Durham University Business School

Department of Economics & Finance

Monetary Policy and Welfare with Heterogeneous Firms and Endogenous Entry

Dudley Cooke & Tatiana Damjanovic

Working Paper No. 2, 2021

Durham Business School working papers series ISSN 1749-3641

Department Economics and Finance
Durham University Business School
Mill Hill Lane
Durham DH1 3LB, UK
Tel: +44 (0)191 3345200
https://www.dur.ac.uk/business/research/economics/

Monetary Policy and Welfare with Heterogeneous Firms and Endogenous Entry*

Dudley Cooke†

University of Exeter

Tatiana Damjanovic‡

Durham University

Abstract: This paper studies the welfare consequences of monetary policy in a sticky-wage New Keynesian model with heterogenous firms and endogenous entry. Cross-sectional dispersion in price-markups and labor shares is generated by a translog demand structure and aggregate fluctuations in these variables are driven by firm entry and selection. We show that when the distribution of firm-level productivity is Pareto, selection is such that the aggregate price-markup and labor share are fixed. If firm entry is static, the divine coincidence appears, and wage stability is optimal. If firm entry is dynamic, or selection is weakened, optimal stabilization policy accounts for the size distribution of firms. We calculate the welfare loss of ignoring firm entry and selection to be $0.1 - 0.3$ percent of steady state consumption.

JEL Classification: E32, E52, L11

Keywords: Firm Entry, Heterogenous Firms, Optimal Monetary Policy, Translog Preferences

---

*We thank Hugo Hopenhayn, Omar Licandro, Julian Neira, and Rish Singhania for comments and suggestions. Different versions of this paper were presented at the 2019 CEF (Ottawa) and the 2019 MMF (London).

†Department of Economics, University of Exeter, Streatham Court, Rennes Drive, Exeter EX4 4PU, United Kingdom. Email: d.cooke@exeter.ac.uk

‡Department of Economics, Durham University, Mill Hill Lane, Durham DH1 3LB, United Kingdom. Email: tatiana.damjanovic@durham.ac.uk.

1. Introduction

Recently, Autor *et al.* (2020) have documented that a reallocation of sales towards highly productive firms contributed to the falling aggregate U.S. labor share. Similarly, de Loecker *et al.* (2020) have argued that a reallocation of market share from low to high markup firms has resulted in a rising aggregate U.S. price-markup. The aggregate labor share and price-markup play a prominent role in the analysis of monetary policy, and ultimately, they influence how policy should be used for the purposes of stabilization.[1] In this paper, we consider the link between the reallocation of resources across heterogenous firms and optimal monetary stabilization policy.[2] In particular, we analyze welfare in a sticky-wage New Keynesian model, where productivity, labor shares, and price-markups are heterogenous, and aggregate fluctuations in these variables are driven by endogenous firm entry and selection.[3]

In the model we study, more productive firms have a greater market share, a lower labor share, and charge higher markups because consumer preferences are characterized by a translog expenditure function.[4] As the minimum level of productivity required to remain in the market - an endogenous variable, which depends on macroeconomic conditions - rises, labor shares rise and price-markups fall across firms. How these variables change in the aggregate rests on two forces. First, they are influenced by firm entry, which is based on an evaluation of expected profits, against the payment of a one-time entry cost. Second,

[1]Cantore *et al.* (2020) and Nekarda and Ramey (2020) both discuss the role of the labor share and price-markup in New Keynesian models, the latter emphasizing the role of wage rigidities.

[2]Evidence that monetary policy generates a reallocation of resources at business cycle frequencies is presented in Hamano and Zanetti (2020). Fasani *et al.* (2020) provide evidence on the link between monetary policy uncertainty shocks and firm exit.

[3]Our analysis also rationalizes an idea, made popular, for example, by Haldane (2018), that firm-level productivity (heterogeneity) and industry concentration matter for monetary policy decisions.

[4]We assume the goods market is characterized by monopolistic competition. The use of translog preferences is discussed in more detail below.

they depend on selection (firm exit), whereby a fraction of firms that enter do not generate ex-post profits, because they are not sufficiently productive to do so. The distribution of productivity controls the strength of selection, and determines how resources are reallocated across firms, influencing the welfare effects of monetary policy, in a sticky-wage New Keynesian model.

We reach a number of conclusions on the extent to which monetary policy should be concerned with the reallocation of resources across heterogeneous firms in the short-run. Firm profits, which are the driver of entry and exit decisions, are of overriding importance. We demonstrate this analytically by assuming the firm entry decision is static with instantaneous zero-profit.[5] Doing so leads to an equivalence result. The textbook New Keynesian model, with sticky-prices, generates the divine coincidence (Blanchard and Gali, 2007) - whereby it is possible to stabilize both price inflation and the welfare-relevant output gap. This idea carries over to a model with sticky-wages and flexible-prices where the focus naturally falls on wage inflation and employment. We show that if the aggregate price-markup is rendered fixed by the distribution of firm-level productivity, the divine coincidence appears, and firm entry is not a relevant concern for the policymaker.

To explain our result, consider a positive innovation to technology, which generates a rise in expected profit, and encourages firm entry. Firm entry occurs alongside a rise in the productivity cut-off, which reflects a selection effect. A higher cut-off makes it harder to generate positive profit and price-markups (labor shares) of incumbent firms fall (rise). At the aggregate level, when the distribution of productivity is Pareto, there is no movement in the price-markup or labor share. When productivity is Pareto, the lowest productivity firm has zero net markup, whereas the highest productivity firm has an infinite markup. If the distribution of markups and labor shares within this range are also Pareto their average

[5] In general, we model firm entry as forward-looking, with fluctuating profits, and include a time-to-build lag, following Bilbiie *et al.* (2012).

values are constant. Thus, despite a change in the composition of firms, which raises average productivity, and magnifies the effect of the shock, there are no implications for optimal monetary stabilization policy.

We identify how firm entry and exit influence optimal policy, for all other productivity distributions, using the difference between the constrained-efficient level of employment and its flexible-wage counterpart. In general, optimal policy requires lowering the target value for wage inflation in the current period, below its long-run value, when the growth rate of the welfare relevant employment gap is positive, a result which we link to selection. For a given change in aggregate technology, the weaker is selection, the greater is the change in the aggregate labor share, and the stronger is the required drop in current wage inflation. Changes in the labor share thus represent a weakening of the strong reallocation effects which occur with a Pareto distribution. Since our translog-Pareto specification is equivalent to a constant elasticity of substitution specification, which features fixed firm-level and aggregate markups, our analysis demonstrates why accounting for the distribution of markups is important for understanding the welfare implications of monetary policy.[6]

In a next step, we numerically analyze the full model, which features a forward-looking entry decision, and where the volatility of firm entry is calibrated to match U.S. data. We allow for a log-normal distribution of idiosyncratic shocks - in addition to considering the case when productivity is Pareto distributed - and aggregate intertemporal preference shocks.[7] For the same long-run exit rate and labor share across the two productivity distributions, the log-normal distribution leads to a lower level of concentration among firms. Lower concentration is consistent with a weaker selection effect. Aggregate intertemporal preference shocks play a role in our quantitative analysis because we consider dynamic entry. Changes in

[6]With a constant elasticity of substitution specification, although it is also necessary to assume a fixed cost of production to generate firm exit (selection), the key point is that the aggregate markup is fixed.

[7]Using a log-normal distribution allows for many smaller firms which are typically observed in the data.

intertemporal preferences act directly on the discount factor and affect the value of firm creation, thus, similar to a model with physical capital, a positive innovation to preferences, generates a rise current consumption, at the expense of current investment (at the extensive margin).

We study the welfare consequences of monetary policy by transposing the optimal targeting rule from a model with a given mass of homogenous firms into our model with firm entry; in effect, asking how costly it is to run an otherwise-optimal policy, but ignore firm dynamics. Ignoring firm entry and exit generates a $0.1 - 0.3$ percent loss of steady state consumption when the aggregate price-markup and labor share are rendered fixed by selection.[8] When the aggregate markup is countercyclical, the business cycle is smoother, and the welfare loss is reduced. Thus, selection leads to differences in the business cycle, which translate into welfare losses from adopting a policy that ignores firm dynamics. In general, the presence of long-lived firms means the divine coincidence cannot hold, in contrast with static entry, where the divine coincidence appears when the markup is fixed. This is an important point, because the difference between the static and dynamic versions of our model is that, in the latter, firm entry is a form of investment in which up-front costs incurred to start a business generate expected future profits.

The normative analysis we present owes a considerable debt to research that focuses on the short-run implications for monetary policy when there is endogenous firm entry. For example, Bergin and Corsetti (2008) and Bilbiie *et al.* (2014) analyze the short-run, and in the latter case, the short- and long-run, implications of firm entry for monetary policy when there are sticky-prices. Our analysis differs from theirs in that we focus on the reallocation of resources across heterogenous firms. We also derive an explicit quadratic approximation for welfare which allows for analytical results. To the extent that our model generates a link

[8] We also consider the welfare loss from adopting a standard interest rate setting rule, in which case, we find a slightly higher loss.

between monetary policy and productivity, our results are also related to, for example, Moran and Queralto (2018), who focus on monetary policy shocks and medium-run movements in TFP, and Ikeda and Kurozumi (2019), who focus on slowdowns in TFP after the 2008-09 financial crisis, and analyze welfare maximizing monetary policy rules.[9]

Our analysis addresses issues that are often ignored when studying monetary policy. Of particular importance is the extent to which the reallocation of resources across firms affects agents' welfare. Firm entry and exit in our analysis is akin to product turnover - see Broda and Weinstein (2010) - and the role of product turnover for the design of monetary policy in the long-run, i.e., the optimal inflation target, is considered in Adam and Weber (2019). Abstracting from endogenous entry, they suggest that sticky-price models featuring heterogeneous firms and systematic firm-level productivity trends imply long-run inflation should be set at between 1 and 3 percent. In a related analysis, Miyakawa *et al.* (2020) provide empirical evidence on the relationship between inflation and the firm size distribution. They develop an endogenous growth model with firm heterogeneity and nominal rigidity to analyze the reallocation effects of monetary policy on welfare.

Finally, we make use of translog preferences, which enables us to understand the relationship between firm-level and aggregate labor shares and price-markups, the subject of recent empirical scrutiny. Translog preferences have been used successfully to explain real and monetary-induced business cycle fluctuations, both in the closed and open economy. For example, absent firm entry, Bergin and Feenstra (2000) use a translog expenditure function to generate persistence arising from monetary shocks when there is nominal price rigidity. Bilbiie *et al.* (2012) show that such preferences, when combined with forward-looking firm entry, perform well in matching key business cycle moments, when there are technology shocks and flexible-prices. Lewis and Poilly (2012) estimate an endogenous-entry model with

[9]Both papers introduce nominal price rigidities to the expanding varieties framework of Comin and Gertler (2006).

translog preferences and nominal price and wage rigidity and identify a competition effect generated through counter-cyclical price-markups in the transmission of monetary shocks.[10] In the open economy, Rodríguez-López (2010) develops a model with endogenous entry, firm heterogeneity, and one-period nominal wage rigidity to study exchange rate pass-through.

The rest of the paper is as follows. In section 2, we develop a sticky-wage New Keynesian model with heterogenous firms and endogenous entry. We study the macroeconomic effects of movements in the productivity cut-off generated through selection. In section 3, we characterize the conditions under which our model produces the same welfare implications as when there is a given mass of homogenous firms. In section 4, we analyze the model numerically, and present a quantitative evaluation of the welfare loss from adopting policies that ignore firm entry and exit. A final section concludes.

2. Model Economy

In this section, we develop a general equilibrium model with heterogenous firms and sticky-wages. We focus on explaining the relationship between the productivity cut-off - the minimum level of productivity for a firm to successfully enter the market - and aggregate variables; in particular, the labor share and price-markup. In this regard, we provide an explicit link between the distribution of idiosyncratic productivity, the strength of firm selection (the decision to produce or exit the market), and the cyclical behavior of the aggregate labor share and price-markup.

2.1. The Model Economy

[10] Lewis and Poilly (2012) further argue that a model with strategic interaction between oligopolistic firms cannot generate an empirically relevant competition effect and is statistically equivalent to a CES specification. Lewis and Stevens (2015) undertake a Bayesian estimation of a medium-scale DSGE model with a tranlsog expenditure function and endogenous firm entry.

The economy is populated with a measure $n_t > 0$ of productive firms and a measure one of households. Each firm has a constant returns to scale technology and supplies a differentiated good. New firms are created each period by paying a sunk cost. Each household consumes a basket of goods and supplies a differentiated labor type. At the beginning of the period, new firms are created, and draw idiosyncratic productivity, $z \geq 0$, from a distribution, $G(z)$. Firms with a low level of productivity, $z \in [z_{\min}, z_t^\star)$, exit, and those with a sufficiently high level of productivity, $z \in [z_t^\star, z_{\max}]$, produce using labor, subject to the demand for their good. Households supply labor, subject to demand for their labor type, after choosing the nominal wage.[11] They receive dividends, net of tax labor-income, and a lump-sum transfer. At the end of the period a fraction of firms exit for exogenous reasons.

Households (Preferences) Each household has a symmetric translog expenditure function over a set of differentiated goods. The representative household has preferences,

$$\ln(e_t) = \ln c_t + \frac{1}{n_t}\int_{i\in\Delta} \ln p_t(i)\, di + \frac{1}{2n_t}\left\{1 + \int_{i\in\Delta}\int_{j\in\Delta} \ln p_t(i)\left[\ln p_t(j) - \ln p_t(i)\right] dj di\right\} \quad (1)$$

where e_t is the minimum expenditure required to obtain a composite of goods c_t and the term $p_t(i)$ is the price of good i.[12] The set of differentiated goods available to the household is denoted Δ, where n_t is the measure of Δ, and the term $\frac{1}{2n_t}$ captures a variety effect.

