Measured Productivity with Endogenous Markups and Economic Profits

Anthony Savagar *†

June 23, 2019

Abstract

I show that sluggish variation in firm entry accounts for significant variation in measured TFP through profit and markup channels. I develop a model of dynamic firm entry, oligopolistic competition and returns to scale in order to decompose TFP into technical change, economic profit and markup channels. 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. The speed of firm adjustment (*'business dynamism'*) will determine importance of each bias. **JEL: E32, D21, D43, L13, C62, Endogenous markups, Dynamic Firm Entry, Endogenous Productivity, Endogenous Entry Costs**

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 procyclical 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 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 procyclical, whereas markups are

---

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

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 procyclical 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 nonparametric 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

---

[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; Kim 2004.

[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.

procylical productivity movements through excess capacity utilization.
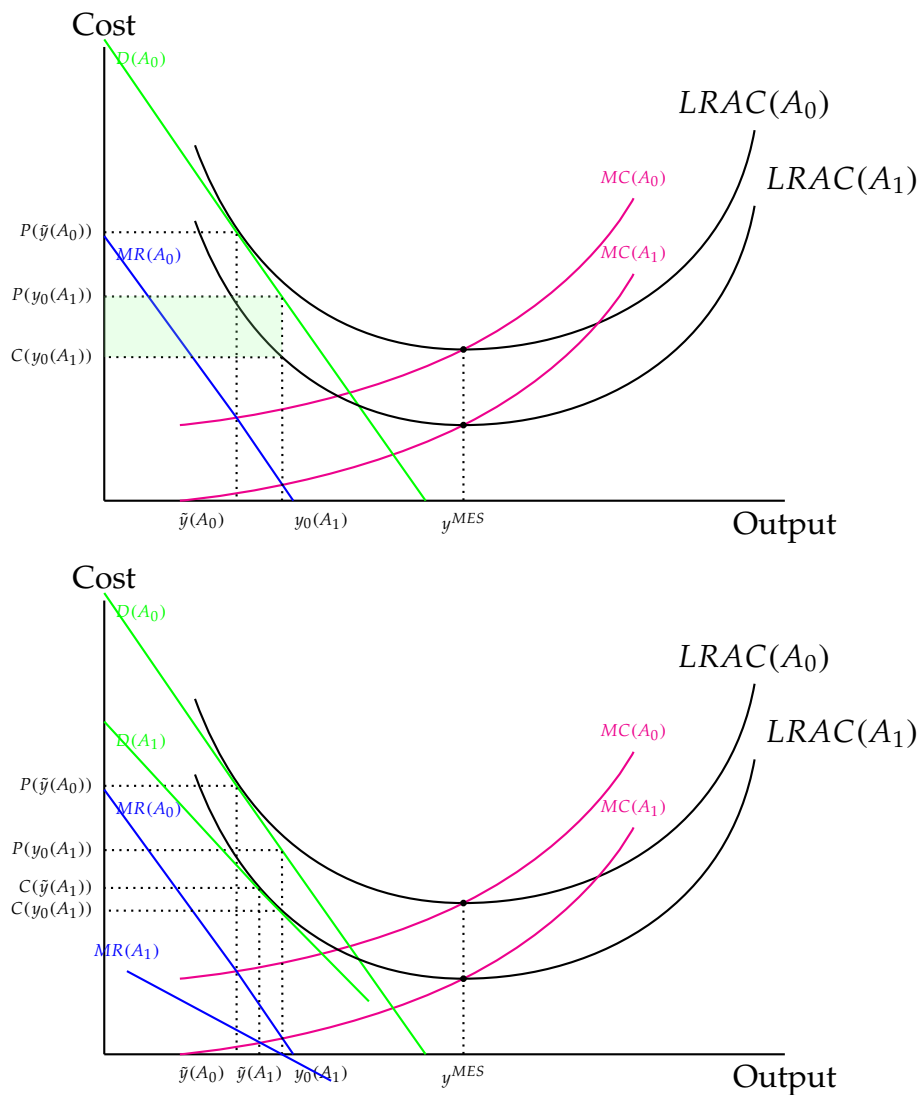
1 Cost Curves Intuition

Figure 1: Cost Curve Explanation

Figure 1 shows the cost curves and equilibria of a firm with increasing marginal costs (MC) and a U-shaped long-run average cost (LRAC) due to a fixed overhead cost. Demand (D) and marginal revenue (MR) curves are downard sloping due to product differentiation. The two diagrams represent the dynamic effect on an incumbent firm's costs and demand and in turn productivity resulting from an improvement in technology $A_0 < A_1$ 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 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 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 im-

---

[5]This is the steady state outcome under imperfect competition due to downward sloping demand curves from product differentiation. The level of output is less than the firm's minimum efficient scale (MES) $\tilde{y}(A_0) < y^{MES}$ 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.

[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.

provement. 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.

2 Model

2.1 Household

The representative household chooses future consumption $\{C(t)\}_0^\infty \in \mathcal{R}$ and labor supply $\{L(t)\}_0^\infty \in [0,1]$ to maximise lifetime utility $U : \mathcal{R}^2 \rightarrow \mathcal{R}$. Instantaneous utility $u : \mathcal{R} \times [0,1] \rightarrow \mathcal{R}$ is jointly concave and differentiable in both of its arguments. It is strictly increasing in in consumption $u_C > 0$, strictly decreasing in labor $u_L < 0$ and separable $u_{CL} = 0$. The household owns capital and takes equilibrium rental rate, wage rate and firm profits $K, r, w, \Pi \in \mathcal{R}_+$ as given. Capital investment $I \in \mathcal{R}$ equals the flow of capital as there is no depreciation. The household solves

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

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

$$I(t) = \dot{K} \tag{3}$$

---

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

5

Optimal paths satisfy the intertemporal consumption Euler equation (4), intratemporal labor-consumption trade-off (5) and resource constraint (2).[9]

$$\dot{C}(t) = -\frac{u_C}{u_{CC}}(r(t) - \rho), \tag{4}$$

$$w(t) = -\frac{u_L(L(t))}{u_C(C(t))} \tag{5}$$

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.

$$\lim_{t \to \infty} K(t)\lambda(t)e^{-\rho t} = 0, \tag{6}$$

$$K(0) = K_0 \tag{7}$$

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

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 aggregate output production function and industry output production function are given by

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

$$Q_\jmath = N^{1+\vartheta} \left[\frac{1}{N} \sum_{\imath=1}^N y_{\jmath\imath}^{\frac{\theta_F-1}{\theta_F}} \right]^{\frac{\theta_F}{\theta_F-1}}, \quad \theta_F > \theta_I \geq 1 \tag{9}$$

where $\vartheta \in [0, \infty)$ is a variety effect; $\theta_F \in (1, \infty)$ is between-firm substitutability; $\theta_I \in [1, \infty)$ is between-industry substitutability. Solving the aggregate and firm-level maximization problems yields aggregate and industry inverse demands that combine

---

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

to give firm-level inverse demand:

$$P_{ji} = \left(\frac{y_{ji}}{Q_J}\right)^{-\frac{1}{\theta_F}} \left(\frac{Q_J}{Y}\right)^{-\frac{1}{\theta_I}} N^{-\frac{1}{\theta_F} + \vartheta \frac{\theta_F - 1}{\theta_F}} P \tag{10}$$

Firms are symmetric in their production technology, firm ι in industry j produces output:

$$y_{ji}(t) := \max\{AF(k_{ji}(t), \ell_{ji}(t)) - \phi, 0\} \tag{11}$$

where $F : \mathfrak{R}_+^2 \supseteq (k, \ell) \to \mathfrak{R}_+$ is homogeneous of degree $v \in (0, 1)$ (hod-v). $\phi \in \mathfrak{R}_+$ is an overhead cost denominated in output.[10] 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 $F_{kk}F_{\ell\ell} - F_{k\ell}^2 > 0$.[11] $A \in [1, \infty)$ is a scale parameter reflecting the production technology. The individual firm solves

$$\max_{y_{J,\iota}} \pi_{ji} = P_{ji}y_{ji} - w\ell_{ji} - rk_{ji} \tag{12}$$

subject to inverse demand (10) which will lead to an endogenous markup.

2.2.1 Factor Market Equilibrium

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

$$\frac{AF_k(k, \ell)}{\mu(N)} = \frac{r}{P_{ji}} \tag{13}$$

$$\frac{AF_\ell(k, \ell)}{\mu(N)} = \frac{w}{P_{ji}} \tag{14}$$

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

---

[10]As in Hornstein 1993; Devereux, Head, and Lapham 1996a; Rotemberg and Woodford 1996; Cook 2001; Kim 2004; 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.

[11]Homogeneity of degree v and cross-derivative symmetry implies

$$vF(k, l) = F_k k + F_l l$$
$$(v - 1)F_l = F_{ll}l + F_{lk}k = F_{ll}l + F_{kl}k$$
$$(v - 1)F_k = F_{kl}l + F_{kk}k = F_{lk}l + F_{kk}k$$

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 hold.[12] I provide uniqueness conditions later.

A firm's maximized profit gives a relationship between firm-level real operating profit, output and markup[13]

$$\frac{\pi_{Jl}}{P_{Jl}} = y_{Jl} - \frac{1}{\mu_{Jl}}\nu(y_{Jl} + \phi) = \left(1 - \frac{\nu}{\mu_{Jl}}\right)y_{Jl} - \frac{\nu}{\mu_{Jl}}\phi = \left(1 - \frac{\nu}{\mu_{Jl}}\right)(y_{Jl} + \phi) - \phi \quad (16)$$

Dividing by y_{Jl} gives a relationship between the profit share and fixed cost share $\mu_{Jl}(1 - s_\pi) = \nu(1 + s_\phi)$.

2.2.2 Returns to Scale (RTS)

Returns to scale are defined as the inverse elasticity costs to output which is the ratio of average cost to marginal cost[14]

$$\text{RTS} \equiv \frac{\text{AC}}{\text{MC}} \quad \begin{cases} \in (1, \infty), & \text{increasing returns} \\ = 1, & \text{constant returns} \\ \in (0, 1), & \text{decreasing returns} \end{cases} \quad (17)$$

The production function exhibits two alternative representations of returns to scale defined as $\frac{\text{AC}}{\text{MC}}$. 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[15]

$$\text{RTS} = \nu\left(1 + s_\phi\right) = \mu\left(1 - s_\pi\right) \gtreqless 1 \quad (18)$$

where $s_\phi \equiv \frac{\phi}{y}$ and $s_\pi \equiv \frac{\pi}{P_{Jl}y}$ are the fixed cost share in output and profit share in revenue. The 'cost-based measure', 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', RTS $= \mu(1 - s_\pi)$, shows markups, profit

---

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

[13]From the profit function (91) and cost function (76) a firm's maximized profits are

$$\pi_{Jl} = P_{Jl}y_{Jl} - C(r, w, y_{Jl}) = P_{Jl}y_{Jl} - \text{MC}_{Jl}\,\nu(y_{Jl} + \phi) \quad (15)$$

This also implies that RTS $\equiv \frac{\text{AC}}{\text{MC}} = \mu(1 - s_\pi)$ where $s_\pi \equiv \frac{\pi}{Py}$ is the revenue share of profits.

