University of Kent

School of Economics Discussion Papers

Measured Productivity with Endogenous Markups and Economic Profits

Anthony Savagar

October 2018

KDPE 1812

Measured Productivity with Endogenous Markups and Economic Profits

Anthony Savagar *†

October 17, 2018

Abstract

I develop a model of dynamic firm entry, oligopolistic competition and returns to scale in order to decompose TFP fluctuations into technical change, economic profit and markup fluctuations. I show that economic profits cause short-run upward bias in measured TFP, but this subsides to upward bias from endogenous markups as firm entry adjusts. I analyze dynamics analytically through a nonparametric DGE model that allows for a perfect competition equilibrium such that there are no biases in measured TFP. Given market power, simulations show that measured TFP is 40% higher than technology in the short-run, due solely to profits, and 20% higher in the long-run due solely to markups. During transition both effects contribute upward bias: initially the profit effect dominates, but by 5 quarters the two effects contribute equally, and by 10 quarters only the markup effect persists. The speed of firm adjustment (*'business dynamism'*) will determine these timings and therefore the importance of each bias. **JEL: E32, D21, D43, L13, C62, Endogenous markups, Dynamic Firm Entry, Endogenous Productivity, Endogenous Entry Costs**

---

*Thanks to: Huw Dixon, Patrick Minford, Vivien Lewis, Jo Van Biesebroeck, Akos Valentinyi, Harald Uhlig, Stephen Millard, Martin Kaae Jensen, David Collie, Jonathan Haskel, Nobu Kiyotaki, Esteban Rossi-Hansberg, Fernando Leibovici, Sungki Hong, Mehdi Sahneh, Mathan Satchi and funders: ESRC, Julian Hodge Institute and RES. I appreciate the hospitality of KU Leuven and St Louis Fed.

†asavagar@gmail.com Project files on https://github.com/asavagar and supplementary appendix on www.asavagar.com.

Non-technical Summary

A standard way to measure productivity is to take output growth and subtract input growth. Typically input growth is growth in capital and growth in labor weighted by their share in production. Whatever remains after subtracting is known as total factor productivity (TFP).

Under specific circumstances, this productivity measure also reflects underlying technology growth. This is important for economists because they struggle to measure underlying technology at an aggregate level, but it is a key variable to understand the behaviour of the economy.

However when we diverge from some basic assumptions, such as perfect competition, the TFP measure no longer reflects underlying technology. Therefore using our TFP measure to represent technology could lead to incorrect conclusions.

In this paper I show that when we recognize the slow adjustment of firms to arbitrage profits and the effect that slowly entering firms have on competition, the relationship between our measure of TFP and technology becomes much more complex. In fact, when we observe changing TFP, it will compose changing technology, changing profits and changing markups. This decomposition allows us to understand how we can get a true measure of underlying technology from our calculated TFP measure. It emphasizes that the composition of profits and markups vary in importance as firm entry takes place.

Rising markups, increasing economic profits and declining business dynamism are topical empirical issues in macroeconomics.[1] From a theoretical perspective, firm entry is core to each mechanism: entry affects competition thus markups; entry arbitrages profits; and entry rates measure business dynamism. In this paper, I develop a model of dynamic firm entry, endogenous markups and endogenous entry costs in order to understand how these emerging trends affect our understanding of *measured TFP*, typically acquired as a Solow residual.

I decompose measured TFP into profit, markup and pure technology components. Crucially, I focus on the dynamic evolution of each component as entry transitions following a permanent technology improvement, rather than providing a static analysis once entry has adjusted. There are three stages: the short run, when firms are fixed; transition, when firms are entering; and the long run, when entry has ceased (profits are zero). I show that both markups and profits cause measured TFP to be an upward biased measure of pure technology, but their importance differs as entry adjusts. In the short run, upward bias in measured TFP is driven solely by profits, whereas in the long run upward bias is driven solely by markups. During transition both effects contribute positively, but the profit bias decreases in importance whilst the markup bias increases in importance. Numerically I show that measured TFP exceeds underlying technology by 40% on impact as profits rise and remains 20% higher in the long-run once profits have been arbitraged but competition has decreased markups. I show that the positive profit effect dominates the positive markup effect for 5 quarters, but after 10 quarters the profit effect disappears leaving only the long-run markup effect. These speeds will vary depending on the speed of entry adjustment, so-called business dynamism, in a given economy. Business dynamism is determined by prospective entrants' sensitivity to endogenously procylical sunk costs.

I extend a continuous-time Ramsey-Cass-Koopmans setup to include endogenous labor, endogenous markups and dynamic firm entry due to endogenous entry costs. Additionally the firm-level production function has

[1] De Loecker and Eeckhout 2017; Eggertsson, Robbins, and Wold 2018; Decker et al. 2018.

a U-shaped average cost curve due to increasing marginal costs and a fixed overhead cost. The fixed overhead cost implies there are increasing returns to scale in steady-state under imperfect competition, but with perfect competition there are constant returns at minimum average cost. Imperfect competition, due to product differentiation, leads to endogenous markups as entry expands the number of competing products which weakens price setting ability. Dynamic entry introduces a new state variable, number of firms, in addition to capital, and implies markups adjust slowly. Entry occurs to arbitrage profits which are non-zero whilst entry takes place, but zero in long-run steady-state when entry ceases.

Given a permanent positive aggregate technology shock: profits, entry, employment, investment, entry costs and productivity are procylical, whereas markups are countercyclical. The shock initially affects a fixed number of incumbent firms with a fixed amount of capital. Consequently the incumbents increase their output which increases profits as markups are unchanged due to no entry. The change in output occurs through the direct effect of a better technology and because labor adjusts instantaneously. Scale economies from overhead costs increase incumbents' productivity inline with the output expansion. However, by raising output and gaining monopoly profits, prospective entrants evaluate that paying a sunk cost to enter and receive profits outweighs the opportunity cost of investing at the market rate, hence entry occurs. Entry reduces the output and profits that incumbents temporarily gained which reduces productivity as scale declines, but with greater competition, lower markups mean that incumbents must produce more output in the long run to break-even at zero-profit steady-state. Hence there is a countervailing long-run scale effect increasing productivity. Therefore entry has opposing effects on measured productivity. It decreases the profit bias, but increases the markup bias.

Related Literature Recent progress to understand aggregate productivity through firm entry focuses on static entry and its selection effects due to heterogeneity in firm productivity, whereas I focus on the intertemporal effects of homogeneous incumbent firms bearing shocks and subsequently adjusting to new entrants.[2] The interaction between imperfect competition, increasing returns to scale and technology shocks is an established explanation for procylical productivity.[3] Also, in a static entry setup the positive effect of endogenous markups on measured TFP is understood (Portier 1995; Jaimovich and Floetotto 2008). The contribution of this paper is to focus on the *dynamic* effect of firm entry on productivity, specifically to distinguish the contributions of profits and markups intertemporally. This focus on productivity differs from emerging literature on dynamic firm entry based on Bilbiie, Ghironi, and Melitz 2012.[4] Additionally, the work generalizes the cost structure of the firm to allow for perfect competition and presents a non-parametric tractable analysis. To facilitate this, I use an endogenous sunk cost setup, based on Datta and Dixon 2002, which generates dynamic entry and allows for a tractable analysis in continuous time. Savagar and Dixon 2017 study a similar dynamic firm entry model with fixed markups and interpret short-run procylical productivity movements through excess capacity utilization.

1 Cost Curves Intuition

Figure 1 shows the cost curves and equilibria of a firm with increasing marginal costs and a U-shaped average cost due to a fixed overhead cost. The two diagrams represent the dynamic effect on an incumbent firm's costs and in turn productivity resulting from an improvement in technology $A_0 < A_1$

[2] Static entry literature, such as Da-Rocha, Tavares, and Restuccia 2017; Baqaee and Farhi 2017, focuses on producer heterogeneity, allocation and selection effects, thus between-firm productivity is the interest whereas, in this paper, with dynamic entry the interest is within-firm productivity over the cycle as firms adjust their production in response to entry.

[3] Hall 1990; Caballero and Lyons 1992; Hornstein 1993; Devereux, Head, and Lapham 1996a; Basu and Fernald 2001.

[4] Etro and Colciago 2010; Lewis and Poilly 2012 investigate model performance under different forms of strategic interaction and endogenous markup. Jaimovich and Floetotto 2008 focus on static entry in their main paper, but simulate for dynamic entry in the appendix. The dynamic entry results cause weaker measured TFP amplification than in the static case. My tractable analytic analysis helps to explain these simulated results, and provides new explanations for the changing role of profits and markups intertemporally.

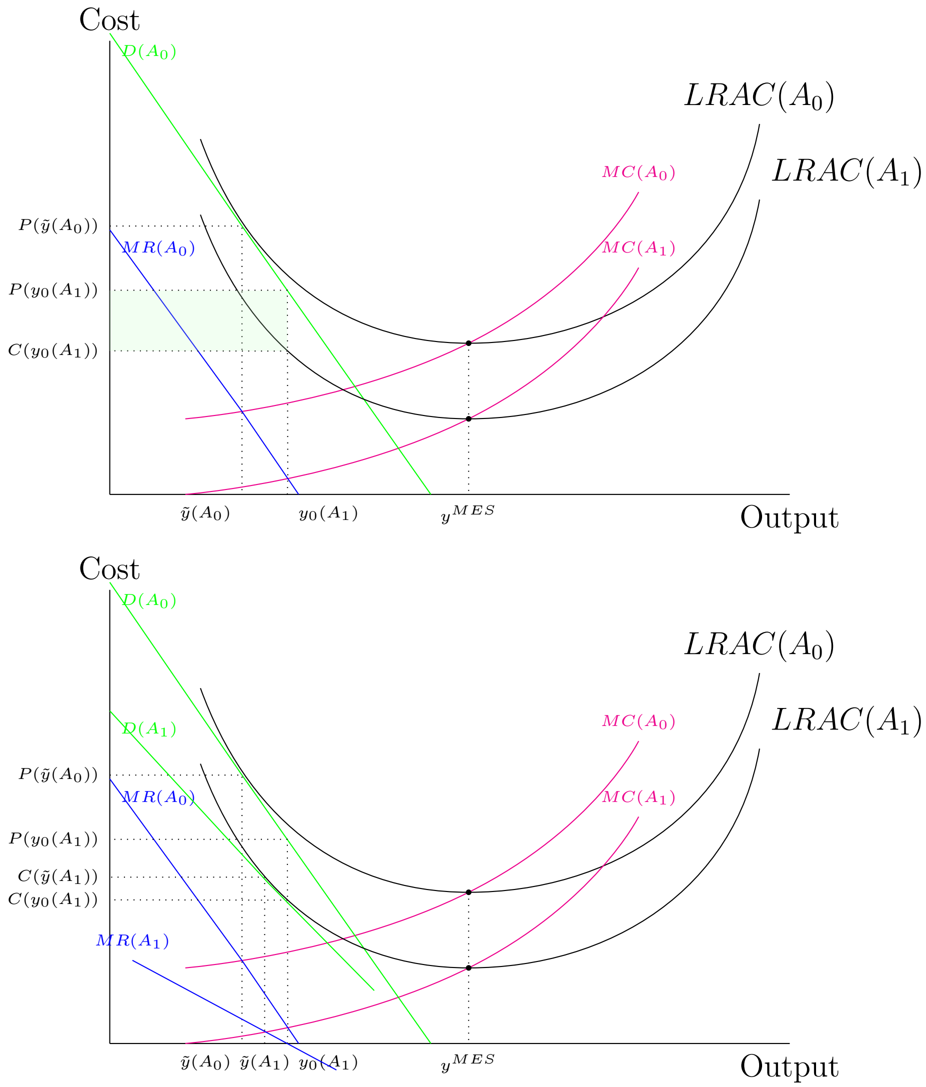

Figure 1: Cost Curve Explanation

when entry is slow and affects price setting ability. The 'true' measure of technological productivity growth that we hope to capture is the shift in the average cost curve, most easily captured at the perfectly competitive minimum average cost level.

Initially the economy is at steady state $\tilde{y}(A_0), p(\tilde{y}(A_0))$ where $MR(A_0) = MC(A_0)$.[5] At time $t = 0$ a permanent technology improvement shifts both

[5] This is the steady state outcome under imperfect competition due to downward slop-

marginal and average cost curves downwards instantaneously. With slow firm adjustment, entry does not take place immediately, firms are quasi-fixed, so that incumbents face demand (price) $D(A_0)$ and $MR(A_0)$ but lower costs $LRAC(A_1)$, such that producing where $MR(A_0) = MC(A_1)$ leads to economic profit indicated by the shaded rectangle. At this profit maximizing level the firm produces $y_0(A_1)$ which shows it expands its scale on impact. This is because it can change labor freely to achieve optimum.[6] The increase in scale implies a movement down the LRAC curve which corresponds to a productivity improvement on top of the parallel shift from technology. This is the productivity increase associated with economic profits. Subsequently (lower graph) entry takes place, arbitraging profit until demand shifts to tangency between $D(A_1)$ and $LRAC(A_1)$. The new steady-state following full entry adjustment corresponds to $\tilde{y}(A_1)$. It shows that following the initial increase in firm scale and corresponding endogenous rise in productivity, there is a decline in scale reducing productivity but not back to the initial level. There is a long-run increase in scale and thus productivity because entry has a second effect: it increases competition among firms making their demand curves more elastic. As the demand curve becomes flatter so tangency is achieved at a point corresponding to an increase in output. This is because with weaker demand (price) firms must produce more units to break even in the long-run zero-profit steady state. Therefore entry has two opposing effects on productivity. The shifting in of the demand curve from business stealing (profit arbitrage) reduces incumbent scale and thus productivity, whereas the greater competition which flattens the demand curve increases scale and thus productivity, and this persists into the long run. If entry did not endogenously affect price setting ability (Dixit and Stiglitz 1977), so

ing demand curves from product differentiation. The level of output is less than the firm's minimum efficient scale (MES) $\tilde{y}(A_0) < y^{\text{MES}}$ where average cost and marginal cost intersect and long run average costs are minimized – the difference is excess capacity. The MES would arise under perfect competition where demand curves are horizontal due to no product differentiation.

[6] See supplementary appendix figure for a full diagram with short-run average cost curves which demonstrates that incumbents move along the SRAC by varying labor instantaneously, whereas in the long-run they move along the LRAC envelope because capital can also be adjusted.

constant markups with dynamic entry, then the demand and revenue curves would shift inwards in parallel (no flattening) such that $MR(A_1) = MC(A_1)$ at the original firm scale $\tilde{y}(A_0)$ implying no long-run scale and thus no long-run productivity effect, despite the short-run profit effect.[7]

Fixed markups and instantaneous entry: Importantly imperfect competition (markup greater than 1) is not sufficient to impart biases on our measure of technological improvement. Beginning at $p(\tilde{y}(A_0))$ if the demand curve shifts in immediately (due to zero-profit i.e.instantaneous entry) and it maintains its slope (due to fixed markup i.e. constant price elasticity of demand) we immediately attain tangency at $p(y_0(A_1))$, which perfectly captures the change in technology as the change $p(\tilde{y}(A_0)) \rightarrow p(y_0(A_1))$ exactly corresponds to the downward shift in the AC curve.[8] The problem in our case is that what we measure captures this distance, plus some extra represented by $p(y_0(A_1)) \rightarrow C(y_0(A_1))$, and corresponding to profit. This is the concept of quasi-fixity (Morrison 1992): the cost of an extra unit of output $C(y_0(A_1))$ diverges from the marginal (revenue) product $p(y_0(A_1))$, unlike in the case discussed earlier in this paragraph where they are instantaneously the same. To overcome this it is better to find an alternative measure of cost (not factor price) that accurately represents MRP i.e. we immediately want $p(y_0(A_1))$ rather than the $C(y_0(A_1))$ we observe. Morrison 1992 uses shadow prices. In steady-state the AC/MC ratio is also the P/MC ratio, hence returns to scale equal markups under zero-profits.

