arises due to short-run non-zero profits and resulting capacity utilization.[25] Unlike endogenous markup biases that are present in both the short-run and the long-run because of changes in the slope of demand curves, capacity utilization biases are present only in the short-run because it delays the shift in the demand curve, which will move once entry and thus business stealing take place to arbitrage profits to zero.

2.3 Firm Entry

The number of firms at time t is determined by two assumptions: an endogenous sunk cost of entry (congestion effect) and an arbitrage condition that equates entry cost with incumbency profits (value of an incumbent).

The congestion effect states that entry sunk cost $q \in \mathbb{R}$ increases with the

edge the bias it creates, but their focus is on endogenous markup bias. They have constant returns to scale $\nu = 1$ which gives $\frac{Y}{F(K,L)} = A\left[1-\zeta\frac{\phi}{\pi+\phi}\right] = A(1-\zeta)\left[1+\frac{\zeta}{1-\zeta}\frac{\pi}{\pi+\phi}\right]$.

[23] This scale-adjusted TFP definition ($Y/F^{\frac{1}{\nu}}$), where the denominator is normalized to make the production function hod-1 as opposed to hod-ν, is widely used with increasing returns and instantaneous entry (Da-Rocha, Tavares, and Restuccia 2017; Barseghyan and DiCecio 2011). Basu and Fernald 2001 give a detailed discussion of scale adjusted productivity, whilst Harrison 1994 and Feenstra 2003 derive a similar measure for regression analysis.

[24] Under $\pi = 0$ productivity is increasing in profits reflecting production to the left-hand side of the minimum AC.

[25] This effect is present in Bilbiie, Ghironi, and Melitz 2012 but is not developed.

flow of entrants $\dot{n}$ in t.

$$q(t) = \gamma \dot{n}, \qquad \gamma \in (0, \infty) \tag{32}$$

The process is symmetric, a prospective firm pays $q(t)$ to enter at t and $-q(s)$ to exit at $s > t$.[26] $\dot{n}$ is the change in the stock of firms so represents net business formation; we define this as 'entry'. γ are dynamic barriers to entry (DBE) that reflect the sensitivity of sunk costs to net entry. They can be interpreted as regulatory costs. When a firm wishes to setup it must access a resource that is in inelastic supply (like a government office), so that if more firms are entering this process is slower and sunk costs are higher. Its bounds capture the limiting cases of entry: $\gamma \to 0$ implies instantaneous free entry because the sunk cost is small so the outcome is similar to the static case, and $\gamma \to \infty$ implies fixed number of firms because the sunk cost is so high that it prohibits entry. The congestion effect assumption has been used in the industrial organization literature (Das and Das 1997), and it is growing in usage in macroeconomics. Recent examples in macroeconomics are Lewis 2009, Berentsen and Waller 2015 and in trade Bergin and Lin 2012.

The second assumption is entry arbitrage. It states the gain from entry equals return from investing the cost of entry at the market rate.

$$r(C, K, n)q(n) = \pi(C, K, n) + \dot{q}(n) \tag{33}$$

The arbitrage condition is a continuous time Bellman equation. It implies that the return to investing in a firm is equal to the return of that investment at the market rate r. And an implication of this is that the value of a firm is equal to present discounted value of future profits as in Bilbiie, Ghironi, and Melitz 2012 and Datta and Dixon 2002. The assumption that the value of the firm is equal to the sunk cost is known as free entry.[27]

[26] If there is exit $\dot{n} < 0$, so $q < 0$ and $-q > 0$ this means an incumbent pays a dismantling fee to exit, for example redundancy payments or legal fees. Sunk cost symmetry is not a necessary feature of the model. It eases exposition as we only focus on deterministic shocks in a single direction. To generalize the process for asymmetric costs, γ must differ for entry and exit.

[27] The continuous time value function (CTB) $\rho V = \pi + \dot{V}$ can be derived from an ex-

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

$$\gamma \ddot{n} - \gamma r(n)\dot{n} + \pi(n) = 0 \tag{34}$$

To interpret this second-order ODE consider that if profits are high, then to maintain equilibrium the speed of net business formation $\dot{n}$ is high which translates to higher sunk entry cost thus discouraging future entry so net business formation decelerates $\ddot{n} < 0$ to maintain equilibrium. By defining entry as the net change in stock of firms $(e(t) \equiv \dot{n})$, this second-order ODE is separable into two first-order ODEs. Hence our model of industry dynamics, which determines the number of firms, is defined by two ODEs, and requires two boundary conditions for uniqueness

$$\dot{n} = e \tag{35}$$

$$\dot{e} = -\frac{\pi}{\gamma} + re, \qquad \gamma > 0 \tag{36}$$

$$\lim_{t\to\infty} n(t)q(t)u_C(t)\exp(-\rho t) = 0 \tag{37}$$

$$n(0) = n_0 \tag{38}$$

The endogenous sunk cost causes a non-instantaneous adjustment path to steady state, which provides an analytical framework to understand short-run dynamics. It creates an incentive to delay entry as congestion effects will fall in the future. Contrarily, take an exogenous entry cost $q = \gamma \geqslant 0$. The second-order ODE becomes static $\pi = r\gamma$ and entry will adjust instantaneously to equate net present value of operating profits with the opportunity cost (sunk cost invested at market rate).

ponential discounting problem, see Stokey 2008, Ch. 3. Therefore the arbitrage condition can be derived from stating that incumbent firm value is the integral of future discounted profits from which the CTB follows. Then imposing the free entry assumption that value of incumbency equals sunk costs $V = q$ would give (33).

2.3.1 Sunk Entry Costs in General Equilibrium

To understand firm dynamics in general equilibrium, the aggregate investment in firms must be accounted for in terms of aggregate output (a market clearing condition). Integrating the sunk cost of entry across all entrants in a period gives the aggregate cost of entry in terms of output

$$Z = \gamma \int_0^e i \, di = \gamma \frac{e^2}{2} \tag{39}$$

This will appear as a quadratic adjustment cost in entry in the aggregate equation of motion for capital when we substitute out aggregate profits.

$$\Pi = n\pi - Z \tag{40}$$

Aggregate profits are all firms' operating profits (27) less the aggregate sunk cost. Substituting Π into the household resource constraint (13) gives $C + \dot{K} + Z = rK + wL + n\pi$, which states expenditure equates to income. Expenditure is divided between consumption, investment in capital[28], and adjustment costs paid to setup firms Z. Income is earned from capital, labor and operating profits from firm ownership. Substituting factor prices and profits into the right-hand side gives aggregate output Y and rearranging gives the aggregate equation of motion for capital

$$\dot{K} = Y - C - \gamma \frac{e^2}{2} \tag{41}$$

2.4 Model Summary

General equilibrium determines prices, consumption, entry and labor given the current capital stock and number of firms. Labor is defined as $L(C, K, n)$ through the static intratemporal condition (15), so by substitution the model reduces to a dynamical system of four ordinary differential equations (ODEs) (14), (36), (41), (35) in four variables (C, e, K, n). Additionally there are two initial conditions (16, 38) and two transversality conditions (17, 37) which

[28] $I = \dot{K}$ since there is no depreciation.

provide the four boundary conditions necessary for a solution to the four dimensional dynamical system. Appendix A outlines the model equilibrium conditions recursively.

Proposition 2 (Instantaneous Entry Reduced Form)**.** *Without dynamic barriers to entry ($\gamma = 0$), entry adjusts instantaneously, profits are always zero, and output per firm is fixed*

$$y(t) = \frac{\nu(1-\zeta)\phi}{1-(1-\zeta)\nu}, \qquad \forall t \in (0, \infty)$$

The model reduces to a 2d-system with the dynamic properties of a Ramsey-Cass-Koopmans model with endogenous labor.

Proof. If barriers to entry are zero $\gamma = 0$, (34) implies $\pi(t) = 0$, which from (28) gives y. The equilibrium conditions reduce to two differential equations $\dot{K}, \dot{C}$, where the quadratic sunk entry cost in $\dot{K}$ is zero, as studied in Turnovsky 2000 with perfect competition. □

2.4.1 Labor Market

Given wage $w(L, K, n)$ at market equilibrium (25), the intratemporal condition (15) defines optimal labor supply statically as a function of consumption, capital and number of firms $L(C, K, n)$

$$AF_l(k, l)(1-\zeta) = -\frac{u_L(L)}{u_C(C)} \tag{42}$$

The intratemporal condition shows that the marginal rate of substitution between consumption and labor equates to the wage (negative because labor decreases utility). From the implicit function theorem, we can determine that labor supply increases in capital ($L_K > 0$) and number of firms ($L_n > 0$) and decreases in consumption ($L_C < 0$).[29] Labor is decreasing in consumption because a rise in consumption causes the marginal utility of consumption to fall (consumption is a normal good) therefore marginal disutility of labor must decrease to maintain the marginal rate of substitution, hence labor

[29] See Appendix B for derivations.

decreases which reduces disutility. Capital causes an increase in the labor supply through an increase in the marginal product of labor and hence real wage. The firm entry effect is more novel:

Proposition 3. *Firm entry increases labor supply $L_n > 0$.*

Proof. See Appendix B □

The result arises because production has increasing marginal costs $\nu \in (0, 1)$. Entry increases labor because an additional firm decreases employment per firm and therefore raises the marginal product of labor and hence wage, this dominates other general equilibrium channels. Additionally, with constant marginal costs (and for existence $0 < \zeta$), entry does not affect labor supply $L_n|^{\nu=1} = 0$. With constant returns the capital-labor ratio (hence MPL) is unaffected by entry, so labor supply is unresponsive.[30]

2.4.2 Business Stealing

Slow firm entry means that profits and output per firm are not instantaneously fixed, unlike the instantaneous entry case of Proposition 2. Instead they vary in the short-run, eventually reaching a zero profit, fixed output per firm level in the long run. Given the general equilibrium behaviour of the labor market, we can understand the general model predictions for output per firm (intensive margin) and therefore operating profits over the transitioning period.