The demand for good i is,

$$c_t(i) = \left[\frac{s_t(i)}{\rho_t(i)}\right] y_t \quad (2)$$

where $s_t(i) \equiv \frac{1}{n_t}\left[1 + \int_{j\in\Delta} \ln p_t(j)\, dj\right] - \ln p_t(i)$ is the expenditure share, $\rho_t(i) \equiv p_t(i)/p_t$ is the price of good i relative to the consumer-based price index, denoted p_t, and y_t is total output, or GDP.[13]

[11] Monetary non-neutrality arises in our model because of sticky-wages. Evidence on the role of sticky-wages is presented in Olivei and Teneyro (2007, 2010) and in Nekarda and Ramey (2020).

[12] Translog preferences are consistent with a class of preferences introduced by Diewert (1976) known as the quadratic mean of order r (QMOR) expenditure function. Translog imposes $r = 0$.

[13] To derive the demand curve we apply Shephard's lemma to the expenditure function.

Firms Each successful entrant produces a differentiated good under conditions of monopolistic competition. Given $l_t(i)$ workers, each firm produces,

$$y_t(i) = a_t z_t(i) l_t(i) \tag{3}$$

where a_t is a technology common to all firms and $z_t(i)$ is firm-level productivity. Firm i maximizes profit, $\vartheta_t(i) = \rho_t(i) y_t(i) - w_t l_t(i)$, subject to the demand for their good and their production function.[14] The optimal price chosen by the firm with productivity level z is,

$$\rho_t(z) = \Omega_t \times \frac{w_t}{a_t z} \tag{4}$$

where $\Omega_t \equiv \Omega\left(\frac{z}{z_t^\star}\exp\right)$ measures the gross firm-level markup and the inverse labor share.[15] Equation (4) determines the minimum level of productivity, $z = z_t^\star$, with which a firm can enter the market successfully. Such a firm has zero net markup (i.e., $\rho_t^\star = w_t/a_t z_t^\star$) and a unit labor share.[16] The market share of a firm can also be expressed in terms of the markup, $s_t \equiv \Omega_t - 1$, and equation (4) therefore shows that the least productive firm has zero market share.

We first use equation (4) to aggregate over the price index, generating,

$$w_t = a_t z_t^\star e^{H_t/2} \tag{5}$$

where $H_t \equiv N_t \int_{z_t^\star}^{z_{\max}} s_t^2 dG(z)$ is the Herfindahl index - i.e., the sum of the squares of the market shares of firms, which is a measure of concentration - and N_t is the mass of

[14] There is no index on the wage as all firms will face the same per-unit wage cost.

[15] The term $\Omega\left(\frac{z}{z_t^\star}\exp\right)$ is also the Lambert-W function.

[16] This is different to the zero-profit cut-off under CES preferences, which requires a fixed operating cost. However, as we will go on to show, a comparison can be made between settings, which depends on the assumed distribution of firm-level productivity.

entrants.[17] We also use equation (4) to derive an expression for aggregate firm profits, $\vartheta_t \equiv \int_{z_t^\star}^{z_{\max}} \vartheta_t(z)\, dG(z)$, in terms of market share,

$$\vartheta_t = z_{1,t} y_t \quad ; \quad z_{1,t} \equiv \int_{z_t^\star}^{z_{\max}} \frac{s_t^2}{1+s_t} dG(z) \tag{6}$$

where the aggregator, $z_{1,t}$, is a function of the cut-off, $z_t^\star$. Finally, the sum of market shares of firms, which is given by $N_t \int_{z_t^\star}^{z_{\max}} s(z)\, dG(z) = 1$, can be written as,

$$N_t = \frac{1}{z_{2,t}} \quad ; \quad z_{2,t} \equiv \int_{z_t^\star}^{z_{\max}} s_t dG(z) \tag{7}$$

which implies the mass of entrants is uniquely related to the productivity-cutoff. Since the mass of goods available to the household, n_t, is equal to the mass of entrants, N_t, multiplied by the probability of successful entry, $\int_{z_t^\star}^{z_{\max}} dG(z)$, equation (7) shows how the mass of goods is also uniquely related to the productivity-cutoff.

A large number of ex-ante identical firms have the option of paying $f_t > 0$ units of output to enter the market. Each entrant obtains a productivity level $z \in [z_{\min}, z_{\max}]$ which is the realization of a random variable drawn independently across firms. Firm i enters if,

$$\sum_{t=0}^{\infty} (1-\delta)^{t+1} Q_{t,t+1} \vartheta_t > f_t \tag{8}$$

where $Q_{t,t+1}$ is a stochastic discount factor ($Q_{t,t} \equiv 1$) and $\delta < 1$ is an exogenous rate of firm destruction. Firms enter the market endogenously until the present discounted value of post-entry profits are zero, net of entry costs. The timing of entry and production is such that the total mass of firms during period t is,

$$N_t = (1-\delta)(N_{t-1} + N_{e,t-1}) \tag{9}$$

which allows for a time-to-build lag in entry.

[17] This condition replaces the more standard expression - in, for example, the case of Dixit-Stiglitz aggregation - that would imply a fixed markup over marginal costs. We provide a derivation in Appendix A1.

Households Each household maximizes expected discounted utility from the composite of goods, c_t, and disutility from its labor type, $l_t(j)$ according to,

$$\sum_{t=0}^{\infty} \beta^t d_t \left[\ln(c_t) - \psi \int_0^1 l_t(j)\, dj \right] \tag{10}$$

subject to the demand for it's labor type, $l_t(j) = [w_t(j)/w_t]^{-\varepsilon} L_t$, and the following flow budget constraint,

$$\begin{aligned} c_t + (N_t + N_{e,t})\, x_t f_t + \frac{b_t}{p_t} &= (1-\tau_L)[w_t(j)\, l_t(j)] + (\vartheta_t + f_t)\, N_t x_{t-1} \\ &\quad + \frac{(1+i_{t-1})\, b_{t-1}}{p_t} - \frac{\chi}{2}\left[\frac{w_t(j)}{w_{t-1}(j)}\pi_t - 1\right]^2 + T_t \end{aligned} \tag{11}$$

where β is the discount factor and d_t captures frictions that affect intertemporal preferences.[18] In equation (11), free entry is imposed, b_t are state contingent securities, τ_L (T_t) is a labor-income (lump-sum) tax (transfer), and $\frac{\chi}{2}\left[\frac{w_t(j)}{w_{t-1}(j)}\pi_t - 1\right]^2$ measures the real cost of wage adjustment, where $\pi_t \equiv p_t/p_{t-1}$ is the gross rate of price inflation.[19]

The household's first-order-condition with respect to bonds and equity implies,

$$1 = E_t\left[Q_{t+1}\frac{1+i_t}{\pi_{t+1}}\right] \quad \text{and} \quad f_t = E_t(1-\delta)Q_{t+1}(\vartheta_{t+1} + f_{t+1}) \tag{12}$$

where $Q_{t+1} = \beta\frac{d_{t+1}}{d_t}\left(\frac{c_{t+1}}{c_t}\right)^{-1}$. The household's first-order-condition with respect to their nominal wage can be expressed as a wage Phillips curve,

$$[\varepsilon\psi c_t + w_t(1-\tau_L)(1-\varepsilon)]\, L_t = \chi\left[(\pi_t^w - 1)\,\pi_t^w - E_t Q_{t+1}\left(\pi_{t+1}^w - 1\right)\pi_{t+1}^w\right] \tag{13}$$

where we impose a symmetric equilibrium in labor markets, such that, $l_t(j) = L_t$ and $w_t(j) = w_t$, for all j. The wage Phillips Curve is forward-looking, in that when setting their

[18]Our specification of utility is similar to assuming indivisible labor, as originally emphasized in Hansen (1985) and Rogerson (1988). The role of indivisibility is discussed very recently in Christiano *et al.* (2020).

[19]We model sticky-wages using a quadratic adjustment cost. Our choice is mostly for the purposes of exposition and a Calvo-wages version of our model produces almost identical results.

wage, suppliers of labor consider the future nominal wage they expect to charge. Without this element (as $\chi \to 0$), equation (13) implies $(1-\tau_L) w_t = \frac{\varepsilon}{\varepsilon-1}\psi c_t$, which relates the net of tax labor-income to a fixed markup, $\frac{\varepsilon}{\varepsilon-1}$, multiplied by the ratio of marginal disutility of labor to the marginal utility of consumption.

Equilibrium The government budget constraint is,

$$T_t = \tau_L p_t w_t L_t - \mathcal{G} \tag{14}$$

where $\tau_L w_t L_t$ is labor-tax income and $\mathcal{G} > 0$ is government expenditure. The resource constraint for the economy is,

$$y_t = c_t + f_t N_{e,t} + \mathcal{G} + \frac{\chi}{2}\left(\pi_t^w - 1\right)^2 = \vartheta_t N_t + w_t L_t \tag{15}$$

where the right-hand side of this condition is total income.[20] Finally, the evolution of the real wage is governed by,

$$\frac{w_t}{w_{t-1}} = \frac{\pi_t^w}{\pi_t} \tag{16}$$

which states that growth in real wages depends on the growth in nominal wages versus the growth in nominal prices.

An equilibrium in our model is a set of endogenous processes $\{y_t, c_t, L_t, w_t, \pi_t, \pi_t^w\}$ and $\{N_t, N_{e,t}, \vartheta_t, z_t^\star\}$ satisfying equations in (5)-(7), (9), (12) and (13), and (15) and (16), for given policy $i_t \geq 0$, with given government expenditure, $\mathcal{G} > 0$, labor-income tax, τ^L, and exogenous processes for a_t and d_t.

2.2. Selection and the Labor Share

In this section, we study the macroeconomic effects of movements in the productivity cut-off. We show how movements in the cut-off affect the aggregate labor share and price-markup.

[20] When the resource constraint is satisfied, the labor market clearing condition, $L_t = N_t \int_{z_t^\star}^{z_{\max}} l_t(z)\, dG(z)$, holds.

We also link the labor share and market concentration with total factor productivity. For further analysis, we will use Lemma 1 and Lemma 2, below, which enable us to establish the relationship between the productivity cut-off and these aggregate variables.

Lemma 1 (Aggregation) Let $G(z)$ be the distribution function of z, bounded by $z_{\max} \leq +\infty$. For any function, $J(z^\star) = \int_{z^\star}^{z_{\max}} j\left(\frac{z}{z^\star}\right) dG(z)$, where $j(1) \geq 0$ and $j' \geq 0$, then, (i), $J'(z^\star) < 0$, and (ii), $\lim\limits_{z^\star \to z_{\max}} J(z^\star) = 0$.

Proof See Appendix. ■

Lemma 1 shows that the aggregators, $z_{1,t}$ and $z_{2,t}$, defined in equations (6) and (7), are decreasing functions of $z_t^\star$. The immediate economic implication of this result is that the mass of entrants, N_t, and the productivity cut-off have a positive relationship. The intuition for this result is that, as more firms enter the market, competition raises the minimum productivity required to produce, which increases the probability of exit, $\int_{z_t^\star}^{z_{\max}} dG(z)$. To determine how changes in the cut-off influence the mass of operating firms, n_t, it is necessary to determine the strength of selection.

Lemma 2 (Selection) Make the change of variables, $u = z/z^\star$. Let $g(u)$ be a density function with elasticity, $\epsilon(u) = -\frac{ug'(u)}{g(u)}$, which is weakly increasing. Let $j_1(u)$ and $j_2(u)$ be positive functions such that $\frac{j_1(u)}{j_2(u)}$ is strictly increasing. The ratio $\frac{J_1(z^\star)}{J_2(z^\star)}$ is a decreasing function of $z^\star$, where $J_i(z^\star) = \int_{z^\star}^{z_{\max}} j_i(u) g(z) dz$, for $i = 1, 2$.

Proof See Appendix. ■

Recall that the mass of operating firms is $n_t = \frac{1}{z_{2,t}} \int_{z_t^\star}^{z_{\max}} dG(z)$. Using Lemma 2 we can determine that the mass of operating firms is increasing with the cut-off. Thus, if selection is relatively strong, a rise in the mass of entrants, which occurs alongside a relatively large fall in the probability of successful entry, results in the mass of operating firms being relatively insensitive to macroeconomic conditions. We can now also determine the aggregate price-markup. The average price of firms that successfully enter is, $\rho_t \equiv \frac{1}{1-G(z_t^\star)} \int_{z_t^\star}^{z_{\max}} \rho_t(z)\, dG(z)$, which, from the pricing equation implies the average markup is the inverse of the mass of operating firms. Thus, the strength of selection, and the underlying distribution of firm-level productivity, affect the cyclical behavior of the aggregate price-markup.

Finally, Lemma 2 allows us to understand how selection affects the inverse aggregate labor share, which, with firm heterogeneity, differs to the aggregate price-markup.[21] Consider the profit share in total output, which is, $\frac{\vartheta_t N_t}{y_t} = \frac{z_{1,t}}{z_{2,t}}$. Using the resource constraint, the inverse labor share of total output is,

$$\frac{y_t}{w_t L_t} - 1 = \frac{1}{\frac{z_{2,t}}{z_{1,t}} - 1} \tag{17}$$

Applying Lemma 2, the aggregate profit share and inverse labor share are increasing in the productivity cut-off. Thus, the profit share, inverse labor share, and price-markup move in the same direction, but they differ because firms differ in their market share. We can further unpack the labor share by recalling that the aggregated price index is $w_t = a_t z_t^\star e^{H_t/2}$. This condition shows that the wage depends on the productivity cut-off in two distinct ways: directly, because a higher cut-off is consistent with a higher real wage, and indirectly, because the real wage is increasing in the cut-off through the Herfindahl index (market concentration), $H_t \equiv \frac{1}{z_{2,t}} \int_{z_t^\star}^{z_{\max}} s_t^2 dG(z)$. We then determine aggregate output (and total factor productivity, TFP), as $y_t = \left(z_t^\star \frac{e^{H_t/2}}{1-\frac{z_{1,t}}{z_{2,t}}} \right) a_t L_t$. Thus, TFP, which only depends on the cut-off, can be linked back to firm entry, by Lemma 1; see equation (7), and changes in TFP, induced by entry and exit, can be mitigated by (weakening) selection, due to the presence of $e^{H_t/2} / \left(1 - \frac{z_{1,t}}{z_{2,t}}\right)$, which reflect market concentration and the labor share.