[14]This follows from the definition of scale elasticity $\varepsilon_{Cy} \equiv C_y\frac{y}{C} = \text{MC} / \text{AC}$ and for homothetic production functions scale elasticity of the cost function equals returns to scale of the prodution function (Silberberg 1990, Ch. 8 Syverson 2019)

[15]See Appendix for derivation.

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 $v = 1$ implying 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 (39), except with constant returns to scale (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 $s_\pi = \left(1 - \frac{1}{\mu}\right) = \text{LI} = \frac{1}{\epsilon} \in (0, 1)$. Hence the profit-based RTS measure in (18) is also 1. However in BGM the profit share is always positive 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 $\phi > 0$, $v = 1$, 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.

2.2.3 Symmetric Equilibrium Profits, Prices, Aggregates

We already derived the symmetric equilibrium elasticity and markup (34). 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 $\frac{\partial Q_J}{\partial y_{ji}} = \frac{Q_J}{y_{ji}}\frac{1}{N}$.[16] Under symmetry, intermediate variables are equivalent

$$\forall (j, i) \in [0, 1] \times [1, N(t)] : y_{ji} = y, \; k_{ji} = k, \; \ell_{ji} = \ell, N_J = N$$

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

---

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

through how resources are divided.

$$k = \frac{K}{N} \tag{19}$$

$$\ell = \frac{L}{N} \tag{20}$$

$$F(k, \ell) = F\left(\frac{K}{N}, \frac{L}{N}\right) = N^{-\nu} F(K, L) \tag{21}$$

$$F_k(k, \ell) = F_K\left(\frac{K}{N}, \frac{L}{N}\right) = N^{1-\nu} F_K(K, L) \tag{22}$$

$$F_\ell(k, \ell) = F_\ell\left(\frac{K}{N}, \frac{L}{N}\right) = N^{1-\nu} F_L(K, L) \tag{23}$$

$$y = A N^{-\nu} F(K, L) - \phi \tag{24}$$

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

$$Q_J = N^{1+\vartheta} \left[\frac{1}{N} \sum_{\iota=1}^{N} y_{J\iota}^{\frac{\theta_F-1}{\theta_F}} \right]^{\frac{\theta_F}{\theta_F-1}} \implies Q_J = N^{1+\vartheta} y \quad (= Y) \tag{25}$$

$$P_J = N^{\frac{1}{\theta_F-1} - \vartheta} \left(\sum_{1}^{N} P_{J\iota}^{1-\theta_F} \right)^{\frac{1}{1-\theta_F}} \implies P_J = N^{-\vartheta} P_{J\iota} \quad (= P) \tag{26}$$

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

$$\varrho(N) = \frac{P_{J\iota}}{P} = N^{\vartheta} \tag{27}$$

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

$$P = P_J = P_{J\iota} = 1 \tag{28}$$

$$Y = Ny \tag{29}$$

$$= N^{1-\nu} A F(K, L) - N\phi \tag{30}$$

---

[17]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 $P_{J\iota}$) 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.

Under symmetry inverse price elasticity of demand (95) under Cournot becomes[18]

$$\epsilon^{-1} = \frac{1}{\theta_F} + \left[\frac{1}{\theta_I} - \frac{1}{\theta_F} \right] \frac{1}{N} \tag{33}$$

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

$$\mu(N) = \frac{1}{1 - \epsilon^{-1}} = \frac{1}{1 - \frac{1}{\theta_F} - \frac{1}{N} \left(\frac{1}{\theta_I} - \frac{1}{\theta_F} \right)}, \qquad \theta_F > \theta_I \geq 1 \tag{34}$$

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

$$\varepsilon_{\mu N}^C = 1 - \frac{\frac{\theta_F - 1}{\theta_F}}{\frac{\theta_F - 1}{\theta_F} - \frac{1}{N} \left(\frac{1}{\theta_I} - \frac{1}{\theta_F} \right)} = 1 - \left(\frac{\theta_F - 1}{\theta_F} \right) \mu < 0 \tag{35}$$

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 $\varepsilon_{\mu N}|^{\theta_I = 1} = -\frac{1}{N-1}$.[19] Appendix B.7 illustrates these markup properties graphically and for the Bertrand case.

---

[18] Taking the derivative of (9) gives

$$\frac{\partial Q_{Jl}}{\partial y_{jl}} = \frac{Q_J}{\left(\sum_{l=1}^{N} y_{jl}^{\frac{\theta_F - 1}{\theta_F}} \right)} y_{jl}^{-\frac{1}{\theta_F}} \implies \frac{\partial Q_{Jl}}{\partial y_{jl}} \frac{y_{jl}}{Q_{jl}} = \frac{y_{jl}^{\frac{\theta_F - 1}{\theta_F}}}{\left(\sum_{l=1}^{N} y_{jl}^{\frac{\theta_F - 1}{\theta_F}} \right)} \tag{31}$$

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

$$\frac{\partial Q_J}{\partial y_{jl}} = \frac{Q_J}{y_{jl}} \frac{1}{N} \implies \frac{y_{jl}}{Q_J} \frac{\partial Q_J}{\partial y_{jl}} = \frac{1}{N} \tag{32}$$

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

2.2.4 Symmetric Equilibrium profit-output-markup Relationship

With normalized pricing and symmetry, factor market equilibrium becomes

$$r = \frac{AF_k(k,\ell)}{\mu(N)} = \frac{AN^{1-\nu}F_K(K,L)}{\mu(N)} \tag{36}$$

$$w = \frac{AF_\ell(k,\ell)}{\mu(N)} = \frac{AN^{1-\nu}F_L(K,L)}{\mu(N)} \tag{37}$$

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

$$rk + w\ell = \frac{\nu}{\mu(N)}(y + \phi) \tag{38}$$

Consequently, from $\pi = y - rk - w\ell$, optimized profits and output per firm are[20]

$$\pi(y, \mu(N)) = \left(1 - \frac{\nu}{\mu(N)}\right)(y + \phi) - \phi \tag{39}$$

$$y(\pi, \mu(N)) = \left(1 - \frac{\nu}{\mu(N)}\right)^{-1}(\pi + \phi) - \phi \tag{40}$$

This implies a correspondence between output per firm, profits and markups.

2.2.5 Aggregation and TFP

The aggregate reduced-form production function is[21]

$$Y = \left(\frac{A}{\pi + \phi}\right)^{\frac{1}{\nu}}\left(1 - \frac{\nu}{\mu}\right)^{\frac{1}{\nu}-1}\left(\pi + \frac{\nu}{\mu}\phi\right)F(K,L)^{\frac{1}{\nu}} \tag{41}$$

TFP relates aggregate output to inputs and is therefore defined as[22]

$$\text{TFP} \equiv \frac{Y}{F(K,L)^{\frac{1}{\nu}}} = \frac{y}{F(k,\ell)^{\frac{1}{\nu}}} \tag{42}$$

$$\text{TFP}(t) \equiv \left(\frac{A}{\pi(t) + \phi}\right)^{\frac{1}{\nu}}\left(1 - \frac{\nu}{\mu(t)}\right)^{\frac{1}{\nu}-1}\left(\frac{\nu}{\mu(t)}\phi + \pi(t)\right) \tag{43}$$

---

[20]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.

[21]See Appendix F for full derivation. The result uses $y(\pi, \mu)$, $Y(N, y)$, and $Y(N, K, L)$ to get $Y(\pi, \mu, K, L)$. First substitute out $N = \frac{Y}{y}$ from (29), then collect terms in Y, lastly replace $y(\pi, \mu(N))$ using (40).

[22]See Kim 2004; Barseghyan and DiCecio 2011 for similar derivations of reduced-form aggregate output with non-constant marginal costs and overhead costs. 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).

TFP varies endogenously with profits and markups. These two endogenous components relate 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

$$\text{TFP}(t) = A^{\frac{1}{\nu}} \frac{y(t)}{(y(t) + \phi)^{\frac{1}{\nu}}} \tag{44}$$

This shows given fixed parameters ϕ, ν, variations in measured TFP can arise through two variables: exogenous technology A and output per firm y. The elasticity of TFP with respect to firm output depends on returns to scale[23]

$$\varepsilon_{\text{TFP}\,y} = 1 - \frac{1}{\text{RTS}} \tag{46}$$

Through (40) output per firm depends on profits and markups which in turn depend on A. Therefore through (44), exogenous shocks to A will have a direct effect on measured TFP but also an effect due to y which by (46) affects measured TFP due to RTS.

2.2.6 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 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.[24] 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. Under free-entry this equates to the entry cost $q \in \mathfrak{R}$

$$V(t) = \int_{s=t}^{\infty} \pi(t)e^{-r(s-t)}dt \tag{47}$$

$$V(t) = q(t) \tag{48}$$

---

[23]Output per firm y has an effect on measured TFP providing there are not constant returns to scale (RTS = 1):

$$\frac{\partial \text{TFP}}{\partial y} = \frac{\text{TFP}}{y}\left(1 - \frac{1}{\nu\left(1 + \frac{\phi}{y}\right)}\right) = \frac{\text{TFP}}{y}\left(1 - \frac{1}{\text{RTS}}\right) \tag{45}$$

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

We assume that the entry cost increases in the rate of entry $\dot{N}$, and its sensitivity to entry depends on the exogenous congestion parameter ζ [25]

$$q(t) = \zeta \dot{N}, \qquad \zeta \in (0, \infty) \tag{49}$$

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.[26] 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. Therefore when combined with the free entry condition and sunk cost assumption

$$r(t)q(t) = \dot{q}(t) + \pi(t) \tag{50}$$

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.

$$\dot{N}(t) \equiv E(t) \tag{51}$$

$$\dot{E}(t) = -\frac{\pi(t)}{\zeta} + r(t)E(t), \quad \zeta \in (0, \infty) \tag{52}$$

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

$$N(0) = N_0 \tag{53}$$

$$\lim_{t \to \infty} e^{-\rho t} u_C N(t) q(t) = 0 \tag{54}$$

The rate of entry increases $\dot{E} > 0$ if the outside option $r(t)E(t)$ exceeds the profit from entering $\frac{\pi(t)}{\zeta}$. 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

---

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

[26]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.

congestion. The result is an increase in the amount of entry.