Proposition 4. *Output per firm and operating profit are decreasing in consumption ($y_C < 0$), increasing in capital ($y_K > 0$) and decreasing in number of firms ($y_n < 0$).*

Proof. Appendix C □

[30] Nontrivial entry effects on labor follow Haltiwanger, Jarmin, and Miranda 2013 that firm births contribute substantially to net job creation. Most papers on firm entry disregard this channel by assuming constant marginal costs ($\nu = 1$) and no perfect competition ($\zeta = 0$).

Entry always decreases incumbents' intensive margin, which is equivalent to decreasing operating profits by relationship (28).[31] This implies that business stealing prevails at the intensive margin, despite the counteracting labor supply effect of Proposition 3.[32] The extensive margin effect is less clear:

Proposition 5. *An entrant has an ambiguous effect on aggregate output (the extensive margin). Whether entry increases, decreases or maximizes aggregate output depends on the trade-off between the negative business stealing effect (Proposition 4) versus the positive labor supply effect (Proposition 3).*

$$Y_n = y + ny_n \tag{43}$$

An entrant contributes its own output y, but also has a negative effect on the intensive margin of all n incumbents.

$$ny_n = -\nu A n^{-\nu} F(K, L) + A n^{1-\nu} F_L L_n < 0 \tag{44}$$

The first term is the amount of resources the entrant steals from incumbents weighted by the efficiency gain of incumbents employing remaining inputs with lower marginal cost. The second effect is a labor supply increase that exists because of increasing marginal costs (Proposition 3). The entrant's own contribution y can be written as profits plus variable costs $y = \pi + (1 - \zeta)\nu A n^{-\nu} F(K, L)$. From (44) the amount it steals is $\nu A n^{-\nu} F(K, L)$ but in order to cover the new overhead ϕ that the entrant has incurred it resells the stolen output with a markup $1 - \zeta$. Hence the entrant steals $\nu A n^{-\nu} F(K, L)$ but then only adds $(1 - \zeta)\nu A n^{-\nu} F(K, L)$. The deadweight loss from the transfer in business is $-\zeta \nu A n^{-\nu} F(K, L)$ giving:

$$Y_n = \pi - \zeta \nu A n^{-\nu} F(K, L) + A n^{1-\nu} F_L L_n \tag{45}$$

[31] This would not be the case with love-of-variety where aggregate demand externalities play a role Acemoglu 2009, Ch. 12, such that profits can increase in entry.

[32] Mankiw and Whinston 1986 state *"[business stealing] exists when the equilibrium strategic response of existing firms to new entry results in their having a lower volume of sales–that is, when a new entrant "steals business" from incumbent firms. Put differently, a business-stealing effect is present if the equilibrium output per firm declines as the number of firms grows."*

Therefore the aggregate entry effect is the entrant's profit, less the deadweight loss from business stealing, plus the general equilibrium labor supply effect from higher wages.[33] By trading off the opposing effects of entry on business stealing and labor efficiency, an optimal ($Y_n = 0$) amount of entry can be achieved. In steady state we shall show profits are zero. Hence by (45), a judicious choice of parameters (ν, ζ) can maximize steady state aggregate output (which equals consumption with no depreciation).[34]

3 Steady State Behaviour

In steady state capital, number of firms and consumption are stationary $\dot{K} = \dot{n} = \dot{C} = \dot{e} = 0$. Therefore the dynamical system is stationary when aggregate supply equals demand $Y^* = C^*$ (as zero depreciation); net entry is zero $e^* = 0$; capital returns equal the discount factor $r^* = \rho$ and operating profits are zero $\pi^* = 0$.[35] Substituting out π^*, r^*, Y^* and using per firm definitions gives steady state conditions in terms of (C^*, K^*, n^*, e^*). Labor is a function of the system variables $L^*(C^*, K^*, n^*)$ through the static intratemporal condition (42), repeated here for steady state

$$(1-\zeta)AF_l\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = -\frac{u_L(L^*)}{u_C(C^*)} \tag{46}$$

[33]The expression generalizes Mankiw and Whinston 1986, eq. 2 to the aggregate economy with endogenous labor. They focus on a partial equilibrium industry setting with constant marginal costs. From (27) rewriting the result as $Y_n = (1-\nu)AF - \phi + An^{1-\nu}F_L L_n$ shows that with constant returns (and strict imperfect competition for existence $\zeta > 0$) the first and third terms are zero, so an entrant always decreases aggregate output $Y_n|^{\nu=1} = -\phi$ by the overhead cost it incurs.

[34]The supplementary appendix develops this idea, and gives an intuitive example for a parameterized model. This extends Etro and Colciago 2010 discussion of excessive entry ('dynamic inefficiency') in a similar model with endogenous markups, but without the offsetting labor supply effect from increasing marginal costs, which is what makes an optimal level attainable here in the absence of endogenous markups.

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

Therefore in steady state

$$\dot{C} = 0: \qquad (1-\zeta)AF_k\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = \rho \tag{47}$$

$$\dot{e} = 0: \qquad F\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) = \frac{\phi}{A(1-(1-\zeta)\nu)} \tag{48}$$

$$\dot{K} = 0: \qquad n^*\left[AF\left(\frac{K^*}{n^*}, \frac{L^*}{n^*}\right) - \phi\right] = C^* \tag{49}$$

$$\dot{n} = 0: \qquad e^* = 0 \tag{50}$$

The system determines (C^*, K^*, n^*, e^*). The entry arbitrage condition (48) implies profits are zero, which determines steady-state variable production and therefore firm output.

$$y^* = \frac{\nu(1-\zeta)\phi}{1-(1-\zeta)\nu} \tag{51}$$

Given y^*, the aggregate resource constraint (49) determines steady-state consumption in terms of n^* since $C^* = n^*y^*$, which gives labor $L^*(C^*(n^*), K^*, n^*)$ in K^*, n^* terms through the intratemporal condition (46). Thus (47) and (48) are in K^*, n^* terms and can be solved simultaneously.

Output per firm (51) is increasing in both fixed cost ϕ and returns to scale ν and is decreasing in market power ζ. Increasing market power raises marginal revenue products of inputs, so less needs to be produced in order to cover fixed costs and attain zero profits. A perfect competition ($\zeta = 0$) steady-state output exists because firms face a fixed cost and increasing marginal cost which leads to U-shaped average cost. In the next section we show that the perfectly competitive output coincides with maximization of measured productivity, and serves as our full capacity efficiency benchmark.

Steady state measured TFP is decreasing in market power ζ because it allows firms to suppress output more, so they exploit fixed cost returns to scale less.

$$\text{TFP}^* = A^{\frac{1}{\nu}}\nu(1-\zeta)\left(\frac{1-\nu(1-\zeta)}{\phi}\right)^{\frac{1-\nu}{\nu}} \tag{52}$$

3.1 Capacity Utilization and Efficient Benchmark

The microproduction literature refers to capacity utilization as temporary or sub-equilibrium changes in production that arise due to quasi-fixity of inputs. Input quasi-fixity causes disparities between shadow prices and actual prices that are captured by positive profit, which subsequntely cause adjustment of quasi-fixed inputs. In our model capital per firm is quasi-fixed because both capital and firms do not respond at time 0 to shocks. In models without dynamic firm entry, capital per firm is not quasi-fixed, despite quasi-fixed capital, because number of firms adjusts instantaneously.

Definition 1. *Capacity utilization is production relative to a full-capacity, efficiency benchmark*

$$CU(t) \equiv \frac{y(t)}{y^e} \tag{53}$$

Excess capacity is $EC(t) \equiv 1 - CU(t)$.

We take the efficiency benchmark to be production that maximizes measured TFP.

Proposition 6 (Efficiency Benchmark)**.** *The level of output that maximizes measured TFP is*

$$y^e = \frac{\nu\phi}{1-\nu} \tag{54}$$

which implies maximimum attainable productivity is

$$\mathrm{TFP}^e = A^{\frac{1}{\nu}}\nu\left(\frac{1-\nu}{\phi}\right)^{\frac{1-\nu}{\nu}} \tag{55}$$

These outcomes are attained under perfect competition.

Proof. The relationship between TFP and output is convex reflecting the U-shaped cost curve. Rearrange TFP $= \frac{y}{F(k,l)^{\frac{1}{\nu}}} = \frac{A^{\frac{1}{\nu}}y}{(y+\phi)^{\frac{1}{\nu}}}$, then take the derivative:

$$\text{TFP}_y = \left(\frac{A}{y+\phi}\right)^{\frac{1}{\nu}} \left[1 - \frac{y}{\nu(y+\phi)}\right] \tag{56}$$

Equating to zero and rearranging for output gives (54), then substitution gives (55). Equivalence with the perfect competition outcome follows from evaluating (51) and (52) with $\zeta = 0$. $\square$

If firms were to produce the efficient scale under imperfect competition they would earn positive profits $\pi^e = \zeta y^e$. However given these positive profits, firms continue to enter to arbitrage them to zero, and the resulting situation is smaller firms each with excess capacity.

Lemma 1 (Capacity Utilization and Excess Capacity)**.** *In zero-profit steady state firms competing under monopolistic competition $\zeta \in (0,1)$ have excess capacity.*

$$CU^* = \frac{y^*}{y^e} = 1 - \frac{\zeta}{1-(1-\zeta)\nu} < 1 \tag{57}$$

where excess capacity is $EC^* = \frac{\zeta}{1-(1-\zeta)\nu}$. Under perfect competition $\zeta = 0$, there is full capacity $CU^* = 1$ and no excess capacity $EC^* = 0$.