[21] As Edmond *et al.* (2019) discuss, the average inverse labor share and price-markup differ with heterogeneity because the latter uses a cost-based, arithmetic mean, whereas the former is defined by an output-weighted, harmonic mean.

3. Optimal Monetary Policy

In this section, we do two things. First, we provide a benchmark characterization of optimal policy with a given mass of homogenous firms. This generates a result that choosing zero wage inflation (wage stability) is optimal because it automatically closes the welfare-relevant employment gap. We then consider the policy problem when endogenous firm entry is static by imposing an instantaneous zero-profit condition. In this case, we derive an equivalence result, and characterize the conditions under which the introduction of heterogenous firms and endogenous entry delivers the same welfare implications as our benchmark case. We show that this result depends on selection because it requires a distribution of firm-level productivity which renders both the aggregate labor share and price-markup fixed.

3.1. Optimal Monetary Policy with a Given Mass of Homogenous Firms

To characterize optimal stabilization policy with a given mass of homogenous firms, we suppose $f_t \to 0$ and $N_t \to 1$, and introduce a price-markup, such that the inverse labor-share is equal to, $\frac{y_t}{w_t L_t} - 1 = \frac{\varepsilon_p}{\varepsilon_p - 1} > 1$. The Ramsey problem is one of choosing consumption, labor, and wage inflation, subject to the resource constraint and wage Phillips curve. The Ramsey policy problem is converted into a linear-quadratic problem as described in the following Proposition.[22]

Proposition 1 The linear-quadratic policy problem with a given mass of homogenous firms is,

$$\min_{\{\widehat{L}, \widehat{\pi}_t^w\}} \mathbb{L} = -\frac{1}{2s_c} E_0 \sum_{t=0}^{\infty} \beta^t \left[\frac{1}{s_c} \left(\widehat{L}_t - \widehat{L}_t^{\star} \right)^2 + \frac{\varepsilon_w}{\xi} \left(\widehat{\pi}_t^w \right)^2 \right] \tag{18}$$

[22]Our approach follows Benigno and Woodford (2005). For simplicity, in the text, we also further assume the labor-income tax, τ_L, generates an efficient steady-state. That is, $1 - \tau_L = \left(\frac{\varepsilon_w}{\varepsilon_w - 1}\right)\left(\frac{\varepsilon_p}{\varepsilon_p - 1}\right)$. In the Appendix we discuss the more general case.

subject to the linear wage Phillips curve,

$$\widehat{\pi}_t^w = \xi \left(\widehat{L}_t - \widehat{L}_t^n \right) - \beta \widehat{\pi}_{t+1}^w \quad ; \quad \widehat{L}_t^n = \widehat{L}_t^\star = (1 - s_c)\,\widehat{a}_t \tag{19}$$

where a circumflex denotes the deviation of a variable from it's Ramsey steady-state value and the variables $\widehat{L}_t^n$ and $\widehat{L}_t^\star$ are the flexible-wage and constrained-efficient levels of employment. The parameters $s_c \equiv \frac{c}{c+\mathcal{G}}$ and $\xi \equiv \frac{\varepsilon_w}{\chi}\frac{c}{s_c}$ are the share of government spending in total consumption and the slope of the wage Phillips curve.

Proof See Appendix. ■

There are a number of standard, and yet important, comments regarding the policy problem in Proposition 1. First, the loss function is derived under the assumption that the steady state is efficient. Second, when the steady-state is efficient, then $\widehat{L}_t^n = \widehat{L}_t^\star$, which eliminates policy trade-off between wage inflation and employment: by the Phillips curve, choosing $\widehat{\pi}_t^w = 0$ is consistent with $\widehat{L}_t = \widehat{L}_t^n$, which further implies $\mathbb{L} = 0$, a result commonly referred to as the divine-coincidence. As Blanchard and Gali (2007) show, the relevant statistic when evaluating welfare is the gap between the flexible-wage and constrained-efficient allocations (here, of employment). Finally, the divine-coincidence result we derive applies to preference (demand) and technology (supply) shocks.

In addition to delivering a simple result with regard to optimal stabilization policy, it is also straightforward to determine the response of output and price inflation under optimal policy. This is a relevant exercise for us to undertake because we want to emphasize the (potentially) time-varying aggregate labor share. The production function without entry is, $\widehat{y}_t = \widehat{a}_t + \widehat{L}_t$, which implies, $\widehat{L}_t - \widehat{L}_t^\star = \widehat{y}_t - \widehat{y}_t^\star$. The path of price inflation is governed by, $\widehat{\pi}_t = \widehat{\pi}_t^w - (\widehat{a}_t - \widehat{a}_{t-1})$, with $\widehat{\pi}_t^w = 0$ under optimal policy. In other words, a positive realization of technology will lead to no change in wage inflation and an immediate fall in price inflation. This is optimal because there are no distortions in the goods market. The drop in price inflation will take place alongside a rise in output, equal to $\widehat{y}_t^\star = s_c \widehat{a}_t$, where $s_c < 1$, which depend on extent of government spending in total consumption.

3.2. An Equivalence Result

In this section, we consider optimal stabilization policy when there is endogenous entry with instantaneous zero-profit. The assumption of static entry allows us to develop intuition for our results because we isolate how much the reallocation of resources across firms matters for the welfare effects of monetary policy. When entry decisions are forward-looking, the dynamic effects of entry act as a separate, independent distortion because, in effect, the entire path of profits can be taxed or subsidized by changes in policy.

With instantaneous zero-profit, we replace the general free entry condition, given by equation (8), with the following static relation,

$$\int_{z_t^\star}^{z_{\max}} \vartheta_t(z)\, dG(z) = f \tag{20}$$

where $f > 0$.[23] To determine optimal policy, we need to relate the benefit of consumption to the costs of supplying labor (employment). Firm entry affects this relationship in the following sense. Define effective output as total output, y_t, less the costs of entry, $fN_{e,t}$. With static entry, since profit is equal to the cost of entry, period-by-period, $\vartheta_t N_t = fN_{e,t}$. Effective output is therefore,

$$F(a_t L_t) \equiv y_t - fN_{e,t} = (a_t L_t)\, z_t^\star e^{H_t/2} \tag{21}$$

We use the function $F(a_t L_t)$ to measure how effective output responds to changes in employment.[24] Since the productivity cut-off is itself a function of $a_t L_t$, the elasticity of effective

[23]The resource constraint is replaced with, $y_t = c_t + fN_t + \mathcal{G} + \frac{\eta}{2}(\pi_t^w - 1)^2$. Other than this change the model is as specified in section 2.1.

[24]It is possible to relate $F(a_t L_t)$ to the inverse labor share, since, $\frac{y_t}{w_t L_t} - 1 = \frac{fN_{e,t}}{F(a_t L_t)}$, which reflects the ratio of investment, at the extensive margin, to effective output.

output with respect to inputs depends on the elasticity of the cut-off to changes in employment. This elasticity features in our explanation as to why the distribution of productivity affects the relative weight attached to employment when evaluating welfare. Using the definition of effective output in equation (21), we present the main analytical result of this section in the following Proposition.

Proposition 2 With heterogenous firms, endogenous entry, and an instantaneous zero-profit condition, at an efficient steady-state, there is no policy trade-off between wage inflation and employment, if the following condition holds,

$$-\frac{LF_{LL}}{F_L} = 1 - \frac{LF_L}{F} \tag{22}$$

where F_L (F_{LL}) is the first (second) derivative of $F(L)$.

Proof See Appendix. ■

Proposition 2 characterizes the conditions under which the introduction of heterogenous firms and endogenous entry delivers the same welfare implications as a standard sticky-wage New Keynesian model. To understand how this condition arises, consider the constrained-efficient and flexible-wage levels of employment,

$$\widehat{L}_t^{\star} \equiv \frac{\eta+\alpha}{1-(\eta+\alpha)}\widehat{a}_t \quad ; \quad \eta \equiv \left(\frac{L}{\frac{LF_L}{F}}\frac{d}{dL}\right)\frac{LF_L}{F} \tag{23}$$

and,

$$\widehat{L}_t^{n} \equiv \frac{\alpha}{1-\alpha}\widehat{a}_t \quad ; \quad \alpha \equiv \left(1-\frac{1}{s_c}\right)\frac{LF_L}{F} \tag{24}$$

respectively. The parameter α (η) is the elasticity (of the elasticity) of the effective output function. In general, with entry and exit decisions, we have $\frac{LF_L}{F} > 1$, and any shock that raises employment, generates firm entry, which magnifies the effect of the shock. With firm heterogeneity, changes in the productivity cut-off, which reflect both entry and exit (selection), further magnify the effect of shocks.[25] Selection also matters in determining how the constrained-efficient level of output responds. However, it is clear that if $\eta = 0$, the constrained-efficient and flexible-wage levels of employment coincide. Put differently, when $\widehat{L}_t^n - \widehat{L}_t^\star = \left(\frac{1}{1-\alpha}\frac{\eta}{\alpha+\eta-1}\right)\widehat{a}_t$ is equal to zero, magnification through entry and selection is efficient, and the divine coincidence appears. As we will go on to show, only in this case, are the aggregate price-markup and labor share fixed.

Before further developing the intuition for our results, we note that the solution to the linear-quadratic policy problem under commitment is,

$$\widehat{\pi}_t^w = -\frac{1}{\varepsilon_w}\frac{1-(\eta+\alpha)}{1-\alpha}\left[\Delta\left(\widehat{L}_t - \widehat{L}_t^\star\right)\right] \tag{25}$$

Optimal policy suggests lowering the target value for wage inflation in the current period below its long-run value when the growth rate of the welfare-relevant employment gap is positive. Again, what is of interest in equation (25) is the presence of $\eta \neq 0$. It reflects the distortion created by variation in the price-markup and labor share induced by entry and exit. The greater the change in the aggregate labor share resulting from a movement in aggregate technology, the greater is the required drop in wage inflation. This is because, changes in the labor share reflect a weakening of the reallocation of resources across firms. The parameter $\eta \neq 0$ therefore captures the welfare implications of selection.

3.3. Special Case: Pareto Distribution and a Fixed Labor Share

In this section, we consider a special case in which both the aggregate labor share and price-markup are fixed due to the strength of selection. We assume that firm-level productivity is drawn from an unbounded Pareto distribution, specified as,

$$G(z) = 1 - z^{-\kappa} \tag{26}$$

[25] This type of magnification effect is well-established with homogenous firms, see, Devereux *et al.* (1996) and Bilbiie *et al.* (2012). The difference, with respect to our analysis, is that the strength of magnification also operates through the distribution of idiosyncratic productivity.

In this case, the solution for effective output is,

$$F(a_tL_t) = \omega \times (a_tL_t)^{\kappa/(\kappa-1)} \tag{27}$$

where $\omega \equiv \left[f\left(\frac{1}{z_{p,2}} - \frac{1}{z_{p,1}}\right)\right]^{-1/(\kappa-1)} e^{(H_p/2\zeta)^{\kappa/(\kappa-1)}}$ is a positive constant.[26] Equation (27) has two implications. First, the effective output function exhibits a type of labor-augmenting technology, where $\kappa/(\kappa-1) > 1$. In other words, any positive movement in labor will translate into a more than one-for-one increase in effective output. Since, in this case, the labor share is fixed, it must be that entry costs rise proportionally with effective output, where the mass of new entrants is given by, $N_t = (z_t^\star)^k / z_{1,p}$.[27] Moreover, any increase in the cut-off raises total factor productivity, with total output given by, $y_t = \left(\frac{e^{H_p/2}}{1-\frac{z_{p,1}}{z_{p,2}}}\right) z_t^\star a_t L_t$.

The second implication of equation (27) is that $\eta = 0$. We can see immediately that $\frac{LF_L}{F} = 1 + \frac{1}{\kappa-1} > 1$ and $\frac{LF_{LL}}{F_L} = \frac{1}{\kappa-1} > 0$ are consistent with the restriction in Proposition 2. This generates $\widehat{L}_t^n = \widehat{L}_t^\star$. As we discuss above, $\frac{LF_L}{F} > 1$ (magnification) is the result of firm entry and selection. To understand the role of $\frac{LF_{LL}}{F_L}$ we can appeal to Lemma 2, which shows why the unbounded Pareto distribution is a special case. Whereas the majority of aggregators move with productivity cut-off, in the unbounded Pareto case, they are fixed. Moreover, for an unbounded Pareto distribution, the elasticity of the density function, $\epsilon(u)$, is a constant, equal to one plus the shape parameter, κ.[28]

Finally, we impose $s_c = 1$, which simplifies our analysis further, and implies $\widehat{L}_t^n = \widehat{L}_t^\star = 0$ and $\widehat{F}_t^n = \widehat{F}_t^\star = \frac{\kappa}{\kappa-1}\widehat{a}_t$. This tells us that the extent of magnification is governed by the shape parameter of the productivity distribution. Since, in this particular setting, optimal policy targets the flexible-wage level of employment, we can determine the impact on price

[26]The term ω is a function of $z_{p,1}$, $z_{p,2}$, and H_p , which are all themselves constants. For example, define $u \equiv \frac{z}{z_t^\star}$. Then, $z_{1,t} \equiv (z_t^\star)^{-k} \int_1 [\Omega_t(\exp u) - 1] ku^{-k-1}du = (z_t^\star)^{-k} z_{p,1}$. The same applies for $z_{p,2}$ and H_p.

[27]As before, with a fixed labor share, there will be no change in the mass of operating firms, and $n_p = 1/z_{1,p}$.