The aggregate cost of entry (total investment in firms) $Z(t) \in \mathfrak{R}$ is

$$Z(t) = \zeta \int_0^{E(t)} i \, di = \zeta \frac{E(t)^2}{2} \tag{55}$$

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.

$$\Pi(t) = N(t)\pi(t) - Z(t) \tag{56}$$

Plugging aggregate profits into the income identity of the household leads to the aggregate resource constraint that output is split between consumption, investment in capital and investment in firms [27]

$$Y = C + I + Z \tag{61}$$

2.3   Competitive Equilibrium

Definition 1 (Competitive Equilibrium)**.** Competitive equilibrium is the equilibrium paths of aggregate quantities and prices $\{C(t), L(t), K(t), N(t), E(t), w(t), r(t)\}_{t=0}^{\infty}$, with prices strictly positive, such that $\{C(t), L(t)\}_{t=0}^{\infty}$ solve the household problem. $\{K(t)\}_{t=0}^{\infty}$ satisfies the law of motion for capital. Labor and capital $\{L(t), K(t)\}_{t=0}^{\infty}$ maximise firm profits given factor prices. The flow of entry causes the arbitrage condition on entry to hold (price of entry equals net present value of incumbency). State variables $\{K(t), N(t)\}_{t=0}^{\infty}$ satisfy transversality. Factor prices are set according to factor market equilibrium (14) and ensure goods and factor markets clear.

---

[27]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 $Z(E(t))$.

$$\Pi(t) = N(t)\pi(t) - Z(E(t)) = Y(t) - w(t)L(t) - r(t)K(t) - Z(E(t)) \tag{57}$$

This income can be spent on consumption and investment in capital

$$I(t) + C(t) = w(t)L(t) + r(t)K(t) + \Pi(t) \tag{58}$$

$$\dot{K}(t) + C(t) = Y(t) - Z(E(t)) \tag{59}$$

$$\dot{K} = Y(t) - \frac{\zeta E(t)^2}{2} - C(t) \tag{60}$$

2.4 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 3 are a system of four ODEs in consumption, entry, capital and number of firms (C, E, K, N). In the appendix, I show that the system is characterized by saddle path dynamics such that K, N are fixed state variables on impact of a shock, whereas C, E jump instantaneously to the saddle path.

2.4.1 Steady-State Solutions

A steady-state equilibrium occurs when the dynamic equations are stationary. Therefore consumption, entry, capital and number of firms are constant over time. 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 $\{A, v, \phi, \rho, \theta_F, \theta_I\}$[28]

$$\dot{C} = 0 \iff \tilde{r}(\tilde{C}, \tilde{K}, \tilde{N}) = \rho \tag{62}$$

$$\dot{E} = 0 \iff \tilde{\pi}(\tilde{C}, \tilde{K}, \tilde{N}) = 0 \tag{63}$$

$$\dot{K} = 0 \iff \tilde{Y}(\tilde{C}, \tilde{K}, \tilde{N}) = \tilde{C} \tag{64}$$

$$\dot{N} = 0 \iff \tilde{E} = 0 \tag{65}$$

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 1 (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.*

$$\tilde{y}(\mu(\tilde{N})) = \frac{v\phi}{\mu(\tilde{N}) - v} \tag{66}$$

$$T\tilde{F}P(\mu(\tilde{N})) = v \left[\frac{A}{\mu(\tilde{N})} \left(\frac{\mu(\tilde{N}) - v}{\phi} \right)^{1-v} \right]^{\frac{1}{v}} \tag{67}$$

We can understand why output is increasing in entry by considering the profit representation (39). In steady-state profit is zero (63), therefore 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. Equivalently, the fixed cost share in output decreases as

---

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

the markup falls with entry

$$\tilde{s}_\phi = \frac{\tilde{\mu}}{\nu} - 1 \tag{68}$$

The TFP result (67) follows because there are increasing returns in steady-state with imperfect competition, this is clear from (18) under zero-profits, in fact the returns to scale in steady-state equate to the long-run markup

$$\tilde{\text{RTS}} = \nu(1 + \tilde{s}_\phi) = \tilde{\mu} \geq 1 \tag{69}$$

therefore an increase in output translates to an increase in TFP.[29] In the special case of perfect competition, when there are many firms $\mu_{N \to \infty} = 1$, and demand curves are flat, the steady-state output (66) is equivalent to a firm's minimum efficient scale (MES) (83), implying TFP is maximized, and firms are operating at the minimum of their average cost curve, where AC = MC implying constant returns RTS = 1. This result would not exist with flat marginal cost curves $\nu = 1$, where a steady-state would not be well-defined.

3 Measured TFP Dynamics

From our definition of measured TFP (44) take the logarithmic derivative

$$\hat{\text{TFP}}(t) \approx \frac{1}{\nu}\hat{A}(t) + \left(1 - \frac{1}{\tilde{\text{RTS}}}\right)\hat{y}(t) \tag{70}$$

Hat notation represents deviation from steady-state: $\hat{x} = \frac{\dot{x}}{x}$. In steady-state the markup is equivalent to returns to scale $\tilde{\mu} = \tilde{\text{RTS}} \geq 1$. Therefore the coefficient on firm-level output variations is the price-cost margin *i.e.* Lerner index: $\left(1 - \frac{1}{\tilde{\mu}}\right) = 1 - \frac{1}{\text{P}/\text{MC}} = \frac{\text{P}-\text{MC}}{\text{P}} \in (0, 1)$. The result shows that variations in measured TFP are composed of variations in pure technology $\hat{A}$, but also variations in output per firm $\hat{y}$ (firm intensive margin) interacted with returns to scale. Since $\tilde{\text{RTS}} = \tilde{\mu} > 1$ with imperfect competition, measured TFP fluctuations overestimate pure technology fluctuations because it also 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:

---

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

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

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 (40) output per firm is a function of profits and markups, which in turn implies measured TFP is a function of technology A, markups μ and profits π as in (43). Log-linearizing the output per firm expression (40) gives

$$\hat{y} = \tilde{\mu} \left(\frac{1}{\nu\phi} \hat{\pi} - \frac{1}{\tilde{\mu} - \nu} \hat{\mu} \right)$$

and substituting this into (70) gives[30]

$$\hat{\text{TFP}}(t) \approx \hat{A} \frac{1}{\nu} - \hat{\mu}(t) \left(\frac{\tilde{\mu} - 1}{\tilde{\mu} - \nu} \right) + \hat{\pi}(t) \left(\frac{\tilde{\mu} - 1}{\nu\phi} \right) \tag{71}$$

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.[31] 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 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 (42) (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 2. *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 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-

---

[30]This result can be verified through direct linearization of (43).

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

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.[32] 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 $[\mu(0) - \tilde{\mu}] = 0$.[33] 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 3. *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 importance shrinks. Therefore both profits and markups are intertemporally having a measured productivity increasing effect 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.[34] The markup effect continuously grows in importance, 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

---

[32]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 verifies Jaimovich and Floetotto 2008, eq. 24. Therefore the coefficient here is smaller given $\tilde{\mu}$.

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

[34]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.
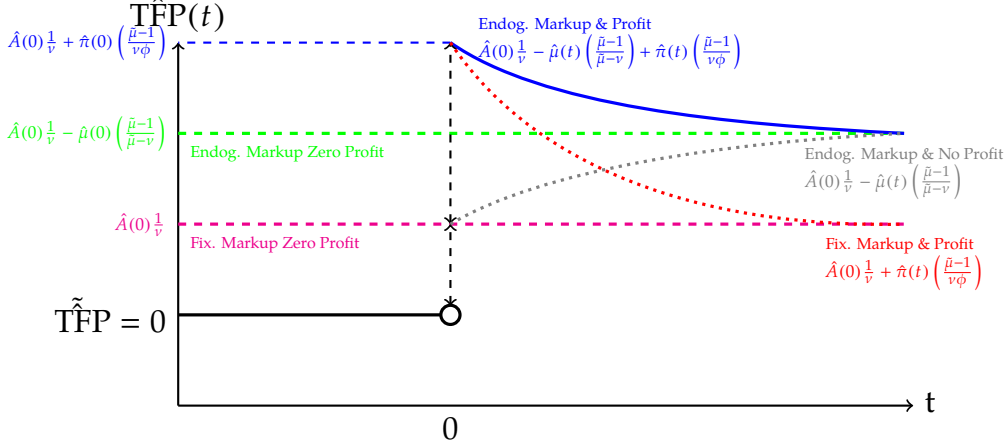
Figure 2: Endogenous Productivity Decomposition, *Positive Shock* $\tilde{A} \to A(0)$

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.

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

$$-\hat{\mu}(t')\left(\frac{\tilde{\mu}-1}{\tilde{\mu}-\nu}\right) = \hat{\pi}(t')\left(\frac{\tilde{\mu}-1}{\nu\phi}\right) \tag{72}$$

$$\frac{\hat{\pi}(t')}{\hat{\mu}(t')} = -\frac{\nu\phi}{\tilde{\mu}-\nu} = -\tilde{y} \tag{73}$$

This is the ratio of coefficients in (71). It shows that the relative importance of each effect depends on the fixed cost and the steady-state markup.[35] In the parametric numerical exercise we estimate $t' = 5$ quarters, implying the length of time for which the profit bias dominates the markup bias.

3.1 Parametric Example

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

$$U(C,L) = \frac{C^{1-\sigma}-1}{1-\sigma} - \xi\frac{L^{1+\eta}}{1+\eta} \tag{74}$$

---

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

$$y = Ak^\alpha \ell^\beta - \phi \tag{75}$$

Isoelastic utility implies there is constant elasticity of marginal utility with respect to consumption and labor $U_{CC}\frac{C}{U_C} = -\sigma$ and $U_{LL}\frac{L}{U_L} = \eta$.[36] Frisch elasticity of labor supply is given by $\frac{1}{\eta}$, hence we calibrate inverse elasticity $\eta \in (0, \infty)$.[37] 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 1 summarizes the calibration used for simulation. The time

Table 1: Parameter Values

Capital Share	α	0.283
Labor Share	β	0.567
Fixed Cost	ϕ	2.5
Entry Congestion	ζ	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 $\ln C$

interval is a quarter and correspondingly $\rho = 0.02$ implies an annualized discount rate of 8%.[38] 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 ζ 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.[39] The intra- and inter- industry substitutability parameters are chosen in line with literature (see table 2). 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.[40] These parameter values

---

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

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

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

[39]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 of firm convergence.