The remainder of the paper formalizes this intuition in a DGE model taking special care to disentangle the two opposing effects from entry.

[7] Savagar and Dixon 2017 focus on this so-called excess capacity utilization effect in the absence of endogenous markups, and relate it to microproduction theory (Morrison 2012) and capital-utilization literature.

[8] Just as in the trivial perfect competition case at minimum AC

2 Model

2.1 Household

The representative household chooses future consumption $\{C(t)\}_0^\infty \in \Re$ and labor supply $\{L(t)\}_0^\infty \in [0,1]$ to maximise lifetime utility $\mathrm{U} : \Re^2 \rightarrow \Re$. Instantaneous utility $u : \Re \times [0,1] \rightarrow \Re$ is jointly concave and differentiable in both of its arguments. It is strictly increasing in in consumption $u_C > 0$, strictly decreasing in labor u_L and separable $u_{CL} = 0$. The household owns capital and takes equilibrium rental rate, wage rate and firm profits $K, r, w, \Pi \in \Re_+$ as given. The household solves

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

$$\text{s.t.} \quad \dot{K}(t) = rK(t) + wL(t) + \Pi(t) - C(t) \tag{2}$$

Optimal paths satisfy the intertemporal consumption Euler equation (3), intratemporal labor-consumption trade-off (4) and resource constraint (2).[9]

$$\dot{C}(t) = -\frac{u_C}{u_{CC}}(r(t) - \rho), \tag{3}$$

$$w(t) = -\frac{u_L(L(t))}{u_C(C(t))} \tag{4}$$

To complete the solution for the boundary value problem, we impose a transversality condition on the upper boundary and an initial condition on the lower boundary.

$$\lim_{t\to\infty} K(t)\lambda(t)e^{-\rho t} = 0, \tag{5}$$

$$K(0) = K_0 \tag{6}$$

$\lambda(t) = u_C$ is the co-state variable which represents the marginal utility of consumption.

[9] See appendix for the Hamiltonian problem and 6 associated Pontryagin conditions.

2.2 Firm

There is a nested CES aggregator composing aggregate output Y from a continuum of industries $\jmath \in [0,1]$ producing $Q_\jmath$ containing a finite number of N firms producing $y_{\jmath\imath}$. At the aggregate level and industry level there is perfect competition (prices P and $P_\jmath$ are given). But at the firm level there is oligopolistic (Cournot) competition. The firm can affect own-price $P_{\jmath\imath}$ through a direct effect (standard monopolistic competition) and through its effect on the industry level output. The result is endogenous demand elasticity that becomes more elastic with more firms, hence a markup that decreases in entry.

The program faced by the final output producer is

$$\max_{Q_\jmath,\ \jmath\in[0,1]} \left\{ PY - \int_0^1 Q_\jmath p_\jmath d\jmath \right\} \tag{7}$$

subject to the aggregate output production function

$$Y = \left[\int_0^1 Q_\jmath^{\frac{\theta_I - 1}{\theta_I}} d\jmath \right]^{\frac{\theta_I}{\theta_I - 1}}, \quad \theta_I \geq 1 \tag{8}$$

The first-order condition gives inverse demand for industry $\jmath$

$$P_\jmath = \left(\frac{Q_\jmath}{Y} \right)^{-\frac{1}{\theta_I}} P \tag{9}$$

The program faced by the industry is

$$\max_{y_{\jmath\imath},\ \imath\in(1\ldots N)} \left\{ p_\jmath Q_\jmath - \sum_1^N y_{\jmath\imath} p_{\jmath\imath} \right\} \tag{10}$$

subject to the industry output production function where $\vartheta \in [0, \infty)$ is a variety effect[10]

$$Q_\jmath = N^{1+\vartheta}\left[\frac{1}{N}\sum_{\imath=1}^{N} y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}\right]^{\frac{\theta_F}{\theta_F-1}}, \quad \theta_F > \theta_I \geq 1 \tag{11}$$

This leads to industry-level inverse demand

$$P_{\jmath\imath} = \left[\frac{y_{\jmath\imath}}{Q_\jmath}\right]^{-\frac{1}{\theta_F}} N^{-\frac{1}{\theta_F}+\vartheta\frac{\theta_F-1}{\theta_F}} P_\jmath \tag{12}$$

Combining the aggregate and industry level inverse demands gives the firm-level inverse demand

$$P_{\jmath\imath} = \left(\frac{y_{\jmath\imath}}{Q_\jmath}\right)^{-\frac{1}{\theta_F}} \left(\frac{Q_\jmath}{Y}\right)^{-\frac{1}{\theta_I}} N^{-\frac{1}{\theta_F}+\vartheta\frac{\theta_F-1}{\theta_F}} P \tag{13}$$

Firms are symmetric in their production technology, firm $\imath$ in industry $\jmath$ produces output:

$$y_{\jmath\imath}(t) := \max\{AF(k_{\jmath\imath}(t), \ell_{\jmath\imath}(t)) - \phi, 0\} \tag{14}$$

where $F : \Re^2_+ \supseteq (k,\ell) \to \Re_+$ is homogeneous of degree $\nu \in (0,1)$ (hod-ν). $\phi \in \Re_+$ is an overhead cost denominated in output.[11] Inada's conditions hold so that marginal products of capital and labor are strictly positive which rules out corner solutions $F_k, F_\ell > 0$, and the Hessian of F satisfies concavity properties $F_{k\ell} = F_{\ell k} > 0$, $F_{kk}, F_{\ell\ell} < 0$ and $F_{kk}F_{\ell\ell} - F^2_{k\ell} > 0$.[12] $A \in [1,\infty)$ is

[10] In comparable setups Jaimovich and Floetotto 2008 remove the variety effect $\vartheta = 0$ whereas Atkeson and Burstein 2008 leave it in $\vartheta = \frac{1}{\theta_F - 1}$, $\theta_F > 1$.

[11] As in Hornstein 1993; Devereux, Head, and Lapham 1996a; Rotemberg and Woodford 1996; Cook 2001; Jaimovich 2007 the fixed overhead parameter implies profits will be zero in steady-state despite market power. The overhead cost means that marginal costs do not measure returns to scale. We focus on the case where marginal costs are increasing, but average costs are decreasing, so there are (locally) increasing returns to scale.

[12] Homogeneity of degree ν and cross-derivative symmetry implies

$$\begin{aligned} \nu F(k,l) &= F_k k + F_l l \\ (\nu-1)F_l &= F_{ll} l + F_{lk} k = F_{ll} l + F_{kl} k \\ (\nu-1)F_k &= F_{kl} l + F_{kk} k = F_{lk} l + F_{kk} k \end{aligned}$$

a scale parameter reflecting the production technology.

2.2.1 Dual Cost Function

The firm's cost minimization problem yields the cost function dual to the intermediate producer's production function. The traditional Lagrange cost minimization problem yields

$$\mathrm{C}(r, w, y_{\jmath\imath}) = \mathrm{MC}\, \nu(y_{\jmath\imath} + \phi) \tag{15}$$

where the Lagrane multiplier is the MC. However the MC is not independent of output, unless there are constant marginal costs. For homothetic functions (of which homogeneous functions are a subset) the cost function can be rearranged to isolate output effects. A hod-ν production function exhibits unit-cost function form[13]

$$\mathrm{C}(r, w, y_{\jmath\imath}) = \left(\frac{y_{\jmath\imath} + \phi}{A}\right)^{\frac{1}{\nu}} \mathrm{C}(r, w, 1) \tag{16}$$

where the unit cost function $C(r, w, 1)$ is independent of output. Given this, the cost function representation (16) implies the marginal cost is positive and increasing if $\nu \in (0, 1)$, but flat with $\nu = 1$

$$\mathrm{MC} = \frac{\partial\, \mathrm{C}(r, w, y_{\jmath\imath})}{\partial y} = \frac{1}{\nu} \frac{\mathrm{C}(r, w, y_{\jmath\imath})}{y + \phi} > 0 \tag{17}$$

$$\frac{\partial\, \mathrm{MC}}{\partial y} = \frac{1 - \nu}{\nu} \frac{\mathrm{MC}}{y_{\jmath\imath} + \phi} > 0 \quad \text{if} \quad \nu \in (0, 1) \tag{18}$$

[13]The outline is that the production function can be rearranged for constant-output factor demands $k(y, k/\ell, w, r), \ell(y, \ell/k, w, r)$ and substituted into $rk + w\ell$ which gives the multiplicative form but with k/ℓ ratios in the second part. Then the Lagrangean cost minimization FOCs imply that this ratio is independent of y, as is the case for all homothetic functions.

From (16), we can see the average cost curve is U-shaped with increasing marginal costs $\nu \in (0,1)$ and an overhead cost ϕ

$$\mathrm{AC} = \frac{\mathrm{C}(r,w,y_{\jmath\imath})}{y_{\jmath\imath}} = \frac{\mathrm{C}(r,w,1)(y_{\jmath\imath}+\phi)^{\frac{1}{\nu}}}{y_{\jmath\imath}} \tag{19}$$

$$\frac{\partial\,\mathrm{AC}}{\partial y} = \frac{\mathrm{AC}}{y}\left(\frac{1}{\nu(1+s_\phi)} - 1\right), \qquad \text{where } \; s_\phi \equiv \frac{\phi}{y} \tag{20}$$

The individual firm solves

$$\max_{y_{\jmath,\imath}} \pi_{\jmath\imath} = P_{\jmath\imath} y_{\jmath\imath} - C(r,w,y_{\jmath\imath}) \tag{21}$$

subject to inverse demand (13) and cost function (16) and taking factor prices as given. This implies that at optimal $\mathrm{MR}_{\jmath\imath} = \mathrm{MC}_{\jmath\imath}$. Marginal revenue depends on both the price gained from increasing output but, due to price setting ability, also the negative effect on price level captured by the price elasticity of demand $\epsilon_{\jmath\imath}$

$$\epsilon_{\jmath\imath} \equiv -\frac{\partial y_{\jmath\imath}}{\partial P_{\jmath\imath}} \frac{P_{\jmath\imath}}{y_{\jmath\imath}} \tag{22}$$

$$\mathrm{MR}_{\jmath\imath} = \frac{\partial P_{\jmath\imath}}{\partial y_{\jmath\imath}} y_{\jmath\imath} + P_{\jmath\imath} = P_{\jmath\imath}\left(\frac{\epsilon_{\jmath\imath}-1}{\epsilon_{\jmath\imath}}\right) \tag{23}$$

Therefore, since $\mathrm{MR}_{\jmath\imath} = \mathrm{MC}_{\jmath\imath}$, a firm prices at a markup $\mu_{\jmath\imath} \in (1,\infty)$ of price over marginal cost

$$\mu_{\jmath\imath} \equiv \frac{P_{\jmath\imath}}{\mathrm{MC}_{\jmath\imath}} = \frac{\epsilon_{\jmath\imath}}{\epsilon_{\jmath\imath}-1} \tag{24}$$

2.2.2 Endogenous Price Elasticity of Demand

At the firm-level an individual firm will maximize profits subject to inverse demand (13) and its production function (14). Crucially the firm has a degree of price setting ability over $P_{\jmath\imath}$. Under quantity competition (Cournot), it chooses $y_{\jmath\imath}$.[14] Differentiating (13) with respect to $y_{\jmath\imath}$ and multiplying by $-\frac{y_{\jmath\imath}}{P_{\jmath\imath}}$

[14] See Appendix A.1 for Bertrand case.

gives inverse price elasticity of demand

$$\epsilon_{\jmath\imath}^{-1} = -\frac{\partial P_{\jmath\imath}}{\partial y_{\jmath\imath}} \frac{y_{\jmath\imath}}{P_{\jmath\imath}} = \frac{1}{\theta_F} + \left[\frac{1}{\theta_I} - \frac{1}{\theta_F}\right] \frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} \frac{y_{\jmath\imath}}{Q_\jmath} \tag{25}$$

There are two components to the inverse price elasticity of demand: a regular direct effect that occurs with monopolistic competition and a second endogenous effect that arises because firms are 'large' in their industry and thus can affect industry output $Q_\jmath$. Therefore by choosing their own production $y_{\jmath\imath}$, a firm has a direct effect $\frac{1}{\theta_F}$ and an indirect effect through $\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}}$. It is the indirect effect that causes endogenous markups, and in its absence (when firms are atomistic $\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} = 0$) there is standard monopolistic competition from the direct effect. Later, when we impose symmetric equilibrium, we shall see $\frac{y_{\jmath\imath}}{Q_\jmath}\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} = \frac{1}{N}$ which makes the price elasticity of demand and in turn markup dependent on number of firms.

2.2.3 Factor Market Equilibrium

An optimizing firm's conditional demand for hours worked and capital are given by the following factor market equilibrium[15]

$$\frac{AF_k(k,\ell)}{\mu(N)} = \frac{r}{P_{\jmath\imath}} \tag{26}$$

$$\frac{AF_\ell(k,\ell)}{\mu(N)} = \frac{w}{P_{\jmath\imath}} \tag{27}$$

The marginal revenue product of capital (MRPK) $\frac{AF_k}{\mu(N)}$ equates to the real price of capital and the MRPL $\frac{AF_\ell}{\mu(N)}$ equates to the real price of labor. As markups increase the marginal revenue from an additional unit of production is less. The presence of the endogenous markup (μ depending on N), causes the MRPs to be nonmonotone functions of N. This can cause multiple equilibria: different numbers of firm cause the factor market relationship to

[15] Where the Lagrange multiplier is marginal cost, a firm's cost minimization problem yields $r = \mathrm{MC}\, AF_k$ and $w = \mathrm{MC}\, AF_\ell$ leading to cost function (15) by $C = rk + w\ell$. Dividing factor prices by $P_{\jmath\imath}$ yields the factor market equilibrium in markup terms.

hold.[16] I provide uniqueness conditions later.

2.2.4 Returns to Scale (RTS)

Returns to scale are defined as the average cost to marginal cost ratio

$$\mathrm{RTS} \equiv \frac{\mathrm{AC}}{\mathrm{MC}} \quad \begin{cases} \in (1,\infty), & \text{increasing returns} \\ = 1, & \text{constant returns} \\ \in (0,1), & \text{decreasing returns} \end{cases} \tag{28}$$

The production function exhibits two alternative representations of returns to scale defined as $\frac{AC}{MC}$. They arise from the cost function and profit definition respectively, the former depends on market structure, whereas the latter relates to technical parameters of the production function[17]

$$\mathrm{RTS} = \nu\,(1+s_\phi) = \mu\,(1-s_\pi) \gtreqless 1 \tag{29}$$

where $s_\phi \equiv \frac{\phi}{y}$ and $s_\pi \equiv \frac{\pi}{y}$ are fixed cost and profit shares in output. The *'cost-based measure'*, $\mathrm{RTS} = \nu(1+s_\phi)$, highlights that entry affects RTS through output only, by affecting the fixed cost share, as the other variables are fixed parameters, whereas the *'profit-based measure'*, $\mathrm{RTS} = \mu(1-s_\pi)$, shows markups, profit and output determine RTS.

It is useful to contrast my work in relation to RTS with related papers. In Bilbiie, Ghironi, and Melitz 2012 there are no overhead costs $\phi = 0$ and marginal costs are constant $\nu = 1$ implying $\mathrm{RTS} = 1$ from the cost-based measure. To demonstrate RTS from the profit-based measure consider their equilibrium profit condition is $\pi = \left(1 - \frac{1}{\mu}\right) y$, analogous to (50), except with constant returns to scale ($\mathrm{RTS} = 1$) the implied profit share is in a one-one mapping with markups and, in fact, equal to the Lerner index measure of market power $s_\pi = \left(1 - \frac{1}{\mu}\right) = \mathrm{LI} = \frac{1}{\epsilon} \in (0,1)$. Hence the profit-based RTS measure in (29) is also 1. However in BGM the profit share is always posi-

[16] Linnemann 2001 investigate this in a model with instantaneous entry, endogenous markups and no capital.