When there is excess capacity in steady-state this implies firms have locally increasing returns to scale, as they do no fully utilize their overhead cost. They produce below their most efficient scale on the left-hand side of their U-shaped average cost curve.

Lemma 2. *In excess capacity steady state $\zeta \in (0,1)$, there are locally increasing returns to scale. TFP is increasing in output per firm:*

$$\text{TFP}_y|^* = \zeta \left(\frac{A(1-(1-\zeta)\nu)}{\phi}\right)^{\frac{1}{\nu}} > 0, \tag{58}$$

Returns to scale are locally constant at the full capacity, perfectly competitive scale, $\zeta = 0 \implies \text{TFP}_y|^* = 0$.

Proof. Evaluate (56) at steady state (51). $\square$

The intuition of these results (efficiency, excess capacity, and increasing returns) can be understood diagrammatically. Figure 2 shows the U-shaped cost curves facing an incumbent firm. The long-run cost function plots minimum cost for each level of output, given labor and capital can adjust, whereas the short-run cost curve plots minimum cost given only labor can adjust, capital is fixed at the cost-minimizing capital for that output. y^* is less than y^e which represents excess capacity, and y^e minimizes long-run average costs which represents full capacity, efficient scale. The slope of the short-run cost function equals the slope of the long-run cost function at y^*, and it is downward sloping which represents increasing returns to scale because costs fall with output. The tangent at y^e is horizontal which implies locally constant returns to scale.

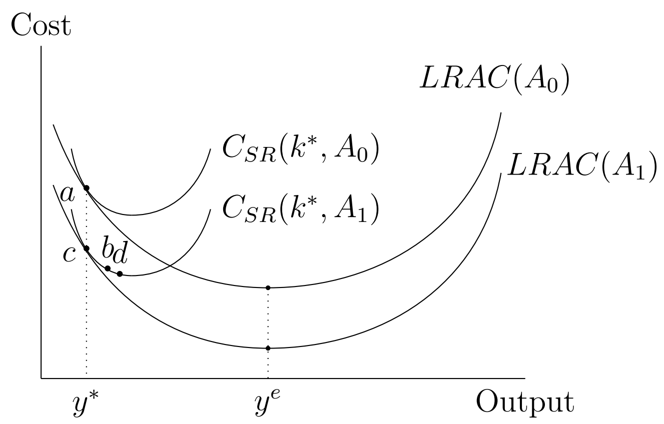

Figure 2: Long-run and short-run average cost curves

4 Technology and Capacity Utilization

In this section we analyze the effect of a technology shock on capacity utilization and productivity. We begin by giving a graphical description, and then we formalize this intuition in the model.

4.1 Graphical Explanation

Figure 2 plots the effect of a positive technology shock on an incumbent firms' costs, and consequently productivity response.[36] The economy begins in steady state at a with technology A_0. A technology improvement to A_1 instantly shifts the long-run and short-run cost curves downwards, but capital per firm remains at its initial steady state level k^* in the short-run whereas labor per firm $l(0, A_1)$ can adjust.[37] Therefore the SRAC curve cannot move along its LRAC envelope (as this requires a change in the firm's capital k), but production can vary along the given SR curve as labor changes instantaneously. Therefore a short-run capacity utilization mechanism arises as l changes causing a movement along the SRAC curve to a position like d. In the long run, entry occurs so that k and l adjust to return the incumbent to producing its fixed long-run level y^* at lower cost point c. Therefore the true change in costs, thus productivity, in the long-run is a to c. But in the short-run there will be a temporary movement to d as other firms adjust. This short-run effect is capacity utilization.

Under perfect competition or instantaneous entry the capacity utilization effect will not arise. Under perfect competition, production is always at minimum average cost, so the level shift is captured accurately because the tangent at minimum is horizontal implying no additional variation in costs. Under instantaneous entry, then capital per firm is no longer quasi-fixed. In this situation, profits are instantaneously zero as the downward shift in cost curve is accompanied by an outward shift in demand (inward shift in inverse demand) to arbitrage profits. The technology shock is only felt through factor demands and the immediate entry means the extensive margin of aggregate output $Y = ny^*$ adjusts immediately whereas the intensive margin y^* is unchanged (output per firm never deviates from y^*).

[36]The diagram abstracts from some complexities of the mathematics, for example we have plotted a parallel shift, but it conveys the intuition well.

[37]If neither k nor l adjust, there is still a scale effect change to b which is what we control for by making the denominator homogeneous of degree 1 in the measured TFP definition. That is, with fixed k, l there would be some movement of the SRAC along the LRAC, due to scale effects since $\nu < 1$ but we adjust for this, see footnote 23.

4.2 Model Derivation

In the long-run, free-entry, zero-profit steady state, a firm only produces enough to cover its fixed cost ϕ, so a positive technology shock allows a firm to combine fewer inputs to cover ϕ. Therefore the intensive margin is fixed, but the extensive margin will adjust. That is, technology does not affect the average firm size, but it will affect the number of firms and thus aggregate output.[38]

Proposition 7 (Long-run Effect of Technology)**.** *Long-run firm size, efficient scale, and therefore capacity utilization are independent of technology*

$$y_A^* = 0, \qquad y_A^e = 0, \qquad CU_A^* = 0$$

Since the intensive margin is fixed, the extensive margin adjusts by number of firms to accomodate technology

$$Y_A^* = C_A^* = n_A^* y^*$$

Proof. Take derivatives of (51), (54), (57) and $Y^* = n^* y^*$. □

To maintain a long-run fixed intensive margin per firm inputs adjust.

Corollary 1. *Labor per firm always decreases, whereas the effect on capital per firm is is ambiguous.*

$$l_A^* < 0 \tag{59}$$

$$k_A^* \lesseqgtr 0 \iff \frac{F_l}{F_{kl}} \gtreqless \frac{(1-\zeta)\phi}{\rho(1-(1-\zeta)\nu)} \tag{60}$$

Proof. Appendix C □

[38] Constant average firm size arises because fixed costs are unaffected by technology. Generalizing production to $y = AF(k,l) - A^\kappa \phi$ with $\kappa \in (-1,1)$ would give $y^* = A^\kappa \frac{\nu(1-\zeta)\phi}{1-(1-\zeta)\nu}$. Therefore κ determines whether firm size increases or decreases in response to A. We focus on $\kappa = 0$ as our main interest is transitional dynamics, so having y^* irresponsive to A focusses on short-run variations in y. Under the $\kappa \in (-1,1)$ setup, our results remain. They depend on differences between short-run and long-run changes in capacity, rather than only the short-run, but these are complex to track.

The relative effect of labor on production to labor on marginal product of capital determines capital per firm response to a technology shock.[39]

Theorem 1 (Endogenous Intertemporal Productivity)**.** *When firms have market power and entry is slow to adjust, a technology shock causes endogenous fluctuations in measured TFP as incumbents vary capacity utilization. The necessary and sufficient conditions are*

1. *Imperfect competition* $\zeta \in (0,1)$ *ensures there are locally increasing returns to scale.*

2. *Dynamic barriers to entry* $\gamma > 0$ *ensure slow firm entry so there are short-run variations in incumbent's capacity utilization.*

Proof. At t a change in technology will affect measured productivity directly and through a change in capacity utilization[40]

$$\mathrm{TFP}_A(t) = \frac{\partial \mathrm{TFP}(t)}{\partial A} + \mathrm{TFP}_y(t) y_A(t) \tag{61}$$

In the long-run there is no capacity utilization effect $y_A^* = 0$ (Proposition 7). Therefore only the first term is present. However, in the short run, beginning at steady state, both the long-run and capacity utilization effects remain[41]

$$\mathrm{TFP}_A(0)|^* = \mathrm{TFP}_A^* + \mathrm{TFP}_y^* y_A(0)|^* \tag{62}$$

The necessary conditions ensure the capacity utilization term $\mathrm{TFP}_y^* y_A(0)|^*$ is nonzero. Condition 1 follows from Lemma 2. Condition 2 follows from Proposition 2. □

[39] In our Cobb-Douglas production, CES preferences example $k_A^* = 0$. Therefore fixity of y^* after an increase in technology, follows solely from a decrease in l^*. Furthermore with logarithmic consumption utility long-run aggregate labor supply is irresponsive to technology $L_A^* = 0$, so firm entry is solely responsible for the fall in labor per firm $l_A^* = -\frac{L^*}{n_A^*}$.

[40] Analogously the result can be interpreted through profits $\pi_A = y_A(1-(1-\zeta)\nu)$, and $\mathrm{TFP}_A = \frac{\partial \mathrm{TFP}}{\partial A} + \mathrm{TFP}_\pi \pi_A$.

[41] Note the limit of the derivative is the derivative of their limits (e.g. $\lim_{t\to\infty} n_A(t) = n_A^*$), see Caputo 2005, p. 476. So in the limit the response of state variables is the same as the response of their steady state value.

We sketch the intuition of the main result in figure 3. Equation (62) shows that the short-run effect of a technology shock on measured productivity consists of the long-run effect plus short-run excess capacity utilization.[42] The simple direct effect TFP_A^* captures that improved technology shifts the production function which increases measured productivity both in the short run and the long run. The *excess capacity utilization effect* $\text{TFP}_y^* y_A(0)|^*$ captures the short-run capacity response $y_A(0)$ interacted with returns to scale $\text{TFP}_y(t)$.