[40]Also see Da-Rocha, Tavares, and Restuccia 2017 for the same approach to decreasing returns.
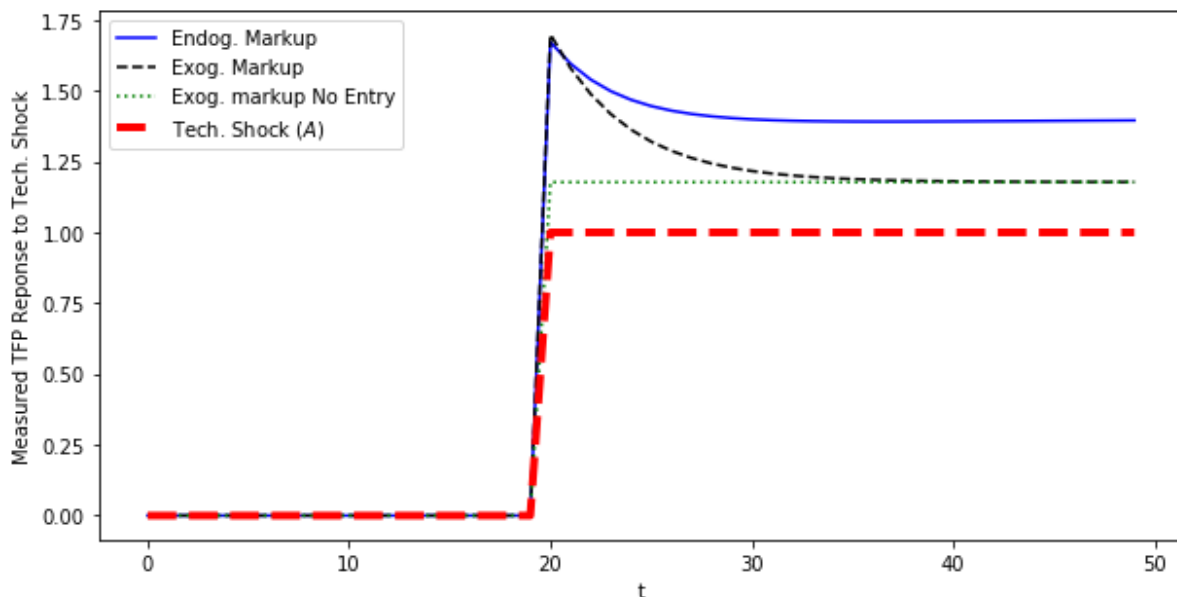
Figure 3: Endogenous Vs Exogenous Responses

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 $A_1 = 1.1$.

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.[41] 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 procyclical sunk costs. When $\zeta \to 0$ an economy exhibits strong business dynamism and will more accurately reflect underlying technology as firms acquire positive profits for a shorter period of time.

4 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

---

[41]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 $\nu = 1$.

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 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 the nominal cost function

$$C(r, w, y_{jt}) = rk_{jt} + w\ell_{jt} = \text{MC}_{jt}\, v(y_{jt} + \phi) \tag{76}$$

where the Lagrange multiplier is the MC.[42] 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-v production function exhibits unit-cost function form[43]

$$C(r, w, y_{jt}) = \left(\frac{y_{jt} + \phi}{A}\right)^{\frac{1}{v}} C(r, w, 1) \tag{78}$$

where the unit cost function $C(r, w, 1)$ is independent of output.[44] Given this, the cost function representation (78) implies the marginal cost is positive and increasing if $v \in (0, 1)$, but flat with $v = 1$

$$\text{MC} = \frac{\partial\, C(r, w, y_{jt})}{\partial y} = \frac{1}{v}\frac{C(r, w, y_{jt})}{y + \phi} > 0 \tag{79}$$

$$\frac{\partial\, \text{MC}}{\partial y} = \frac{1 - v}{v}\frac{\text{MC}}{y_{jt} + \phi} > 0 \quad \text{if} \quad v \in (0, 1) \tag{80}$$

From (78), we can see the average cost curve is U-shaped with increasing marginal costs $v \in (0, 1)$ and an overhead cost ϕ

$$\text{AC} = \frac{C(r, w, y_{jt})}{y_{jt}} = \frac{C(r, w, 1)(y_{jt} + \phi)^{\frac{1}{v}}}{y_{jt}} \tag{81}$$

---

[42]Where the Lagrange multiplier is marginal cost, a firm's cost minimization problem yields $r = \text{MC}\, AF_k$ and $w = \text{MC}\, AF_\ell$ leading to cost function (76) by $C = rk + w\ell$. Dividing factor prices by P_{jt} yields the factor market equilibrium in markup terms. The real cost function, where factor prices are in real terms, is

$$\frac{C(r, w, y_{jt})}{P_{jt}} = \frac{r}{P_{jt}}k_{jt} + \frac{w}{P_{jt}}\ell_{jt} = \frac{\text{MC}_{jt}}{P_{jt}}v(y_{jt} + \phi) = \frac{v}{\mu_{jt}}(y_{jt} + \phi) \tag{77}$$

[43]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.

[44]I have assumed an output denomiated overhead cost such that $\pi_{jt} = P_{jt}AF(k_{jt}, \ell_{jt}) - P_{jt}\phi - w\ell_{jt} - rk_{jt}$, whereas denominating the overhead cost in terms of a factor of production, e.g. $\pi_{jt} = P_{jt}AF(k_{jt}, \ell_{jt}) - w\phi - w\ell_{jt} - rk_{jt}$, would give a unit cost function $C(r, w, y_{jt}) = \left(y_{jt}^{\frac{1}{v}} + \phi\right)C(r, w, A, 1)$.

$$\frac{\partial\,\mathrm{AC}}{\partial y} = \frac{\mathrm{AC}}{y}\left(\frac{1}{\nu(1+s_\phi)}-1\right), \qquad \text{where}\ \ s_\phi \equiv \frac{\phi}{y} \tag{82}$$

A.1 Minimum Efficient Scale (MES)

From the U-shaped AC curve slope (82), 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)

$$y^{\mathrm{MES}} = \frac{\nu\phi}{1-\nu} \tag{83}$$

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 (46) 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.[45] 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.

B Endogenous Markup Details

B.1 Aggregate and Industry Demands

The program faced by the final output producer is

$$\max_{Q_J,\ J\in[0,1]}\left\{PY - \int_0^1 Q_J p_J dJ\right\} \tag{84}$$

subject to the aggregate output production function

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

---

[45]In the steady state analysis (Section 2.4.1) we show that the minimum efficient scale arises under perfect competition (i.e. no markups $\mu = 1$).

The first-order condition gives constant-elasticity inverse demand for industry j

$$P_J = \left(\frac{Q_J}{Y}\right)^{-\frac{1}{\theta_I}} P \tag{85}$$

The program faced by the industry is

$$\max_{y_{Ji},\, i \in (1...N)} \left\{ p_J Q_J - \sum_1^N y_{Ji} p_{Ji} \right\} \tag{86}$$

subject to the industry output production function where $\vartheta \in [0, \infty)$ is a variety effect[46]

$$Q_J = N^{1+\vartheta} \left[\frac{1}{N} \sum_{i=1}^N y_{Ji}^{\frac{\theta_F-1}{\theta_F}}\right]^{\frac{\theta_F}{\theta_F-1}}, \quad \theta_F > \theta_I \geq 1 \tag{9}$$

This leads to industry-level inverse demand[47]

$$P_{Ji} = \left[\frac{y_{Ji}}{Q_J}\right]^{-\frac{1}{\theta_F}} N^{-\frac{1}{\theta_F} + \vartheta \frac{\theta_F-1}{\theta_F}} P_J \tag{90}$$

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

$$P_{Ji} = \left(\frac{y_{Ji}}{Q_J}\right)^{-\frac{1}{\theta_F}} \left(\frac{Q_J}{Y}\right)^{-\frac{1}{\theta_I}} N^{-\frac{1}{\theta_F} + \vartheta \frac{\theta_F-1}{\theta_F}} P \tag{10}$$

B.2   Firm Profit Maximization Problem

The individual firm solves

$$\max_{y_{J,i}} \pi_{Ji} = P_{Ji} y_{Ji} - C(r, w, y_{Ji}) \tag{91}$$

---

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

[47]The elasticity of demand depends on own-effect on industry quantities

$$\frac{\partial P_{Ji}}{\partial y_{Ji}} = -\frac{1}{\theta_F} \frac{P_{Ji}}{y_{Ji}} + \frac{1}{\theta_F} \frac{P_{Ji}}{Q_J} \frac{\partial Q_J}{\partial y_{Ji}} \tag{87}$$

$$\frac{\partial P_{Ji}}{\partial y_{Ji}} \frac{y_{Ji}}{P_{Ji}} = -\frac{1}{\theta_F} + \frac{1}{\theta_F} \frac{y_{Ji}}{Q_J} \frac{\partial Q_J}{\partial y_{Ji}} \tag{88}$$

$$\varepsilon_{P_{Ji} y_{Ji}} = -\frac{1}{\theta_F} + \frac{1}{\theta_F} \varepsilon_{Q_{Ji} y_{Ji}} \tag{89}$$

subject to inverse demand (10) and cost function (78) and taking factor prices as given. This implies that at optimal $\text{MR}_{Jı} = \text{MC}_{Jı}$. 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 $\epsilon_{Jı}$

$$\epsilon_{Jı} \equiv -\frac{\partial y_{Jı}}{\partial P_{Jı}} \frac{P_{Jı}}{y_{Jı}} \tag{92}$$

$$\text{MR}_{Jı} = \frac{\partial P_{Jı}}{\partial y_{Jı}} y_{Jı} + P_{Jı} = P_{Jı} \left(\frac{\epsilon_{Jı} - 1}{\epsilon_{Jı}} \right) \tag{93}$$

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

$$\mu_{Jı} \equiv \frac{P_{Jı}}{\text{MC}_{Jı}} = \frac{\epsilon_{Jı}}{\epsilon_{Jı} - 1} \tag{94}$$

B.2.1   Endogenous Price Elasticity of Demand

At the firm-level an individual firm will maximize profits subject to inverse demand (10) and its production function (11). Crucially the firm has a degree of price setting ability over $P_{Jı}$. Under quantity competition (Cournot), it chooses $y_{Jı}$.[48] Differentiating (10) with respect to $y_{Jı}$ and multiplying by $-\frac{y_{Jı}}{P_{Jı}}$ gives inverse price elasticity of demand

$$\epsilon_{Jı}^{-1} = -\frac{\partial P_{Jı}}{\partial y_{Jı}} \frac{y_{Jı}}{P_{Jı}} = \frac{1}{\theta_F} + \left[\frac{1}{\theta_I} - \frac{1}{\theta_F} \right] \frac{\partial Q_J}{\partial y_{Jı}} \frac{y_{Jı}}{Q_J} \tag{95}$$

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_J. Therefore by choosing their own production $y_{Jı}$, a firm has a direct effect $\frac{1}{\theta_F}$ and an indirect effect through $\frac{\partial Q_J}{\partial y_{Jı}}$. It is the indirect effect that causes endogenous markups, and in its absence (when firms are atomistic $\frac{\partial Q_J}{\partial y_{Jı}} = 0$) there is standard monopolistic competition from the direct effect. Later, when we impose symmetric equilibrium, we shall see $\frac{y_{Jı}}{Q_J} \frac{\partial Q_J}{\partial y_{Jı}} = \frac{1}{N}$ which makes the price elasticity of demand and in turn markup dependent on number of firms.

