Supply Shocks in a Heterogeneous-Firm New Keynesian Model: The Entry Multiplier*

Florin O. Bilbiie† Marc J. Melitz‡

January 2026

Abstract

We study productivity shocks in a New Keynesian model with endogenous entry, selection, and nominal rigidities. Adjustment along the extensive margin fundamentally alters the transmission of TFP shocks. Under sticky prices, productivity disturbances generate a large "entry multiplier": firm entry-exit responds much more strongly than under flexible prices, even when output is allocatively efficient to first order. Introducing wage stickiness breaks this neutrality. Adverse TFP shocks reduce profits, trigger exit, and generate a negative output gap while remaining inflationary. Productivity shocks therefore behave as true supply shocks, without resorting to ad hoc cost-push or markup disturbances. Unlike standard markup shocks, TFP shocks in our framework imply procyclical profits and entry, consistent with the data. When wages are sufficiently sticky, expansionary monetary policy raises entry, closes the negative output gap, and improves welfare. The model remains analytically tractable and isomorphic to the standard New Keynesian framework, with entry and selection appearing as simple wedges.

OLD: Due to its impact on nominal firm profits, price rigidity amplifies the response of entry and exit to supply shocks. When those supply shocks are negative, such as those following supply chain disruptions, this "entry multiplier" substantially magnifies the associated welfare losses—especially when wages are also rigid. This is in stark contrast to the benchmark New Keynesian model (NK), which predicts a *positive* output gap in response to that same shock under the same monetary policy. Endogenous entry thus radically changes the consequences of nominal rigidities. In addition to the aggregate-demand amplification of supply disruptions, our model also reconciles the response of hours worked across the NK and RBC models. And unlike the standard NK model, our model can also be used to evaluate how monetary expansions can alleviate or even eliminate the *negative* output gap induced by supply disruptions.

JEL Codes: E3, E4, E5, E6

Keywords: entry; Aggregate Demand and Supply; Variety; Sticky Prices; Sticky Wages; Monetary Policy; Recessions.

*We thank the Editor Nir Jaimovich and three anonymous referees, as well as Susanto Basu, Cristiano Cantore, Vasco Carvalho, Larry Christiano, Giancarlo Corsetti, Martin Eichenbaum, Andrea Ferrero, Francesco Furlanetto, Jordi Galí, Pierre-Olivier Gourinchas, Basile Grassi, Masahige Hamano, Antoine Lepetit, Virgiliu Midrigan, Sara Moreira, Anthony Savagar, Petr Sedlacek, Riccardo Silvestrini, Vincent Sterk, Paolo Surico, Mathias Thoenig, Michael Woodford, Francesco Zanetti, and participants in numerous seminars and conferences for useful comments and discussions.

†University of Cambridge and CEPR; florin.bilbiie@gmail.com.

‡Harvard University, CEPR, and NBER; mmelitz@harvard.edu

1 Introduction

Notation:

- Fixed level: X°
- Steady state: X for X_t
- Flex price: X_t^* for X_t
- Log deviation:
$$\hat{X}_t = \ln X_t - \ln X$$
- Gap notation:
$$\tilde{X}_t = \hat{X}_t - \hat{X}_t^*$$
- Wage stickiness: κ_w for κ, θ_w for θ, $\psi_w$for ψ
- Averages: $\bar{X}_t$ for X_t (arithmetic or CES average)
- Cutoff: q_{0t} for q_t
- Wages: w_t real and W_t nominal

How should productivity shocks affect inflation, economic slack, and the conduct of monetary policy? In the standard New Keynesian framework, transitory TFP shocks are not "true" supply shocks: under nominal rigidities, *adverse* productivity shocks raise inflation and generate a *positive* output gap, calling for *contractionary* policy. To reconcile this implication with stagflationary episodes, the literature typically introduces ad hoc cost-push or markup shocks, which shift marginal costs independently of productivity. While useful for matching inflation dynamics, these shocks—other than lacking microfoundations—generate counterfactual implications for profits, entry, and firm dynamics.