[28]Feenstra (2018) contains a discussion on the importance of bounded and unbounded Pareto distributions and our paper offers a natural extension of his analysis using the elasticity of the density function.

inflation and total output, as we did above, when the mass of firms was fixed. In particular, we now find, $y_t = \left(\frac{\kappa}{\kappa-1}\right) \widehat{a}_t$ and $\widehat{\pi}_t = -\left(\frac{\kappa}{\kappa-1}\right)(\widehat{a}_t - \widehat{a}_{t-1})$, such that optimal policy requires a steeper drop in price inflation to offset the greater expansion in total output generated through firm entry and exit.

4. Numerical Analysis

In this section, we undertake a quantitative analysis of the model with dynamic entry and evaluate the welfare implications of monetary policy. In particular, we calibrate the model - using a standard, historical interest rate rule - to match the volatility of establishment entry at the business cycle frequency, in addition to standard long-run targets, and compute the welfare loss from adopting stabilization policies that abstract from firm entry and exit.

4.1. Functional Forms

Idiosyncratic shocks are drawn from either a log-normal or a Pareto distribution. The density function of the log-normal distribution is,

$$g(z) = \frac{1}{z\sqrt{2\pi}} \exp\left[-\frac{(\ln z - \mu)^2}{2\sigma^2}\right] \tag{28}$$

where $\mu = -\frac{\sigma^2}{2}$ and σ are the location and scale parameters. The density function of the Pareto distribution is $g(z) = \kappa z^{-(1+\kappa)}$. As we show above, these functional forms determine the extent to which selection operates. Selection determines the cyclical properties of a number of objects in our model, the most important being the labor share of income and price-markup.

When monetary policy is set exogenously, we assume it is consistent with a standard interest rate setting rule, which takes the form,

$$1 + i_t = (1 + i_{t-1})^{\rho_i} \left[\frac{1}{\beta}\pi_t^{\phi_\pi} \left(\frac{y_t}{y}\right)^{\phi_y}\right]^{1-\rho_i} \tag{29}$$

where ρ_i is an interest rate smoothing parameter and ϕ_π (ϕ_y) is the weight attached to inflation (output).

Finally, we specify a cost function for entry (congestion) as,

$$f_t = f\left(\frac{N_{e,t}}{N_{e,t-1}}\right)^{\gamma} \tag{30}$$

By choosing a value for γ, given aggregate shocks to technology and preferences, we can pin-down the volatility of firm entry.[29]

4.2. Model Calibration

Our calibration proceeds in two stages. First, we choose parameter values which match long-run targets and values for fiscal instruments. Table 1 presents the parameters used to characterize the steady-state and their respective targets.

===== Table 1 Here =====

We calculate the real return over the post-1979 period as the average FEDFUNDS rate minus the rate of CPI inflation. This implies a real return of 1.44 percent and we set β to match this statistic. We assume government spending constitutes 17 percent of GDP and the average labor-income tax is 20 percent. The annualized rate of firm exit, using data from the Bureau of Labor Statistics (BLS), over the period 1993:Q2-2018:Q4, for which data is available, is just below 12 percent. Finally, we assume the wage-markup is 30 percent

[29]A similar specification to equation (30) is used in Bergin and Corsetti (2019). In Loualiche (2019), the entry cost is convex in the entry rate, $N_{e,t}/N_{t-1}$, which is akin to the specification of capital adjustment costs adopted in RBC models.

and the average time to reset the wage is five quarters (Gali and Monacelli, 2016), which we map into the Rotemberg wage adjustment cost parameter.[30]

Next, we choose the labor share, $\frac{wL}{y}$ and the endogenous failure rate of firms, $\int_0^{z_t^\star} dG(z)$. The post-1979 US labor share is 60 percent (BLS). We note that studies such as Bilbiie *et al.* (2012), who also focus on the business cycle implications of firm entry, assume an aggregate price-markup roughly consistent figure.[31] With a Pareto distribution, the labor share is independent of the failure rate, which we assume to be 15 percent, and we set $\kappa = 2.5$. With a log-normal distribution, there are many more small firms, and with the same failure rate, we set $\sigma = 0.423$ to match the labor share. Our choice over the labor share determines the costs of firm entry (investment) as a fraction of GDP to be around 25 percent. To compare between distributions, notice, with a Pareto distribution, it is possible to map the failure rate into the productivity premium explicitly. Define $z^h \equiv \int_{z^\star}^{\infty} \frac{z}{1-G(z^\star)} dG(z)$ and $z^l \equiv \int_1^{z^\star} \frac{z}{G(z^\star)} dG(z)$ as the two relevant productivity averages. The productivity premium is $\left(\frac{z^h}{z^l} - 1\right) \times 100$, where $\frac{z^h}{z^l} = z^\star \frac{1-(z^\star)^{-\kappa}}{1-(z^\star)^{1-\kappa}}$ and $(z^\star)^{-\kappa} = 1 - G(z^\star)$. Given our calibration, the productivity premium is 72.3 percent, which further implies a (non-targeted) Herfindahl index of 0.47.[32] With a log-normal distribution, the productivity premium (Herfindahl index) rises (falls) to 126.7 percent (0.42).

In the second step, we consider parameters that affect dynamics. First, we assume that

[30]A log-linearized Calvo wage-setting model implies a difference equation for wage inflation of the form $\widehat{\pi}_t^w = \frac{(1-\phi)(1-\beta\phi)}{\phi} \left(\widehat{L}_t - \widehat{L}_t^n\right) + \beta \widehat{\pi}_{t+1}^w$, where ϕ is the constant probability that a household must keep its wage unchanged in any given period. The Rotemberg adjustment cost model we used gives, $\widehat{\pi}_t^w = \xi(1-\alpha)\left(\widehat{L}_t - \widehat{L}_t^n\right) + \beta \widehat{\pi}_{t+1}^w$, where $\xi \equiv \left(\frac{\varepsilon_w}{\chi} \frac{c}{s_c}\right) \frac{\alpha}{1-\frac{1}{s_c}}$. Equivalence imposes $\chi = \varepsilon \psi L c \frac{\phi}{(1-\phi)(1-\beta\phi)}$. Under our parameterization, the Calvo probability is set at $\phi = 0.8$.

[31]We choose to target the labor share because the difference between the aggregate labor share and price-markup in our model depends on weighting, as we discuss above.

[32]The figures we report on concentration are indicative and the problems associated with measuring industry concentration are discussed in Syverson (2019).

aggregate technology (a) and intertemporal preferences (d) follow independent AR(1) processes,

$$\lambda_t = \Lambda \lambda_{t-1} + u_t \quad ; \quad u_t \sim N(0, I) \tag{31}$$

where $\lambda_t = [\ln(a_t), \ln(d_t)]^T$ and $u_t = \left[u_t^a, u_t^d\right]^T$ is the vector of shocks. Since it is not our intention to estimate the contribution of these shocks to the economy, we take standard values from the literature. Consistent with the RBC literature, we set the persistence of technology to 0.979 and the standard deviation of innovations to technology at 0.72 percent. For preference shocks, we set persistence at 0.68 and the standard deviation at 3 percent, values which are reported in Justiniano *et al.* (2013). Likewise, we choose values for the policy rule that reflect standard a parameterization; specifically, $\{\rho_R, \phi_\pi, \phi_y\} = \{0.75, 1.5, 0.125\}$.[33]

We simulate our model and match the volatility of firm entry with U.S. data from the BLS. Over the 1993:Q2-2018:Q4 period, the standard deviation of firm entry was 7.47 percent.[34] In all cases, we choose γ such that the our model matches this statistic, and leave other moments non-targeted. Table 2 reports key business cycle statistics generated by our model for technology and preference shocks for both productivity distributions (log-normal and Pareto).

===== Table 2 Here =====

The statistics produced by the model are, broadly speaking, in line with U.S. data. However, for both shocks, GDP is a little too volatile (a point we discuss below in more detail), so

[33] The inflation and output weights are 1.5 and $0.125 = 0.5/4$ respectively, following Taylor (1993). The value for the coefficient on the lagged interest rate is 0.75, consistent with the estimates of Clarida *et al.* (2000). These values are also the mean values of the prior distributions chosen by Smets and Wouters (2007).

[34] All standard deviations reported in the paper are HP-filtered with smoothing parameter set at $1,600$.

that establishment entry, relative to GDP, is somewhat less volatile than in the data. The volatility of consumption and employment differ, conditional on the shock, with technology (preference) shocks generating to little (much) volatility. Again, it is worth restating, that in this exercise we only aim to target the standard deviation of firm entry, as this is the focus of our analysis.

4.2. Impulse Responses

In this section, we use impulse responses to provide intuition for the model we have developed. Figure 1 presents the impulse responses of selected variables to a one-time shock to technology and preferences, of 1 percent, with a log-normal distribution for idiosyncratic productivity shocks.

===== **Figure 1 Here** =====

We start with impulse responses to the technology shock (bold-blue line). Output (GDP), consumption, and firm entry all rise on impact. The response of entry is mitigated by the extent of congestion (the parameter $\gamma > 0$) and this is reflected in the increase in the mass of operating firms. At this point, we make a comparison with Bilbiie *et al.* (2012), who consider a similarly parameterized technology shock, in a model with dynamic firm entry and translog preferences (see Figure 2, page 326). Such a comparison is meaningful because, absent heterogeneity, our model is the sticky-wage version of their model.[35] Absent congestion, with fixed entry costs, our model generates a relatively larger rise in entry and output. That is because, whilst a model with homogenous firms also generates magnification, such that, an exogenous rise in technology, boosts output by virtue of increased firm entry,

[35]In Bilbiie *et al.* (2012), entry costs are in specified in normalised units of labor. A positive change in technology leads to rise in wages and a relatively small rise in entry costs.

in our model, these firms are also, on average, more productive. That is, in addition to nominal rigidities, firm heterogeneity (and selection) is a source of business cycle volatility.

Now consider the intertemporal preference shock (dashed-red line). Since this shock is equivalent to a decrease in the rate of time preference, consumption is immediately shifted forward, which generates a strong but short-lived rise in consumption, and contributes to raise output. However, the same force acts to reduce firm entry, which is also forward-looking, and this has a countervailing effect on output. The reason entry falls is straightforward. A reduction in the rate of time preference, all else equal, is equivalent to a reduction in the value of investing in firm creation - see equation (8). Quantitatively, the drop in firm creation leads to a relatively shallow fall in the mass of operating firms, and the net result of the change in consumption and firm entry is a rise in output initially, then, in the transition back to the steady-state, a period of low output, because the fall in firm entry is relatively persistent.

4.3. Welfare Analysis

In this section, we consider the welfare implications of alternative monetary policies. Welfare loss is computed in terms of consumption units that households would be willing to forgo to continue under Ramsey-optimal policy. The alternative policies we consider are, (i), the interest rate rule used to calibrate the model - equation (29) - and, (ii), the Ramsey-optimal policy from a sticky-wage model without entry and exit. We focus primarily on the latter policy because we want to gauge the welfare implications of abstracting from firm dynamics.[36]

Table 3 presents the welfare losses of the alternative monetary policies and standard deviations of key macro-variables, under Ramsey-optimal policy, for both productivity distribu-

[36] Although we consider the Ramsey-optimal policy generated from a model with a given mass of homogenous firms, this is very similar to a zero wage inflation policy because, absent other shocks, or frictions, there is very little welfare loss from running a zero wage inflation (wage stability) policy in the standard model.

tions.

===== Table 3 Here =====

The first result of Table 3 is that welfare losses associated with ignoring entry and exit are higher when productivity is Pareto distributed. This result is surprising, because, analytically, we showed that when the aggregate labor share is fixed, the welfare implications of monetary policy are the same as when there is a given mass of homogenous firms. We argued this was the result of a strong selection effect and the reallocation of resources across firms. However, in deriving that result, we assumed entry was static, and imposed an instantaneous zero-profit condition. Thus, our quantitative results emphasize that there are two forces at work: namely, a selection effect (via a constant labor share or price-markup) and an entry effect (via profits). The difference between static and dynamic entry is simply that profits fluctuate over time in the latter (see the impulse response in Figure 1) and prospective entrants evaluate the entire path of future profits when making their entry decision. Thus, fluctuations in entry generated through future profits significantly outweigh those associated with shutting down movements in the labor share, which we associate with a reallocation effect.

We also appeal to the analysis of Bilibiie *et al.* (2014) to provide an alternative interpretation of our results. They report welfare gains for translog preferences (using a sticky-price model with homogenous firms) that are generally lower than when preferences are of a constant elasticity of substitution (CES) form. In their model, the difference between translog and CES preferences is that the former leads to a time-varying price-markup, which is the same for all firms. We also compare cases in which the price-markup (or labor share) can move, but in our case, this is not due to a change in consumer preferences; rather, the distribution of idiosyncratic productivity differs. With heterogenous firms, the reallocation of resources is strongest when firm-level productivity is drawn from a Pareto distribution. As we argued above, translog-Pareto, is very much like a CES specification, where fixed costs are required to generate selection upon entry.

A second result we observe in Table 3 is that welfare losses are around four times greater when there are preference shocks as opposed to technology shocks, at 0.08 and 0.32 percent, respectively, for the case of the Pareto distribution.[37] We explain the differences in the following way. Under Ramsey-optimal policy, firm entry is relatively more volatile than GDP when there are preference shocks (σ_{N_e}/σ_y is around 50 percent higher with preference shocks than with technology), whereas consumption is less volatile: $\sigma_c/\sigma_y = 0.39$ for preference shocks versus $\sigma_c/\sigma_y = 0.69$ for technology shocks. This is the opposite pattern to what we observe when entry and exit is ignored by policy; either because there is an otherwise-optimal policy, that abstracts from firm dynamics, or there is an interest rate setting rule, as specified in equation (29). The Ramsey planner, in attempting to suppress volatility in consumption, does so at the cost of additional volatility in firm entry, which translates into a reallocation of resources across firms. Despite this additional reallocation, the reduction in consumption volatility that occurs under Ramsey policy, outweighs the added cost of volatility in firm entry.