---

[48]See Appendix B.3 for Bertrand case.

B.3 Bertrand Derivation of Markup

For the Bertrand case we recast each first order condition to obtain conditional demand as a function of prices. Therefore (85), (90), (10) become (96), (97), (98).

$$Q_J = \left(\frac{P_J}{P}\right)^{-\theta_I} Y \tag{96}$$

$$y_{Ji} = \left(\frac{P_{Ji}}{P_J}\right)^{-\theta_F} \left(\frac{1}{N}\right)^{1-\vartheta(\theta_F-1)} Q_J \tag{97}$$

$$y_{Ji} = \left(\frac{P_{Ji}}{P_J}\right)^{-\theta_F} \left(\frac{P_J}{P}\right)^{-\theta_I} \left(\frac{1}{N}\right)^{1-\vartheta(\theta_F-1)} Y \tag{98}$$

The corresponding price indices follow from substituting the two FOCs (96), (97) into their corresponding constraints (production functions) (8), (9), giving aggregate price index and industry price index

$$P = \left(\int_0^1 P_J^{1-\theta_I} dJ\right)^{\frac{1}{1-\theta_I}} \tag{99}$$

$$P_J = N^{\frac{1}{\theta_F-1}-\vartheta} \left(\sum_1^N P_{Ji}^{1-\theta_F}\right)^{\frac{1}{1-\theta_F}} \tag{100}$$

Under Bertrand firms maximize profits subject to conditional demand (98) by choosing P_{Ji}. Therefore

$$\frac{\partial y_{Ji}}{\partial P_{Ji}} = -\theta_F \frac{y_{Ji}}{p_{Ji}} + (\theta_F - \theta_I) \frac{y_{Ji}}{P_J} \frac{\partial P_J}{\partial P_{Ji}} \tag{101}$$

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

$$\frac{\partial P_J}{\partial P_{iJ}} = \frac{P_J}{\sum_1^N P_{Ji}^{1-\theta_F}} P_{Ji}^{-\theta_F} \tag{102}$$

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

$$\epsilon \equiv -\frac{\partial y_{Ji}}{\partial P_{Ji}} \frac{P_{Ji}}{y_{Ji}} \tag{103}$$

Then (98) becomes

$$\epsilon = \theta_F - (\theta_F - \theta_I)\frac{1}{N} \tag{104}$$

Therefore for Bertrand the markup is

$$\mu(N)^{\mathrm{B}} = \frac{\theta_I + \theta_F(N-1)}{(\theta_F - 1)N - (\theta_F - \theta_I)} = \frac{1 - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\theta_I}{1 - \frac{1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\theta_I}, \qquad \theta_F > \theta_I \geq 1 \quad (105)$$

$$\varepsilon^{\mathrm{B}}_{\mu N} = \frac{N\theta_F}{\theta_I + \theta_F(N-1)}\left(1 - \frac{\theta_F - 1}{\theta_F}\mu^{\mathrm{B}}\right) < 0 \quad (106)$$

B.4 Cournot and Bertrand first-order Equivalence

To a first-order approximation the relationship between the markup and number of firms is the same under price and quantity competition. For Bertrand we can rearrange the markup to

$$\theta_I\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\frac{1}{N} = 1 - \frac{1}{\theta_F} - \frac{1}{\mu} \quad (107)$$

For Cournot we can rearrange the markup to

$$\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\frac{1}{N} = 1 - \frac{1}{\theta_F} - \frac{1}{\mu} \quad (108)$$

If we take logarithms and then differentiate with respect to time both relationships give

$$-\hat{N} = \left(1 - \frac{1}{\theta_F} - \frac{1}{\tilde{\mu}}\right)^{-1}\frac{\hat{\mu}}{\tilde{\mu}} \quad (109)$$

B.5 Cournot Markup as a Function of Bertrand

The Cournot and Bertrand markups are related as follows

$$\mu^{\mathrm{C}} = \frac{\mu^{\mathrm{B}}}{\theta_I\left(\frac{1}{\theta_I} - \left[1 + \mu^{\mathrm{B}}\left(\frac{1}{\theta_I} - 1\right)\right]\frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\right)} \quad (110)$$

This follows from rewriting the Cournot markup as Bertrand

$$\mu^{\mathrm{C}} = \left[(\mu^{\mathrm{B}})^{-1}(1 - \Upsilon\theta_I) + \Upsilon\theta_I - \Upsilon\right]^{-1}, \quad \text{where } \Upsilon = \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right) \quad (111)$$

and rearranging.

B.6 Markup Calibration Survey

Table 2 surveys the intersectoral θ_I and intrasectoral θ_F parameters used in the literature.

Table 2: Endogenous Markup Literature Survey

	θ_I	θ_F	μ	$\varepsilon_{\mu N}$
Cournot				$\varepsilon^C = 1 - \frac{\theta_F-1}{\theta_F}\mu^C$
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				$\varepsilon^B = \frac{N\theta_F}{\theta_I+\theta_F(N-1)}\left(1-\frac{\theta_F-1}{\theta_F}\mu^B\right)$
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 $\theta_F = \frac{1}{1-\tau} = \frac{1}{1-0.949} = 19.6$.
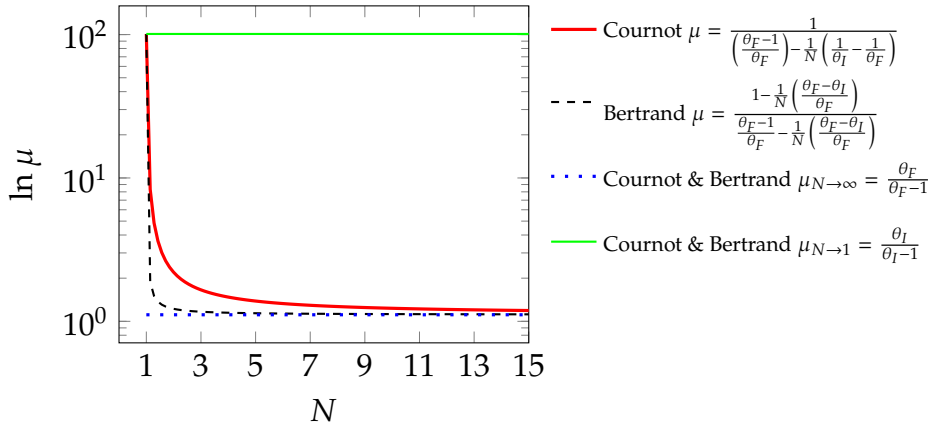[4] Lewis and Poilly 2012, p.680 plot their competition effect for a domain of substitutabilities $\theta_I, \theta_F \in (1.0, 4.0)^2$.
† θ_F does not affect the markups elasticity to number of firms when there is unitary elasticity across industries.
‡ Estimated value to 2 decimal places.

B.7 Graphical Illustration of Markup Properties

General Markups $\theta_I = 1.01,\ \theta_F = 10$

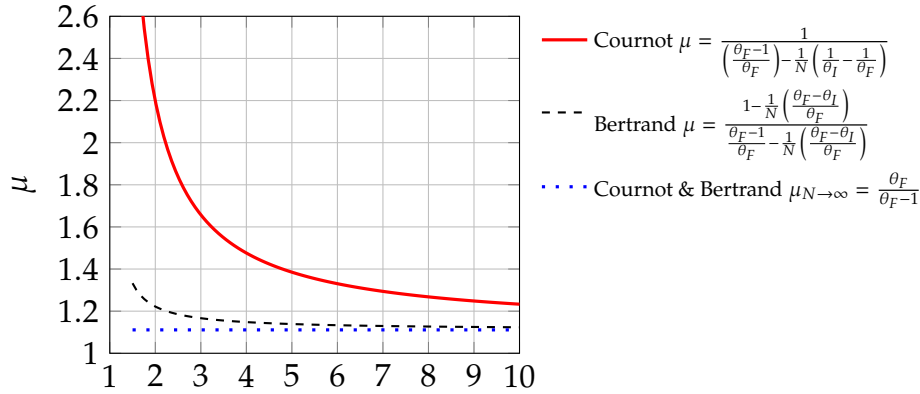Empirically Relevant Markups $\theta_I = 1.01,\ \theta_F = 10$
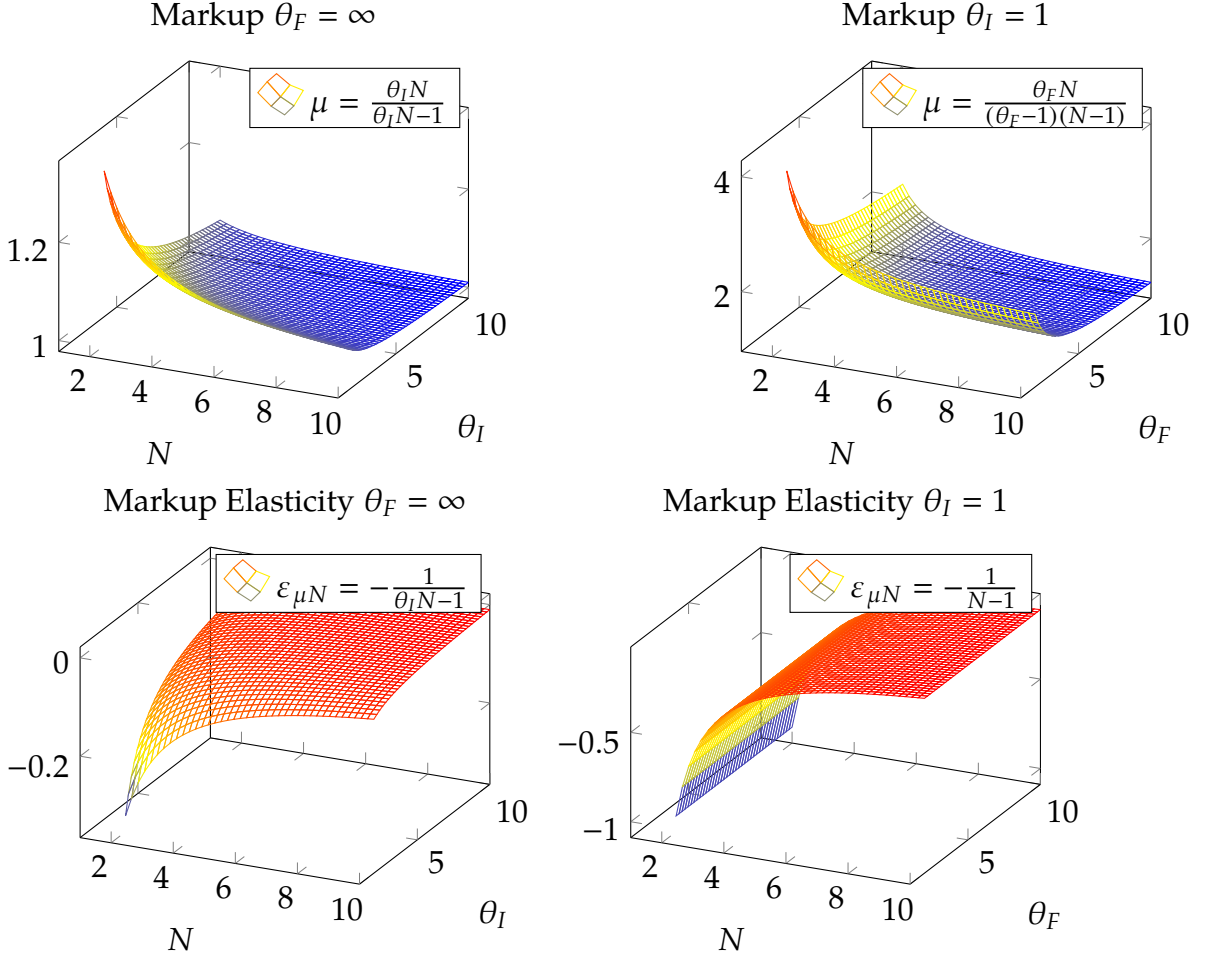