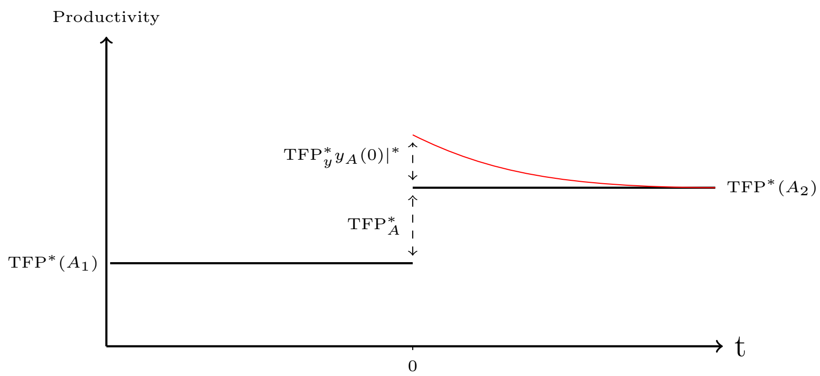

Figure 3: Endogenous Productivity

Corollary 2. *Given firm response $y_A(0)|^*$, greater imperfect competition implies greater excess capacity, greater returns to scale, and greater productivity fluctuations.*

Proof. Increasing imperfect competition ζ, decreases (57), increases (58), and consequently increases $|\text{TFP}_y^* y_A(0)|^*|$. $\square$

From figure 2, Corollary 2 formalizes that a less competitive firm produces further from minimum LRAC (more excess capacity), faces a steeper slope (stronger returns to scale) and thus a given change in output affects costs more, hence productivity fluctuates more.

[42]The result is still present with constant returns $\nu = 1$. The returns to scale component is simpler $\text{TFP}_y^* = \zeta\nu$ so $\text{TFP}_A(0)|^* = \text{TFP}_A^* + \zeta A y_A(0)|^*, \quad \zeta > 0$. But perfect competition outcomes do not exist in this setting, so there is no efficiency benchmark to measure excess capacity against.

Figure 3 shows the case of excess capacity utilization $(y_A(0))|^* > 0)$ creating an overshooting effect. However, a positive technology shock can create capacity widening $(y_A(0)|^* < 0)$ and undershooting if technological advancement strongly decreases labor supply due to a strong income effect. From (62):

$$y_A(0)|^* \gtreqless 0 \implies \text{TFP}_A(0)|^* \gtreqless \text{TFP}^*_A$$

In general output per firm response is[43]

$$y_A(t) = \frac{\partial y}{\partial A} + y_L \frac{\partial L}{\partial A} + y_n n_A + y_K K_A + y_C C_A \tag{63}$$

where each output response coefficient internalizes the labor effect, as given in Proposition 4. Assuming the primitive variables are monotonically increasing following an increase in technology $(n_A, K_A, C_A > 0)$, then technology has a positive direct effect $\frac{\partial y}{\partial A} > 0$, a positive MPL effect (substitution effect) $y_L \frac{\partial L}{\partial A} > 0$, a positive effect from capital accumulation $y_K = \frac{\partial y}{\partial K} + y_L L_K > 0$, a negative business stealing effect from entry $y_n = \frac{\partial y}{\partial n} + y_L L_n < 0$ and a negative effect from consumption crowding out (income effect) $y_C = y_L L_C < 0$. Proposition 7 implies that the positive and negative effects on output per firm in (63) cancel out in the long-run ensuring $y^*_A = 0$.[44] However in the short run, absence of negative business stealing effect may lead to overshooting.

Proposition 8. *A necessary and sufficient condition for overshooting is that the direct output effect and labor substitution effect collectively dominate the labor income effect. A sufficent condition is that the substitution effect dominates the income effect.*

Proof. If capital and number of firms are quasi-fixed (state variables), they do not respond to the shock in the short-run. Therefore the output response

[43] Which follows from $y_A = \frac{\partial y}{\partial A} + \frac{\partial y}{\partial n} n_A + \frac{\partial y}{\partial K} K_A + y_L L_A$ and substitution of $L_A = \frac{\partial L}{\partial A} + L_n n_A + L_K K_A + L_C C_A$.

[44] If $n^*_A, K^*_A, C^*_A > 0$, then a combination of the negative business stealing $(y_n < 0)$ and income effects $(y_C < 0)$ reduce output after any initial overshooting. In a logarithmic utility case, income and substitution effects exactly equate so $L^*_A = 0$ and business stealing is solely responsible for reducing output per firm.

(63) depends on the direct effect, and labor's immediate jump which consists of positive substitution and negative income effect.

$$y_A(0)|^* = \frac{\partial y}{\partial A} + y_L L_A(0) = \frac{\partial y}{\partial A} + y_L \left(\frac{\partial L}{\partial A} + L_C C_A(0) \right) \tag{64}$$

The necessary and sufficient condition is

$$\frac{\partial y}{\partial A} + y_L \frac{\partial L}{\partial A} > -y_L L_C C_A(0)$$

and the sufficient condition is $\frac{\partial L}{\partial A} > -L_C C_A(0)$ implying $L_A(0) > 0$. □

Therefore if technological advancement initially increases labor supply, then measured productivity overshoots its long-run level. Through profits, we can also understand the dual-interpretation of this: if the initial rise in A is offset by the fall in $L(0)$, then through (27) operating profits fall, which is analagous to capacity widening through (28).

5 Functional Forms and Simulations

In this section we specify functional forms and parameterizations. As in the baseline RBC model we assume isoelastic utility and Cobb-Douglas production.

$$U(C, L) = \frac{C^{1-\sigma} - 1}{1 - \sigma} - \xi \frac{L^{1+\eta}}{1 + \eta} \tag{65}$$

$$F(k, l) = k^\alpha l^\beta = K^\alpha L^\beta n^{-(\alpha+\beta)} \tag{66}$$

where α and β are capital and labor shares. σ is a curvature parameter, and $\eta \geqslant 0$ is inverse Frisch elasticity.[45] The specific dynamical system and intratemporal condition for this specification are given in appendix D.

[45] $\eta = \frac{1}{FE}$ where $FE = \frac{dL}{dw}\frac{w}{L}$ is the Frisch elasticity which captures elasticity of hours worked to the wage rate, given a constant marginal utility of wealth. So it captures the substitution effect of a change in the wage rate on labor supply. $\eta = 0$ is indivisible labor, assuming a higher Frisch elasticity of labor supply, i.e. $\eta \to 0$, strengthens results as hours respond more strongly. Mertens and Ravn 2011 estimate it as $\eta = 0.976$.

5.1 Steady State

Solving for steady state gives

$$n^* = \left[\frac{\beta}{\xi \nu^\sigma} \left\{ \left(A \left(\frac{\alpha}{\rho} \right)^\alpha \right)^{1+\eta} (1-\zeta)^{\alpha(1+\eta)+\beta(1-\sigma)} \right. \right.$$

$$\left. \left. \left(\frac{1-(1-\zeta)\nu}{\phi} \right)^{1-\nu+\eta(1-\alpha)+\sigma\beta} \right\}^{\frac{1}{\beta}} \right]^{\frac{1}{\eta+\sigma}} \tag{67}$$

$$L^* = n^* \left[\frac{1}{A} \left(\frac{\rho}{\alpha(1-\zeta)} \right)^\alpha \left(\frac{\phi}{1-(1-\zeta)\nu} \right)^{1-\alpha} \right]^{\frac{1}{\beta}} \tag{68}$$

$$K^* = n^* \frac{\phi\alpha(1-\zeta)}{(1-(1-\zeta)\nu)\rho} \tag{69}$$

$$C^* = n^* \frac{\phi(1-\zeta)\nu}{1-(1-\zeta)\nu} \tag{70}$$

$$e^* = 0 \tag{71}$$

The steady-state is defined in terms of parameters $\Omega = \{\alpha, \beta, \phi, \gamma, \xi, \rho, \eta, \zeta, \sigma\}$, where all except dynamic barriers to entry γ enter the steady state. The number of firms is decreasing in fixed costs ϕ, discount factor ρ, labor weight in utility ξ. It is convex in ζ, such that an n^*-maximizing ζ exists.

In the long run, output per firm and capital per firm are independent of technology $k^*_A = y^*_A = 0$. Labor per firm decreases to maintain fixed scale given better technology. Therefore technological improvement causes firms to maintain a fixed capital stock, but reduce employment, and in aggregate the number of firms increases which expands aggregate output $Y^*_A = n^*_A y^*$, but average firm size does not change.

The steady state expression for output per firm (51) gives fixed costs as a proportion of variable cost $\frac{\phi}{y^*} = \frac{1-(1-\zeta)\nu}{\nu(1-\zeta)}$.[46] For a calibration of $\nu = 0.8$ then our model implies overheads as a proportion of output in steady state vary from 0.25 with perfect competition ($\zeta = 0$) to 0.56 when the Lerner

[46] Variable costs are equivalent to output in steady state since $y = \pi + rk + wl$, and $\pi^* = 0$.

Index is $\zeta = 0.2$.[47] With logarithmic utility in consumption ($\sigma = 1$), then labor elasticity η does not affect firms n^*, and in turn technology will not affect long-run labor. Table 1 summarizes the parameter values we use for simulation exercises [48] This calibration implies $CU = 0.8$ with $\zeta = 0.05$

ζ	α	β	ϕ	γ	A	σ	ρ	η
$\{0.05, 0.2\}$	0.3	0.5	0.3	50.0	$\{1.0, 1.01\}$	1.0	0.025	0.0

Table 1: Parameter Values for Numerical Exercises

and $CU = 0.44$ with $\zeta = 0.2$. ξ is chosen such that steady-state labor is normalized to one ($L^* = 1$). Dynamic barriers to entry η only affect model dynamics, not steady state outcomes, we choose this parameter to be large enough so that number of firms adjusts more slowly than capital. This implies that capital per firm increases following a positive shock before firm adjustment catches up to revert it to its long-run level which is unchanged.