5. Conclusion

This paper studies the welfare consequences of monetary policy with heterogenous firms and endogenous entry. We characterize the conditions under which the introduction of these two features delivers the same welfare implications as a sticky-wage New Keynesian model with a given mass of homogenous firms. First, equivalence requires that firm-level productivity be drawn from a Pareto distribution in which case the aggregate labor share and price-markup

[37]For all shock specifications, we re-calibrate congestion, captured by the parameter $\gamma > 0$, such that the volatility of firm entry is identical, under the historical rule.

are fixed. This mechanism works through selection. Second, equivalence requires static firm entry with an instantaneous zero-profit condition. When either the aggregate labor share and price-markup are time-varying, or prospective entrants are forward-looking, and entry is dynamic, the size distribution of firms is a relevant concern for the policymaker.

Appendix A

Here we present proof of Propositions and details omitted from the main text.

Appendix A.1 (Derivation of Resource Constraint and Price Index)

Here we describe the aggregate conditions we use to solve the model. From the main text, the demand curve is, $c_t(i) = [s_t(i)/\rho_t(i)]\, y_t$, and firm-level profits are, $\vartheta_t(i) = [\rho_t(i)\, y_t(i) - w_t]\, l_t(i) = s_t(z)\left(1 - \frac{1}{\Omega_t}\right) y_t$, where $s_t(z) = \Omega_t - 1$ is market share. Given these conditions, aggregate profit is, $y_t \int_{z_t^\star}^{z_{\max}} (\Omega_t - 1)\left(1 - \frac{1}{\Omega_t}\right) dG(z)$. The sum of market shares is unity, so, $\int_{z_t^\star}^{z_{\max}} s(z)\, dG(z) = 1 \iff N_t \int_{z_t^\star}^{z_{\max}} (\Omega_t - 1)\, dG(z)$. Finally, labor market clearing is, $L_t = N_t \int_{z_t^\star}^{z_{\max}} l_t(z)\, dG(z)$. Since firm-level revenue is, $\rho_t(z)\, y_t(z) = \rho_t(z)\, a_t z l_t(z) = s_t(z)\, y_t$, and prices are, $\rho_t(z) = \Omega_t \frac{w_t}{a_t z}$, we express the wage bill in terms of output, $w_t l_t(z) = \frac{\Omega_t - 1}{\Omega_t} y_t$, and, $w_t L_t = y_t N_t \int_{z_t^\star}^{z_{\max}} \frac{\Omega_t - 1}{\Omega_t} dG(z)$. Using this condition in aggregate profits,

$$f_t N_t = N_t y_t \int_{z_t^\star}^{z_{\max}} (\Omega_t - 1)\, dG(z) - N_t y_t \int_{z_t^\star}^{z_{\max}} \frac{\Omega_t - 1}{\Omega_t} dG(z) \Rightarrow w_t L_t = y_t - f_t N_t$$

which we use as the labor market clearing expression. Finally, since $w_t L_t = c_t + \mathcal{G}$, from the aggregate household budget constraint, goods market clearing is,

$$y_t = c_t + f_t N_{e,t} + \mathcal{G} + \frac{\chi}{2}\left(\pi_t^w - 1\right)^2 \tag{32}$$

which is presented as equation (15) in the text.

The price index is given by,

$$\ln P_t = \frac{1}{2n_t} + \frac{1}{n_t} \int \ln p_t(i)\, di + \frac{1}{2n_t}\left[\int \ln p_t(i)\, di\right]^2 - \frac{1}{2} \int \left[\ln p_t(i)\right]^2 di \tag{33}$$

Insert $\frac{1}{n_t} \int \ln p_t(i)\, di = \frac{1}{1 - G_t} \int \ln p_t(z)\, dG(z)$ and $n_t = N_t(1 - G_t)$, where $G_t \equiv G(z_t^\star)$, into this condition, and note, $\int [\ln p_t(i)]^2\, di = N_t \int_{z_t^\star} [\ln p_t(z)]^2\, dG(z)$. Then recall, $\rho(z_t^\star) = \frac{mc_t}{z_t^\star}$

and $\rho_t(z) = \Omega_t \frac{mc_t}{z}$, where $\Omega_t \equiv \Omega_t(u_t)$ and $u_t \equiv \exp \frac{z_t}{z_t^\star}$ implies, $\ln p_t(z) = \ln p(z_t^\star) + 1 - \ln u_t + \ln \Omega_t$. Inserting this into (33) and simplifying,

$$[\ln p(z_t^\star) + 1 - \Omega_t]^2 = [\ln p(z_t^\star)]^2 + 2\ln p(z_t^\star) - 2\Omega_t \ln p(z_t^\star) + (1-\Omega_t)^2$$

where $\int \ln p(z_t^\star)\, dG(z) = \ln p(z_t^\star) \int dG(z) = (1-G_t)\ln p(z_t^\star)$. Plugging-in for the mass of entrants generates,

$$2\ln \frac{w_t}{z_t^\star a_t} = N_t \int_{z_t^\star}^{z_{\max}} (\Omega_t - 1)^2\, dG(z) \tag{34}$$

which appears as equation (5) in the main text.

Appendix A.2 (Proof of Lemmas 1 and 2)

In what follows we drop all t-subscripts The productivity level z, obtained by each new entrant, is the realization of a random variable drawn independently across firms from a distribution, $G(z)$. We denote by $z^\star$ the minimum level of productivity required to produce a good.

Lemma 1 (Aggregation) Let $G(z)$ be the distribution function of z, bounded by $z_{\max} \leq +\infty$. For any function, $J(z^\star) = \int_{z^\star}^{z_{\max}} j\left(\frac{z}{z^\star}\right) dG(z)$, where $j(1) \geq 0$ and $j' \geq 0$, then, (i), $J'(z^\star) < 0$, and (ii), $\lim\limits_{z^\star \to z_{\max}} J(z^\star) = 0$.

Proof Totally differentiating, $J'(z^\star) = -j(1)g(z^\star) - \frac{1}{(z^\star)^2}\int_{z^\star}^{z_{\max}} zj'\left(\frac{z}{z^\star}\right) dG(z) < 0$. Consider the integral $J(a) = \int_a^{z_{\max}} j\left(\frac{z}{a}\right) dG(z)$, where $a > 0$ is any number. Since the integral exists, for any $\varepsilon > 0$, there should exist $k(\varepsilon, a) > a$, such that $\int_{k(\varepsilon,a)}^{z_{\max}} j\left(\frac{z}{a}\right) dG(z) < \varepsilon$. Since $j' \geq 0$ and $k(\varepsilon, a) > a$ then $j\left(\frac{z}{a}\right) > j\left(\frac{z}{k(\varepsilon,a)}\right)$ and $\int_{k(\varepsilon,a)}^{z_{\max}} j\left(\frac{z}{k(\varepsilon,a)}\right) dG(z) < \int_{k(\varepsilon,a)}^{z_{\max}} j\left(\frac{z}{a}\right) dG(z) < \varepsilon$. $\blacksquare$

We will say that the distribution function $G(z)$ has an increasing elasticity of the probability density function, $g(z)$, if $\varepsilon_g(z) \equiv -\frac{g'(z)z}{g(z)}$ is an increasing function.

Lemma 2 (Selection) Make the change of variables, $u = z/z^{\star}$. Let $g(u)$ be a density function with elasticity, $\epsilon(u) = -\frac{ug'(u)}{g(u)}$, which is weakly increasing. Let $j_1(u)$ and $j_2(u)$ be positive functions such that $\frac{j_1(u)}{j_2(u)}$ is strictly increasing. The ratio $\frac{J_1(z^{\star})}{J_2(z^{\star})}$ is a decreasing function of $z^{\star}$, where $J_i(z^{\star}) = \int_{z^{\star}}^{z_{\max}} j_i(u)\, g(z)\, dz$, for $i = 1, 2$.

Proof We want to show that the ratio $\frac{J_1(z^{\star})}{J_2(z^{\star})}$ is a decreasing function of $z^{\star}$, where $J_i(z^{\star}) = \int_{z^{\star}}^{z_{\max}} j_i\left(\frac{z}{z^{\star}}\right) g(z)\, dz$. Consider $J_i(z^{\star}) = z^{\star} \int_1^{z_{\max}/z^{\star}} j_i(u)\, g(z^{\star}u)\, du$, which we differentiate as,

$$J_i'(z) = \frac{J_i(z^{\star})}{z^{\star}} - \frac{z_{\max}}{z^{\star}} j_i\left(\frac{z_{\max}}{z^{\star}}\right) g(z_{\max}) + \int_1^{z_{\max}/z^{\star}} j_i(u)\,(uz^{\star})\, g'(z^{\star}u)\, du \tag{35}$$

Now consider the ratio,

$$\begin{aligned}
\mathcal{J}(z^{\star}) \quad &\equiv \quad [J_2(z^{\star})]^2 \frac{d}{dz^{\star}} \frac{J_1(z^{\star})}{J_2(z^{\star})} = J_1'(z^{\star})\, J_2(z^{\star}) - J_1(z^{\star})\, J_2'(z^{\star}) \\
&= \quad \left(\frac{J_1(z^{\star})}{z^{\star}} - \frac{z_{\max}}{z^{\star}} j_1\left(\frac{z_{\max}}{z^{\star}}\right) g(z_{\max}) + \int_1^{z_{\max}/z^{\star}} j_1(u)\,(uz^{\star})\, g'(z^{\star}u)\, du \right) J_2(z^{\star}) \\
&\quad - \left(\frac{J_2(z_t^{\star\star})}{z^{\star}} - \frac{z_{\max}}{z^{\star}} j_2\left(\frac{z_{\max}}{z^{\star}}\right) g(z_{\max}) + \int_1^{z_{\max}/z^{\star}} j_2(u)\,(uz^{\star})\, g'(z^{\star}u)\, du \right) J_1(z^{\star})
\end{aligned}$$

therefore,

$$\begin{aligned}
\mathcal{J}(z^{\star}) \quad \equiv \quad & \int_1^{z_{\max}/z^{\star}} j_2(u)\, g(uz^{\star})\, du \int_1^{z_{\max}/z^{\star}} j_1(u)\, g(uz^{\star}) \left[\frac{uz^{\star} g'(uz^{\star})}{g(uz^{\star})}\right] du \\
& - \int_1^{z_{\max}/z^{\star}} j_1(u)\, g(uz^{\star})\, du \int_1^{z_{\max}/z^{\star}} j_2(u)\, g(uz^{\star}) \left[\frac{uz^{\star} g'(uz^{\star})}{g(uz^{\star})}\right] du \\
& + z_{\max} g(z_{\max}) \left[\int_1^{z_{\max}/z^{\star}} j_1(u)\, g(uz^{\star})\, j_2\left(\frac{z_{\max}}{z^{\star}}\right) du \right] \\
& - z_{\max} g(z_{\max}) \left[\int_1^{z_{\max}/z^{\star}} j_2(u)\, g(uz^{\star})\, j_1\left(\frac{z_{\max}}{z^{\star}}\right) du \right]
\end{aligned} \tag{36}$$

We consider the terms in the first two lines and the final two lines of equation (36) separately. The sign of the first two lines can be written as,

$$\frac{\int_1^{z_{\max}/z^\star} \epsilon(uz_t^\star) j_2(u)\, g(uz_t^\star)\, du}{\int_1^{z_{\max}/z^\star} j_2(u)\, g(uz_t^\star)\, du} - \frac{\int_1^{z_{\max}/z^\star} \epsilon(uz_t^\star) j_1(u)\, g(uz_t^\star)\, du}{\int_1^{z_{\max}/z^\star} j_1(u)\, g(uz_t^\star)\, du} \tag{37}$$

where $\epsilon(z) = -\frac{zg'(z)}{g(z)}$. Now consider CDFs defined as,

$$G_i(u) = \int_{-\infty}^{u} g_i(y)\, dy = \frac{\int_1^{y} j_i(u)\, g(uz_t^\star)\, du}{\int_1^{z_{\max}/z^\star} j_i(u)\, g(uz_t^\star)\, du} \quad \text{for } i = 1, 2\ ;\ 1 < y < z_{\max}/z^\star$$

Damjanovic (2005) shows (formula 3) that if $\frac{j_1(u)}{j_2(u)}$ is an increasing function of u then, $G_1(u) < G_2(u)$. This further implies that for any weakly increasing function $\epsilon(u)$,

$$\int_{-\infty}^{+\infty} \epsilon(u) g_1(u) du > \int_{-\infty}^{+\infty} \epsilon(u) g_2(u) du \tag{38}$$

and (37) is negative if $\epsilon(u)$ is increasing.

The sign of the second two lines in (36) can be written as,

$$\int_1^{z_{\max}/z^\star} \left[j_1(u)\, j_2\left(\frac{z_{\max}}{z^\star}\right) - j_2(u)\, j_1\left(\frac{z_{\max}}{z^\star}\right)\right] g(uz^\star)\, du \tag{39}$$

However, as j_1/j_2 is an increasing function, for any $u < \frac{z_{\max}}{z^\star}$, then $\frac{j_1(u)}{j_2(u)} < \frac{j_1\left(\frac{z_{\max}}{z^\star}\right)}{j_2\left(\frac{z_{\max}}{z^\star}\right)}$, and the integral in equation (39) is negative. In this case, so is the final term in equation (36). ■

Recall, we have the following definitions (with $s = \Omega - 1$),

$$z_1 \equiv \int_{z^\star}^{z_{\max}} \frac{(\Omega - 1)^2}{\Omega} dG(z) \quad ; \quad z_2 \equiv \int_{z^\star}^{z_{\max}} (\Omega - 1)\, dG(z) \tag{40}$$

where $\Omega = \Omega\left(\frac{z}{z_t^\star}\exp\right)$ is the Lambert-W function. Make the further definition, $z_3 \equiv z_2 - z_1 = \int_{z^\star}^{z_{\max}} \frac{\Omega-1}{\Omega} dG(z)$, such that, $\left(\frac{1}{z_1} - \frac{1}{z_2}\right)^{-1} = z_2 \frac{z_1}{z_3}$. Applying lemma 2, $\exp\left(\frac{H}{2}\right)$ and $\frac{z_1}{z_3}$ are decreasing functions. Since $z_2\left[\frac{z_1}{z_3}\exp\left(\frac{H}{2}\right)\right]$ is a product of positive and decreasing functions.