Figure 4: Markup Comparison

Figure 5: Markup Properties

C   Returns to Scale Derivations

There are two ways to identify returns to scale. The first is based on market structure, namely markups and profits, whereas the second is based on technical properties of the production function and its dual cost function, namely fixed costs and homogeneity parameter.

C.1   Market Structure Approach

First, from the profit definition, without imposing restrictions on the cost function.

$$\pi_{ji} = P_{ji} y_{ji} - C(r, w, y_{ji}) = (P_{ji} - AC_{ji}) y_{ji} \tag{112}$$

Divide by $P_{ji} y_{ji}$ and multiply AC_{ji} by MC_{ji} / MC_{ji} to express as a markup

$$s_\pi = 1 - \frac{AC_{ji}}{P_{ji}} = 1 - \frac{AC_{ji}}{MC_{ji}} \frac{MC_{ji}}{P_{ji}} \tag{113}$$

$$\text{RTS}_{Ji} = \mu_{Ji}\left(1 - \frac{\pi_{Ji}}{P_{Ji}y_{Ji}}\right) = \mu_{Ji}\left(1 - s_\pi\right), \quad \text{where,} \; s_\pi \equiv \frac{\pi_{Ji}}{P_{Ji}y_{Ji}} \tag{114}$$

The profit share s_π is the share of operating profits in sales revenue.

C.2 Production Function / Cost Function Approach

Second, from the cost function that occurs under cost-minimizing factor prices (input demands). The Lagrange cost minimization problem yields factor demands in nominal terms as $w = \lambda A F_\ell$ and $r = \lambda A F_k$, where the Lagrange multiplier is the MC, plugging these into the cost function $C = rk + w\ell$ gives:[49]

$$C_{Ji} = \text{MC}_{Ji}\, v(y_{Ji} + \phi) \tag{115}$$

Divide by $y_{Ji} \, \text{MC}_{Ji}$

$$\text{RTS}_{Ji} = v(1 + s_\phi), \quad \text{where,} \; s_\phi = \frac{\phi}{y_{Ji}} \tag{116}$$

Remember the fixed cost is output denominated, so prices cancel $s_\phi = \frac{P_{Ji}\phi}{P_{Ji}y_{Ji}} = \frac{\phi}{y_{Ji}}$. We could also obtain this by dividing the explicit expressions for AC and MC, (81) and (79), that we obtain from the cost function rewritten in unit cost form.

D   Model Summary

Table 3 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.

---

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

Static	
(34) Markup (Cournot)	$\mu(N) = \left[1 - \frac{1}{\theta_F} - \frac{1}{N}\left(\frac{1}{\theta_I} - \frac{1}{\theta_F}\right)\right]^{-1}$
(5) Labor Supply	$w = -\frac{u_L(L)}{u_C(C)}$
(37) Labor demand	$w = \frac{AF_\ell(k,\ell)}{\mu(N)}$
(36) Capital Rental	$r = \frac{AF_k(k,\ell)}{\mu(N)}$
(11) Output per Firm	$y = AF(k,\ell) - \phi$
(39) Operating Profit per Firm	$\pi = (y + \phi)\left(1 - \frac{v}{\mu(N)}\right) - \phi$
(20) Aggregate Labour	$L = N\ell$
(19) Aggregate Capital	$K = Nk$
(29) Aggregate Output	$Y = Ny$
(42) TFP	$\text{TFP} = \frac{y}{F(k,\ell)^{\frac{1}{v}}}$
(49) Entry Cost	$q = \zeta E$
(48) Firm Value (free entry)	$V = q$
(55) Entry Investment	$Z = \frac{\zeta}{2}E^2$
(61) Aggregate Accounting	$Y = C + I + Z$
Dynamic	
(4) Consumption Euler	$\dot{C} = -\frac{u_C(C(t))}{u_{CC}(C(t))}(r(t) - \rho)$
(50) Entry arbitrage	$\dot{V} = r(t)V(t) - \pi(t)$
(3) Capital Investment	$\dot{K} = I(t)$
(51) Entry Definition	$\dot{N} = E(t)$
Boundary Values	
(7) Capital Initial Condition	$K(0) = K_0$
(53) Firms Initial Condition	$N(0) = N_0$
(6) Capital Transversality	$\lim_{t\to\infty} e^{-\rho t} u_C(C(t))K(t) = 0$
(54) Firms Transversality	$\lim_{t\to\infty} e^{-\rho t} u_C(C(t))q(t)N(t) = 0$

Table 3: Model Summary

In the dynamical system we typically use the *capital accumulation* (60) $\dot{K} = Y(t) - \frac{\zeta}{2}E(t)^2 - C(t)$ and *entry arbitrage* (52) $\dot{E} = r(t)E(t) - \frac{\pi(t)}{\zeta}$ representations, implying the state of the dynamical system is C, E, K, N space.

E Labor Responses

Labor market equilibrium occurs where labor demand (14) equals labor supply (5), and it allows us to understand endogenous labor behaviour.[50]

---

[50]Related literature (Devereux, Head, and Lapham 1996b; Jaimovich 2007) terms (14) the aggregate labor demand function.

Proposition 4. *In general equilibrium the response of aggregate labor is negative to consumption and positive to capital and entry*

$$\frac{\partial L}{\partial K} = \Phi^{-1}\frac{F_{LK}}{F_L} \qquad\qquad\qquad > 0 \qquad\qquad (117)$$

$$\frac{\partial L}{\partial C} = \Phi^{-1}\frac{u_{CC}}{u_C} \qquad\qquad\qquad < 0 \qquad\qquad (118)$$

$$\frac{\partial L}{\partial N} = \Phi^{-1}\frac{1}{N}\left(1 - v - \varepsilon_{\mu N}\right) \qquad > 0 \qquad\qquad (119)$$

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 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

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* $1 - v > 0$ *and countercyclical markups* $\varepsilon_{\mu N} < 0$ *augment the labor response to entry. With constant marginal costs* $v = 1$ *and fixed markups* $\varepsilon_{\mu N} = 0$, *entry does not affect labor* $\frac{dL}{dN} = 0.$

The first effect is from increasing marginal costs ($1 - v > 0$), so that as entry divides inputs (labor and capital) across more units, the marginal product of labor increases $\text{MPL} = AN^{1-v}F_L(K,L)$, 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.[51]

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

$$\frac{dL}{dA} = \frac{\partial L}{\partial A} + \frac{\partial L}{\partial C}\frac{dC}{dA} + \frac{\partial L}{\partial K}\frac{dK}{dA} + \frac{\partial L}{\partial N}\frac{dN}{dA} \gtreqqless 0, \quad \text{where } \frac{\partial L}{\partial A} = \Phi^{-1}\frac{1}{A} > 0 \qquad (120)$$

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

---

[51]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.

effect influence the impact labor response to technology.[52] 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.[53]

E.1 Labor Responses Proof

Proof of Proposition 4. 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

$$\Xi(L, C, K, N, A) \equiv u_L(L) + u_C(C)w(L, K, N) = 0 \tag{5}$$

$$w(L, K, N) = \frac{A}{\mu(N)}N^{1-\nu}F_L(K, L) \tag{14}$$

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

$$0 = \frac{\partial \Xi}{\partial L}\frac{dL}{d\varpi} + \frac{\partial \Xi}{\partial C}\frac{dC}{d\varpi} + \frac{\partial \Xi}{\partial K}\frac{dK}{d\varpi} + \frac{\partial \Xi}{\partial N}\frac{dN}{d\varpi} + \frac{\partial \Xi}{\partial \varpi} \tag{121}$$

$$\frac{dL}{d\varpi} = -\left(\frac{\partial \Xi}{\partial L}\right)^{-1}\left[\frac{\partial \Xi}{\partial C}\frac{dC}{d\varpi} + \frac{\partial \Xi}{\partial K}\frac{dK}{d\varpi} + \frac{\partial \Xi}{\partial N}\frac{dN}{d\varpi} + \frac{\partial \Xi}{\partial \varpi}\right] \tag{122}$$

$$\frac{dL}{d\varpi} = \frac{\partial L}{\partial C}\frac{dC}{d\varpi} + \frac{\partial L}{\partial K}\frac{dK}{d\varpi} + \frac{\partial L}{\partial N}\frac{dN}{d\varpi} + \frac{\partial L}{\partial \varpi} \tag{123}$$

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

$$\frac{\partial L}{\partial C} = -\frac{\frac{\partial \Xi}{\partial C}}{\frac{\partial \Xi}{\partial L}} = \Phi^{-1}\frac{u_{CC}}{u_C} \qquad\qquad = \frac{dL}{dC} \tag{124}$$

$$\frac{\partial L}{\partial K} = -\frac{\frac{\partial \Xi}{\partial K}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial K} \qquad\qquad = \frac{dL}{dK} \tag{125}$$

---

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

[53]With logarithmic consumption utility, the income and substitution effects equate in the long run so that long-run labor supply is irresponsive to technology.