5.2 Dynamics

We solve the four dimensional system locally for trajectories of the variables over t. The system is nonlinear, so we linearize it to the form $\dot{X} = \mathbf{J}(X - X^*)$ where $X = [C, e, K, n]'$. We then analyse the Jacobian matrix $\mathbf{J} : \mathbb{R}^4 \to \mathbb{R}^4$ where each element is a respective derivative evaluated at steady state.

$$\begin{bmatrix} \dot{C} \\ \dot{e} \\ \dot{K} \\ \dot{n} \end{bmatrix} = \left. \begin{bmatrix} \frac{C}{\sigma} r_C & 0 & \frac{C}{\sigma} r_K & \frac{C}{\sigma} r_n \\ -\frac{\pi_C}{\gamma} & \rho & -\frac{\pi_K}{\gamma} & -\frac{\pi_n}{\gamma} \\ Y_C - 1 & 0 & Y_K & Y_n \\ 0 & 1 & 0 & 0 \end{bmatrix} \right|^* \begin{bmatrix} C - C^* \\ e - e^* \\ K - K^* \\ n - n^* \end{bmatrix} \tag{72}$$

[47] $\zeta = 0.2$ implies a price-over-marginal-cost markup of $\mu = \frac{1}{1-\zeta} = \frac{\theta}{\theta-1} = 1.2$ and intersector substitutability $\theta = 6$. In general estimates of markups in value added data range from 1.2 to 1.4, and in gross output they vary between 1.05 and 1.15, see Basu and Fernald 2001 and Morrison 1992

[48] In this deterministic model $\sigma \in (0, \infty) \setminus \{1\}$, it is a curvature parameter as there is no risk. $\sigma \to \infty$ implies infinite risk aversion, consumption has little effect on utility. $\sigma \to 0$ is risk neutrality, a % change in consumption has the same % change on utility. The $\sigma \to 1$ case implies log utility $\ln(C)$.

Proposition 9. *The economy is locally asymptotically unstable. A 2d-stable manifold exists in capital and number of firms.*

Proof. Appendix D. □

Proposition 9 formalizes quasi-fixity of capital and number of firms $n_A(0) = K_A(0) = 0$ which we use to prove our main result (Theorem 1). With saddle dynamics variables on the system's stable manifold are predetermined. They do not respond on impact of a shock (Caputo 2005, p.426), instead jump variables (C, e) move instantanously to put the system on the stable manifold, and subsequently the state variables (K, n) converge to the long-run steady state.[49]

Figure 4 illustrates Theorem 1 numerically.[50] It shows that after a once-

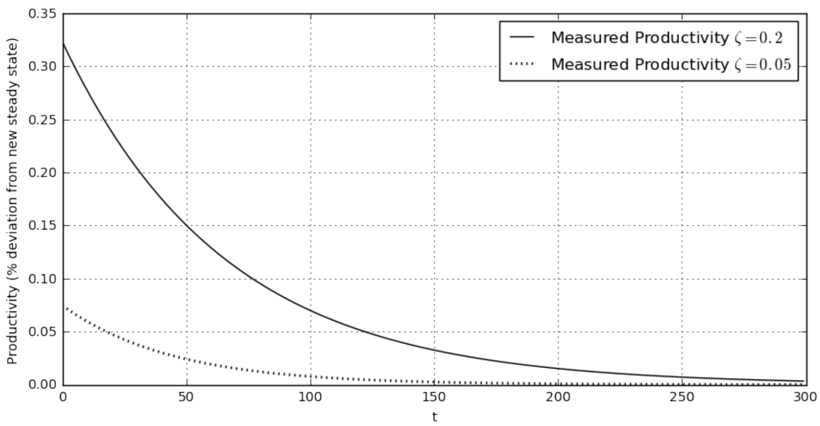

Figure 4: Measured Productivity

and-for-all 1% technology improvement $A = 1$ to $A = 1.01$ measured productivity TFP overshoots its new long-run level TFP$^*(A = 1.01)$ by 0.32%, and as market power decreases $\zeta = 0.05$ the effect is weaker (Corollary 2). Figure 5 shows the transmission of the positive deterministic shock through the model's underlying variables.[51] As explained by our theoretical discus-

[49]The supplementary appendix defines the stable manifold. It shows that K, n do not respond on impact $t = 0$, whereas C, e respond instantaneously. The simulations in figure 5 illustrate this.

[50]

[51]In Appendix E we superimpose these dynamics on a single graph for relative comparison of rates of convergence and magnitudes.

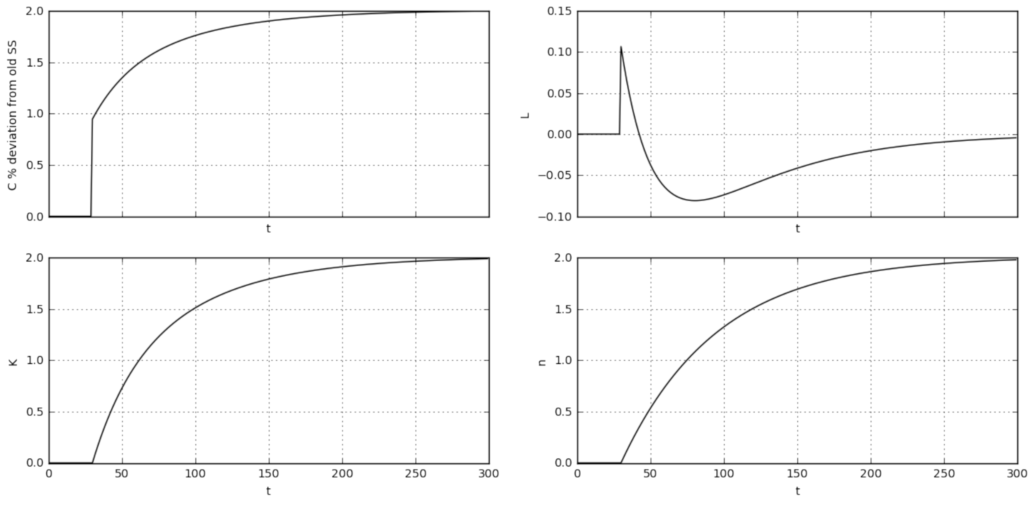

Figure 5: Positive Shock Dynamics

sion, the positive technology shock causes capital and number of firms to begin at their initial pre-shock steady state and start increasing over time as they converge to the new steady state with improved technology. It is this slow response of number of firms to the shock which leads to the productivity overshooting shown in figure 4. And the slower the response $\gamma \to \infty$ of firms the more persistent the endogenous effect.[52] We can also see that the number of firms increases at a decreasing rate, reflecting that entry is initially high to arbitrage large profits, but diminishes over time as congestion effects increase. In the long run, entry is zero as the number of firms is fixed at its new long-run level. Consumption jumps on impact, which implies labor jumps too. The short-run rise in labor can be understood through the static labor condition

$$L(C,K,n) = \left(\frac{(1-\zeta)AK^{\alpha}\beta n^{1-(\alpha+\beta)}}{\xi C^{\sigma}} \right)^{\frac{1}{1+\eta-\beta}} \tag{73}$$

Given K, n are initially fixed at their old steady state level, the increase in A on impact offsets the increase in C on impact creating an increase in L.

[52]This can be fomalized by showing the regulatory parameter (dynamic barrier to entry) γ strictly decreases the system's eigenvalues, and hence more entry regulation slows recovery after a shock as it inhibits firm dynamics. Working paper available on request.

Subsequently, as K, n are able to adjust their increase is weaker than the increase in C and hence L decreases. After some point, the increasing C becomes weaker than the increasing K, n which leads to the hump-shape, and eventual increase in L back to its steady state.

6 Capital Utilization Vs Capacity Utilization

We have shown that slow firm entry causes variations in capacity utilization which creates endogenous productivity fluctuations. However RBC literature emphasizes that *'capital utilization'* can account for endogenous variations in measured TFP, which amplify exogenous technology shocks.[53] In this section we include capital utilization in conjunction with our capacity utilization mechanism, and find they are of equal importance.

In the absence of endogenous labor, RBC literature uses the terms *capacity utilization* and *capital utilization* interchangeably.[54] We define capacity utilization as production relative to a full capacity benchmark (Definition 1). Whereas, capital utilization is the endogenous depreciation of capital based on its usage. Capital utilization nests a new functional in the firm production function.[55] The utilization function $\mathfrak{u}(t) : K \times n \to (0, 1)$ reflects the intensity of capital usage, so the capital utilization production function is

$$y = AF(\mathfrak{u}k, l) - \phi \tag{74}$$

In addition to the modified production function, capital utilization will affect depreciation in the budget constraint. This creates a difference between the household's market return from lending capital $r(t)$ and the firm's cost of renting capital $R(t)$, which are equivalent in the no depreciation case. The relationship is that the return to lending capital is the rental paid by firms

[53] For example, Kydland and Prescott 1988, Shapiro 1993, Bils and Cho 1994, Burnside, Eichenbaum, and Rebelo 1995, King and Rebelo 1999

[54] For example, Greenwood, Hercowitz, and Huffman 1988, and Benhabib, Nishimura, and Shigoka 2008.

[55] King and Rebelo 1999 were early adopters of the preciser term capital utilization, and Basu and Fernald 2001 also emphasize the distinction.

less depreciation $r(t) = R(t) - \delta(\mathfrak{u}, t)$.