This paper offers an alternative mechanism. We show that once firm entry is endogenous, productivity shocks themselves acquire the properties of genuine supply disturbances. The key channel is an *entry multiplier*: under nominal rigidities, small productivity shocks induce large movements in the extensive margin of production, with first-order effects on aggregate demand, inflation, and welfare. As a result, adverse TFP shocks generate inflation and a negative output gap—without invoking exogenous markup shocks—and monetary policy plays a fundamentally different role.

The negative supply impulse that we consider first is a classic negative productivity shock: a downward shift in the production function, such as the one associated with severe restrictions on the availability of inputs. Such a shock can be regarded as a metaphor for the aftermath of the

COVID recession, supply chain disruptions, the Ukraine war, and the China shock.[1]

Our first result isolates the entry multiplier in its most distilled form. In a model with endogenous entry but flexible wages, we show that TFP shocks generate substantially larger responses of firm entry under sticky prices than under flexible prices. This amplification operates entirely through the extensive margin. Nominal rigidities induce changes in profits that trigger entry dynamics, setting off a feedback loop to (endogenous) aggregate productivity. Consider a negative shock. Firms wish to increase their price to reflect their increased marginal cost. With sticky prices they cannot, so they are "stuck" with their suboptimal price. This induces further losses and triggers further exit, engendering an additional (endogenous) aggregate productivity decrease that amplifies the initial impulse.[2]

At the same time, output coincides with its flexible-price counterpart to first order: despite large movements in entry, there is no output gap. This "envelope" or neutrality result—derived analytically—provides a clean benchmark: nominal rigidities dramatically magnify entry responses, yet leave allocative efficiency unaffected absent wage stickiness.

Introducing sticky wages fundamentally changes this conclusion. Our second main result is that adverse TFP shocks now generate a negative output gap, that is, a recession relative to the flexible-price-and-wage allocation. Lower productivity reduces profits and expected profitability; under sticky prices, firms cannot restore margins quickly, leading to exit and a persistent contraction in the mass of producers. Sticky wages prevent hours from adjusting efficiently, so actual output falls by more than potential. In this environment, TFP shocks behave as true supply shocks: they are inflationary and simultaneously create slack.

This mechanism overturns a central prediction of the standard New Keynesian model. There, negative TFP shocks raise inflation but imply a positive output gap, observationally indistinguishable from demand expansions. In our framework, the comovement is reversed. Inflation and slack move in opposite directions, as in stagflationary episodes, but without relying on exogenous markup shocks. Instead, entry and exit act as an endogenous cost-push wedge in the Phillips curve—a result we characterize analytically. Importantly, unlike standard cost-push shocks, TFP shocks in our model generate procyclical profits and entry, consistent with the data.

Our third main result concerns monetary policy. When wages are sufficiently sticky relative to

[1]Such a shock can be regarded as a metaphor for the recent COVID-19 crisis, which also highlighted the importance of the extensive margin, given the associated very sharp responses in entry and exit of businesses and varieties; an older working paper version of this paper reviews empirical evidence for these entry dynamics in the COVID crisis.

[2]This mechanism captures an intuition that is more general than the inability to reset prices. It applies more generally to profitability shocks induced by nominal rigidities. Thus, this is a reduced form for frictions that impinge upon intensive-margin adjustments, with negative consequences for profitability.

prices, a monetary expansion raises profits, stimulates entry, and closes the negative output gap induced by adverse TFP shocks. In this region, policy can fully restore efficient output and improve welfare. This stands in sharp contrast to the standard New Keynesian prescription, where contractionary policy is required in response to both adverse TFP shocks (which generate positive gaps) and stagflationary markup shocks. In our framework, expansionary policy is appropriate precisely because productivity disturbances reduce demand through endogenous capacity destruction.