$$\frac{\partial L}{\partial N} = -\frac{\frac{\partial \Xi}{\partial N}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial N} \qquad\qquad = \frac{dL}{dN} \tag{126}$$

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

$$\frac{\partial \Xi}{\partial L} = u_L \Phi \qquad\qquad < 0 \tag{127}$$

$$\frac{\partial w}{\partial C} = 0 \qquad\qquad = 0 \tag{128}$$

$$\frac{\partial w}{\partial L} = w\frac{F_{LL}}{F_L} \qquad\qquad < 0 \tag{129}$$

$$\frac{\partial w}{\partial K} = w\frac{F_{LK}}{F_L} \qquad\qquad > 0 \tag{130}$$

$$\frac{\partial w}{\partial N} = \frac{w}{N}\left(1 - v - \varepsilon_{\mu N}\right) \qquad\qquad > 0 \tag{131}$$

$\square$

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 $\omega = A$, causes all endogenous variables to respond in addition to its partial effect from holding C, K, N constant.

$$\frac{dL}{dA} = \frac{\partial L}{\partial C}\frac{dC}{dA} + \frac{\partial L}{\partial K}\frac{dK}{dA} + \frac{\partial L}{\partial N}\frac{dN}{dA} + \frac{\partial L}{\partial A} = \frac{dL}{dC}\frac{dC}{dA} + \frac{dL}{dK}\frac{dK}{dA} + \frac{dL}{dN}\frac{dN}{dA} + \frac{\partial L}{\partial A} \tag{132}$$

where

$$\frac{\partial L}{\partial A} = -\frac{\frac{\partial \Xi}{\partial A}}{\frac{\partial \Xi}{\partial L}} = \frac{1}{w}\Phi^{-1}\frac{\partial w}{\partial A} \qquad\qquad \neq \frac{dL}{dA} \tag{133}$$

$$\frac{\partial w}{\partial A} = \frac{w}{A} \qquad\qquad > 0 \tag{134}$$

F   Reduced-form Aggregate Production Function

Our aim is to explain aggregate output as a function of inputs K, L. The derivation uses $Y = Ny$, $y = AF(k, \ell) - \phi$ and $\pi = \left(1 - \frac{v}{\mu}\right)(y + \phi) - \phi$.

$$y = N^{-v}AF(K, L) - \phi \tag{135}$$

$$\text{Use,} \quad y + \phi = (\pi + \phi)\left(1 - \frac{v}{\mu}\right)^{-1} \tag{136}$$

$$(\pi + \phi)\left(1 - \frac{v}{\mu}\right)^{-1} = N^{-v}AF(K, L) \tag{137}$$

$$\text{Use,} \quad N = \frac{Y}{y} \tag{138}$$

$$Y^\nu = y^\nu AF(K,L)(\pi+\phi)^{-1}\left(1-\frac{\nu}{\mu}\right) \tag{139}$$

$$\text{Use,} \quad y = \left(\pi+\frac{\nu}{\mu}\phi\right)\left(1-\frac{\nu}{\mu}\right)^{-1} \tag{140}$$

$$Y = \left(\frac{A}{\pi+\phi}\right)^{\frac{1}{\nu}}\left(1-\frac{\nu}{\mu}\right)^{\frac{1}{\nu}-1}\left(\pi+\frac{\nu}{\mu}\phi\right)F(K,L)^{\frac{1}{\nu}} \tag{141}$$

G   Dynamics and existence

The model consists of a four-dimensional dynamical system in $\mathbf{X} = [C, E, K, N]^\intercal$. It is a first-order system of ordinary differential equations in nonlinear form $\dot{\mathbf{X}} = g(\mathbf{X})$ given by:

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

$$\dot{E} = r(t)E(t) - \frac{\pi(t)}{\zeta}$$

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

$$\dot{N} = E(t)$$

Linearization of the nonlinear form gives $\dot{X} \approx \mathbf{J}(\tilde{\mathbf{X}})(\mathbf{X}(t)-\tilde{\mathbf{X}})$.[54] $\mathbf{J}(\tilde{C}, \tilde{E}, \tilde{K}, \tilde{N})$ is the Jacobian matrix. Each element of the Jacobian matrix is a respective partial derivative evaluated at steady state.[55]

$$\begin{bmatrix} \dot{C} \\ \dot{E} \\ \dot{K} \\ \dot{N} \end{bmatrix} \approx \begin{bmatrix} -\frac{u_C}{u_{CC}}r_C & 0 & -\frac{u_C}{u_{CC}}r_K & -\frac{u_C}{u_{CC}}r_N \\ -\frac{\pi_C}{\zeta} & \rho & -\frac{\pi_K}{\zeta} & -\frac{\pi_N}{\zeta} \\ Y_C-1 & 0 & Y_K & Y_N \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} C(t)-\tilde{C} \\ E(t) \\ K(t)-\tilde{K} \\ N(t)-\tilde{N} \end{bmatrix} \tag{142}$$

This *matrix differential equation* has open-loop solution in terms of initial values K_0, N_0; steady states $\tilde{C}, \tilde{E}, \tilde{K}, \tilde{N}$; time t and the exogenous parameters of the model given by

---

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

[55]The derivatives are partial since they 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 $\{t, [K_0 - \tilde{K}], [N_0-\tilde{N}]\}$ and structural model parameters 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.

(Caputo [2005], p. 494-495):

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

and the time derivative, which allows us to substitute out fluctuations $\hat{X} = \dot{X}/\tilde{X}$ in our log-linearized analysis, is

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

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 eigenvalues $\lambda_1 < \lambda_2 < 0$. The eigenvectors and eigenvalues consist of model parameters and are derived from the Jacobian matrix. 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

$$C(t) = \tilde{C} + \frac{(\hat{K} - v_{3,2}\hat{N})v_{1,1}e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})v_{1,2}e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{143}$$

$$E(t) = \frac{(\hat{K} - v_{3,2}\hat{N})\lambda_1 e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})\lambda_2 e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{144}$$

$$K(t) = \tilde{K} + \frac{(\hat{K} - v_{3,2}\hat{N})v_{3,1}e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})v_{3,2}e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{145}$$

$$N(t) = \tilde{N} + \frac{(\hat{K} - v_{3,2}\hat{N})e^{\lambda_1 t} + (v_{3,1}\hat{N} - \hat{K})e^{\lambda_2 t}}{v_{3,1} - v_{3,2}} \tag{146}$$

The right-hand side only depends on t and exogenously given parameters, which in the Cobb-Douglas production with isoelatic utility cases means they are a function of:

$$(t; \ K_0, N_0, \tilde{K}, \tilde{N}, , \tilde{C}, \tilde{E}, \zeta, \phi, \nu, \theta_I, \theta_F, \eta, \sigma, \xi, \rho)$$

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 perfect 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 $t = 0$.

G.1 Eigenvalues

To compute the eigenvalues of the system, we solve for the roots of the characteristic polynomial corresponding to the Jacobian:

$$c(\lambda) = \det(\mathbf{J} - \lambda\mathbf{I}) = \lambda^4 - M_1\lambda^3 + M_2\lambda^2 - M_3\lambda + M_4 \tag{147}$$

where coefficients M_k denote the sum of principal minors of dimension k, therefore $M_1 = \text{Tr}(\mathbf{J})$ and $M_4 = \det(\mathbf{J})$. The trace is unambiguously positive, whereas endogenous markups ($\varepsilon_{\mu N} < 0$) make the determinant ambiguous (with exogenous markups the determinant is strictly positive).[56] Since $\text{Tr}(\mathbf{J}) > 0$ and $\det(\mathbf{J}) > 0$, there 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.[57]

Characteristic Polynomal

The supplementary appendix provides an extended version of this section. Below we show that $M_1 = \text{Tr}(\mathbf{J}) > 0$ and $M_4 = \det(\mathbf{J}) > 0$, thus the sequence of signs of coefficients of the characteristic polynomial is

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

Leaving two coefficients unspecified, there is a minimum of two sign changes and a maximum of four sign changes. Hence, by Descartes' rule of signs (for positive roots), there are either four, two or zero positive roots.[58] 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 $-, -$. The four solutions of this quartic polynomial are saddle stable if there are two positive (unstable) and two

---

[56] I provide necessary conditions for a positive determinant and this also ensures steady-state existence. The Jacobian determinant of the dynamical system is equivalent to the determinant of the Jacobian of the steady-state conditions (62) - (65). Hence the conditions for it to be positive ensure steady-state is well-defined.

[57] A supplementary appendix provides full derivation of dynamic solutions.

[58] 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.

negative (stable). We denote these 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.

Jacobian Trace

The trace of the Jacobian matrix is

$$\mathrm{Tr}(\mathbf{J}) = -\Phi^{-1}\frac{F_{KL}}{F_K}\rho + \rho + \left(1 + \Phi^{-1}\frac{F_{KL}}{F_K}\right)\rho\tilde{\mu} \tag{148}$$

$$= \rho\left(1 + \tilde{\mu} + (\tilde{\mu} - 1)\Phi^{-1}\right) \tag{149}$$

$$= \rho(1 + \tilde{\mu})\left(1 + \left(\frac{\tilde{\mu} - 1}{\tilde{\mu} + 1}\right)\Phi^{-1}\right) > 0 \tag{150}$$

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

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.

$$\det \mathbf{J} = \frac{\tilde{N}}{\varepsilon_{u_C C}}\frac{\phi}{\zeta}\frac{\nu}{\mu}\left\{\Theta + \Gamma\varepsilon_{\mu_N N}\right\} \tag{151}$$

where

$$\Theta = \left(\frac{L}{\varepsilon_{u_L L} - \varepsilon_{F_L L}}\right)\left(\frac{\rho}{N}\right)^2\mu(\varepsilon_{u_C C} - \varepsilon_{u_L L})\left(\frac{F_{KL}}{F_L} - \frac{F_{KK}}{F_K}\right)\frac{F_L}{F_K} < 0 \tag{152}$$

$$\Gamma = y\left(1 + \frac{y}{\phi}\right)\Phi^{-2}\left(\frac{\rho}{N}\right)^2\mu$$

$$\left(\left[\Phi + \frac{F_{LK}^2}{F_L F_{KK}}\right]\frac{F_{KK}F_L}{F_K^2}\left\{\frac{\varepsilon_{u_C C}}{y} - \frac{\Phi N}{\rho\mu}\frac{F_K}{F_L} + 1\right\} - \left(\Phi + \frac{F_{LK}}{F_K}\right)^2\right) \gtreqless 0 \tag{153}$$

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

$$\det \mathbf{J} \gtreqless 0 \iff \Theta + \Gamma\varepsilon_{\mu_N N} \lesseqgtr 0 \tag{154}$$

Hence this provides a necessary condition for a positive determinant

$$\Theta + \Gamma \varepsilon_{\mu_N N} < 0 \tag{155}$$

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

$$\Gamma \geq 0 \tag{156}$$

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

$$\Gamma \varepsilon_{\mu_N N} < -\Theta \tag{157}$$

$$\varepsilon_{\mu_N N} > -\frac{\Theta}{\Gamma} \tag{158}$$

Hence the negative markup effect must be sufficiently small.