[17] See Appendix for derivation.

tive if $\mu > 1$, this is due to the absence of a fixed overhead cost which in our model wipes out any excess profit, and the absence of sunk costs in the long run as there is no congestion. In their work the one-off, wage-denominated entry cost is fixed so fulfills the role of eliminating profits. In the long-run it equates to profits, so net of the entry cost profits are zero. Therefore the presence of constant returns, will remove any of the endogenous productivity effects I focus on in this paper, even though their dynamic entry setup creates the same short-run intensive margin (variation in output per firm) adjustment and long-run intensive-extensive margin (variation in output per firm and aggregate output) adjustment. In Jaimovich and Floetotto 2008 $\phi > 0,\ \nu = 1$, implying flat marginal cost and globally downward sloping average cost, so there are globally increasing returns and equilibrium only exists with imperfect competition $\mu > 1$ – there is no Walrasian benchmark at *minimum efficient scale*. In their work the equilibrium profit condition is $\pi = \left(1 - \frac{1}{\mu}\right) y - \frac{1}{\mu}\phi$ such that $s_\pi = \left(1 - \frac{1}{\mu}\right) - \frac{1}{\mu} s_\phi$ and entry is instantaneous such that the profit-share is always zero and returns to scale always equal the endogenous markup, which always reflect the fixed cost share in output.

2.2.5 Symmetric Equilibrium Profits, Prices, Aggregates

We already derived the symmetric equilibrium elasticity and markup (45). In deriving the markup, symmetry provided the crucial step to link price setting ability to number of competitors. That is, it determined that own output effect on industry output is declining in number of competitors $\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} = \frac{Q_\jmath}{y_{\jmath\imath}} \frac{1}{N}$.[18] Under symmetry, intermediate variables are equivalent

$$\forall (\jmath, \imath) \in [0,1] \times [1, N(t)] : y_{\jmath\imath} = y,\ k_{\jmath\imath} = k,\ \ell_{\jmath\imath} = \ell, N_\jmath = N$$

Perfect factor markets imply aggregate capital and hours are divided evenly among firms, such that the number of firms behaves as a quasi-input deter-

[18] For Betrand own price on industry price is declining in number of competitors $\frac{\partial P_\jmath}{\partial P_{\jmath\imath}} = \frac{P_\jmath}{P_{\jmath\imath}} \frac{1}{N}$.

mining output through how resources are divided.

$$k = \frac{K}{N}, \tag{30}$$

$$\ell = \frac{L}{N} \tag{31}$$

$$F(k,\ell) = F\left(\frac{K}{N}, \frac{L}{N}\right) = N^{-\nu}F(K,L) \tag{32}$$

$$F_k(k,\ell) = F_K\left(\frac{K}{N}, \frac{L}{N}\right) = N^{1-\nu}F_N(K,L) \tag{33}$$

$$F_\ell(k,\ell) = F_\ell\left(\frac{K}{N}, \frac{L}{N}\right) = N^{1-\nu}F_L(K,L) \tag{34}$$

$$y = AN^{-\nu}F(K,L) - \phi \tag{35}$$

Integrating a symmetric quantity over the $[0,1]$ interval implies industries are representative of aggregate $Y = Q_\jmath$ and $P = P_\jmath \; \forall \jmath$. The aggregation of firm level to industry level will depend on any variety effect (ϑ)

$$Q_\jmath = N^{1+\vartheta}\left[\frac{1}{N}\sum_{\imath=1}^{N} y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}\right]^{\frac{\theta_F}{\theta_F-1}} \implies Q_\jmath = N^{1+\vartheta}y \quad (=Y) \tag{36}$$

$$P_\jmath = N^{\frac{1}{\theta_F-1}-\vartheta}\left(\sum_{1}^{N} P_{\jmath\imath}^{1-\theta_F}\right)^{\frac{1}{1-\theta_F}} \implies P_\jmath = N^{-\vartheta}P_{\jmath\imath} \quad (=P) \tag{37}$$

Given a variety effect $\vartheta > 0$, in symmetric equilibrium the relative price depends on the number of firms (which each produce a variety)

$$\varrho(N) = \frac{P_{\jmath\imath}}{P} = N^{\vartheta} \tag{38}$$

In Atkeson and Burstein 2008, Bilbiie, Ghironi, and Melitz 2012, Table 1, p.312 and Etro and Colciago 2010, Eq. C.1, p1230 the variety effect is $\vartheta = \frac{1}{\theta_F-1}$, but is ignored for data-consistent variable interpretations.[19] Here we

[19]In both Etro and Colciago 2010; Bilbiie, Ghironi, and Melitz 2012, the implication of variety effects is to carefully interpret different methods to deflate variables: either data-consistent (divide by $P_{\jmath\imath}$) or welfare-consistent (divide by P). Empirically relevant variables net out the effect of changes in the number of firms (varieties/products), whereas welfare-consistent variables always assume variations in the number of varieties are left in the deflator.

follow Jaimovich and Floetotto 2008, and assume no variety effect $\vartheta = 0$, thus normalizing final good price to 1 we have

$$P = P_\jmath = P_{\jmath\imath} = 1 \tag{39}$$

$$Y = Ny \tag{40}$$

$$= N^{1-\nu} AF(K, L) - N\phi \tag{41}$$

Under symmetry inverse price elasticity of demand (25) under Cournot becomes[20]

$$\epsilon^{-1} = \frac{1}{\theta_F} + \left[\frac{1}{\theta_I} - \frac{1}{\theta_F}\right]\frac{1}{N} \tag{44}$$

which is also the Lerner index (LI $\in (0, 1)$) of market power LI $= \frac{\mathrm{P-MC}}{\mathrm{P}} = 1 - \frac{1}{\mathrm{P/MC}} = \epsilon^{-1}$. Therefore the symmetric Cournot markup, with $N \geq 1$, is

$$\mu(N) = \frac{1}{1-\epsilon^{-1}} = \frac{1}{1 - \frac{1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)}, \qquad \theta_F > \theta_I \geq 1 \tag{45}$$

Markups decrease as number of competitors increase because they dilute incumbents' market share, making price elasticity of demand more elastic. The markup is bounded above by the single firm industry case (monopolistic competition) $\lim_{N\to 1} = \frac{\theta_I}{\theta_I - 1}$ and bounded below by $\lim_{N\to\infty} = \frac{\theta_F}{\theta_F - 1}$, which is the perfect competition case if goods are homogeneous at the firm level $\lim_{\substack{N\to\infty \\ \theta_F\to\infty}} \mu(N) = 1$. The elasticity of the markup with respect to number of

[20] Taking the derivative of (11) gives

$$\frac{\partial Q_{\jmath\imath}}{\partial y_{\jmath\imath}} = \frac{Q_\jmath}{\left(\sum_{\imath=1}^N y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}\right)} y_{\jmath\imath}^{-\frac{1}{\theta_F}} \implies \frac{\partial Q_{\jmath\imath}}{\partial y_{\jmath\imath}}\frac{y_{\jmath\imath}}{Q_{\jmath\imath}} = \frac{y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}}{\left(\sum_{\imath=1}^N y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}\right)} \tag{42}$$

and under symmetry $\left(\sum_{\imath=1}^N y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}\right) = N y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}}$ so

$$\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} = \frac{Q_\jmath}{y_{\jmath\imath}}\frac{1}{N} \implies \frac{y_{\jmath\imath}}{Q_\jmath}\frac{\partial Q_\jmath}{\partial y_{\jmath\imath}} = \frac{1}{N} \tag{43}$$

firms $\varepsilon_\mu \equiv \frac{\partial \mu}{\partial N}\frac{N}{\mu}$ measures the responsiveness of markups to entry

$$\varepsilon^{\mathrm{C}}_{\mu N} = 1 - \frac{\frac{\theta_F - 1}{\theta_F}}{\frac{\theta_F - 1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)} = 1 - \left(\frac{\theta_F - 1}{\theta_F}\right)\mu < 0 \tag{46}$$

The case that gives the most responsive markups is when $N \to 1, \theta_I = 1$: a small number of firms competing in a unique industry. The corresponding markup is $\mu = \frac{\theta_F N}{(\theta_F - 1)(N-1)}$ and its elasticity is $\varepsilon_{\mu N}|^{\theta_I = 1} = -\frac{1}{N-1}$.[21] Appendix A.4 illustrates these markup properties graphically and for the Bertrand case.

With normalized pricing and symmetry, factor market equilibrium becomes

$$r = \frac{AF_k(k,\ell)}{\mu(N)} = \frac{AN^{1-\nu}F_K(K,L)}{\mu(N)} \tag{47}$$

$$w = \frac{AF_\ell(k,\ell)}{\mu(N)} = \frac{AN^{1-\nu}F_L(K,L)}{\mu(N)} \tag{48}$$

Through these factor prices and Euler's homogeneous function theorem, variable costs can be expressed as a function of output per firm and markups

$$rk + w\ell = \frac{\nu}{\mu(N)}(y + \phi) \tag{49}$$

In turn operating profits and output per firm are in a one-one mapping, where operating profits under symmetry are $\pi = y - rk - w\ell$.[22]

$$\pi(y, \mu(N)) = \left(1 - \frac{\nu}{\mu(N)}\right)(y + \phi) - \phi = \left(1 - \frac{\nu}{\mu(N)}\right)y - \frac{\nu}{\mu(N)}\phi \tag{50}$$

$$y(\pi, \mu(N)) = \frac{\pi + \phi}{1 - \frac{\nu}{\mu(N)}} - \phi = \frac{\pi + \frac{\nu}{\mu(N)}\phi}{1 - \frac{\nu}{\mu(N)}} \tag{51}$$

[21] Between firm substitutability $\theta_F > \theta_I = 1$ can be specified arbitrarily, it does not affect the elasticity.

[22] Operating profit, often called dividends, is the profit of an incumbent firm in a given period which excludes the one-time sunk entry cost. The sunk entry cost is included in aggregate profits.

2.2.6 Measured TFP

In symmetric equilibrium aggregate output can be shown to be the product of an efficiency term, which we call *measured TFP*, and aggregate inputs of capital and labor. Beginning from aggregate output in symmetric equilibrium (41) and accounting for the gross overhead cost $N\phi$ in terms of inputs, we can derive[23]

$$Y = \left(\frac{A}{\pi + \phi}\right)^{\frac{1}{\nu}} \left(1 - \frac{\nu}{\mu}\right)^{\frac{1}{\nu}-1} \left(\pi + \frac{\nu}{\mu}\phi\right) F(K, L)^{\frac{1}{\nu}} \tag{52}$$

Hence measured TFP relates aggregate output to inputs[24]

$$\text{TFP} \equiv \frac{Y}{F(K, L)^{\frac{1}{\nu}}} = \frac{y}{F(k, \ell)^{\frac{1}{\nu}}} \tag{53}$$

This term is influenced by endogenously varying markups and profits

$$\text{TFP}(t) \equiv \left(\frac{A}{\pi(t) + \phi}\right)^{\frac{1}{\nu}} \left(1 - \frac{\nu}{\mu(t)}\right)^{\frac{1}{\nu}-1} \left(\frac{\nu}{\mu(t)}\phi + \pi(t)\right) \tag{54}$$

The two endogenous components of measured TFP trace back to the two key mechanisms in our model. Profits are endogenous due to dynamic entry and markups are endogenous due to competition effects. A useful representation of our measured TFP definition is

$$\text{TFP}(t) = A^{\frac{1}{\nu}} \frac{y(t)}{(y(t) + \phi)^{\frac{1}{\nu}}} \tag{55}$$

This shows given fixed parameters ϕ, ν, variations in measured TFP can arise through two variables: exogenous technology A and output per firm y. Output per firm y has an effect on measured TFP providing there are not

[23]See Appendix D for full derivation. First substitute out $N = \frac{Y}{y}$ from (40), then collect terms in Y, lastly replace $y(\pi, \mu(N))$ using (51).

[24]See Barseghyan and DiCecio 2011 for a similar derivation of measured TFP with non-constant marginal costs and an overhead cost (albeit wage denominated). The normalization of the denominator to be homogeneous of degree 1 with scale economies is common (see Da-Rocha, Tavares, and Restuccia 2017, p.22 who also have an output denominated fixed cost).

constant returns to scale (RTS = 1):

$$\frac{\partial \, \mathrm{TFP}}{\partial y} = \frac{\mathrm{TFP}}{y}\left(1 - \frac{1}{\nu\left(1+\frac{\phi}{y}\right)}\right) = \frac{\mathrm{TFP}}{y}\left(1 - \frac{1}{\mathrm{RTS}}\right) \tag{56}$$

Through (51) output per firm depends on profits and markups which in turn depend on A. Therefore through (55), exogenous shocks to A will have a direct effect on measured TFP but also an effect due to y which by (56) affects measured TFP due to RTS.

2.2.7 Minimum Efficient Scale (MES)

From the U-shaped AC curve slope (20), average costs are minimized when the share of fixed costs in output is $s_\phi = \frac{1-\nu}{\nu}$. The corresponding output is at the firms' *minimum efficient scale* (MES) $y^{\mathrm{MES}} = \frac{\nu\phi}{1-\nu}$. Returns to scale are also constant RTS $= \nu(1+s_\phi) = 1$ at this level of output as average and marginal costs intersect at the minimum. We can also observe from (56) that the minimum efficient scale maximizes measured TFP.

The MES is an indicator of market structure. It is the socially optimal firm size that gives the lowest production costs per unit of output.[25] In symmetric equilibrium the overall size of the market is equivalent to aggregate output or industry output $Y = Q = Ny$, which in steady state equates to consumption $\tilde{Y} = \tilde{C}$. If the minimum efficient scale is small relative to the overall size of the market (demand for the good), there will be a large number of firms $N = \frac{Y}{y^{\mathrm{MES}}}$. The firms in this market will behave in a perfectly competitive manner due to the large number of competitors. Hence ϕ will be the main determinant of market structure, and in turn the endogenous markup.

2.2.8 Firm Entry

An endogenous sunk entry cost and an entry arbitrage condition determine the level of entry and consequently the number of firms operating at time

[25]In the steady state analysis (Section 3.2.1) we show that the minimum efficient scale arises under perfect competition (i.e. no markups $\mu = 1$).

t. The sunk entry cost exhibits a congestion effect, and it is this dynamic sunk cost that prevents instantaneous adjustment of firms to steady state.[26] A prospective entrant's post-entry value is equal to the present discounted value of future operating profits (dividends), which is also the value of an incumbent firm

$$V(t) = \int_{s=t}^{\infty} \pi(t) e^{-r(s-t)} dt \tag{57}$$

A free entry assumption implies that firm value equates to the sunk entry cost $V(t) = q(t)$. We assume that the sunk entry cost $q \in \Re$ increases in the rate of entry $\dot{N}$, and its sensitivity to entry depends on the exogenous congestion parameter γ[27]

$$q(t) = \gamma \dot{N}, \qquad \gamma \in (0, \infty) \tag{58}$$

The congestion effect assumption is a particular form of endogenous sunk cost that captures that resources needed to setup a firm are in inelastic supply, and therefore a greater rate of entry increases the entry cost. For example if firms need to register documents with a Government office before operating, then as more firms enter, the queue increases and the cost increases.[28] Lewis 2009 provides empirical evidence on the importance of entry congestion in replicating empirical dynamics to aggregate shocks.

By taking the derivative, the value function can be represented as the well-known arbitrage condition $rV = \pi + \dot{V}$ that equates an assets opportunity cost to its dividends and change in underlying value.[29] Therefore when

[26]The entry adjustment costs theory is analogous to capital adjustment cost models which recognize that investment in capital is more costly for larger investment.