We assume that the rate of capital depreciation $\delta(\mathfrak{u}, t) \in (0, 1)$ is an increasing convex function of the rate of utilization $\mathfrak{u} \in (0, 1)$ given by

$$\delta = z\mathfrak{u}^{\vartheta} \tag{75}$$

where $\vartheta > 1$ and $z \in R_+$.[56] Therefore $z = 0$ and $\mathfrak{u} = 1$ cause production (74) and depreciation (75) to collapse to the no utilization setup. Since utilization is endogenous it cannot be exogenously set to full utilization $\mathfrak{u} \to 1$, but the exogenous convexity parameter can be made large to achieve this effect $\lim_{\vartheta \to \infty} \mathfrak{u} = 1$. The intuition is that ϑ is the elasticity of $\delta_{\mathfrak{u}}$.[57] When it is large the marginal cost of utilization (replacement rate) $\delta_{\mathfrak{u}}$ responds elastically to utilization which encourages full utilization. It is calibrated to 1.1 in King and Rebelo 1999, 1.4 in Wen 1998, and between 1.25 and 2.0 in Benhabib, Nishimura, and Shigoka 2008. We use $\vartheta = 1.4, z = 1.0$ and the table 1 numerical values.

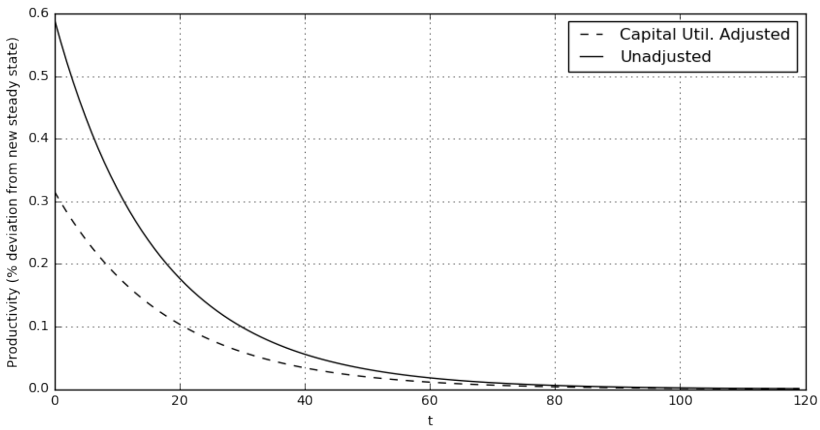

Figure 6: Capacity Utilization and Capital Utilization Amplification

Figure (6) shows that overshooting is 0.6% which comprises 0.31% from capacity utilization and 0.29% from capital utilization. Clearly the adjusted

[56]This parametric restriction imposes a convex cost structure on capital utilization so there is an interior solution for $\mathfrak{u}$ in steady state. If $\vartheta \leqslant 1$ then the optimal rate of capital utilization is always $\mathfrak{u} = 1$ i.e. full utilization.

[57]$\mathfrak{u}\frac{\delta_{\mathfrak{u}\mathfrak{u}}}{\delta_{\mathfrak{u}}} = \vartheta - 1$, King and Rebelo 1999 use notation $\vartheta - 1 = \xi$ and calibrate to 0.1.

overshooting is the same as figure 4 with no capital utilization. This shows that King and Rebelo 1999 suggested 'modified Solow Residual' works well at eradicating the capital utilization bias, leaving only the capacity utilization bias. The unadjusted measure fails to account for $\mathfrak{u}$ in the denominator of the measured productivity definition $\text{TFP}^{\text{unadj.}} = y/F(k,l)^{\frac{1}{\nu}}$ whereas the adjusted measure ensures the denominator is correctly specified $\text{TFP}^{\text{adj.}} = y/F(\mathfrak{u}k,l)^{\frac{1}{\nu}}$, where for both definitions y includes utilization as in (74).

7 Summary

This paper shows that with imperfect competition, non-instantaneous firm entry causes endogenous productivity dynamics over the business cycle because incumbent firms vary their excess capacity in the short-run absence of entry. Crucially it is the short-run absence of entry that creates procyclical productivity, and subsequent entry degrades productivity through business stealing. This is distinct from most literature that focuses on the long-run pro-competitive effect of entry on markups, heterogeneous firm composition, or aggregate returns to scale, and thus free-entry outcomes rather than intertemporal effects.

Our methodological contribution is to offer a tractable theory of endogenous firm entry over the business cycle with imperfect competition and endogenous sunk costs. Our static analysis shows that imperfect competition causes excess capacity and locally increasing returns. Our dynamic analysis shows that this excess capacity varies in the short-run in response to shocks, but returns in the long-run when entry has adjusted. We are the first authors to prove local stability results in a popular class of macroeconomic models that contain two-state variables: capital and number of firms. Our results follow from analysis of this two dimensional stable manifold.

Quantitatively we show that the endogenous productivity movements that slow firm entry creates are at least as important as the popular capital utilization mechanism of traditional RBC papers.

A supplementary appendix offers greater details and extensions to the results of this appendix.

A Model Equations

Collecting the model equilibrium conditions recursively gives

$$k = \frac{K}{n} \tag{18}$$

$$l = \frac{L}{n} \tag{19}$$

$$y = AF(k, l) - \phi \tag{22}$$

$$Y = ny \tag{23}$$

$$\pi = y(1 - (1 - \zeta)\nu) - \nu(1 - \zeta)\phi \tag{28}$$

$$r = (1 - \zeta)AF_k(k, l) \tag{24}$$

$$w = (1 - \zeta)AF_l(k, l) \tag{25}$$

$$w = -\frac{u_L(L)}{u_C(C)} \tag{15}$$

$$\dot{C} = \frac{C}{\sigma(C)}(r - \rho), \qquad \sigma(C) = -\frac{Cu_{CC}}{u_C} \tag{14}$$

$$\dot{e} = re - \frac{\pi}{\gamma}, \qquad \gamma > 0 \tag{36}$$

$$\dot{K} = Y - \frac{\gamma}{2}e^2 - C \tag{41}$$

$$\dot{n} = e \tag{35}$$

$$K(0) = K_0 \tag{16}$$

$$n(0) = n_0 \tag{38}$$

$$\lim_{t\to\infty} K(t)u_C(t)\exp(-\rho t) = 0 \tag{17}$$

$$\lim_{t\to\infty} n(t)q(t)u_C(t)\exp(-\rho t) = 0 \tag{37}$$

B Labor Market

The supplementary appendix gives an extensive discussion of the labor market. Treating labor as an implicit function, take the derivative with respect to C, K, n of the general intratemporal condition at factor market equilibrium as in (42)

$$u_L(L) + u_C(C)w(C, K, n) = 0 \tag{76}$$

then substitute in the total derivatives of wage, which are

$$\begin{aligned} w(C, K, n) &= (1-\zeta)An^{1-\nu}F_L(K, L) & (77)\\ w_L &= (1-\zeta)An^{1-\nu}F_{LL}(K, L) < 0 & (78)\\ w_C &= w_L L_C & (79)\\ w_K &= \frac{\partial w}{\partial K} + w_L L_K & (80)\\ w_n &= \frac{\partial w}{\partial n} + w_L L_n & (81) \end{aligned}$$

and collect terms in labor response of the left-hand side.

$$\begin{aligned} L_C &= \frac{-u_{CC}w}{u_{LL} + u_C w_L} = \frac{u_L}{u_{LL} + u_C w_L}\frac{u_{CC}}{u_C} < 0 & (82)\\ L_K &= \frac{-u_C}{u_{LL} + u_C w_L}\frac{\partial w}{\partial K} = \frac{u_L}{u_{LL} + u_C w_L}\frac{F_{LK}}{F_L} > 0 & (83)\\ L_n &= \frac{-u_C}{u_{LL} + u_C w_L}\frac{\partial w}{\partial n} = \frac{u_L}{u_{LL} + u_C w_L}\frac{1-\nu}{n} > 0, \quad \nu \in (0,1) & (84) \end{aligned}$$

The consistent denominator is $\hat{\mathcal{H}}_{LL} = u_{LL} + u_C w_L$, which is the intratemporal condition differentiated with respect to L. It is negative which reflects that utility is decreasing in labor.

Proof of Proposition 3. Follows from (84) as $\nu \in (0, 1)$. □

The remark that $L_n|^{\nu=1} = 0$ follows trivially, remembering imperfect competition must hold for constant returns existence when there is an overhead cost.

C Additional Proofs

Proof of Lemma 1. Total operating profits are aggregate output less total variable costs[58]

$$n\pi = ny - n(rk + wl) \tag{85}$$
$$= Y - n^{1-\nu}(1-\zeta)\nu AF(K,L) \tag{86}$$

substitute out $n = \frac{Y}{y}$ and collect Y

$$Y = \left(1 - \frac{\pi}{y}\right)^{-\frac{1}{\nu}} \left(\left(\frac{1}{y}\right)^{1-\nu} AF(K,L)(1-\zeta)\nu\right)^{\frac{1}{\nu}} \tag{87}$$
$$= y\left(\frac{(1-\zeta)\nu A}{y-\pi}\right)^{\frac{1}{\nu}} F(K,L)^{\frac{1}{\nu}} \tag{88}$$

use $y = \frac{(1-\zeta)\nu\phi + \pi}{1-(1-\zeta)\nu}$ (which comes from $\pi = y - (1-\zeta)\nu AF(k,l)$, substitute out $y = AF(k,l) - \phi$ then rearrange for $F(k,l) = \frac{\pi+\phi}{A(1-(1-\zeta)\nu)}$) thus

$$\frac{Y}{F(K,L)^{\frac{1}{\nu}}} = \left(\frac{A}{\pi+\phi}\right)^{\frac{1}{\nu}} (1-(1-\zeta)\nu)^{\frac{1}{\nu}-1}[(1-\zeta)\nu\phi + \pi] \tag{89}$$

□

Proof of Proposition 4. In each case take the total derivative of y and then substitute in the labor effect from B.