We characterize analytically the degree of wage stickiness required for entry to respond positively to monetary expansions. When this condition holds, monetary policy operates through the extensive margin: higher demand raises profitability, induces entry, and expands productive capacity. This channel is absent in representative-firm models and reverses the standard prediction that supply-driven inflation necessarily calls for tightening.

Finally, we show that introducing firm heterogeneity and endogenous selection affects magnitudes but not mechanisms. The model remains mathematically isomorphic to the standard New Keynesian framework, with entry and selection appearing as tractable wedges in the Phillips curve and resource constraints. This structure allows us to derive closed-form solutions for the entry multiplier, the output gap, and welfare effects, while preserving the familiar logic of New Keynesian policy analysis. The simple NK model with heterogeneous firms and selection solved analytically may be of independent interest to some researchers.

Taken together, our results reframe the role of productivity shocks in monetary economies. Endogenous entry transforms TFP shocks into genuine supply disturbances—inflationary, contractionary, and welfare-relevant—eliminating the need for ad hoc markup shocks and restoring coherence between inflation dynamics, slack, profits, and policy prescriptions.

␣

XX–OLD VERSION–XX
When and how do nominal rigidities *amplify* supply disruptions? That is, when do negative supply shocks generate an aggregate-demand recession, understood as a greater fall in the level of output and income relative to the benchmark case without nominal rigidities? We show that this is an inherent feature of a business cycle model with *endogenous* entry *and product variety*—whereas the standard "New Keynesian" model with no entry (hereafter NK) predicts a *positive* output gap in response to a negative productivity shock. The endogenous responses of entry associated with those nominal rigidities thus plays a key role in delivering this important business cycle comovement, and in shaping the design of monetary policy in response to it. The key intuition is that nominal rigidities also *distort the extensive margin*. We start with the simplest pared-down model in order to distill our main channel. We then show how when both price and wage rigidities are present, this leads to negative output gaps in response to negative

TFP shocks, overturning an implication of the standard NK model; furthermore, we show that in such a supply-driven recession, expansionary monetary policy closes the output gap and is welfare-improving; this too is in contrast to the standard NK model, where the policy response to a negative TFP shock (or a recession-inducing cost-push shock) is always contractionary.

The negative supply impulse that we consider first is a classic negative productivity shock: a downward shift in the production function, such as the one associated with severe restrictions on the availability of inputs. Such a shock can be regarded as a metaphor for the aftermath of the COVID recession, supply chain disruptions, the Ukraine war, and the China shock. Finally, we study the role of selection (firm heterogeneity) for these aggregate dynamics, both analytically and numerically. The simple NK model with heterogeneous firms and selection may be of independent interest to some researchers.[a] We first show that the response of entry to such supply shocks is amplified in a model with sticky prices, when firms cannot optimally adjust prices in response to productivity shocks: a supply-side phenomenon we dub *the* entry *multiplier*. Nominal rigidities induce changes in profits that trigger entry dynamics, setting off a feedback loop to (endogenous) aggregate productivity. Consider a negative shock. Firms wish to increase their price to reflect their increased marginal cost. With sticky prices they cannot, so they are "stuck" with their suboptimal price. This induces further losses and triggers further exit, engendering an additional (endogenous) aggregate productivity decrease that amplifies the initial impulse.[b] The endogenous fall in aggregate productivity is driven by the inefficient equilibrium level of variety.

In our benchmark economy where output is a constant elasticity of substitution aggregate of intermediate inputs, henceforth CES, entry responds proportionately with the supply shock when prices are flexible. This is the well-known market size effect on entry with constant markups going back to Krugman (1980). When prices are sticky, however, the same supply shock leads to a more than proportionate response of entry. This is the entry "multiplier" under sticky prices. While this channel is present and operates in existing models of monetary policy with entry and nominal rigidities (see i.a. Bilbiie, Ghironi, and Melitz 2007; Bergin and Corsetti 2008; Bilbiie, Fujiwara, and Ghironi 2014; Bilbiie 2021), it has not been identified, isolated, and analytically characterized before. This is our paper's first contribution.[c] This amplification is important in and of itself, but also most critically for its consequences: the further amplification of the output response leading to a demand-determined recessions. This transmission channel has not been previously analyzed.