[27]Its bounds are the two well-known cases: less sensitivity to congestion $\lim_{\gamma \to 0} q(t)$ implies instantaneous free entry, and more congestion sensitivity $\lim_{\gamma \to \infty} q(t)$ implies fixed number of firms.

[28]See Aloi, Dixon, and Savagar 2018 for empirical evidence, from OECD Doing Business data, that links number of procedures to start-up with length of time to create a firm.

[29]See Stokey 2008 for this derivation, where the arbitrage equation is referred to as a stationary Hamilton-Jacobi-Bellman equation.

combined with the free entry condition and sunk cost assumption

$$r(t)q(t) = \dot{q}(t) + \pi(t) \tag{59}$$

The return to paying a sunk costs q to enter and receiving profits equals the return from investing the cost of entry at the market rate $r(t)$. Since the endogenous sunk cost is itself dynamic, the arbitrage condition is a second-order ODE in number of firms. If we define net entry, this second-order ODE is separable into two first-order ODEs.

$$\dot{N}(t) \equiv E(t) \tag{60}$$

$$\dot{E}(t) = -\frac{\pi(t)}{\gamma} + r(t)E(t), \quad \gamma \in (0, \infty) \tag{61}$$

The second-order differential equation requires two boundary conditions for a unique solution

$$N(0) = N_0 \tag{62}$$

$$\lim_{t \to \infty} e^{-\rho t} u_C N(t) q(t) = 0 \tag{63}$$

The rate of entry increases $\dot{E} > 0$ if the outside option $r(t)E(t)$ exceeds the profit from entering $\frac{\pi(t)}{\gamma}$. This is because households invest in the more attractive outside option, as opposed to setting up firms, hence the entry cost falls because there is less congestion. The result is an increase in the amount of entry.

The aggregate cost of entry $Z(t) \in \Re$ is

$$Z(t) = \gamma \int_0^{E(t)} i \, di = \gamma \frac{E(t)^2}{2} \tag{64}$$

In general equilibrium sunk entry costs are accounted for in the aggregate profits of the household's income constraint. Aggregate profits are each firm's operating profits less the aggregate sunk cost of entry.

$$\Pi(t) = N(t)\pi(t) - Z(t) \tag{65}$$

This leads to the aggregate resource constraint[30]

$$\dot{K} = Y - Z(E) - C \tag{70}$$

3 Model Analysis

Table 1 summarizes the model. The core of the model is a four dimensional dynamical system that determines consumption, entry, capital and number of firms (C, E, K, N). Labor supply L does not enter the system as an independent variable because it can be defined in terms of C, K, N through labor market equilibrium.

3.1 Endogenous Labor

Labor market equilibrium occurs where labor demand (27) equals labor supply (4), and it allows us to understand endogenous labor behaviour.[31]

Proposition 1. *In general equilibrium the response of aggregate labor is negative to consumption and positive to capital and entry*

$$\frac{\partial L}{\partial K} = \Phi^{-1} \frac{F_{LK}}{F_L} \qquad > 0 \tag{71}$$

$$\frac{\partial L}{\partial C} = \Phi^{-1} \frac{u_{CC}}{u_C} \qquad < 0 \tag{72}$$

[30] This follows from the income identity. A representative household's income is earned from wages w on labor L, rental r of capital K and total profits Π, which equal aggregate dividends (operating profits) $N(t)\pi(t)$ less entry costs $Z(E(t))$.

$$\Pi(t) = N(t)\pi(t) - Z(E(t)) = Y(t) - w(t)L(t) - r(t)K(t) - Z(E(t)) \tag{66}$$

This income can be spent on consumption and investment in capital

$$I^K(t) + C(t) = w(t)L(t) + r(t)K(t) + \Pi(t) \tag{67}$$

$$\dot{K}(t) + C(t) = Y(t) - Z(E(t)) \tag{68}$$

$$\dot{K} = Y(t) - \frac{\gamma E(t)^2}{2} - C(t) \tag{69}$$

[31] Related literature (Devereux, Head, and Lapham 1996b; Jaimovich 2007) terms (27) the aggregate labor demand function.

Static	
(45) Markup (Cournot)	$\mu(N) = \left[1 - \frac{1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\right]^{-1}$
(4) Labor Supply	$w = -\frac{u_L(L)}{u_C(C)}$
(48) Labor demand	$w = \frac{AF_\ell(k,\ell)}{\mu(N)}$
(47) Capital Rental	$r = \frac{AF_k(k,\ell)}{\mu(N)}$
(14) Output per Firm	$y = AF(k,\ell) - \phi$
(50) Operating Profit per Firm	$\pi = (y+\phi)\left(1 - \frac{\nu}{\mu(N)}\right) - \phi$
(40) Aggregate Output	$Y = Ny$
(53) Productivity	$\text{TFP} = \frac{y}{F(k,\ell)^{\frac{1}{\nu}}}$
Dynamic	
(3) Consumption Euler	$\dot{C} = -\frac{u_C(C(t))}{u_{CC}(C(t))}(r(t) - \rho)$
(61) Entry Arbitrage	$\dot{E} = r(t)E(t) - \frac{\pi(t)}{\gamma}$
(69) Capital Accumulation	$\dot{K} = Y(t) - \frac{\gamma}{2}E(t)^2 - C(t)$
(60) Entry Definition	$\dot{N} = E(t)$
Boundary Values	
(6) Capital Initial Condition	$K(0) = K_0$
(62) Firms Initial Condition	$N(0) = N_0$
(5) Capital Transversality	$\lim_{t\to\infty} e^{-\rho t} u_C(C(t))K(t) = 0$
(63) Firms Transversality	$\lim_{t\to\infty} e^{-\rho t} u_C(C(t))q(t)N(t) = 0$

Table 1: Model Summary

$$\frac{\partial L}{\partial N} = \Phi^{-1}\frac{1}{N}\left(1 - \nu - \varepsilon_{\mu N}\right) \qquad > 0 \tag{73}$$

where $\Phi \equiv \left[\frac{u_{LL}}{u_L} - \frac{F_{LL}}{F_L}\right] = \left[\frac{\varepsilon_{u_L L} - \varepsilon_{F_L L}}{L}\right] > 0.$

Proof. See Appendix □

Capital increases labor because a rise in capital increases the marginal product of labor ($F_{LK} > 0$) which consequently raises wage and labor supply. Consumption decreases labor supply because additional consumption reduces the marginal utility of consumption ($u_{CC} < 0$) so the value of consumption declines, thus reducing labor to support consumption (leisure becomes more attractive). The effect of entry on labor supply is more complex.

Corollary 1. *Increasing marginal costs* $1-\nu > 0$ *and countercyclical markups* $\varepsilon_{\mu N} < 0$ *augment the labor response to entry. With constant marginal costs*

$\nu = 1$ *and fixed markups* $\varepsilon_{\mu N} = 0$, *entry does not affect labor* $\frac{dL}{dN} = 0$.

The first effect is from increasing marginal costs $(1 - \nu > 0)$, so that as entry divides inputs (labor and capital) across more units, the marginal product of labor increases $\text{MPL} = AN^{1-\nu}F_L(K, L)$, hence wage $w = \frac{\text{MPL}}{\mu}$ and consequently labor increase. The second positive effect occurs because entry decreases the markup between wage and a worker's marginal product.[32]

The total effect of technology on labor incorporates the endogenous adjustments of these variables.

$$\frac{dL}{dA} = \frac{\partial L}{\partial A} + \frac{\partial L}{dC}\frac{dC}{dA} + \frac{\partial L}{dK}\frac{dK}{dA} + \frac{\partial L}{dN}\frac{dN}{dA} \gtreqless 0, \quad \text{where } \frac{\partial L}{\partial A} = \Phi^{-1}\frac{1}{A} > 0 \tag{74}$$

Assuming technology increases consumption, capital and firms, then there is a negative income effect from consumption and a positive substitution effect from the direct, capital and entry effects. Therefore the overall effect is ambiguous. Labor increases if substitution effects dominate income effects. In the short run, when number of firms and capital are fixed, only the positive partial effect and negative consumption effect influence the impact labor response to technology.[33] On impact of a shock to A the response of output y will depend on a direct effect from A and an indirect effect through L but importantly the state variable capital and number of firms are fixed $K = \bar{K}$ and $N = \bar{N}$, thus $y(0) = A(0)\bar{N}^{-\nu}F(\bar{K}, L(A(0))) - \phi$. Labor does not necessarily increase, and nor does it need to in order to expand output. Output expands on impact providing the direct effect outweighs the possibility of a negative labor effect. If the substitution effect, of a higher MPL and in turn wage, dominates the income effect of a rise in consumption such that labor increases on impact, then this is sufficient for an output expansion.[34]

[32] The first effect (returns to scale) is present in Rotemberg 2008; Barseghyan and DiCecio 2011 and the second effect (endogenous markups) is studied in Cook 2001; Jaimovich 2007. The effects can be interpreted through wage behaviour, as in Jaimovich 2007, see appendix.

[33] The instantaneous response of labor to aggregate technology shocks is an unsettled debate (see Basu, Fernald, and Kimball 2006).

[34] With logarithmic consumption utility, the income and substitution effects equate in

3.2 Dynamics and Steady-State

With labor defined implicitly $L(C, K, N)$ by the intratemporal condition, and substituting in factor prices, profits and output, the dynamic equations from Table 1 are a system of four ODEs in consumption, entry, capital and number of firms (C, E, K, N).

3.2.1 Steady-State Solutions

The steady-state $\tilde{X} = (\tilde{C}, \tilde{E}, \tilde{K}, \tilde{N})$ satisfies the following conditions jointly leading to steady-state levels in terms of fixed parameters $\{A, \nu, \phi, \rho, \theta_F, \theta_I\}$[35]

$$\dot{C} = 0 \iff \tilde{r}(\tilde{C}, \tilde{K}, \tilde{N}) = \rho \tag{75}$$

$$\dot{E} = 0 \iff \tilde{\pi}(\tilde{C}, \tilde{K}, \tilde{N}) = 0 \tag{76}$$

$$\dot{K} = 0 \iff \tilde{Y}(\tilde{C}, \tilde{K}, \tilde{N}) = \tilde{C} \tag{77}$$

$$\dot{N} = 0 \iff \tilde{E} = 0 \tag{78}$$

Hence in steady state the interest rate equals the discount factor; profits are zero; and aggregate output equals consumption. The conditions are nonlinear, and steady-state may not be well-defined. Later, I provide conditions for existence. The steady-state conditions imply output per firm, and thus measured TFP are endogenously dependent on the number of firms in steady-state.

Proposition 2 (Endogenous Steady State Output and Productivity)**.** *Steady state output per firm $\tilde{y}$ and measured productivity $T\tilde{F}P$ are endogenously increasing in the number of firms.*

$$\tilde{y}(\mu(\tilde{N})) = \frac{\nu\phi}{\mu(\tilde{N}) - \nu} \tag{79}$$

$$T\tilde{F}P(\mu(\tilde{N})) = \nu \left[\frac{A}{\mu(\tilde{N})} \left(\frac{\mu(\tilde{N}) - \nu}{\phi} \right)^{1-\nu} \right]^{\frac{1}{\nu}} \tag{80}$$

the long run so that long-run labor supply is irresponsive to technology.

[35]The congestion parameter γ does not affect steady state because it pre-multiplies a differential.

With profit represented as in equation (50) since ϕ is fixed, production $F(k,l)$ must increase to cover this overhead and break-even at zero-profit as entry decreases the markup. The TFP result follows because there are increasing returns in steady-state with imperfect competition.[36] In the special case of perfect competition, when there are many firms $\mu_{N\to\infty}=1$, the steady-state levels are equivalent to a firm's minimum efficient scale (MES), implying TFP is maximized, and firm's are operating at the minimum of their average cost curve. This result would not occur if $\nu=1$. A steady-state outcome would not be well-defined. The steady-state output result also implies (by rearranging) that the long-run markup equates to returns to scale as implied in (29) under zero-profits.

3.2.2 Equilibrium Dynamics and Existence

The dynamical system in Table 1 is in nonlinear form $\dot{X}=g(X)$. We linearize it to $\dot{X}\approx \mathbf{J}(X)(X(t)-\tilde{X})$,[37] and analyze the Jacobian matrix $\mathbf{J}(\tilde{C},\tilde{E},\tilde{K},\tilde{N})$ where each element is a respective derivative evaluated at steady state.[38]

$$\begin{bmatrix}\dot{C}\\ \dot{E}\\ \dot{K}\\ \dot{N}\end{bmatrix}\approx\begin{bmatrix}-\frac{u_C}{u_{CC}}r_C & 0 & -\frac{u_C}{u_{CC}}r_K & -\frac{u_C}{u_{CC}}r_N\\ -\frac{\pi_C}{\gamma} & \rho & -\frac{\pi_K}{\gamma} & -\frac{\pi_N}{\gamma}\\ Y_C-1 & 0 & Y_K & Y_N\\ 0 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}C(t)-\tilde{C}\\ E(t)\\ K(t)-\tilde{K}\\ N(t)-\tilde{N}\end{bmatrix}\tag{81}$$

The characteristic polynomial corresponding to this four-dimensional sys-

[36] The rate at which TFP increases in N is

$$\frac{\partial T\tilde{F}P}{\partial N}=-\frac{A\nu}{\mu^2}\frac{(\mu-1)}{(\mu-\nu)^{\nu}\phi^{1-\nu}}\cdot\frac{\partial\mu(\tilde{N})}{\partial N}>0$$

[37] By definition variables are constant at steady state so that evaluating at steady state the differential form is $d\dot{x}=[\dot{x}(t)-\dot{\tilde{x}}]=\dot{x}(t)$.

[38] The derivatives treat $\{C,E,K,N\}$ as independent. Labor is a function of these variables through the intratemporal condition $L(C,K,N)$. The linearized solution has a recursive structure, denoting the solution paths $C^*(t),E^*(t),K^*(t),N^*(t)$, they will all be a function of $\{t,[K_0-\tilde{K}],[N_0-\tilde{N}]\}$ in open-loop form or $\{[K(t)-\tilde{K}],[N(t)-\tilde{N}]\}$ in closed-loop form. At time zero $C(0)$ and $E(0)$ will respond, whereas $K(0)$ and $N(0)$ remain fixed and subsequently move after $C(0),E(0)$ adjust.

tem is

$$c(\lambda) = \det(\mathbf{J} - \lambda\mathbf{I}) = \lambda^4 - M_1\lambda^3 + M_2\lambda^2 - M_3\lambda + M_4 \tag{82}$$

where coefficients M_k denote the sum of principal minors of dimension k, therefore $M_1 = \mathrm{Tr}(\mathbf{J})$ and $M_4 = \det(\mathbf{J})$.[39] The trace is unambiguously positive, whereas endogenous markups ($\varepsilon_{\mu N} < 0$) make the determinant ambiguous (with exogenous markups the determinant is strictly positive). In the Appendix I provide necessary conditions for a positive determinant and this also ensures steady-state existence.[40] Since $\mathrm{Tr}(\mathbf{J}) > 0$ and $\det(\mathbf{J}) > 0$, there are are either two or four positive eigenvalues. By Descartes' rule of signs restrictions on M_2 and M_3 can rule out the global instability case (four positive eigenvalues), so that we focus on the saddle-point stable case of two positive and two negative eigenvalues.