$$y = An^{-\nu}F(K,L) - \phi \tag{90}$$
$$y_C = An^{-\nu}F_L L_C = An^{-\nu}F_L \frac{u_L}{u_{LL} + u_C w_L}\frac{u_{CC}}{u_C} = -An^{-\nu}F_L \frac{u_{CC} w}{u_{LL} + u_C w_L} < 0 \tag{91}$$
$$y_K = An^{-\nu}(F_K + F_L L_K) = An^{-\nu}\left(F_K + \frac{u_L}{u_{LL} + u_C w_L} F_{LK}\right) > 0 \tag{92}$$

An entrant's effect on the intensive margin y_n is more complex because labor

[58] The fixed cost ϕ denominated in terms of output. It could be denomiated in terms of wages so it would appear in the variable costs component not y.

opposes the business stealing effect. First substitute in the labor response L_n

$$y_n = -\nu A n^{-\nu-1} F(K,L) + A n^{-\nu} F_L L_n \tag{93}$$

$$= A n^{-\nu-1}\left[-\nu F(K,L) + \frac{u_L}{u_{LL}+u_C w_L}(1-\nu)F_L\right] \tag{94}$$

Then, by Euler's homogeneous function theorem, use that $\nu F = F_K K + F_L L$ and $(\nu-1)F_L = F_{LL}L + F_{LK}K$ and using the relationship $u_C w_L = u_C w \frac{F_{LL}}{F_L} = -u_L \frac{F_{LL}}{F_L}$ since $u_C w = -u_L$. Hence

$$y_n = A n^{-\nu-1}\frac{u_L}{u_{LL}+u_C w_L}\left[-\nu F(K,L)\frac{u_{LL}}{u_L} + K\left(\frac{F_{LL}}{F_L}F_K - F_{LK}\right)\right] < 0 \tag{95}$$

$\square$

Proof of Corollary 1. Use Cramer's rule to determine the effect of a change in technology on k^*, l^*. From (47) and (48), technology decreases per firm marginal product of capital, and variable production

$$F_k|_A^* = \frac{-1}{A^2}\frac{\rho}{(1-\zeta)} < 0$$

$$F_A^* = \frac{-1}{A^2}\frac{\phi}{(1-(1-\zeta)\nu)} < 0$$

In general

$$F_{kk}k_A + F_{kl}l_A = F_{kA}$$

$$F_k k_A + F_l l_A = F_A$$

$$\underbrace{\begin{bmatrix} F_{kk} & F_{kl} \\ F_k & F_l \end{bmatrix}}_{\mathbf{H}}\begin{bmatrix} k_A \\ l_A \end{bmatrix} = \begin{bmatrix} F_{kA} \\ F_A \end{bmatrix}$$

$$\begin{bmatrix} k_A \\ l_A \end{bmatrix} = \frac{1}{det(\mathbf{H})}\begin{bmatrix} F_l & -F_{kl} \\ -F_k & F_{kk} \end{bmatrix}\begin{bmatrix} F_{kA} \\ F_A \end{bmatrix}$$

Since $det(\mathbf{H}) = F_{kk}F_l - F_{kl}F_k < 0$ and at steady state the effect of a change in technology on marginal product of capital and production is negative ($F_k|_A^*, F_A^* < 0$) then

$$l_A^* = \frac{1}{det(\mathbf{H})}\left(-F_k F_k|_A^* + F_{kk}F_A^*\right) < 0 \tag{96}$$

$$k_A^* = \frac{1}{det(\mathbf{H})}\left(F_l F_k|_A^* - F_{kl}F_A^*\right) \gtreqless 0 \iff \frac{F_l}{F_{kl}} \gtrless \frac{(1-\zeta)\phi}{\rho(1-(1-\zeta)\nu)} \tag{97}$$

□

D Parameterized Model

Under the functional forms we have assumed the intratemporal condition is

$$L(C, K, n) = \left(\frac{(1-\zeta)AK^{\alpha}\beta n^{1-(\alpha+\beta)}}{\xi C^{\sigma}}\right)^{\frac{1}{1+\eta-\beta}} \tag{98}$$

Hence substituting out $L(C, K, n)$ gives a 4d dynamical system in (C, e, K, n)

$$\dot{C} = \frac{C}{\sigma}\left[(1-\zeta)A\alpha K^{\alpha-1}L^{\beta}n^{1-(\alpha+\beta)} - \rho\right] \tag{99}$$

$$\begin{aligned}\dot{e} = {} & (1-\zeta)A\alpha K^{\alpha-1}L^{\beta}n^{1-(\alpha+\beta)}e \\ & - \frac{1}{\gamma}\left(AK^{\alpha}L^{\beta}n^{-(\alpha+\beta)}(1-(1-\zeta)\nu) - \phi\right)\end{aligned} \tag{100}$$

$$\dot{K} = n\left[AK^{\alpha}L^{\beta}n^{-(\alpha+\beta)} - \phi\right] - \frac{\gamma}{2}e^2 - C \tag{101}$$

$$\dot{n} = e \tag{102}$$

The corresponding Jacobian matrix evaluated at steady state is

$$
\mathbf{J} = \left.\begin{bmatrix} \dot{C}_C & \dot{C}_e & \dot{C}_K & \dot{C}_n \\ \dot{e}_C & \dot{e}_e & \dot{e}_K & \dot{e}_n \\ \dot{K}_C & \dot{K}_e & \dot{K}_K & \dot{K}_n \\ \dot{n}_C & \dot{n}_e & \dot{n}_K & \dot{n}_n \end{bmatrix}\right|^* \tag{103}
$$

$$
= \begin{bmatrix} -\frac{\rho\beta}{1+\eta-\beta} & 0 & \frac{-\rho^2\nu(1-\nu+\eta(1-\alpha))}{(1+\eta-\beta)\sigma\alpha} & \frac{\phi(1-\zeta)\nu\rho(1-\nu)(1+\eta)}{(1+\eta-\beta)(1-(1-\zeta)\nu)\sigma} \\ \frac{(1-(1-\zeta)\nu)\beta\sigma}{(1+\eta-\beta)\gamma n^*(1-\zeta)\nu} & \rho & \frac{-(1-(1-\zeta)\nu)\rho(1+\eta)}{(1+\eta-\beta)\gamma n^*(1-\zeta)} & \frac{\phi(\nu(1+\eta)-\beta)}{(1+\eta-\beta)\gamma n^*} \\ \frac{-\beta\sigma}{(1+\eta-\beta)(1-\zeta)\nu} - 1 & 0 & \frac{\rho(1+\eta)}{(1+\eta-\beta)(1-\zeta)} & \frac{\phi[-\zeta\nu(1+\eta-\beta)+\beta(1-\nu)]}{(1+\eta-\beta)(1-(1-\zeta)\nu)} \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{104}
$$

where n^* is defined in (67). The quartic characteristic polynomial associated with the Jacobian matrix is (Jacobson 2012, p. 196)

$$
c(\lambda) = \det(\mathbf{J} - \lambda\mathbf{I}) = \lambda^4 - M_1\lambda^3 + M_2\lambda^2 - M_3\lambda + M_4 \tag{105}
$$

where M_k denotes the sum of principal minors of dimension k, and $M_1 = tr(\mathbf{J})$ and $M_4 = \det\mathbf{J}$.

$$
M_1 = \frac{1}{1-\zeta}\left[2\rho + \frac{\zeta(1+\eta)}{1+\eta-\beta}\right] > 0 \tag{106}
$$

$$
\begin{aligned} M_2 = \frac{\rho^2}{1+\eta-\beta}\Bigg[&\frac{-\phi(\alpha+\eta\nu)}{\gamma n^*\rho^2} + \frac{1+\eta}{1-\zeta} \\ &- \frac{\beta\sigma(1+(1-\zeta)\alpha) + \nu(1-\zeta)(1-\nu+(1-\alpha)\eta)}{(1-\zeta)\alpha\sigma}\Bigg] \lesseqgtr 0 \end{aligned} \tag{107}
$$

$$
\begin{aligned} M_3 = &\frac{-\rho(1+\eta)\phi\nu}{\gamma n^*(1+\eta-\beta)} + \frac{\rho\beta\phi}{\gamma n^*(1+\eta-\beta)} \\ &+ \frac{-\rho^3\left[\beta\sigma + \nu(1-\zeta)(1-\nu+(1-\alpha)\eta)\right]}{(1-\zeta)(1+\eta-\beta)\sigma\alpha} < 0 \end{aligned} \tag{108}
$$

$$
M_4 = \frac{\rho^2\phi\beta\nu(\eta+\sigma)}{(1+\eta-\beta)\gamma\sigma\alpha n^*} > 0 \tag{109}
$$

Proof of Proposition 9. We show the characteristic polynomial has four solutions, and that two must be positive (unstable) and two negative (stable). Denote these solutions (eigenvalues) $\lambda_1 \leqslant \lambda_2 < 0 < \lambda_3 \leqslant \lambda_4$. Since the determinant is positive $\lambda_1\lambda_2\lambda_3\lambda_4 > 0$. This rules out zero eigenvalues and restricts possibilities to (1) Two positive, two negative (2) All negative (3) All positive. The trace is positive so $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 > 0$ which rules out all negative. Both trace and determinant are positive, but M_3 is negative, which by Descartes' Rule of Signs implies (1) Two positive, two negative eigenvalues, is the only option. $\square$

E Relative Dynamics

Superimposing all the dynamics on a single graph shows that number of firms adjusts slower than capital which adjusts slower than consumption. It also shows that labor's deviation is small and the overshooting in measured TFP ($\mathcal{P}$) is pronounced relative to the shift change in technology A.