Appendix A.3 (Proof of Proposition 1)

Consider the planning problem,

$$\max_{\{c_t,L_t\}} E_0 \sum_{t=0}^{\infty} \beta^t d_t \left(\ln c_t - \psi L_t\right) + \beta^t \lambda_{1,t} \left(a_t L_t - c_t - \mathcal{G}\right)$$

$$\Rightarrow \quad c_t : \beta^{-t}\frac{\partial J}{\partial c_t} c_t = d_t - \lambda_{1,t} c_t = 0 \quad \text{and} \quad L_t : \beta^{-t}\frac{\partial J}{\partial L_t} L_t = -\psi d_t L_t + \lambda_{1,t} a_t L_t = 0$$

$$\Rightarrow \quad c_t = \frac{a_t}{\psi}$$

Plugging this into the steady-state conditions for the decentralized economy (labor supply and firm pricing; $\varepsilon_w \psi L_t = (1-\tau_L)\, c_t^{-1} (\varepsilon_w - 1) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right) a_t L_t$) generates:

$$\tau_L = 1 - \left(\frac{\varepsilon_w}{\varepsilon_w - 1}\right)\left(\frac{\varepsilon_p}{\varepsilon_p - 1}\right) a_t$$

Setting the labor-income tax at this level will generate an efficient steady state.

Now consider the Ramsey problem,

$$\begin{aligned}
&\max_{\{c_t,L_t,\pi_t^w\}} E_0 \sum_{t=0}^{\infty} \beta^t d_t \left(\ln c_t - \psi L_t\right) \\
&+\beta^t \mu_{1,t} \left\{ \begin{array}{c} \chi \left[d_t c_t^{-1} \left(\pi_t^w - 1\right) \pi_t^w\right] - \chi \beta E_t \left[d_{t+1} c_{t+1}^{-1} \left(\pi_{t+1}^w - 1\right) \pi_{t+1}^w\right] \\ -\varepsilon_w \psi d_t L_t - (1-\tau_L)\, d_t c_t^{-1} (1-\varepsilon_w) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right) a_t L_t \end{array} \right\} \\
&+\beta^t \mu_{2,t} \left[a_t L_t - c_t - \mathcal{G} - \frac{\chi}{2}\left(\pi_t^w - 1\right)^2\right]
\end{aligned}$$

The first-order condition with respect to π_t^w, evaluated at the steady-state, implies $\pi = 1$ is optimal. The remaining first-order conditions are,

$$\begin{aligned}
d_t \quad = \quad & \chi \left[d_t c_t^{-1} \left(\pi_t^w - 1\right) \pi_t^w\right] \left(\mu_{1,t} - \mu_{1,t-1}\right) + \mu_{1,t} (1-\tau_L)\, d_t c_t^{-1} (1-\varepsilon_w) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right) a_t L_t \\
& -\mu_{2,t} c_t
\end{aligned}$$

and,

$$\psi d_t L_t = -\mu_{1,t} \varepsilon_w \psi d_t L_t - \mu_{1,t} (1-\tau_L)\, d_t c_t^{-1} (1-\varepsilon_w) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right) a_t L_t + \mu_{2,t} a_t L_t$$

which at the steady-state yield,

$$\mu_1 \varepsilon_w \psi L + \mu_2 c = 1 \quad \text{and} \quad \mu_2 = \psi > 0 \tag{41}$$

where $(1-\tau_L)\, c^{-1} (\varepsilon_w - 1) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right) = \varepsilon_w \psi$ has been used. Solving for μ_1:

$$\mu_1 = (1 - \psi c) \frac{1}{\varepsilon_w \psi L} = \frac{\Phi}{\varepsilon_w \psi L} \quad ; \quad \Phi \equiv 1 - (1 - \tau_L) \left(\frac{\varepsilon_w - 1}{\varepsilon_w}\right) \left(\frac{\varepsilon_p - 1}{\varepsilon_p}\right)$$

where Φ is a measure of distortions in the economy created by monopolistic competition. An efficient steady-state is one in which $\Phi = 0$ and $\mu_1 = 0$.

A second-order approximation to the Ramsey-problem implies the welfare-based loss function is:

$$\mathbb{L} = Q_{u,t} + \mu_1 Q_{1,t} + \mu_2 Q_{2,t}$$

where,

$$\begin{aligned} Q_{u,t} &= -\frac{1}{2} (\widehat{c}_t)^2 + \widehat{d}_t \widehat{c}_t - \psi L \left(\widehat{d}_t \widehat{L}_t\right) \quad ; \quad Q_{2,t} = L \left(\widehat{a}_t \widehat{L}_t\right) - \frac{\eta}{2} (\widehat{\pi}_t^w)^2 \\ Q_{1,t} &= -\varepsilon_w \psi L \left[\widehat{d}_t \widehat{L}_t - (\widehat{c}_t)^2 + \widehat{d}_t \widehat{c}_t - \widehat{d}_t \widehat{L}_t + \widehat{c}_t \widehat{a}_t + \widehat{c}_t \widehat{L}_t - \widehat{a}_t \widehat{L}_t\right] \end{aligned}$$

We then use $\widehat{c}_t = \frac{L}{c}\left(\widehat{a}_t + \widehat{L}_t\right)$ and note $\left(\widehat{a}_t + \widehat{L}_t\right)^2 = \widehat{L}_t^2 + 2\widehat{a}_t \widehat{L}_t + tip$ and define $s_c \equiv \frac{c}{y} < 1$. Recalling equations (41) it is immediate that the multiplier on the cross term $\widehat{d}_t \widehat{L}_t$ is zero, even when the steady-state is not efficient, and as such, we can write,

$$\mathbb{L} = -\frac{1}{2} \left[\xi_L \left(\widehat{L}_t - \widehat{L}_t^\star\right)^2 + \eta \psi (\widehat{\pi}_t^w)^2\right] \quad \text{and} \quad \widehat{L}_t^\star \equiv \frac{\xi_{aL}}{\xi_L} \widehat{a}_t$$

where,

$$\xi_L \equiv \frac{1}{s_c} \left[\frac{1}{s_c} + 2\Phi \left(1 - \frac{1}{s_c}\right)\right] \quad \text{and} \quad \xi_{aL} \equiv \frac{1 - (1 - s_c)\,\Phi}{s_c} - \xi_L$$

The linear wage Phillips curve is, $\widehat{\pi}_t^w = \left[\frac{\varepsilon_w}{\chi} (1 - \Phi)\, y\right] (\widehat{c}_t - \widehat{a}_t) + \beta \widehat{\pi}_{t+1}^w$, where $\widehat{c}_t = \frac{1}{s_c}\left(\widehat{a}_t + \widehat{L}_t\right)$, and this implies,

$$\widehat{\pi}_t^w = \frac{\xi}{s_c} \left(\widehat{L}_t - \widehat{L}_t^n\right) + \beta \widehat{\pi}_{t+1}^w \quad ; \quad \widehat{L}_t^n \equiv -(1 - s_c)\, \widehat{a}_t \quad \text{and} \quad \xi \equiv \frac{\varepsilon_w}{\chi} (1 - \Phi)\, y$$

The final expressions we use are (by eliminating the parameter $\chi > 0$ in the loss function):

$$\mathbb{L} = -\frac{1}{2}\frac{1}{s_c}\left[q_L\left(\widehat{L}_t - \widehat{L}_t^{\star}\right)^2 + q_\pi\left(\widehat{\pi}_t^w\right)^2\right] \quad ; \quad \widehat{L}_t^{\star} \equiv -\left[1 - \frac{1-(1-s_c)\Phi}{\frac{1}{s_c} + 2\Phi\left(1-\frac{1}{s_c}\right)}\right]\widehat{a}_t \tag{42}$$

where,

$$q_L \equiv \frac{1}{s_c} + 2\Phi\left(1 - \frac{1}{s_c}\right) \quad ; \quad q_\pi \equiv \frac{\varepsilon_w}{\xi}(1-\Phi)^2 \quad ; \quad \widehat{\pi}_t^w = \frac{\xi}{s_c}\left(\widehat{L}_t - \widehat{L}_t^n\right) + \beta\widehat{\pi}_{t+1}^w \tag{43}$$

and $\widehat{L}_t - \widehat{L}_t^{\star} = \widehat{y}_t - \widehat{y}_t^{\star}$ and $\widehat{L}_t - \widehat{L}_t^n = \widehat{y}_t - \widehat{y}_t^n$. Equations (42) and (43) are used to generate Proposition 1.

Appendix A.4 (Proof of Proposition 2)

Here we generate an equivalence result. We start by writing down the equilibrium conditions for the static economy. The wage Phillips curve and labor demand condition are:

$$\begin{aligned}\chi\left[d_t c_t^{-1}\left(\pi_t^w - 1\right)\pi_t^w\right] &= \chi\beta E_t\left[d_{t+1}c_{t+1}^{-1}\left(\pi_{t+1}^w - 1\right)\pi_{t+1}^w\right] + (1-\tau_L)\, d_t c_t^{-1}(1-\varepsilon_w)\, w_t L_t \\ &\quad + \varepsilon_w \psi d_t L_t\end{aligned}$$

and

$$w_t = a_t z_t^{\star} e^{H_t/2\zeta}$$

The remaining equations are:

$$y_t = c_t + fN_t + \mathcal{G} + \frac{\chi}{2}\left(\pi_t^w - 1\right)^2 = w_t L_t + N_t\vartheta_t \text{ and } \vartheta_t = z_{1,t}y_t = f \ ; \ N_t = \frac{1}{z_{2,t}}$$

These seven equations solve for $w_t, L_t, N_t, \vartheta_t, c_t, y_t, z_t^{\star}$. Combine these expressions to generate an expression for effective output,

$$F(a_t L_t) \equiv y_t - fN_t = w_t L_t = a_t z_t^{\star} e^{H_t/2} L_t$$

Having already solved implicitly for the cut-off, i.e., $z_t^{\star} = z_t^{\star}(a_t L_t)$, we write effective output, $y_t - fN_t$, as a function of $a_t L_t$ alone. As such, by picking L_t, we are picking the cut-off and our choice of L_t determines $z_t^{\star}$ and also N_t.

We return to the planning problem, which is now,

$$\max_{\{c_t,L_t\}} E_0 \sum_{t=0}^{\infty} \beta^t d_t \left(\ln c_t - \psi L_t\right) + \beta^t \lambda_{1,t} \left[F\left(a_t L_t\right) - c_t - \mathcal{G}\right]$$

$$\Rightarrow \quad c_t : \beta^{-t} \frac{\partial J}{\partial c_t} c_t = d_t - \lambda_{1,t} c_t = 0 \quad \text{and} \quad L_t : \beta^{-t} \frac{\partial J}{\partial L_t} L_t = -\psi d_t L_t + \lambda_{1,t} F_L a_t L_t = 0$$

$$\Rightarrow \quad c_t = F_L \frac{a_t}{\psi}$$

where F_L is the partial derivative of $F\left(a_t L_t\right)$ with respect to L_t. Using the equilibrium condition, $\varepsilon_w \psi L_t = \left(1 - \tau_L\right) c_t^{-1} \left(\varepsilon_w - 1\right) F\left(a_t L_t\right)$ we have,

$$\tau_L = 1 - \left(\frac{\varepsilon_w}{\varepsilon_w - 1}\right) \frac{\left(a_t L_t\right) F_L}{F\left(a_t L_t\right)}$$

Again, setting this labor-income tax at this level in the steady-state will deliver efficiency.

The corresponding Ramsey problem is:

$$\begin{aligned}
&\max_{\{c_t,L_t,\pi_t^w\}} E_0 \sum_{t=0}^{\infty} \beta^t d_t \left(\ln c_t - \psi L_t\right) \\
&+\beta^t \mu_{1,t} \left\{ \begin{array}{c} \chi \left[d_t c_t^{-1} \left(\pi_t^w - 1\right) \pi_t^w\right] - \chi \beta E_t \left[d_{t+1} c_{t+1}^{-1} \left(\pi_{t+1}^w - 1\right) \pi_{t+1}^w\right] \\ -\varepsilon_w \psi d_t L_t - \left(1 - \tau_L\right) d_t c_t^{-1} \left(1 - \varepsilon_w\right) F\left(a_t L_t\right) \end{array} \right\} \\
&+\beta^t \mu_{2,t} \left[F\left(a_t L_t\right) - c_t - \mathcal{G} - \frac{\eta}{2} \left(\pi_t^w - 1\right)^2\right]
\end{aligned}$$

An interpretation of the effective output production function, is that, for the basic model, we simply have $F(a_t L_t) = a_t L_t$. As with the basic case zero steady state inflation is Ramsey optimal. The remaining two FOCs are:

$$\begin{aligned}
d_t &= -\chi \left[d_t c_t^{-1} \left(\pi_t^w - 1\right) \pi_t^w\right] \left(\mu_{1,t} - \mu_{1,t-1}\right) + \mu_{1,t} \left(1 - \tau_L\right) d_t c_t^{-1} \left(1 - \varepsilon_w\right) F\left(a_t L_t\right) - \mu_{2,t} c_t \\
\psi d_t L_t &= -\mu_{1,t} \varepsilon_w \psi d_t L_t - \mu_{1,t} \left(1 - \tau_L\right) d_t c_t^{-1} \left(1 - \varepsilon_w\right) F_L L_t + \mu_{2,t} F_L L_t
\end{aligned}$$

where F_L is the partial derivative of $F\left(a_t L_t\right)$ with respect to L_t. At the steady-state (where $a_t = a = 1$ and $F\left(a_t L_t\right) = F$) these conditions reduce to:

$$\mu_1 \varepsilon_w \psi L = 1 - \mu_2 c \quad \text{and} \quad \mu_2 F_L = \psi + \mu_1 \varepsilon_w \psi + \mu_1 \left(1 - \tau_L\right) c^{-1} \left(1 - \varepsilon_w\right) F_L$$

where $\frac{F}{c} = \frac{\varepsilon_w}{\varepsilon_w - 1}\frac{1}{1-\tau_L}\psi L$. Efficiency requires $\mu_1 = 0$, in which case,

$$c = \frac{1}{\mu_2} \quad \text{and} \quad \mu_2 F_L = \psi \quad \Rightarrow \quad F_L = \psi c$$

Plugging this into the steady-state wage Phillips curve, we have, the efficiency condition, $1 = \frac{\varepsilon_w}{\varepsilon_w - 1}\frac{1}{1-\tau_L}\frac{F_L L}{F}$. We therefore define,

$$\Phi = 1 - (1 - \tau_L)\left(\frac{\varepsilon_w - 1}{\varepsilon_w}\right)\frac{F}{F_L L}$$

as we did for the no-entry case, seeing that the fixed price-markup, in that case $\frac{\varepsilon_p - 1}{\varepsilon_p}$, is replaced by $\frac{F}{F_L L}$. We don't provide explicit expressions for the lagrange multipliers because we will consider an efficient steady state where $\mu_1 = 0$ and $\mu_2 = \frac{1}{c}$.