Therefore the system is a saddle with a two dimensional stable manifold defined on K, N. Hence capital and number of firms, so capital per firm, are fixed on impact whereas consumption and entry C, E jump instantaneously on to the stable arm. See Appendix for full derivation of dynamic solutions. The open-loop solution is in terms of initial values K_0, N_0, time t and the exogenous parameters of the model is

$$\mathbf{X}(t) = \tilde{\mathbf{X}} + a\mathbf{V_1}e^{\lambda_1 t} + b\mathbf{V_2}e^{\lambda_2 t}$$

where $\mathbf{X} = [C, E, K, N]^\intercal$ and $\mathbf{V_j} = [v_{1,j}, v_{2,j}, v_{3,j}, v_{4,j}]^\intercal$ $j \in \{1, 2\}$ are the normalized eigenvectors associated with stable roots $\lambda_1 < \lambda_2 < 0$. The constants are $a = \frac{\hat{K} - v_{3,2}\hat{N}}{v_{3,1} - v_{3,2}}$ and $b = \frac{v_{3,1}\hat{N} - \hat{K}}{v_{3,1} - v_{3,2}}$ where $\hat{K} = K_0 - \tilde{K}$ and $\hat{N} = N_0 - \tilde{N}$, so in long-hand, noting $\tilde{E} = 0$, we have

$$C(t) = \tilde{C} + \frac{(\hat{K} - v_{3,2}\hat{N})v_{1,1}e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})v_{1,2}e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{83}$$

[39] See Murata and Shell 2014, pp. 14-16.

[40] The Jacobian determinant of the dynamical system is equivalent to the determinant of the Jacobian of the steady-state conditions (75) - (78). Hence the conditions for it to be positive ensure steady-state is well-defined.

$$E(t) = \frac{(\hat{K} - v_{3,2}\hat{N})\lambda_1 e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})\lambda_2 e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{84}$$

$$K(t) = \tilde{K} + \frac{(\hat{K} - v_{3,2}\hat{N})v_{3,1} e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})v_{3,2} e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{85}$$

$$N(t) = \tilde{N} + \frac{(\hat{K} - v_{3,2}\hat{N})e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{86}$$

Our transition experiments specify the initial capital and number of firms equal to steady state under the old technology. We then study transition towards the new technology, thus a pefect foresight equilibrium. Formally, the model begins at steady-state under the old technology $(K_0, N_0) = (\tilde{K}(A^{\text{old}}), \tilde{N}(A^{\text{old}}))$ and transitions to $(\tilde{K}, \tilde{N}) = (\tilde{K}(A^{\text{new}}), \tilde{N}(A^{\text{new}}))$ under the new technology which comes online in period $t = 0$.

4 Measured Productivity Dynamics

From our definition of measured TFP (55) take the logarithmic derivative

$$\hat{\text{TFP}}(t) \approx \frac{1}{\nu}\hat{A}(t) + \left(\frac{\tilde{\text{RTS}} - 1}{\tilde{\text{RTS}}}\right)\hat{y}(t) \tag{87}$$

Hat notation represents deviation from steady-state. This shows that variations in measured TFP are composed of variations in pure technology, but also variations in output per firm (firm intensive margin). Since $\tilde{\text{RTS}} > 1$ with imperfect competition, measured TFP overestimates pure technology because it captures variations in returns to scale. There is no bias if there are (locally) constant returns to scale $\tilde{\text{RTS}} = 1$ which would occur with perfect competition $\tilde{\mu} = 1$ at minimum efficient scale. Additionally, if firm size is not varying $\hat{y} = 0$, as in status-quo models with fixed markups and instantaneous entry, then there is no bias: measured productivity accurately reflects underlying technology, even though there is imperfect competition in the form of monopolistic competition.

Our focus is to understand how the endogenous variations in firms' intensive margin and therefore measured productivity can be decomposed into profit and markup components. From (51) output per firm could be expressed as a function of profits and markups, which in turn implies measured TFP could be written as a function of technology A, markups μ and profits π in (54). To understand the effect of a change in technology on measured TFP, linearize the expression for measured TFP (54) around $\tilde{A}, \mu(\tilde{N}), \tilde{\pi}$.

$$\hat{\text{TFP}}(t) \approx \hat{A}\frac{1}{\nu} - \hat{\mu}(t)\left(\frac{\tilde{\mu}-1}{\tilde{\mu}-\nu}\right) + \hat{\pi}(t)\left(\frac{\tilde{\mu}-1}{\nu\phi}\right) \tag{88}$$

A is always positively related to TFP but, in a neighborhood of steady state, μ is negatively related and π positively related, providing there is not perfect competition $\tilde{\mu} = 1$ and therefore constant returns as discussed above.[41] Profits increase measured TFP whereas markups decrease measured TFP, but over the cycle profits are procyclical and markups are countercyclical hence both cause upward bias. Therefore a deviation in technology $\hat{A}$ is *not* accurately measured by a deviation in $\hat{\text{TFP}}$ which is what we typically measure from the data using a Solow Residual type approach, either in the crudest Solow Residual sense from acquiring the residuals of a logged regression of (53) (in our model the relevant SR would account for ν), or using the 'modified Solow Residual' of Basu and Fernald 1997. The result shows that the measured TFP series we acquire is composed of variations in technology, but also upward biased by variations in markups and profits. Consequently a pure technology series would purge a measured TFP series of these two extra endogenous components. Our main interest is the dynamic implications of firm entry, that is how important are these two biases at various stages following a technology shock.

Proposition 3. *Following a technology shock, all short-run endogenous variation in measured* TFP *arises from economic profits. The markup effect is zero.*

Consider that the economy begins at steady state and there is a positive shock to technology $\tilde{A} < A(0)$ at time 0, hence $[A(0) - \tilde{A}]\frac{1}{\tilde{A}\nu}$ immediately

[41] Away from steady state, the effects are ambiguous depending on the size of increasing marginal costs, as shown in the supplementary appendix.

increases TFP(0). Under instantaneous entry the markup would immediately decrease $\mu(0) < \tilde{\mu}$ in response to the immediate entry of new firms to bring about zero profits. The double-negative means this will have an additional positive effect $-[\mu(0)-\tilde{\mu}]\left(\frac{\tilde{\mu}-1}{\tilde{\mu}(\tilde{\mu}-\nu)}\right) > 0$ on TFP(0). The conclusion being that with static entry (instantaneous zero profits) measured TFP is an upward biased estimator or technology.[42] Under dynamic entry at $t = 0$ the markup does not move because the number of firms is a state variable which takes an instance to adjust $[\mu(0) - \tilde{\mu}] = 0$.[43] Therefore $\mu(N_0) - \mu(\tilde{N}) = 0$ so there is no instantaneous markup effect on productivity. However, profits will have an effect. At $t = 0$ profits increase $\pi(0) > \tilde{\pi}$ so in addition to the technology effect we observe $[\pi(0) - \tilde{\pi}]\left(\frac{\tilde{\mu}-1}{\nu\phi}\right) > 0$. Hence our theory, which shows that markups do not move instantaneously, allows us to disentangle the profit effect from the markup effect due to timing.

Proposition 4. *Following a technology shock, all long-run endogenous variation in measured* TFP *arises from endogenous markups. The profit effect is zero.*

In the instantaneous entry case, the instantaneous technology and markup effect persist into the long run. Hence there is no distinction. In the dynamic entry case, as $t \to \infty$ markups decrease to their long-run level as number of firms increases so there is a permanent effect equivalent to the instantaneous effect in the static entry case $-[\mu(\infty) - \tilde{\mu}]\left(\frac{\tilde{\mu}-1}{\tilde{\mu}(\tilde{\mu}-\nu)}\right) > 0$ on TFP(∞). The profit effect, observed on impact in the static entry case, disappears as profits return to their initial position so $\pi(\infty) - \tilde{\pi} = 0$.

Transition: At $t = 0$ the only effect is from profits. At $t \to \infty$ the only effect is from markups. In transition there are two positive effects on TFP, but the importance of profits decreases and of markups increases. Each period $\mu(t)$ moves further away from its initial position $\tilde{\mu}$ so its importance grows. Whereas $\pi(t)$ moves closer to its original position $\tilde{\pi} = 0$, so its im-

[42] With $\nu = 1$ and the deviation expressed as a growth rate, the coefficient is $-\left(\frac{\tilde{\mu}-1}{\tilde{\mu}-\nu}\right)|^{\nu=1} = -1$, which is the result of Jaimovich and Floetotto 2008, eq. 24. Therefore the coefficient here is smaller given $\tilde{\mu}$.

[43] As it is a state variable it remains at the given initial condition N_0 which we begin at steady state $N_0 = \tilde{N}$

portance shrinks. Therefore both profits and markups are intertemporally having a measured productivity increasing effect $\hat{\text{TFP}}(t)$, on top of the exogenous shock, but the positive effects change in relative importance over time. It is firm entry that drives this change. Firm entry increases the importance of the positive markup effect, and decreases the importance of the positive profit effect.[44] The markup effect continuously grows in importance,

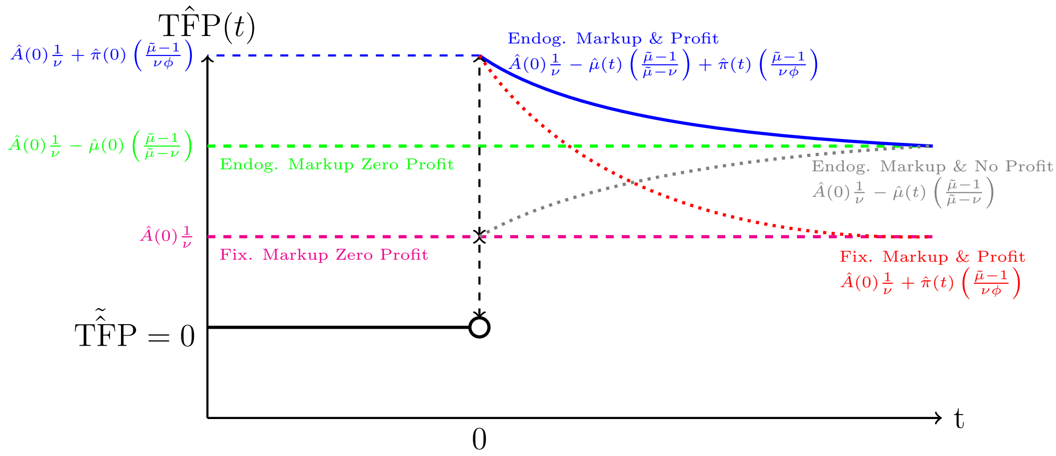

Figure 2: Endogenous Productivity Decomposition, *Positive Shock* $\tilde{A} \to A(0)$

whereas the profit effect continuously shrinks. Figure 2 demonstrates the composition of these effects: the top blue line shows the measured productivity that we observe, whereas the dashed horizontal pink line represents the pure technology effects. The lines in between represent the composition. Specifically the increasing gray dotted line shows the growing importance of markups in influencing measured TFP, whereas the decreasing dotted red line shows the diminishing importance of profit bias. Separately, the middle green horizontal line shows the case that would arise if entry were instantaneous but markups endogenous hence there is no profit effect and all bias arises from markups. This is discussed from a quantitative perspective with Cobb-Douglas production and flat marginal costs in Jaimovich and Floetotto 2008.

[44] Therefore we can identify instantaneous endogenous productivity effects as entirely due to profits and long-run endogenous productivity effects as entirely due to markups. During transition it is a combination of both.

Figure 2 also emphasizes that at some time t' the upward bias from profit and markup effect equate. Before t' the profit effect is the dominant component of measured TFP and after t' the markup effect dominates. Therefore at some time t' the two effects will intersect

$$-\hat{\mu}(t')\left(\frac{\tilde{\mu}-1}{\tilde{\mu}-\nu}\right)=\hat{\pi}(t')\left(\frac{\tilde{\mu}-1}{\nu\phi}\right) \tag{89}$$

$$\frac{\hat{\pi}(t')}{\hat{\mu}(t')}=-\frac{\nu\phi}{\tilde{\mu}-\nu}=-\tilde{y} \tag{90}$$

This is the ratio of coefficients in (88). It shows that the relative importance of each effect depends on the fixed cost and the steady-state markup.[45] In the parametric numerical exercise we estimate $t'=5$ quarters, implying the length of time for which the profit bias dominates the markup bias.

4.1 Parametric Example

The baseline RBC model assumes isoelastic separable subutilties and a Cobb-Douglas production function.

$$U(C,L)=\frac{C^{1-\sigma}-1}{1-\sigma}-\xi\frac{L^{1+\eta}}{1+\eta} \tag{91}$$

$$y=Ak^{\alpha}\ell^{\beta}-\phi \tag{92}$$

Isoelastic utility implies there is constant elasticity of marginal utility with respect to consumption and labor $U_{CC}\frac{C}{U_C}=-\sigma$ and $U_{LL}\frac{L}{U_L}=\eta$.[46] Frisch elasticity of labor supply is given by $\frac{1}{\eta}$, hence we calibrate inverse elasticity $\eta\in(0,\infty)$.[47] Cobb-Douglas production conforms to our assumptions on the production function derivatives. The production function is homogeneous of degree $\nu\equiv\alpha+\beta$, where α and β are capital and labor shares. Table 2 summarizes the calibration used for simulation. The time interval is a quarter

[45] The steady-state markup will also change indirectly with the fixed cost.

[46] The limiting case of $\sigma\rightarrow 1$ implies log utility $\ln(C)$, implying fixed aggregate labor in the long run following a technology shock as income and substitution effects cancel out.

[47] $\eta=0$ implies indivisible labor.

Table 2: Parameter Values

Capital Share	α	0.283
Labor Share	β	0.567
Fixed Cost	ϕ	2.5
Entry Congestion	γ	2.0
Technology	A	1.0
Risk Aversion†	σ	1.0
Discount Rate	ρ	0.02
Labor Weight	ξ	0.01
Labor Elast. (Frisch)	η	0.5
Industry (inter) Subs.	θ_I	1.0
Firm (intra) Subs.	θ_F	8.0

† Unit risk aversion $\sigma = 1$ implies $\ln C$

and correspondingly $\rho = 0.02$ implies an annualized discount rate of 8%.[48] We calibrate the fixed cost ϕ to be a percentage of sales in steady-state $\frac{\phi}{\tilde{y}} \approx 0.10$. We choose the entry congestion parameter γ such that firm convergence is similar to Bilbiie, Ghironi, and Melitz 2012. That is, most convergence has taken place within about 20 quarters (5 years), the approximate length of a cycle.[49] The intra- and inter- industry substitutability parameters are chosen in line with literature (see table 3). Together with the steady-state number of firms these substitutabilities imply that the steady state markup before shock is $\tilde{\mu} = \frac{\theta_F \tilde{N}}{\theta_F \tilde{N} - 1} = 1.42$, which declines following the shock in the endogenous markup case and is fixed at this level for the exogenous markup benchmark. The capital and labor income shares follow Restuccia and Rogerson 2008 who cite that estimates of $\alpha + \beta = 0.85$ and then division between the two is determined by a 1/3 capital to 2/3 labor share.[50] These parameter values imply $\frac{\tilde{K}}{\tilde{Y}} \approx 2.5$ as suggested in US data (Rotemberg and Woodford 1993). The shock shifts technology from $A_0 = 1$ to $A_1 = 1.1$.

[48]Corresponding to a discount factor in the discrete time case of $1.08^{-\frac{1}{4}} = 0.98$ as in Restuccia and Rogerson 2008.

[49]A more sophisticated approach could solve the industry dynamics second-order differential equation in partial equilibrium (r given by $\tilde{r} = \rho$) then calibrate based on half-life if firm convergence.

[50]Also see Da-Rocha, Tavares, and Restuccia 2017 for the same approach to decreasing returns.