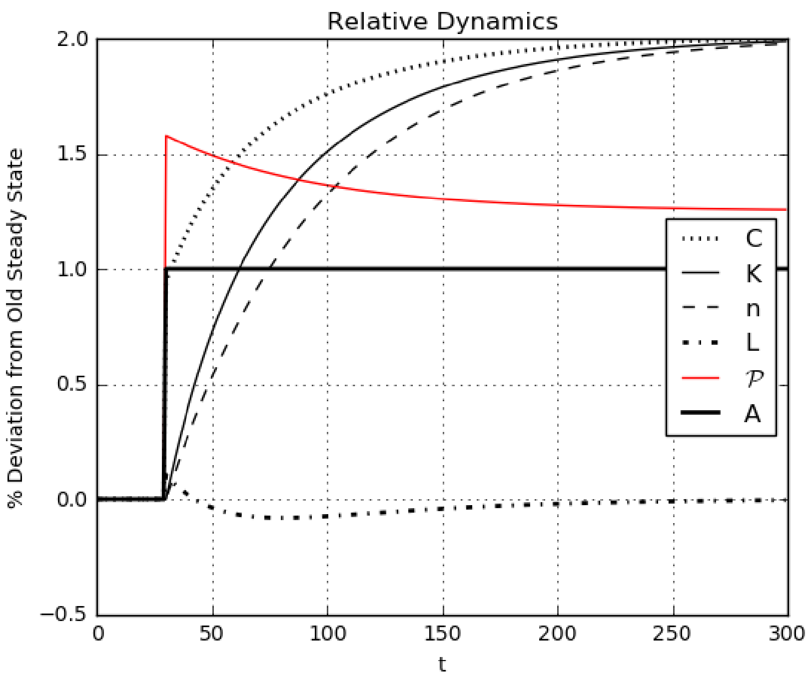

Figure 7

References

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Introduction to Modern Economic Growth v. 1. Princeton University Press.

Atkeson, Andrew and Ariel Burstein (2010). “Innovation, Firm Dynamics, and International Trade”. In: *Journal of Political Economy* 118.3, pp. 433–484.

Barseghyan, Levon and Riccardo DiCecio (2011). “Entry costs, industry structure, and cross-country income and TFP differences”. In: *Journal of Economic Theory* 146.5, pp. 1828–1851.

Basu, Susanto and John Fernald (2001). “Why Is Productivity Procyclical? Why Do We Care?” In: *New Developments in Productivity Analysis*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 225–302.

Benhabib, Jess, Kazuo Nishimura, and Tadashi Shigoka (2008). “Bifurcation and sunspots in the continuous time equilibrium model with capacity utilization”. In: *International Journal of Economic Theory* 4.2, pp. 337–355.

Berentsen, Aleksander and Christopher Waller (2015). “Optimal Stabilization Policy with Search Externalities”. In: *Macroeconomic Dynamics* 19 (03), pp. 669–700.

Bergin, Paul R. and Ching-Yi Lin (2012). “The dynamic effects of a currency union on trade”. In: *Journal of International Economics* 87.2, pp. 191–204.

Berndt, Ernst R and Melvyn A Fuss (1986). “Productivity measurement with adjustments for variations in capacity utilization and other forms of temporary equilibrium”. In: *Journal of econometrics* 33.1, pp. 7–29.

Berndt, Ernst R and Catherine J Morrison (1981). “Capacity utilization measures: underlying economic theory and an alternative approach”. In: *The American Economic Review* 71.2, pp. 48–52.

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.

Bils, Mark and Jang-Ok Cho (1994). "Cyclical factor utilization". In: *Journal of Monetary Economics* 33.2, pp. 319–354.

Blanchard, Olivier Jean and Nobuhiro Kiyotaki (1987). "Monopolistic competition and the effects of aggregate demand". In: *The American Economic Review*, pp. 647–666.

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.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo (1995). "Capital utilization and returns to scale". In: *NBER Macroeconomics Annual 1995, Volume 10*. MIT Press, pp. 67–124.

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

Campbell, Jeffrey R (1998). "Entry, exit, embodied technology, and business cycles". In: *Review of economic dynamics* 1.2, pp. 371–408.

Caputo, M.R. (2005). *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press.

Chatterjee, Satyajit and Russell Cooper (1993). *Entry and Exit, Product Variety and the Business Cycle*. NBER Working Papers 4562. National Bureau of Economic Research, Inc.

Clementi, Gian Luca and Berardino Palazzo (2016). "Entry, exit, firm dynamics, and aggregate fluctuations". In: *American Economic Journal: Macroeconomics* 8.3, pp. 1–41.

Da-Rocha, Jose-Maria, Marina Mendes Tavares, and Diego Restuccia (2017). *Policy Distortions and Aggregate Productivity with Endogenous Establishment-Level Productivity*. Working Paper 23339. National Bureau of Economic Research.

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.

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.

Devereux, Michael B, Allen C Head, and Beverly J Lapham (1996). “Aggregate fluctuations with increasing returns to specialization and scale”. In: *Journal of economic dynamics and control* 20.4, pp. 627–656.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu (2015). “Competition, markups, and the gains from international trade”. In: *American Economic Review* 105.10, pp. 3183–3221.

Etro, Federico and Andrea Colciago (2010). “Endogenous Market Structures and the Business Cycle*”. In: *The Economic Journal* 120.549, pp. 1201–1233.

Feenstra, R.C. (2003). *Advanced International Trade: Theory and Evidence.* Princeton University Press.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W Huffman (1988). “Investment, capacity utilization, and the real business cycle”. In: *The American Economic Review*, pp. 402–417.

Hall, R. E. (1987). “Productivity and the Business Cycle”. In: *Carnegie Rochester Conference Series on Public Policy* 27, pp. 421–444.

Hall, R.E. and M. Lieberman (2009). *Microeconomics: Principles and Applications.* Cengage Learning.

Hall, Robert E. (1990). “Invariance Properties of Solow’s Productivity Residual”. In: *Growth, Productivity, Unemployment: Essays to Celebrate Bob Solow’s Birthday.* MIT Press, pp. 71–112.

Hall, Robert E et al. (1986). “Market Structure and Macroeconomic Fluctuations”. In: *Brookings Papers on Economic Activity* 17.2, pp. 285–338.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda (2013). “Who creates jobs? Small versus large versus young”. In: *Review of Economics and Statistics* 95.2, pp. 347–361.

Harrison, Ann E. (1994). “Productivity, imperfect competition and trade reform : Theory and evidence”. In: *Journal of International Economics* 36.1-2, pp. 53–73.

Harrison, Ann E. and Brian J. Aitken (1999). “Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela”. In: *American Economic Review* 89.3, pp. 605–618.

Hornstein, Andreas (1993). “Monopolistic competition, increasing returns to scale, and the importance of productivity shocks”. In: *Journal of Monetary Economics* 31.3, pp. 299–316.

Hsieh, Chang-Tai and Peter J. Klenow (2014). “The Life Cycle of Plants in India and Mexico”. In: *The Quarterly Journal of Economics* 129.3, pp. 1035–1084.

Hulten, Charles R (1986). “Productivity change, capacity utilization, and the sources of efficiency growth”. In: *Journal of econometrics* 33.1-2, pp. 31–50.

Jacobson, N. (2012). *Basic Algebra I: Second Edition*. Dover Books on Mathematics. Dover Publications.

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.

Jaimovich, Nir and Max Floetotto (2008). “Firm dynamics, markup variations, and the business cycle”. In: *Journal of Monetary Economics* 55.7, pp. 1238 –1252.

King, Robert G. and Sergio T. Rebelo (1999). “Resuscitating real business cycles”. In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and M. Woodford. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 14, pp. 927–1007.

Kydland, Finn E and Edward C Prescott (1988). “The workweek of capital and its cyclical implications”. In: *Journal of Monetary Economics* 21.2-3, pp. 343–360.

Lewis, Vivien (2009). “Business Cycle Evidence On Firm Entry”. In: *Macroeconomic Dynamics* 13 (05), pp. 605–624.

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

Mankiw, N Gregory and Michael D Whinston (1986). “Free entry and social inefficiency”. In: *The RAND Journal of Economics*, pp. 48–58.

Mertens, Karel and Morten Overgaard Ravn (2011). “Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks”. In: *Review of Economic Dynamics* 14.1, pp. 27–54.

Morrison, Catherine J (1992). “Unraveling the productivity growth slowdown in the United States, Canada and Japan: the effects of subequilibrium, scale economies and markups”. In: *The Review of Economics and Statistics*, pp. 381–393.

Morrison, Catherine J. (2012). *Cost Structure and the Measurement of Economic Performance: Productivity, Utilization, Cost Economics, and Related Performance Indicators*. Springer US.

Nadiri, M. Ishaq and Ingmar Prucha (2001). “Dynamic Factor Demand Models and Productivity Analysis”. In: *New Developments in Productivity Analysis*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 103–172.

Nelson, Randy A (1989). “On the measurement of capacity utilization”. In: *The Journal of Industrial Economics*, pp. 273–286.

Rossi, Lorenza and Emilio Zanetti Chini (2016). *Firms' Dynamics and Business Cycle: New Disaggregated Data*. Tech. rep. University of Pavia, Department of Economics and Management.

Rotemberg, Julio J and Michael Woodford (1992). “Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity”. In: *Journal of Political Economy* 100.6, pp. 1153–1207.

Shapiro, Matthew D (1993). “Cyclical productivity and the workweek of capital”. In: *The American Economic Review* 83.2, pp. 229–233.

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

Turnovsky, S.J. (2000). *Methods of Macroeconomic Dynamics*. MIT Press.

Wen, Yi (1998). “Capacity Utilization under Increasing Returns to Scale”. In: *Journal of Economic Theory* 81.1, pp. 7 –36.