A second-order approximation to the Ramsey-problem generates:

$$\mathbb{L} = Q_{u,t} + \mu_1 Q_{1,t} + \mu_2 Q_{2,t} \tag{44}$$

where,

$$\begin{aligned}
Q_{u,t} &= -\frac{1}{2}\left(\widehat{c}_t\right)^2 + \widehat{d}_t\widehat{c}_t - \psi L\left(\widehat{d}_t\widehat{L}_t\right) \\
Q_{1,t} &= -\varepsilon_w \psi L \begin{bmatrix} \left(1 - \frac{LF_L}{F}\right)\widehat{d}_t\widehat{L}_t - \left(\widehat{c}_t\right)^2 + \widehat{d}_t\widehat{c}_t + \left(\frac{F_a}{F}\right)\widehat{c}_t\widehat{a}_t + \left(\frac{LF_L}{F}\right)\widehat{c}_t\widehat{L}_t \\ -\frac{1}{2}\left(\frac{F_{LL}L^2}{F}\right)\left(\widehat{L}_t\right)^2 - \left(\frac{LF_L + L^2F_{LL}}{F}\right)\left(\widehat{a}_t\widehat{L}_t\right) \end{bmatrix} \\
Q_{2,t} &= \frac{1}{2}F_{LL}L^2\left(\widehat{L}_t\right)^2 + \left(LF_L + L^2F_{LL}\right)\left(\widehat{a}_t\widehat{L}_t\right) - \frac{\eta}{2}\left(\widehat{\pi}_t^w\right)^2
\end{aligned}$$

In deriving the above expression we use $F\left(a_t L_t\right) \approx \frac{1}{2}F_{LL}L^2\left(\widehat{L}_t\right)^2 + F_{aL}aL\left(\widehat{a}_t\widehat{L}_t\right)$ since $\left(\widehat{a}_t\right)^2$ is a term independent of policy and we assume $a = 1$ in what follows We then use $F_{aL} = \frac{\partial}{\partial a}\frac{\partial}{\partial L}F(a \times L) = \frac{\partial}{\partial a}\left(aF'(a \times L)\right) = F' + aLF''$, such that, $F_{aL}aL\left(\widehat{a}_t\widehat{L}_t\right) = F_{aL}L\left(\widehat{a}_t\widehat{L}_t\right) = \left(LF_L + L^2F_{LL}\right)\left(\widehat{a}_t\widehat{L}_t\right)$.

In an efficient steady state, in which, $\mu_2 = 1/c$, equation (44) is:

$$\begin{aligned}
\mathbb{L} &= -\frac{1}{2}\left(\frac{F}{c}\frac{LF_L}{F}\right)^2\left(\widehat{L}_t^2 + 2\widehat{a}_t\widehat{L}_t\right) + \left(\frac{F}{c}\frac{LF_L}{F} - \psi L\right)\left(\widehat{d}_t\widehat{L}_t\right) \\
&\quad + \mu_2\left[\frac{1}{2}F_{LL}L^2\left(\widehat{L}_t\right)^2 + \left(LF_L + L^2F_{LL}\right)\left(\widehat{a}_t\widehat{L}_t\right) - \frac{\eta}{2}\left(\widehat{\pi}_t^w\right)^2\right]
\end{aligned}$$

where we use $\widehat{c}_t = \frac{F}{c}\left(\frac{LF_L}{F}\right)\left(\widehat{a}_t + \widehat{L}_t\right)$ and $\left(\widehat{a}_t + \widehat{L}_t\right)^2 = \widehat{L}_t^2 + 2\widehat{a}_t\widehat{L}_t$ and impose $\widehat{a}_t\widehat{d}_t = 0$. Collecting term and simplifying,

$$\begin{aligned}\mathbb{L} \;=\; & -\frac{1}{2}\left(\frac{F}{c}\frac{LF_L}{F}\right)\left[\left(\frac{F}{c}\frac{LF_L}{F}\right) - \frac{LF_{LL}}{F_L}\right]\left(\widehat{L}_t\right)^2 \\ & +\frac{1}{2}\left(\frac{F}{c}\frac{LF_L}{F}\right)\left[1 + \frac{LF_{LL}}{F_L} - \left(\frac{F}{c}\frac{LF_L}{F}\right)\right]\left(2\widehat{a}_t\widehat{L}_t\right) \\ & + \left(\frac{F}{c}\frac{LF_L}{F} - \psi L\right)\left(\widehat{d}_t\widehat{L}_t\right) - \frac{1}{2}\left(\frac{F}{c}\frac{LF_L}{F}\right)\frac{\eta}{LF_L}\left(\widehat{\pi}_t^w\right)^2\end{aligned}$$

First, recall $\frac{F}{c} = \frac{\varepsilon_w}{\varepsilon_w - 1}\frac{1}{1-\tau_L}\psi L$ and the definition of Φ. With an efficient steady state, $\psi L = \frac{F}{c}\left(\frac{\varepsilon_w - 1}{\varepsilon_w}\right)(1-\tau_L) = \frac{F}{c}(1-\Phi)\frac{LF_L}{F} = \frac{F}{c}\frac{LF_L}{F}$, and the coefficient on the term $\left(\widehat{d}_t\widehat{L}_t\right)$ is zero. Second, the linear wage Phillips curve is, $\widehat{\pi}_t^w = \left(\frac{\varepsilon_w L}{\chi}\psi c\right)(\widehat{c}_t - \widehat{w}_t) + \beta\widehat{\pi}_{t+1}^w$. Again using, $\frac{F}{c} = \frac{\varepsilon_w}{\varepsilon_w - 1}\frac{1}{1-\tau_L}\psi L$, this means, $\widehat{\pi}_t^w = \xi(\widehat{c}_t - \widehat{w}_t) + \beta\widehat{\pi}_{t+1}^w$, where ξ determines the slope of the wage Phillips curve, such that, $\chi = \frac{\varepsilon_w}{\xi}(1-\Phi)F_L L$. Since $\Phi = 0$ under an efficient steady-state, this means, $\frac{\eta}{LF_L} = \frac{\varepsilon_w}{\xi}$, is the term attached to inflation loss. Finally, we make two important definitions,

$$\alpha \equiv \left(1 - \frac{F}{c}\right)\frac{LF_L}{F} \quad \text{and} \quad \eta \equiv \frac{LF_L}{F} - \frac{LF_{LL}}{F_L}$$

which allow us to re-write the loss function as:

$$\mathbb{L} = -\frac{1}{2}\left(\frac{F}{c}\frac{LF_L}{F}\right)\left[(\eta - \alpha)\left(\widehat{L}_t - \widehat{L}_t^{\star}\right)^2 + \frac{\varepsilon_w}{\xi}\left(\widehat{\pi}_t^w\right)^2\right] \quad ; \quad \widehat{L}_t^{\star} = \frac{1 - \eta + \alpha}{\eta - \alpha}\widehat{a}_t$$

Since $\widehat{w}_t + \widehat{L}_t = \frac{c}{F}\widehat{c}_t$, the Phillips curve can be written as $\widehat{\pi}_t^w = \xi(1-\alpha)\left(\widehat{L}_t - \widehat{L}_t^n\right) + \beta\widehat{\pi}_{t+1}^w$, where $\widehat{L}_t^n = \left(\frac{\alpha}{1-\alpha}\right)\widehat{a}_t$. We can now verify the main result,

$$\widehat{L}_t^n - \widehat{L}_t^{\star} = \left(\frac{1}{1-\alpha}\frac{1-\eta}{\alpha - \eta}\right)\widehat{a}_t$$

The gap $\widehat{L}_t^n - \widehat{L}_t^{\star}$ is only zero when $\eta = 1$ such that, $\widehat{L}_t^{\star} = \frac{\alpha}{1-\alpha}\widehat{a}_t$. Finally, recall, in the no-entry case, when the steady-state is efficient, $\mathbb{L} = -\frac{1}{2}\frac{1}{s_c}\left[\frac{1}{s_c}\left(\widehat{L}_t - \widehat{L}_t^{\star}\right)^2 + \frac{\varepsilon_w}{\xi}\left(\widehat{\pi}_t^w\right)^2\right]$ and $\widehat{L}_t^{\star} \equiv -(1 - s_c)\widehat{a}_t$.

For completeness, in the general case, we need an expression for $\mu_2 c$ (since we already have $\mu_1 \varepsilon_w \psi L = 1 - \mu_2 c$). We have,

$$\begin{aligned} \mu_2 F_L L &= \psi L + \mu_1 \varepsilon_w \psi L + \mu_1 (1 - \tau_L) c^{-1} (1 - \varepsilon_w) F_L L = \psi L + \varepsilon_w \mu_1 \psi L - (\varepsilon_w \mu_1 \psi L) \frac{F_L L}{F} \\ &= \psi L - (1 - \mu_2 c) \left(1 - \frac{F_L L}{F}\right) \end{aligned}$$

rearranging,and introducing Φ and α, we find:

$$\mu_2 c = \frac{(1 - \Phi) \left(\frac{LF_L}{F} - \alpha\right) - \left(1 - \frac{F_L L}{F}\right)}{\left(\frac{LF_L}{F} - \alpha\right) - \left(1 - \frac{F_L L}{F}\right)}$$

If we now return to equation we can use this expression to eliminate $\mu_1 \varepsilon_w \psi L$ which enters through the term $Q_{1,t}$.

Appendix B

Appendix B.1 (Summary of Dynamic Model)

The following table presents a summary of the dynamic model equations:[38]

Description	Equation
	New Keynesian Block
Resource Constraint	$y_t = c_t + f_t N_{e,t} + \mathcal{G} + \frac{\chi}{2}\left(\pi_t^w - 1\right)^2$
GDP (income)	$y_t = w_t L_t + \vartheta_t N_t$
Price Aggregation	$w_t = a_t z_t^\star e^{H_t/2}$
Wage Phillips Curve	$\left[\varepsilon_w \psi L_t + \left(1-\tau_L\right)\left(1-\varepsilon_w\right) x_t\right] c_t$ $= \chi \left[\left(\pi_t^w - 1\right)\pi_t^w - \beta E_t \frac{c_t/d_t}{c_{t+1}/d_{t+1}}\left(\pi_{t+1}^w - 1\right)\pi_{t+1}^w\right]$
Real Wage	$\frac{w_t}{w_{t-1}} = \frac{\pi_t^w}{\pi_t}$ and $x_t \equiv \frac{w_t L_t}{c_t}$
Consumption Euler Equation	$\frac{\pi_{t+1}}{1+i_t} = \beta E_t \left(\frac{c_t/d_t}{c_{t+1}/d_{t+1}}\right)$
	Heterogeneous Firms Block
Firm Value and Free Entry	$f_t = \left(1-\delta\right)\beta E_t \left(\frac{c_t/d_t}{c_{t+1}/d_{t+1}}\right)\left(\vartheta_{t+1} + f_{t+1}\right)$
Expected Profits	$E_t \vartheta_t = \int_{z_t^\star}^{z_{\max}} \vartheta_t \left[z\left(i\right)\right] dG\left(z\right) = z_{1,t} y_t$
Mass of Entrants	$N_t z_{2,t} = 1$
Firm Dynamics	$N_t = \left(1-\delta\right)\left(N_{t-1} + N_{e,t-1}\right)$
Entry Congestion	$f_t = f\left(\frac{N_{e,t}}{N_{e,t-1}}\right)^\gamma$

To complete the system we either specify an interest rate policy or calculate optimal policy. Thus, for a given nominal interest rate, $i_t \geq 0$, the 12 equations in the table solve for

[38] Variables $z_{1,t} \equiv \int_{z_t^\star}^{z_{\max}} \left(\Omega_t - 1\right) dG\left(z\right)$ and $z_{2,t} \equiv \int_{z_t^\star}^{z_{\max}} \frac{\left(\Omega_t - 1\right)^2}{\Omega_t} dG\left(z\right)$ are aggregators, $H_t \equiv \zeta^2 N_t \int_{z_t^\star}^{z_{\max}} \left(\Omega_t - 1\right)^2 dG\left(z\right)$ is the Herfindahl index, $\Omega_t \equiv \Omega\left(\frac{z}{z_t^\star}\exp\right)$ is the Lambert-W function, and $s_t = \Omega_t - 1$ is market share, all of which are only functions of $z_t^\star$. Variables $\pi_{t+1}^w \equiv \frac{W_{t+1}}{W_t}$ and $\pi_{t+1} \equiv \frac{p_{t+1}}{p_t}$ are nominal wage and price inflation.

$y_t, c_t, L_t, w_t, \pi_t, \pi_t^w, x_t$ and $f_t, N_t, N_{e,t}, \vartheta_t, z_t^\star$ with given government expenditure, $\mathcal{G} > 0$, dividend and labor-income, τ^L, and exogenous processes for a_t and d_t.