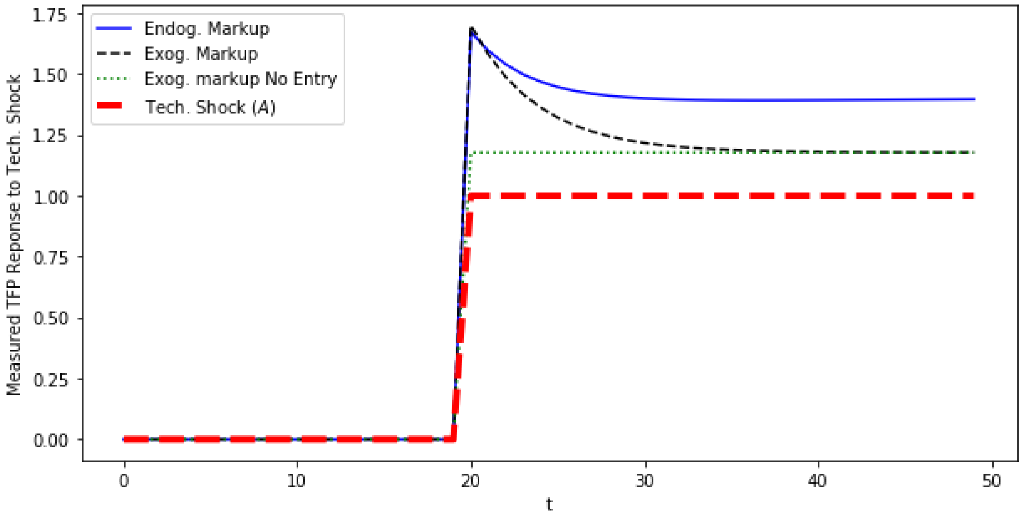

Figure 3: Endogenous Vs Exogenous Responses

Figure 3 shows the permanent shock to technology appearing unexpectedly in period 20. The measured TFP improvement with dynamic entry and endogenous markups (blue thick) exceeds underlying technology improvement by 40% on impact as profits rise and remains 20% higher in the long-run once profits have been arbitraged but competition has decreased markups. This follows from comparing the blue thick with green dotted lines.[51] The positive profit effect dominates the positive markup effect for 5 quarters, but after 10 quarters the profit effect disappears leaving only the long-run markup effect. These speeds will vary depending on the business dynamism of a given economy. That is how fast firm entry is able to adjust to a shock, which is determined by endogenously procylical sunk costs. When $\gamma \to 0$ an economy exhibits strong business dynamism and will more accurately reflect underlying technology as firms acquire positive profits for a shorter period of time.

[51] The green dotted line that captures measured TFP with neither endogenous markups or dynamic entry does not exactly reflect the technology change (dotted red) due to the ν component. The two lines would be equivalent with $\nu = 1$.

5 Summary

The paper investigates the effect of firm entry on measured productivity over the business cycle. I consider that entry is non-instantaneous leading to temporary profits and entry affects the price markups that incumbents charge. Together these mechanisms can explain short-run procylical productivity and long-run persistence following a technology shock.

The theory explains that productivity is exacerbated on impact, since firms cannot adjust immediately so incumbents gain profits and expand output, and in the long run underlying productivity is not regained because subsequent adjustment of firms causes structural changes in competition. Furthermore I show that in highly competitive (low markup) industries the distinction between short-run and long-run productivity is small, so measured productivity quickly and accurately reflects underlying technology. And industries with fast adjustment of firms (strong business dynamism), due to low endogenous sunk costs, will observe measured productivity closer to underlying technology. If business dynamism has declined, such that economic profits are protected for longer, this will have increased the importance of the profit bias in measured TFP and delayed the importance of markup bias.

A Endogenous Markup Details

A.1 Bertrand Derivation of Markup

For the Bertrand case we recast each first order condition to obtain conditional demand as a function of prices. Therefore (9), (12), (13) become (93), (94), (95).

$$Q_\jmath = \left(\frac{P_\jmath}{P}\right)^{-\theta_I} Y \tag{93}$$

$$y_{\jmath\imath} = \left(\frac{P_{\jmath\imath}}{P_\jmath}\right)^{-\theta_F} \left(\frac{1}{N}\right)^{1-\vartheta(\theta_F-1)} Q_\jmath \tag{94}$$

$$y_{\jmath\imath} = \left(\frac{P_{\jmath\imath}}{P_\jmath}\right)^{-\theta_F} \left(\frac{P_\jmath}{P}\right)^{-\theta_I} \left(\frac{1}{N}\right)^{1-\vartheta(\theta_F-1)} Y \tag{95}$$

The corresponding price indices follow from substituting the two FOCs (93), (94) into their corresponding constraints (production functions) (8), (11), giving aggregate price index and industry price index

$$P = \left(\int_0^1 P_\jmath^{1-\theta_I} d\jmath\right)^{\frac{1}{1-\theta_I}} \tag{96}$$

$$P_\jmath = N^{\frac{1}{\theta_F-1}-\vartheta} \left(\sum_1^N P_{\jmath\imath}^{1-\theta_F}\right)^{\frac{1}{1-\theta_F}} \tag{97}$$

Under Bertrand firms maximize profits subject to conditional demand (95) by choosing $P_{\jmath\imath}$. Therefore

$$\frac{\partial y_{\jmath\imath}}{\partial P_{\jmath\imath}} = -\theta_F \frac{y_{\jmath\imath}}{p_{\jmath\imath}} + (\theta_F - \theta_I) \frac{y_{\jmath\imath}}{P_\jmath} \frac{\partial P_\jmath}{\partial P_{\jmath\imath}} \tag{98}$$

From the industry-level price index $P_\jmath = N^{\frac{1}{\theta_F-1}-\vartheta} \left(\sum_1^N P_{\jmath\imath}^{1-\theta_F}\right)^{\frac{1}{1-\theta_F}}$ then

$$\frac{\partial P_\jmath}{\partial P_{\imath\jmath}} = \frac{P_\jmath}{\sum_1^N P_{\jmath\imath}^{1-\theta_F}} P_{\jmath\imath}^{-\theta_F} \tag{99}$$

Thus under symmetry $\frac{\partial P_\jmath}{\partial P_{\imath\jmath}} = \frac{1}{N}\frac{P_\jmath}{P_{\jmath\imath}}$. Hence defining price elasticity of demand

$$\epsilon \equiv -\frac{\partial y_{\jmath\imath}}{\partial P_{\jmath\imath}}\frac{P_{\jmath\imath}}{y_{\jmath\imath}} \tag{100}$$

Then (95) becomes

$$\epsilon = \theta_F - (\theta_F - \theta_I)\frac{1}{N} \tag{101}$$

Therefore for Bertrand the markup is

$$\mu(N)^{\mathrm{B}} = \frac{\theta_I + \theta_F(N-1)}{(\theta_F - 1)N - (\theta_F - \theta_I)} = \frac{1 - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\theta_I}{1 - \frac{1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\theta_I}, \qquad \theta_F > \theta_I \geq 1 \tag{102}$$

$$\varepsilon^{\mathrm{B}}_{\mu N} = \frac{N\theta_F}{\theta_I + \theta_F(N-1)}\left(1 - \frac{\theta_F - 1}{\theta_F}\mu^{\mathrm{B}}\right) < 0 \tag{103}$$

A.2 Cournot Markup as a Function of Bertrand

The Cournot and Bertrand markups are related as follows

$$\mu^{\mathrm{C}} = \frac{\mu^{\mathrm{B}}}{\theta_I\left(\frac{1}{\theta_I} - \left[1 + \mu^{\mathrm{B}}\left(\frac{1}{\theta_I} - 1\right)\right]\frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\right)} \tag{104}$$

This follows from rewriting the Cournot markup as Bertrand

$$\mu^{\mathrm{C}} = \left[(\mu^{\mathrm{B}})^{-1}(1 - \Upsilon\theta_I) + \Upsilon\theta_I - \Upsilon\right]^{-1}, \quad \text{where } \Upsilon = \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right) \tag{105}$$

and rearranging.

A.3 Markup Calibration Survey

Table 3 surveys the intersectoral θ_I and intrasectoral θ_F parameters used in the literature.

A.4 Graphical Illustration of Markup Properties

Table 3: Endogenous Markup Literature Survey

	θ_I	θ_F	μ	$\varepsilon_{\mu N}$
Cournot				$\varepsilon^{\mathrm{C}} = 1 - \frac{\theta_F - 1}{\theta_F} \mu^{\mathrm{C}}$
Atkeson et al., p.2015	1.01	10.0	$\frac{1}{\left(\frac{\theta_F-1}{\theta_F}\right) - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)}$	$1 - \frac{\frac{\theta_F-1}{\theta_F}}{\left(\frac{\theta_F-1}{\theta_F}\right) - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)}$
Jaimovich et al., App p.15[1]	1.001	∞	$\frac{\theta_I N}{\theta_I N - 1}$	$-\frac{1}{\theta_I N - 1}$
Etro et al., p.1215	1.0	$6.0, 20.0, \infty$	$\frac{\theta_F N}{(\theta_F - 1)(N-1)}$	$-\frac{1}{N-1}$ †
Colciago et al., p.1104[2]	1.0	∞	$\frac{N}{N-1}$	$-\frac{1}{N-1}$
Bertrand				$\varepsilon^{\mathrm{B}} = \frac{N\theta_F}{\theta_I + \theta_F(N-1)} \left(1 - \frac{\theta_F - 1}{\theta_F} \mu^{\mathrm{B}}\right)$
Jaimovich et al., p.1248[3]	1.001	19.6	$\frac{\theta_I + \theta_F(N-1)}{(\theta_F - 1)N - (\theta_F - \theta_I)}$	
Etro et al., p.1216	1.0	$6.0, 20.0$	$\frac{1 + \theta_F(N-1)}{(\theta_F - 1)(N-1)}$	
Lewis et al., p.678[4]	1.00‡	2.624	$\frac{\theta_I + \theta_F(N-1)}{(\theta_F - 1)N - (\theta_F - \theta_I)}$	

[1] Portier 1995; Brito, Costa, and Dixon 2013 also use Cournot with homogeneous goods but differentiated industries.
[2] Homogeneous goods and unitary elasticity is common in theoretical papers *e.g.* Dos Santos Ferreira and Dufourt 2006; Opp, Parlour, and Walden 2014.
[3] Jaimovich and Floetotto 2008, p.1248 specify their nested CES aggregators as a p-norm τ and Holder conjugate ω, rather than elasticities, hence $\theta_I = \frac{1}{1-\omega} = \frac{1}{1-0.001} = 1.001$ and $\theta_F = \frac{1}{1-\tau} = \frac{1}{1-0.949} = 19.6$.
[4] Lewis and Poilly 2012, p.680 plot their competition effect for a domain of substitutabilities $\theta_I, \theta_F \in (1.0, 4.0)^2$.
† θ_F does not affect the markups elasticity to number of firms when there is unitary elasticity across industries.
‡ Estimated value to 2 decimal places.

B Returns to Scale Derivations

There are two ways to identify returns to scale. First, from the profit definition, without imposing restrictions on the cost function.

$$P\pi = Py - C(r, w, y) = (P - \mathrm{AC})y \tag{106}$$

Divide by Py and multiply AC by MC / MC to express as a markup

$$s_\pi = 1 - \frac{AC}{P} = 1 - \frac{AC}{MC}\frac{MC}{P} \tag{107}$$

$$\mathrm{RTS} = \mu\left(1 - \frac{\pi}{y}\right) \tag{108}$$

Second from the cost function that occurs under cost-minimizing factor prices (input demands).[52]

$$C = \mathrm{MC}\,\nu(y + \phi) \tag{109}$$

[52] We could divide by prices and begin with the familiar real cost function $\frac{C}{P} = \frac{\nu}{\mu}(y + \phi)$.

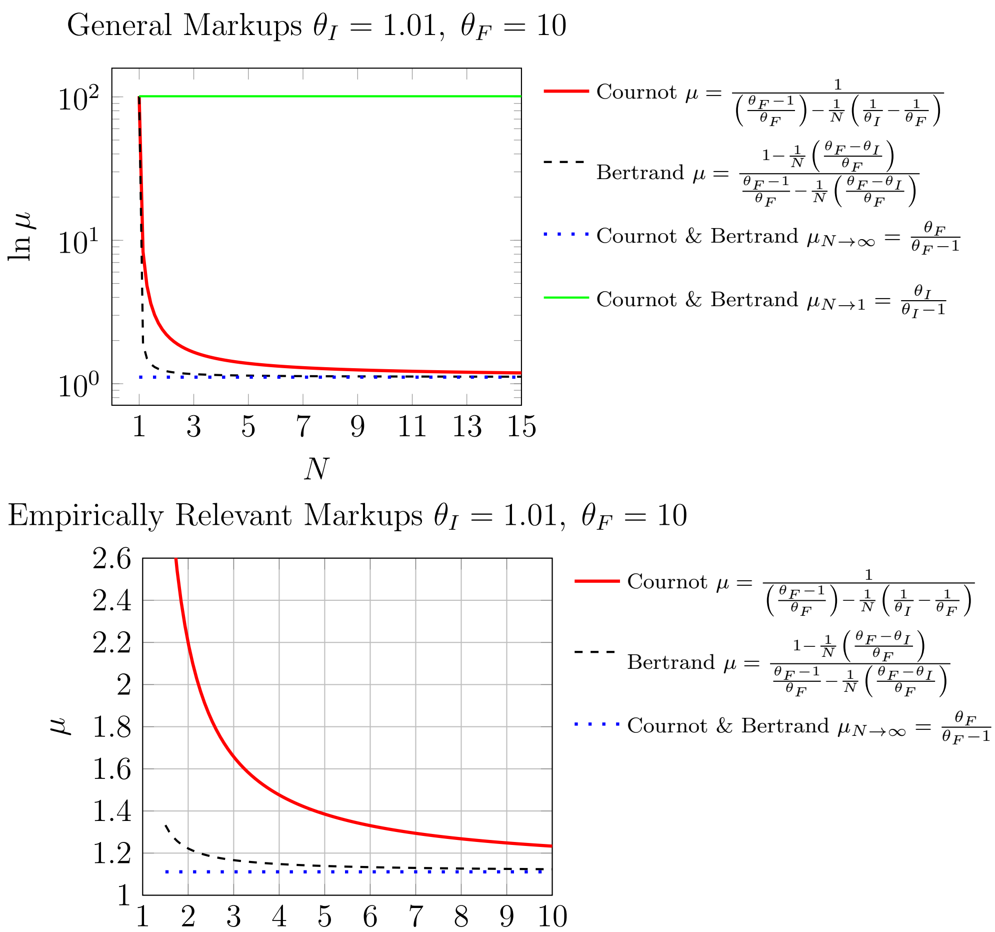

Figure 4: Markup Comparison

Divide by $y\,\mathrm{MC}$

$$\mathrm{RTS} = \nu(1 + s_\phi) \tag{110}$$

We could also obtain this by dividing the explicit expressions for AC and MC, (19) and (17), that we obtain from the cost function rewritten in unit cost form.