G.2 Eigenvectors

To calculate the four eigenvectors corresponding to the eigenvalues λ_j, $j \in \{1,2,3,4\}$ solve $(\mathbf{J} - \lambda_j \mathbf{I})\mathbf{V}_j = 0$ for $\mathbf{V}_j$. By definition, the eigenvalues are chosen such that $\det(\mathbf{J} - \lambda_j \mathbf{I}) = 0$, and a zero determinant implies matrix $\mathbf{J} - \lambda_j \mathbf{I}$ is completely linearly dependent (perfectly coupled). Then the eigenvectors are unique only up to a scalar multiple. Hence choose $v_{4,j} = 1$ as the normalization. The result is normalized eigenvectors. Then by multiplying out $(\mathbf{J} - \lambda_j \mathbf{I})[v_{1,j} \ v_{2,j} \ v_{3,j} \ 1]^{\mathsf{T}} = 0$, it is immediate from row four that $v_{2,j} = \lambda_j$. With $v_{2,j} = \lambda_j$, $v_{4,j} = 1$, we get from row 1 and 3

$$v_{1,j} = \frac{1}{\frac{C}{\sigma} r_C - \lambda_j} \left[\frac{-C}{\sigma} r_n - \frac{C}{\sigma} r_K v_{3,j} \right] \tag{159}$$

$$v_{1,j} = \frac{1}{Y_C - 1} \left[-Y_n - (Y_K - \lambda_j) v_{3,j} \right] \tag{160}$$

Equating and solving

$$v_{3,j} = \frac{\frac{C}{\sigma} r_n (Y_C - 1) - Y_n (\frac{C}{\sigma} r_C - \lambda_j)}{(\frac{C}{\sigma} r_C - \lambda_j)(Y_K - \lambda_j) - \frac{C}{\sigma} r_K (Y_C - 1)} \tag{161}$$

Plug back in

$$v_{1,j} = \frac{\frac{C}{\sigma} (r_K Y_n - r_n (Y_K - \lambda_j))}{(\frac{C}{\sigma} r_C - \lambda_j)(Y_K - \lambda_j) - \frac{C}{\sigma} r_K (Y_C - 1)} \tag{162}$$

So the four normalized eigenvectors are

$$
\begin{bmatrix} v_{1,j} \\ v_{2,j} \\ v_{3,j} \\ v_{4,j} \end{bmatrix} = \begin{bmatrix} \frac{\frac{C}{\sigma}(r_K Y_n - r_n(Y_K - \lambda_j))}{(\frac{C}{\sigma} r_C - \lambda_j)(Y_K - \lambda_j) - \frac{C}{\sigma} r_K(Y_C - 1)} \\ \lambda_j \\ \frac{\frac{C}{\sigma} r_n(Y_C - 1) - Y_n(\frac{C}{\sigma} r_C - \lambda_j)}{(\frac{C}{\sigma} r_C - \lambda_j)(Y_K - \lambda_j) - \frac{C}{\sigma} r_K(Y_C - 1)} \\ 1 \end{bmatrix}, \quad j \in \{1, 2, 3, 4\} \tag{163}
$$

In general derivative terms these results follow Brito and Dixon 2013 where there is perfect competition, and Savagar and Dixon 2017 where there is (constant markup) imperfect competition. However, due to the endogenous markup derivatives with respect to N will change and steady-states that derivatives are evaluated around will also change.

H Parametric Model and Steady State

We begin with isoelastic utility

$$
U(C, L) = \frac{C^{1-\sigma} - 1}{1 - \sigma} - \xi \frac{L^{1+\eta}}{1 + \eta} \tag{164}
$$

The partial derivatives are

$$
U_C = C^{-\sigma}, \quad U_{CC} = -\sigma \frac{U_C}{C}, \quad U_L = -\xi L^\eta, \quad U_{LL} = \eta \frac{U_L}{L} \tag{165}
$$

and Cobb-Douglas production

$$
F(k, \ell) = k^\alpha \ell^\beta = K^\alpha L^\beta N^{-(\alpha+\beta)} = F(K, L) N^{-(\alpha+\beta)} \tag{166}
$$

so derivatives are

$$
F_k = \alpha k^{\alpha-1} \ell^\beta = \alpha K^{\alpha-1} L^\beta N^{1-(\alpha+\beta)}, \tag{167}
$$
$$
F_\ell = k^\alpha \beta \ell^{\beta-1} = K^\alpha \beta L^{\beta-1} N^{1-(\alpha+\beta)} \tag{168}
$$

H.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.

$$
L = \left(\frac{A K^\alpha \beta N^{1-(\alpha+\beta)}}{\mu(N) \xi C^\sigma} \right)^{\frac{1}{1+\eta-\beta}} \tag{169}
$$

Under this parameterization the dynamical system is

$$\dot{C} = \frac{C}{\sigma}\left[\frac{1}{\mu(N)}A\alpha K^{\alpha-1}L^{\beta}N^{1-(\alpha+\beta)} - \rho\right] \tag{170}$$

$$\dot{E} = \frac{1}{\mu(N)}A\alpha K^{\alpha-1}L^{\beta}N^{1-(\alpha+\beta)}E$$

$$- \frac{1}{\zeta}\left(AK^{\alpha}L^{\beta}N^{-(\alpha+\beta)}\left(1 - \frac{v}{\mu(N)}\right) - \phi\right) \tag{171}$$

$$\dot{K} = N\left[AK^{\alpha}L^{\beta}N^{-(\alpha+\beta)} - \phi\right] - \frac{\zeta}{2}E^2 - C \tag{172}$$

$$\dot{N} = E \tag{173}$$

H.2   Parameteric Steady State

The intratemporal condition is

$$\tilde{C} = \left(\frac{\beta A\tilde{k}^{\alpha}\tilde{\ell}^{\beta-1-\eta}\tilde{N}^{\eta}}{\mu(\tilde{N})\xi}\right)^{\frac{1}{\sigma}} \tag{174}$$

The dynamical system is

$$\tilde{C} = \frac{\phi v}{\mu(\tilde{N}) - v}\tilde{N} \tag{175}$$

$$\tilde{E} = 0 \tag{176}$$

$$\alpha\tilde{k}^{\alpha-1}\ell(\tilde{k}, \tilde{N})^{\beta} = \frac{\mu(\tilde{N})\rho}{A} \tag{177}$$

$$\tilde{k}^{\alpha}\tilde{\ell}(\tilde{k}, \tilde{N})^{\beta} = \frac{\phi}{A\left(1 - \frac{v}{\mu(\tilde{N})}\right)} \tag{178}$$

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

$$\tilde{\ell}(\tilde{k}, \tilde{N}) = \left(\frac{A\tilde{k}^{\alpha}\beta\tilde{N}^{-\sigma-\eta}}{\mu(\tilde{N})\xi\left(\frac{\phi v}{\mu(\tilde{N})-v}\right)^{\sigma}}\right)^{\frac{1}{1+\eta-\beta}} \tag{179}$$

Solving (177) and (178) gives $\tilde{k}$ and $\tilde{\ell}$

$$\tilde{k}(\tilde{N}) = \frac{\phi\alpha}{(\mu(\tilde{N}) - v)\rho} \tag{180}$$

$$\tilde{\ell}(\tilde{N}) = \left[\frac{\mu(\tilde{N})}{A}\left(\frac{\rho}{\alpha}\right)^{\alpha}\left(\frac{\phi}{\mu(\tilde{N}) - v}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} \tag{181}$$

43

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

$$\frac{\partial \tilde{\ell}}{\partial \tilde{\mu}} = \frac{\tilde{\ell}}{\tilde{\mu}(\tilde{\mu} - \nu)}\left(\frac{\alpha}{\beta}(\tilde{\mu} - 1) - 1\right) \gtrless 0 \iff \frac{\alpha}{\beta}(\tilde{\mu} - 1) \gtrless 1 \tag{182}$$

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 (179) gives a nonlinear function in terms of $\tilde{N}$.

$$\tilde{N} = \left(\left(\frac{\alpha}{\rho \nu}\right)^\sigma \frac{A\beta}{\mu(\tilde{N})\xi \tilde{k}^{\sigma - \alpha}\tilde{\ell}^{1 + \eta - \beta}}\right)^{\frac{1}{\sigma + \eta}} \tag{183}$$

where $\tilde{k}, \tilde{\ell}$ are defined in (180) and (181) 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 (175, 180, 181). Equation (184) presents equation (183) in $\tilde{N}$ terms only.

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

$$\left.\left.\left(\frac{1 - \frac{\nu}{\mu(\tilde{N})}}{\phi}\right)^{1 - \nu + \eta(1 - \alpha) + \sigma\beta}\right\}^{\frac{1}{\beta}}\right]^{\frac{1}{\eta + \sigma}} \tag{184}$$

H.3   Parametric Labor Responses

$$\frac{dL}{dC} = -\frac{\sigma}{1 + \eta - \beta}\frac{L}{C} \qquad\qquad < 0 \tag{185}$$

$$\frac{dL}{dK} = \frac{\alpha}{1 + \eta - \beta}\frac{L}{K} \qquad\qquad > 0 \tag{186}$$

$$\frac{dL}{dN} = \frac{1 - \nu - \varepsilon_{\mu N}}{1 + \eta - \beta}\frac{L}{N} \qquad\qquad > 0 \tag{187}$$

$$\frac{\partial L}{\partial A} = \frac{1}{1 + \eta - \beta}\frac{L}{A} \qquad\qquad > 0 \tag{188}$$

and the total effect of technology on labor is

$$\frac{dL}{dA} = \left(\frac{L}{1 + \eta - \beta}\right)\left(-\frac{\sigma}{C}\frac{dC}{dA} + \frac{\alpha}{K}\frac{dK}{dA} + \frac{1 - \nu - \varepsilon_{\mu N}}{N}\frac{dN}{dA} + \frac{1}{A}\right) \tag{189}$$

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.

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.

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

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

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

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.

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.

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.

Kim, Jinill (2004). "What determines aggregate returns to scale?" In: *Journal of Economic Dynamics and Control* 28.8, pp. 1577–1594.

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.

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.

Silberberg, E. (1990). *The structure of economics: a mathematical analysis*. McGraw-Hill Pub. Co.

Syverson, Chad (2019). "Macroeconomics and Market Power: Facts, Potential Explanations and Open Questions". In: *Brookings Economic Studies Report*.