Appendix B.2 (Solution for Steady-State)

The solution for the zero inflation steady-state is as follows. Assume is known $z^\star$ and set the probability of successful entry, $G(z^\star)$. We have the following zero inflation steady-state equations:

$$\begin{aligned}
\frac{1}{1-\tau_L}\frac{\varepsilon_w}{\varepsilon_w - 1}\psi L &= c^{-1} wL \;\; ; \;\; wL = y\left(1 - z_1 N\right) \;\; ; \;\; w = z^\star \exp\left(\frac{H}{2}\right) \\
y &= c + fN_e + \mathcal{G} \;\; ; \;\; N = \frac{1}{\zeta z_2} \;\; ; \;\; N_e = \left(\frac{\delta}{1-\delta}\right) N \\
\vartheta &= y z_1 = f\left[\frac{1-\beta\left(1-\delta\right)}{\beta\left(1-\delta\right)}\right]
\end{aligned}$$

which we use to determine $c, L, y, w, N, N_e, \vartheta$. Specifically, we compute $N = \frac{1}{z_2}$ and $N_e = \left(\frac{\delta}{1-\delta}\right) N$. The wage is then given by, $w = z^\star \exp\left(\frac{H}{2}\right)$. We assume $L = 1/3$. Then $y = \frac{wL}{1-z_1 N}$ determines y. Then we find the value of entry costs using $\vartheta = yz_1$ and $f = \vartheta\left[\frac{\beta(1-\delta)}{1-\beta(1-\delta)}\right]$. If we know $\mathcal{G}/y$, then c and $\mathcal{G}$ can be found from $y = c + fN_e + \mathcal{G}$. Finally, $\psi = \left(1 - \tau_L\right) wc\left(\frac{\varepsilon_w - 1}{\varepsilon_w}\right)$ defines the value of ψ (that which delivers $L = 1/3$). We define the inverse labor share as, $\frac{y}{wL} = \frac{1}{1-\frac{z_1}{z_2}}$. Also notice, $\vartheta = \left(y - wL\right)/N$. This must equal yz_1 and also be equal to $f\frac{1-\beta(1-\delta)}{\beta(1-\delta)}$.

References

ADAM, K. and WEBER, H. (2019), "Optimal Trend Inflation", American Economic Review 109, 702-737.

AUTOR, D., KATZ, L. PATTERSON, C. and van REENEN, J. (2020), "The Fall of the Labor Share and the Rise of Superstar Firms", Quarterly Journal of Economics 135, 645-709

BENIGNO, P. and WOODFORD, M. (2005), "Inflation Stabilization and Welfare: The Case of a Distorted Steady-State", Journal of the European Economic Association 3, 1185-1236.

BERGIN, P. and CORSETTI, G. (2008), "The Extensive Margin and Monetary Policy", Journal of Monetary Economics 55, 1222-1237.

BERGIN, P. and FEENSTRA, R. (2000), "Staggered Price Setting, Translog Preferences, and Endogenous Persistence", Journal of Monetary Economics 45, 657-680.

BILBIIE, F., GHIRONI, F. and MELITZ, M. (2012), "Endogenous Entry, Product Variety, and Business Cycles", Journal of Political Economy 120, 304-345.

BILBIIE, F., FUJIWARA, I. and GHIRONI, F. (2014), "Optimal Monetary Policy with Endogenous Entry and Product Variety, Journal of Monetary Economics 64, 1-20.

BLANCHARD, O., and GALI, J. (2007), "Real Wage Rigidities and the New Keynesian Model", Journal of Money, Credit and Banking, Supplement to 39, 35-65.

BRODA, C. and WEINSTEIN, D. (2010), "Product Creation and Destruction: Evidence and Price Implications", American Economic Review 100, 691-723.

CANTORE, C. FERRONI, F. and LEON-LEDESMA, M. (2020), "The Missing Link: Labor Share and Monetary Policy", Journal of the European Economic Association, forthcoming.

COMIN, D. and GERTLER, M. (2006), "Medium-Term Business Cycles", American Economic Review 96, 523-551.

CHRISTIANO, L., TRABANDT, M. and WALENTIN, K. (2020), "Involuntary Unemployment and the Business Cycle", Review of Economic Dynamics (https://doi.org/10.1016/j.red.2020.05.003), forthcoming.

CLARIDA, R., GALI, J. and GERTLER, M., "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", Quarterly Journal of Economics 115, 147-180.

DAMJANOVIC, T. (2005), "Lorenz Dominance for Transformed Income Distributions: A Simple Proof", Mathematical Social Sciences 50, 234-237.

DIEWERT, E. (1976), "Exact and Superlative Index Numbers", Journal of Econometrics 4, 115-145.

EDMOND, C. MIDRIGAN, V. and XU, D. (2019), "How Costly are Markups?", mimeo.

FASANI, S., MUMTAZ, H., and ROSSI, L. (2020), "Monetary Policy Uncertainty and Firm Dynamics", Working Paper No. 903, School of Economics and Finance, Queen Mary University of London

FEENSTRA, R., and WEINSTEIN, D. (2017), "Globalization, Markups and US Welfare", Journal of Political Economy, 125, 1041-1074.

FEENSTRA, R. (2018), "Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity", Journal of International Economics 110, 16-27.

GALI, J. and MONACELLI, T. (2016), "Understanding the Gains from Wage Flexibility: The Exchange Rate Connection", American Economic Review 106, 3829-3868.

HALDANE, A. (2018), "Market Power and Monetary Policy", speech at, Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, 24 August 2018.

HAMANO, M. and ZANETTI, F. (2020), "Monetary Policy, Firm Heterogeneity, and Product Variety", WINPEC Working Paper Series No.E2005

HANSEN, G. (1985), “Indivisible Labor and the Business Cycle”, Journal of Monetary Economics 16, 309-327

HOPENHAYN, H. (1992), “Entry, Exit, and Firm Dynamics in Long Run Equilibrium”, Econometrica 60, 1127-1150.

IKEDA, D. and KUROZUMI, T. (2019), “Slow Post-financial Crisis Recovery and Monetary Policy”, American Economic Journal: Macroeconomics 11, 82-112.

JUSTINIANO, A., PRIMICERI, G. and TAMBALOTTI, A. (2013), “Is There a Trade-Off Between Inflation and Output Stabilization?”, American Economic Journal: Macroeconomics 5, 1-31.

de LOECKER, J., EECKHOUT, J. and UNGER, G., (2020), “The Rise of Market Power and the Macroeconomic Implications”, Quarterly Journal of Economics 135, 561–644.

LEWIS, V. and POILLY, C., (2015), “Firm Entry, Markups and the Monetary Transmission Mechanism”, Journal of Monetary Economics 59, 670-685.

LEWIS, V. and STEVENS, A. (2015), “Entry and Markup Dynamics in an Estimated Business Cycle Model, European Economic Review 74, 14-35.

LOUALICHE, E., (2019), “Asset Pricing with Entry and Imperfect Competition”, mimeo.

MIYAKAWA, D. OIKAWA, K. and UEDA, K., (2020), “Reallocation Effects of Monetary Policy”, mimeo.

MORAN, P. and QUERALTO, A., (2018), “Innovation, Productivity, and Monetary Policy”, Journal of Monetary Economics 93, 24-41.

NEKARDA, C. and RAMEY, V. (2020), “The Cyclical Behavior of the Price-Cost Markup”, mimeo.

OLIVEI, G. and TENREYRO, S. (2007), “The Timing of Monetary Policy Shocks”, American Economic Review 97, 636–663.

OLIVEI, G. and TENREYRO, S. (2010), “Wage-Setting Patterns and Monetary Policy: International Evidence”, Journal of Monetary Economics 57, 785–802.

RODRIGUEZ-LOPEZ, A. (2011), “Prices and Exchange Rates: A Theory of Disconnect”, Review of Economic Studies 78, 1135-1177.

ROGERSON, R. (1988), “Indivisible Labor, Lotteries and Equilibrium”, Journal of Monetary Economics 21, 3-16.

SMETS, R. and WOUTERS, R. (2007), “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach”, American Economic Review 97, 586-606.

TAYLOR, J. (1993), “Discretion versus Policy Rules in Practice”, Carnegie-Rochester Conference Series on Public Policy 39, 195-214.

Table 1: Steady-State Parameter Values

Parameters Set Exogenously				
Statistic	Parameter	Value		Target/Source
Discount factor	β	$(\beta^{-4}-1)\times 100 = 1.44\%$		FRED
Fiscal Policy	$\{\tau_L, \mathcal{G}/y\}$	$\{0.2, 0.17\}$		FRED
Wage Markup	ε_w	$\left(\frac{\varepsilon_w}{\varepsilon_w-1}-1\right)=30\%$		standard
Wage Rigidity	χ	23.43		standard
Exogenous Firm Exit	δ	2.935%		BLS
Calibrated Parameters				
Statistic	Parameter	Value	Target	Source
Failure rate	$\int_0^{z^\star} dG\left(z\right)$	15%	-	see text
Labor Share	$\{\kappa, \sigma\}$	$\{2.50, 0.42\}$	60%	BLS
Entry costs	$\{f_\kappa, f_\sigma\}$	$\{0.98, 0.74\}$	-	see text
Hours worked	$\{\psi_\kappa, \psi_\sigma\}$	2.29	33%	normalization

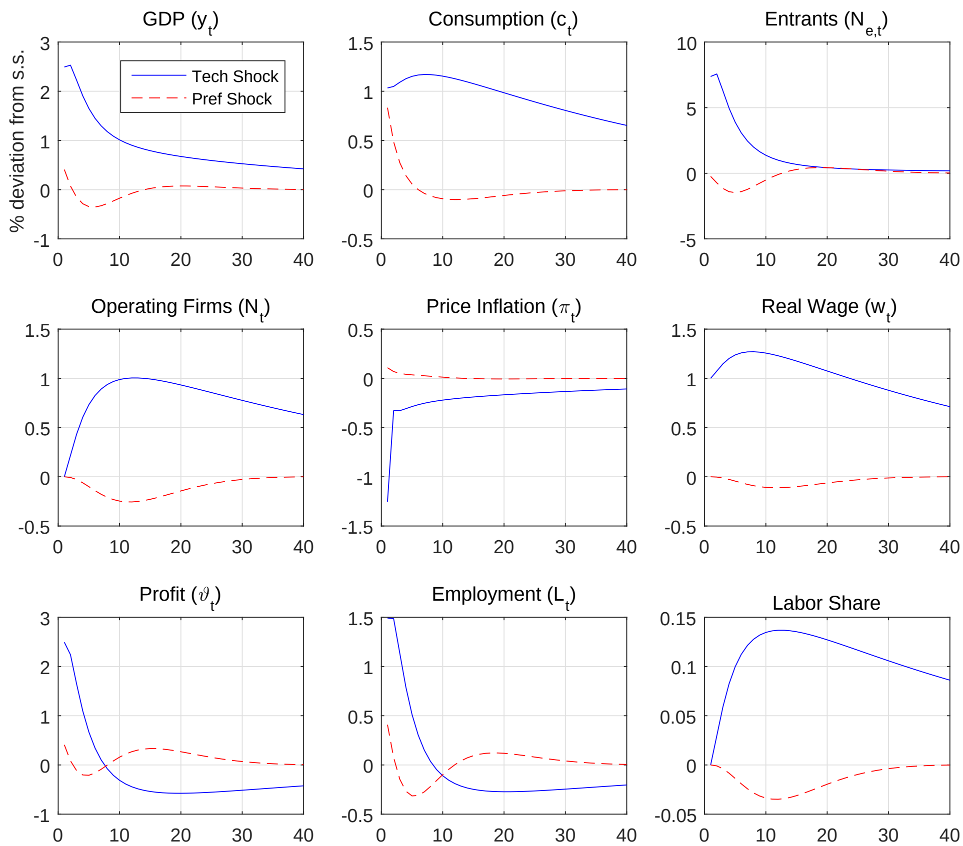
Table 2: Key Business Cycle Statistics for Historical Policy Rule[39]

Business Cycle Moments (Historical Policy Rule)

		Both Shocks		Technology Shocks		Preference Shocks	
	Data	Log-normal	Pareto	Log-normal	Pareto	Log-normal	Pareto
GDP (σ_y)	2.05	2.63	2.65	2.40	2.38	2.32	2.39
σ_x/σ_y							
Firm Entry	3.64	2.86	2.83	3.10	3.14	3.22	3.12
Consumption	0.75	1.13	1.12	0.43	0.39	1.22	1.20
Employment	0.87	0.82	0.85	0.65	0.63	0.97	0.96

[39]Notes: The data target is $\sigma_{N_e} = 7.47$. For a log-normal distribution: both, technology, preference shocks require, $\gamma = \{0.229, 0.011, 0.165\}$. For a Pareto distribution: both, technology, preference shocks require, $\gamma = \{0.229, 0.003, 0.157\}$. All reported statistics are HP-filtered with a value of 1600.

Figure 1: Impulse Responses for Technology and Preference Shocks[40]



[40]Notes: Deviations from steady state (in percent) reported on the vertical axis. Quarters reported on the horizontal axis. The independent shocks to technology and preferences are of 1 percent.

Table 3: Welfare Loss and Business Cycle Moments[41]

Welfare Losses and Business Cycle Moments						
	Both Shocks		Technology Shocks		Preference Shocks	
	Log-normal	Pareto	Log-normal	Pareto	Log-normal	Pareto
Welfare Loss ($\mathbb{L}$)						
Optimal-No Firms	0.07	0.29	0.01	0.08	0.09	0.32
Historical Rule	0.29	0.43	0.08	0.11	0.25	0.42
Ramsey Policy: σ_x/σ_y						
Firm Entry	3.49	3.19	2.29	2.85	3.87	3.53
Consumption	0.58	0.58	0.96	0.69	0.27	0.39
Employment	0.86	0.82	0.5	0.73	0.97	0.39

[41] Notes: Welfare losses (in percent) are calculated as losses in steady-state consumption vs alternative policies, relative to the Ramsey-optimal policy. All reported business cycle statistics are HP-filtered with a value of 1600.