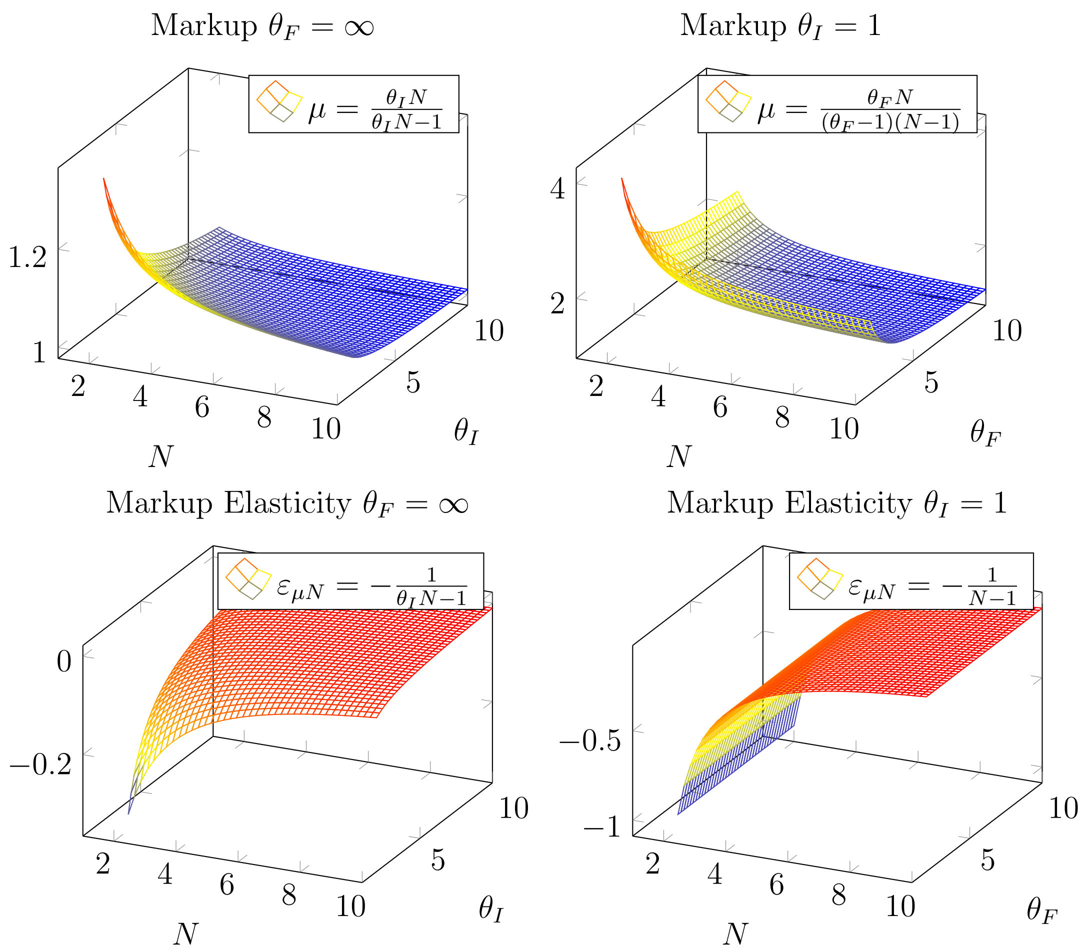

Figure 5: Markup Properties

C Labor Responses

Proof of Proposition 1. The partial derivatives of utility have the following properties: $u_{CC}, u_{LL}, u_L < 0$, $u_C > 0$ and $u_{CL} = u_{LC} = 0$. From the intratemporal condition and wage equation

$$\Xi(L, C, K, N, A) \equiv u_L(L) + u_C(C)w(L, K, N) = 0 \tag{4}$$

$$w(L, K, N) = \frac{A}{\mu(N)} N^{1-\nu} F_L(K, L) \tag{27}$$

take the total derivative, treating $\{C, K, N\}$ independently, with respect to dummy ϖ

$$0 = \frac{\partial \Xi}{\partial L}\frac{dL}{d\varpi} + \frac{\partial \Xi}{\partial C}\frac{dC}{d\varpi} + \frac{\partial \Xi}{\partial K}\frac{dK}{d\varpi} + \frac{\partial \Xi}{\partial N}\frac{dN}{d\varpi} + \frac{\partial \Xi}{\partial \varpi} \tag{111}$$

$$\frac{dL}{d\varpi} = -\left(\frac{\partial \Xi}{\partial L}\right)^{-1}\left[\frac{\partial \Xi}{\partial C}\frac{dC}{d\varpi} + \frac{\partial \Xi}{\partial K}\frac{dK}{d\varpi} + \frac{\partial \Xi}{\partial N}\frac{dN}{d\varpi} + \frac{\partial \Xi}{\partial \varpi}\right] \tag{112}$$

$$\frac{dL}{d\varpi} = \frac{\partial L}{\partial C}\frac{dC}{d\varpi} + \frac{\partial L}{\partial K}\frac{dK}{d\varpi} + \frac{\partial L}{\partial N}\frac{dN}{d\varpi} + \frac{\partial L}{\partial \varpi} \tag{113}$$

Replacing the dummy ϖ with each of the endogenous variables $\{C, K, N\}$, we get that the total and partial derivatives are equivalent and follow the usual multivariate implicit function rule

$$\frac{\partial L}{\partial C} = -\frac{\frac{\partial \Xi}{\partial C}}{\frac{\partial \Xi}{\partial L}} = \Phi^{-1}\frac{u_{CC}}{u_C} \qquad = \frac{dL}{dC} \tag{114}$$

$$\frac{\partial L}{\partial K} = -\frac{\frac{\partial \Xi}{\partial K}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial K} \qquad = \frac{dL}{dK} \tag{115}$$

$$\frac{\partial L}{\partial N} = -\frac{\frac{\partial \Xi}{\partial N}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial N} \qquad = \frac{dL}{dN} \tag{116}$$

where $\Phi \equiv \left[\frac{u_{LL}}{u_L} - \frac{F_{LL}}{F_L}\right] = \left[\frac{\varepsilon_{u_L L} - \varepsilon_{F_L L}}{L}\right]$ and to finish we subsitute in

$$\frac{\partial \Xi}{\partial L} = u_L \Phi \qquad < 0 \tag{117}$$

$$\frac{\partial w}{\partial C} = 0 \qquad = 0 \tag{118}$$

$$\frac{\partial w}{\partial L} = w\frac{F_{LL}}{F_L} \qquad < 0 \tag{119}$$

$$\frac{\partial w}{\partial K} = w\frac{F_{LK}}{F_L} \qquad > 0 \tag{120}$$

$$\frac{\partial w}{\partial N} = \frac{w}{N}\left(1 - \nu - \varepsilon_{\mu N}\right) \qquad > 0 \tag{121}$$

□

Notice that we treat the endogenous variables $\{C, K, N\}$ independently so that total and partial derivatives equate. However a change in an exogenous parameter, such as technology $\varpi = A$, causes all endogenous variables to respond in addition to its partial effect from holding C, K, N constant.

$$\frac{dL}{dA} = \frac{\partial L}{\partial C}\frac{dC}{dA} + \frac{\partial L}{\partial K}\frac{dK}{dA} + \frac{\partial L}{\partial N}\frac{dN}{dA} + \frac{\partial L}{\partial A} = \frac{dL}{dC}\frac{dC}{dA} + \frac{dL}{dK}\frac{dK}{dA} + \frac{dL}{dN}\frac{dN}{dA} + \frac{\partial L}{\partial A} \tag{122}$$

where

$$\frac{\partial L}{\partial A} = -\frac{\frac{\partial \Xi}{\partial A}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial A} \qquad \neq \frac{dL}{dA} \tag{123}$$

$$\frac{\partial w}{\partial A} = \frac{w}{A} \qquad > 0 \tag{124}$$

D Measured TFP Derivation

Our aim is to explain aggregate output as a function of inputs K, L

$$y = N^{-\nu} A F(K, L) - \phi \tag{125}$$

$$\text{Use,} \quad y + \phi = (\pi + \phi)\left(1 - \frac{\nu}{\mu}\right)^{-1} \tag{126}$$

$$(\pi + \phi)\left(1 - \frac{\nu}{\mu}\right)^{-1} = N^{-\nu} A F(K, L) \tag{127}$$

$$\text{Use,} \quad N = \frac{Y}{y} \tag{128}$$

$$Y^{\nu} = y^{\nu} A F(K, L)(\pi + \phi)^{-1}\left(1 - \frac{\nu}{\mu}\right) \tag{129}$$

$$\text{Use,} \quad y = \left(\pi + \frac{\nu}{\mu}\phi\right)\left(1 - \frac{\nu}{\mu}\right)^{-1} \tag{130}$$

$$Y = \left(\frac{A}{\pi + \phi}\right)^{\frac{1}{\nu}}\left(1 - \frac{\nu}{\mu}\right)^{\frac{1}{\nu} - 1}\left(\pi + \frac{\nu}{\mu}\phi\right) F(K, L)^{\frac{1}{\nu}} \tag{131}$$

E Characteristic Polynomal

The characteristic polynomial associated with the Jacobian matrix of the 4d dynamical system can be determined through its principal minors. Principal minors are those that correspond to the leading diagonal of the Jacobian matrix. Where M_k denotes the sum of principal minors of dimension k, the characteristic polynomial can be expressed as follows[53]

$$c(\lambda) = \det(\mathbf{J} - \lambda \mathbf{I}) = \lambda^4 - M_1\lambda^3 + M_2\lambda^2 - M_3\lambda + M_4 \tag{132}$$

where $M_1 = \mathrm{Tr}(\mathbf{J}) > 0$ and $M_4 = \det(\mathbf{J}) > 0$, where we have shown these are both positive. The sequence of signs of coefficients of the polynomial is

$$\{+, -, \pm, \pm, +\}$$

Descartes' rule of signs (for positive roots) states: *If the terms of a single-variable polynomial with real coefficients are ordered by descending variable exponent, then the number of positive roots of the polynomial is either equal to the number of sign differences between consecutive nonzero coefficients, or is less than it by an even number. Multiple roots of the same value are counted separately.*

Leaving two coefficients unspecified, there is a minimum of two sign changes and a maximum of four sign changes. Hence there are either four, two or zero positive roots.[54] The zero positive roots case is ruled out by a positive trace, the four positive root case (implying asymptotic stability) is consistent with positive trace and positive determinant but can be ruled out if the two unspecified signs are of the same sign, either $+, +$ or $-, -$.[55]

The four solutions of this quartic polynomial are saddle stable if there are two positive (unstable) and two negative (stable).[56] We denote these

[53] The proofs in this section follow results developed in Murata and Shell 2014, pp. 14-16; Flaschel et al. 2008, pp. 503-510 (Theorem A.2, p.504); Jacobson 2012, p. 196.

[54] If unspecified coefficients are zero, the possibility of one sign change occurs if $\pm, \pm$ are $0, 0$ or $0, +$ or $-, 0$. However, from the positive trace and determinant, we know there are either four positive roots, or two positive and two negative roots.

[55] Jaimovich 2007 studies an endogenous markups, static entry model that leads to sunspots (global stability) in a one state variable model (only capital) so has two positive roots.

[56] Chatterjee, Sakoulis, and Turnovsky 2003, pp.1087-88 and Bhandari, Haque, and Turnovsky 1990, pp. 413-415 employ a similar strategy to derive sufficient conditions for saddle stability with a quartic polynomial.

eigenvalues

$$\lambda_1 \leq \lambda_2 < 0 < \lambda_3 \leq \lambda_4$$

We derive the general minors and thus coefficents on the quartic polynomial for the general case.

E.1 Jacobian Trace

The trace of the Jacobian matrix is

$$\begin{aligned}
\mathrm{Tr}(\mathbf{J}) &= -\Phi^{-1}\frac{F_{KL}}{F_K}\rho + \rho + \left(1 + \Phi^{-1}\frac{F_{KL}}{F_K}\right)\rho\tilde{\mu} && (133)\\
&= \rho\left(1 + \tilde{\mu} + (\tilde{\mu} - 1)\Phi^{-1}\right) && (134)\\
&= \rho(1+\tilde{\mu})\left(1 + \left(\frac{\tilde{\mu}-1}{\tilde{\mu}+1}\right)\Phi^{-1}\right) > 0 && (135)
\end{aligned}$$

with perfect competition $\tilde{\mu} = 1$ then $\mathrm{Tr}(\mathbf{J}) = 2\rho$.

E.2 Jacobian Determinant

A lengthy derivation shows the determinant of the Jacobian associated with the dynamical system is positive when the endogenous markup effect is sufficiently small, where I provide the sufficient condition.

$$\det \mathbf{J} = \frac{\tilde{N}}{\varepsilon_{u_C C}}\frac{\phi}{\gamma}\frac{\nu}{\mu}\left\{\Theta + \Gamma\varepsilon_{\mu_N N}\right\} \tag{136}$$

where

$$\Theta = \left(\frac{L}{\varepsilon_{u_L L} - \varepsilon_{F_L L}}\right)\left(\frac{\rho}{N}\right)^2 \mu(\varepsilon_{u_C C} - \varepsilon_{u_L L})\left(\frac{F_{KL}}{F_L} - \frac{F_{KK}}{F_K}\right)\frac{F_L}{F_K} < 0 \tag{137}$$

$$\begin{aligned}
\Gamma = {} & y\left(1 + \frac{y}{\phi}\right)\Phi^{-2}\left(\frac{\rho}{N}\right)^2 \mu \\
& \left(\left[\Phi + \frac{F_{LK}^2}{F_L F_{KK}}\right]\frac{F_{KK}F_L}{F_K^2}\left\{\frac{\varepsilon_{u_C C}}{y} - \frac{\Phi N}{\rho\mu}\frac{F_K}{F_L} + 1\right\} - \left(\Phi + \frac{F_{LK}}{F_K}\right)^2\right) \gtreqless 0
\end{aligned} \tag{138}$$

where elasticities are defined as $\varepsilon_{u_L,L} \equiv \frac{u_{LL}}{u_L}L$. Then, remembering $\varepsilon_{u_C C} < 0$

$$\det \mathbf{J} \gtreqless 0 \iff \Theta + \Gamma\varepsilon_{\mu_N N} \lesseqgtr 0 \tag{139}$$

Hence this provides a necessary condition for a positive determinant

$$\Theta + \Gamma\varepsilon_{\mu_N N} < 0 \tag{140}$$

Since $\Theta < 0, \varepsilon_{\mu_N N} < 0$, a sufficient condition is

$$\Gamma \geq 0 \tag{141}$$

Or, if $\Gamma < 0$, less restrictive

$$\Gamma\varepsilon_{\mu_N N} < -\Theta \tag{142}$$

$$\varepsilon_{\mu_N N} > -\frac{\Theta}{\Gamma} \tag{143}$$

Hence the negative markup effect must be sufficiently small.

F Measured Productivity Dynamics

Use growth notation notation $\hat{X} = \frac{dX}{\tilde{X}}$.

From (55)

$$\mathrm{TFP} = A^{\frac{1}{\nu}} \frac{y}{(y+\phi)^{\frac{1}{\nu}}} \tag{144}$$

$$\ln \mathrm{TFP} = \frac{1}{\nu}\ln A + \ln y + \frac{1}{\nu}\ln(y+\phi) \tag{145}$$

$$\frac{1}{\mathrm{TFP}}\frac{d\,\mathrm{TFP}}{dA} = \frac{1}{\nu}\frac{1}{A} + \frac{1}{y}\frac{dy}{dA} - \frac{1}{\nu}\frac{1}{y+\phi}\frac{dy}{dA} \tag{146}$$

$$\text{Local approx.} \quad \times\, dA = [A(t) - \tilde{A}] \tag{147}$$

$$\frac{d\,\mathrm{TFP}}{\mathrm{TFP}} = \frac{1}{\nu}\frac{dA}{\tilde{A}} + \left(1 - \frac{1}{\nu(1+\tilde{s}_\phi)}\right)\frac{dy}{\tilde{y}} \tag{148}$$

$$\hat{\mathrm{TFP}} = \frac{1}{\nu}\hat{A} + \left(\frac{\tilde{\mathrm{RTS}} - 1}{\tilde{\mathrm{RTS}}}\right)\hat{y} \tag{149}$$

From (54)

$$\ln \text{TFP} = \frac{1}{\nu}\left[\ln A - \ln(\pi+\phi)\right] + \left(\frac{1}{\nu}-1\right)\ln\left(1-\frac{\nu}{\mu}\right) + \ln\left(\frac{\nu}{\mu}\phi+\pi\right) \tag{150}$$

$$\begin{aligned}\frac{d\ln \text{TFP}}{dA} &= \frac{1}{\nu}\left[\frac{1}{A} - \frac{1}{\pi+\phi}\frac{d\pi}{dA}\right] \\ &\quad + \left(\frac{1}{\nu}-1\right)\frac{1}{1-\frac{\nu}{\mu}}\frac{\nu}{\mu^2}\frac{d\mu}{dA} + \frac{1}{\frac{\nu}{\mu}\phi+\pi}\left(-\frac{\nu}{\mu^2}\phi\frac{d\mu}{dA}+\frac{d\pi}{dA}\right) \end{aligned} \tag{151}$$

$$\begin{aligned}\frac{1}{\text{TFP}}\frac{d\,\text{TFP}}{dA} &= \frac{1}{\nu}\frac{1}{A} - \left[\left(1-\frac{1}{\nu}\right)\frac{1}{1-\frac{\nu}{\mu}}\frac{\nu}{\mu^2} + \left(\frac{1}{\frac{\nu}{\mu}\phi+\pi}\right)\frac{\nu}{\mu^2}\phi\right]\frac{d\mu}{dA} \\ &\quad + \left[-\frac{1}{\nu}\frac{1}{\pi+\phi} + \frac{1}{\frac{\nu}{\mu}\phi+\pi}\right]\frac{d\pi}{dA} \end{aligned} \tag{152}$$

Use local approx. $\times dA = [A(t) - \tilde{A}]$ and remember $\tilde{\pi} = 0$

$$\hat{\text{TFP}} = \frac{1}{\nu}\hat{A} - \left(\frac{\tilde{\mu}-1}{\tilde{\mu}-\nu}\right)\hat{\mu} + \left(\frac{\hat{\mu}-1}{\nu\phi}\right)\hat{\pi} \tag{153}$$

G Parametric Model and Steady State

We begin with isoelastic utility

$$U(C,L) = \frac{C^{1-\sigma}-1}{1-\sigma} - \xi\frac{L^{1+\eta}}{1+\eta} \tag{154}$$

The partial derivatives are

$$U_C = C^{-\sigma}, \quad U_{CC} = -\sigma\frac{U_C}{C}, \quad U_L = -\xi L^{\eta}, \quad U_{LL} = \eta\frac{U_L}{L} \tag{155}$$

and Cobb-Douglas production

$$F(k,\ell) = k^{\alpha}\ell^{\beta} = K^{\alpha}L^{\beta}N^{-(\alpha+\beta)} = F(K,L)N^{-(\alpha+\beta)} \tag{156}$$

so derivatives are

$$F_k = \alpha k^{\alpha-1}\ell^{\beta} = \alpha K^{\alpha-1}L^{\beta}N^{1-(\alpha+\beta)}, \tag{157}$$
$$F_\ell = k^{\alpha}\beta\ell^{\beta-1} = K^{\alpha}\beta L^{\beta-1}N^{1-(\alpha+\beta)} \tag{158}$$

G.1 Parametric Equilibrium Conditions

L is defined in terms of (C, K, N) through the intratemporal condition, which conforms to our theoretical derivations for optimal labor behaviour.

$$L = \left(\frac{AK^{\alpha}\beta N^{1-(\alpha+\beta)}}{\mu(N)\xi C^{\sigma}}\right)^{\frac{1}{1+\eta-\beta}} \tag{159}$$

Under this parameterization the dynamical system is

$$\dot{C} = \frac{C}{\sigma}\left[\frac{1}{\mu(N)}A\alpha K^{\alpha-1}L^{\beta}N^{1-(\alpha+\beta)} - \rho\right] \tag{160}$$
$$\dot{E} = \frac{1}{\mu(N)}A\alpha K^{\alpha-1}L^{\beta}N^{1-(\alpha+\beta)}e - \frac{1}{\gamma}\left(AK^{\alpha}L^{\beta}N^{-(\alpha+\beta)}\left(1-\frac{\nu}{\mu(N)}\right) - \phi\right) \tag{161}$$
$$\dot{K} = N\left[AK^{\alpha}L^{\beta}N^{-(\alpha+\beta)} - \phi\right] - \frac{\gamma}{2}E^2 - C \tag{162}$$
$$\dot{N} = E \tag{163}$$

G.2 Parameteric Steady State

The intratemporal condition is

$$\tilde{C} = \left(\frac{\beta A\tilde{k}^{\alpha}\tilde{\ell}^{\beta-1-\eta}\tilde{N}^{\eta}}{\mu(\tilde{N})\xi}\right)^{\frac{1}{\sigma}} \tag{164}$$

The dynamical system is

$$\tilde{C} = \frac{\phi\nu}{\mu(\tilde{N})-\nu}\tilde{N} \tag{165}$$
$$\tilde{E} = 0 \tag{166}$$

$$\alpha\tilde{k}^{\alpha-1}\ell(\tilde{k},\tilde{N})^{\beta} = \frac{\mu(\tilde{N})\rho}{A} \tag{167}$$

$$\tilde{k}^{\alpha}\tilde{\ell}(\tilde{k},\tilde{N})^{\beta} = \frac{\phi}{A\left(1-\frac{\nu}{\mu(\tilde{N})}\right)} \tag{168}$$

Substituting $\tilde{C}(\tilde{N})$ into the intratemporal condition

$$\tilde{\ell}(\tilde{k},\tilde{N}) = \left(\frac{A\tilde{k}^{\alpha}\beta\tilde{N}^{-\sigma-\eta}}{\mu(\tilde{N})\xi\left(\frac{\phi\nu}{\mu(\tilde{N})-\nu}\right)^{\sigma}}\right)^{\frac{1}{1+\eta-\beta}} \tag{169}$$

Solving (167) and (168) gives $\tilde{k}$ and $\tilde{\ell}$

$$\tilde{k}(\tilde{N}) = \frac{\phi\alpha}{(\mu(\tilde{N})-\nu)\rho} \tag{170}$$

$$\tilde{\ell}(\tilde{N}) = \left[\frac{\mu(\tilde{N})}{A}\left(\frac{\rho}{\alpha}\right)^{\alpha}\left(\frac{\phi}{\mu(\tilde{N})-\nu}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} \tag{171}$$

Capital per firm is decreasing in market power, whilst labor per firm is ambiguous depending on the capital-labor income ratio $\frac{\tilde{r}\tilde{k}}{\tilde{w}\tilde{l}} = \frac{\alpha}{\beta}$.

$$\frac{\partial\tilde{\ell}}{\partial\tilde{\mu}} = \frac{\tilde{\ell}}{\tilde{\mu}(\tilde{\mu}-\nu)}\left(\frac{\alpha}{\beta}(\tilde{\mu}-1)-1\right) \gtreqless 0 \iff \frac{\alpha}{\beta}(\tilde{\mu}-1) \gtreqless 1 \tag{172}$$

An increase in number of firms, decreases market power and increases capital per firm. This is important to explain why output per firm $\tilde{y}$ increases, which we have shown in general. Rearranging the intratemporal condition (169) gives a nonlinear function in terms of $\tilde{N}$.

$$\tilde{N} = \left(\left(\frac{\alpha}{\rho\nu}\right)^{\sigma}\frac{A\beta}{\mu(\tilde{N})\xi\tilde{k}^{\sigma-\alpha}\tilde{\ell}^{1+\eta-\beta}}\right)^{\frac{1}{\sigma+\eta}} \tag{173}$$

where $\tilde{k},\tilde{\ell}$ are defined in (170) and (171) as functions of $\tilde{N}$. Therefore solving this nonlinear expression numerically gives $\tilde{N}$, which in turn provides $\tilde{C},\tilde{K},\tilde{L}$

by (165, 170, 171). Equation (174) presents equation (173) in $\tilde{N}$ terms only.

$$\tilde{N} = \left[\frac{\beta}{\xi\nu^{\sigma}}\left\{\left(A\left(\frac{\alpha}{\rho}\right)^{\alpha}\right)^{1+\eta}\left(\frac{1}{\mu(\tilde{N})}\right)^{\alpha(1+\eta)+\beta(1-\sigma)}\right.\right.$$
$$\left.\left.\left(\frac{1-\frac{\nu}{\mu(\tilde{N})}}{\phi}\right)^{1-\nu+\eta(1-\alpha)+\sigma\beta}\right\}^{\frac{1}{\beta}}\right]^{\frac{1}{\eta+\sigma}} \tag{174}$$

G.3 Parametric Labor Responses

$$\frac{dL}{dC} = -\frac{\sigma}{1+\eta-\beta}\frac{L}{C} \qquad < 0 \tag{175}$$
$$\frac{dL}{dK} = \frac{\alpha}{1+\eta-\beta}\frac{L}{K} \qquad > 0 \tag{176}$$
$$\frac{dL}{dN} = \frac{1-\nu-\varepsilon_{\mu N}}{1+\eta-\beta}\frac{L}{N} \qquad > 0 \tag{177}$$
$$\frac{\partial L}{\partial A} = \frac{1}{1+\eta-\beta}\frac{L}{A} \qquad > 0 \tag{178}$$

and the total effect of technology on labor is

$$\frac{dL}{dA} = \left(\frac{L}{1+\eta-\beta}\right)\left(-\frac{\sigma}{C}\frac{dC}{dA} + \frac{\alpha}{K}\frac{dK}{dA} + \frac{1-\nu-\varepsilon_{\mu N}}{N}\frac{dN}{dA} + \frac{1}{A}\right) \tag{179}$$

References

Aloi, Marta, Huw Dixon, and Anthony Savagar (2018). *Labor Responses, Regulation and Business Churn.* Studies in Economics 1804. School of Economics, University of Kent.

Atkeson, Andrew and Ariel Burstein (2008). "Pricing-to-market, trade costs, and international relative prices". In: *The American Economic Review* 98.5, pp. 1998–2031.

Baqaee, David Rezza and Emmanuel Farhi (2017). *Productivity and Misallocation in General Equilibrium.* Working Paper 24007. National Bureau of Economic Research.

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.

Basu, Susanto and John G Fernald (1997). "Returns to scale in US production: Estimates and implications". In: *Journal of political economy* 105.2, pp. 249–283.

Basu, Susanto, John G. Fernald, and Miles S. Kimball (2006). "Are technology improvements contractionary?" In: *The American Economic Review* 96.5, pp. 1418–1448.

Bhandari, Jagdeep S., Nadeem Ul Haque, and Stephen J. Turnovsky (1990). "Growth, External Debt, and Sovereign Risk in a Small Open Economy". In: *IMF Staff Papers* 37.2, pp. 388–417.

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.

Brito, Paulo, Luís F. Costa, and Huw Dixon (2013). "Non-smooth dynamics and multiple equilibria in a Cournot-Ramsey model with endogenous markups". In: *Journal of Economic Dynamics and Control* 37.11, pp. 2287 –2306.

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

Chatterjee, Santanu, Georgios Sakoulis, and Stephen J. Turnovsky (2003). "Unilateral capital transfers, public investment, and economic growth". In: *European Economic Review* 47.6, pp. 1077–1103.

Colciago, Andrea and Federico Etro (2010). "Real business cycles with Cournot competition and endogenous entry". In: *Journal of Macroeconomics* 32.4, pp. 1101–1117.

Cook, David (2001). “Time to enter and business cycles”. In: *Journal of Economic Dynamics and Control* 25.8, pp. 1241–1261.

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.

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.

De Loecker, Jan and Jan Eeckhout (2017). *The rise of market power and the macroeconomic implications*. Tech. rep. National Bureau of Economic Research.

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda (2018). *Changing Business Dynamism and Productivity: Shocks vs. Responsiveness*. Working Paper 24236. National Bureau of Economic Research.

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

— (1996b). “Monopolistic competition, increasing returns, and the effects of government spending”. In: *Journal of Money, credit and Banking* 28.2, pp. 233–254.

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

Dos Santos Ferreira, Rodolphe and Frédéric Dufourt (2006). “Free entry and business cycles under the influence of animal spirits”. In: *Journal of Monetary Economics* 53.2, pp. 311–328.

Eggertsson, Gauti B, Jacob A Robbins, and Ella Getz Wold (2018). *Kaldor and Piketty’s Facts: The Rise of Monopoly Power in the United States*. Tech. rep. National Bureau of Economic Research.

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

Flaschel, P., G. Groh, C. Proano, and W. Semmler (2008). *Topics in Applied Macrodynamic Theory*. Dynamic Modeling and Econometrics in Economics and Finance. Springer Berlin Heidelberg.

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.

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.

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.

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.

Linnemann, Ludger (2001). “The price index effect, entry, and endogenous markups in a macroeconomic model of monopolistic competition”. In: *Journal of Macroeconomics* 23.3, pp. 441–458.

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.

Murata, Y. and K. Shell (2014). *Mathematics for Stability and Optimization of Economic Systems*. Economic theory and mathematical economics. Elsevier Science.

Opp, Marcus M, Christine A Parlour, and Johan Walden (2014). "Markup cycles, dynamic misallocation, and amplification". In: *Journal of Economic Theory* 154, pp. 126–161.

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

Restuccia, Diego and Richard Rogerson (2008). "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments". In: *Review of Economic Dynamics* 11.4, pp. 707–720.

Rotemberg, Julio J. (2008). "Comment on Monetary Policy and Business Cycles with Endogenous Entry and Product Variety". In: *NBER Macroeconomics Annual 2007, Volume 22*. University of Chicago Press, pp. 367–375.

Rotemberg, Julio J. and Michael Woodford (1993). *Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets*. Working Paper 4502. National Bureau of Economic Research.

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

Savagar, Anthony and Huw D Dixon (2017). *Firm Entry, Excess Capacity and Endogenous Productivity*. Working Paper. University of Kent.

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

Recent Kent Discussion Papers in Economics

18/11: 'Endogenous Time-Varying Volatility and Emerging Market Business Cycles', Jan-Philipp Dueber

18/10: 'Marriage, Work and Migration: The Role of Infrastructure Development and Gender Norms', Amrit Amirapu, M Niaz Asadullah and Zaki Wahhaj

18/09: 'Population Aging, Government Policy and the Postwar Japanese Economy', Keisuke Otsu and Katsuyuki Shibayama

18/08: 'The Missing Link: Monetary Policy and The Labor Share', Cristiano Cantore, Filippo Ferroni and Miguel A. León-Ledesma

18/07: 'University vice-chancellor pay, performance and (asymmetric) benchmarking', Adelina Gschwandtner and Richard McManus

18/06: 'Constrained public goods in networks', Nizar Allouch and Maia King

18/05: 'The Fall in German Unemployment: A Flow Analysis', Carlos Carrillo-Tudela, Andrey Launov and Jean-Marc Robin

18/04: 'Labor Responses, Regulation and Business Churn in a Small Open Economy', Marta Aloi, Huw Dixon and Anthony Savagar

18/03: 'Will Urban Migrants Formally Insure their Rural Relatives? Family Networks and Rainfall Index Insurance in Burkina Faso', Harounan Kazianga and Zaki Wahhaj

18/02: 'Social Diversity and Bridging Identity', María D. C. García-Alonso and Zaki Wahhaj

18/01: 'Biased Technological Change and Employment Reallocation', Zsófia L. Bárány and Christian Siegel

17/21: 'Strategic Default in Financial Networks', Nizar Allouch and Maya Jalloul

17/20: 'Improving Drinking Water Quality in South Korea: A Choice Experiment', Adelina Gschwandtner, Cheul Jang and Richard McManus

17/19: 'Financial Frictions in Macroeconomic Models', Alfred Duncan and Charles Nolan

17/18: 'Aggregation in Networks', Nizar Allouch

17/17: 'Was Domar Right? Serfdom and Factor Endowments in Bohemia', Alexander Klein and Sheilagh Ogilvie

17/16: 'Disputes, Debt and Equity', Alfred Duncan and Charles Nolan

17/15: 'A Long-Run Perspective on the Spatial Concentration of Manufacturing Industries in the United States', Nicholas Crafts and Alexander Klein

17/14: 'Allocative efficiency of UK firms during the Great Recession', Florian Gerth