We characterize the conditions under which *aggregate-demand amplification* of supply shocks (a magnified response of aggregate output under sticky prices and wages) occurs—relative to the flexible-price benchmark. It is well-known that such a negative supply shock *cannot* drive a demand recession in a standard NK model, wherein (given a standard monetary policy rule) a temporary negative productivity shock implies an output gap *increase*: a smaller output fall under sticky than under flexible prices.[d] Our second contribution is thus to analyze the entry multiplier's ensuing impact on aggregate demand.

With *no* (or *exogenous*) entry and exit—such as the standard NK model—the response of aggregate activity is proportional to the adverse supply shock when prices are flexible: if productivity falls by 1%, consumption and output fall by 1%. In this case, the response is at most proportional, and generally smaller than 1% under sticky prices. In other words, there is aggregate-demand *dampening* of supply shocks, an issue well-known in NK models.[e]

With *endogenous* entry and exit, there is amplification of the aggregate response relative to this no-entry model, even under *flexible* prices. This is due to the "increasing returns" inherent in an expanding-variety model magnifying the effect of productivity shocks, whereby entry variations act as endogenous aggregate productivity.[f] But there is *further* amplification under sticky prices and wages.[g] Thus, endogenous entry radically changes the consequences of nominal rigidities for supply disruptions: price and wage stickiness *dampens* the aggregate response without entry, but it *amplifies* that aggregate response with entry. The requirement for the entry multiplier and for aggregate-demand amplification of *aggregate* supply shocks is that the elasticity of substitution between goods be *higher* than the elasticity of intertemporal substitution in consumption. Virtually all empirical estimates for those two elasticities satisfy this ranking.

Our full model thus features sticky wages in addition to sticky prices, as well as firm heterogeneity and selection. We show that the latter does not interact with nominal rigidity—equilibrium selection is operative and shapes the affects of aggregate shock and policies, but it does so independently of price or wage rigidity. Wage stickiness has a crucial role that is two-folded: first, it implies first-order amplification effects arising from the inefficiency of labor allocation due to cyclical variation sin the wage markup. And second, it allows the model to deliver an entry response to monetary policy that is intuitive and in line with the data: by making markups less countercyclical and profits per firm procyclical, wage stickiness implies an increase in entry in response to a monetary expansion; we provide an analytical expression for the threshold value of stickiness parameters that ensure this comovement. We then show that expansionary monetary policy can be employed to mitigate—and even completely close—the negative output gap induced by supply disruptions.

It should be emphasized that it is really the interaction of these nominal rigidities with entry that gives rise to all these desirable business-cycle properties; indeed, we show—both in the quantitative version and analytically—that a model with the same rigidities but with fixed entry (the standard New Keynesian model with both rigidities) still suffers from the same well-known issues, in particular a positive output gap in response to a negative productivity shock, and a monetary contraction in order to mitigate it and improve welfare.[h] Therefore, we conclude that the New Keynesian framework needs to include endogenous entry as well as both nominal rigidities (in prices and wages) in order to deliver *at the same time* (i) demand recessions in response to negative supply shocks, (ii) realistic dynamics following demand shocks, and (the combination of i and ii) (iii) an expansionary monetary policy as the optimal response to a negative supply shock inducing a negative output gap.

[a]Such a shock can be regarded as a metaphor for the recent COVID-19 crisis, which also highlighted the importance of the extensive margin, given the associated very sharp responses in entry and exit of businesses and varieties; an older working paper version of this paper reviews empirical evidence for these entry dynamics in the COVID crisis.

[b]This mechanism captures an intuition that is more general than the inability to reset prices. It applies more generally to profitability shocks induced by nominal rigidities. Thus, this is a reduced form for frictions that impinge upon intensive-margin adjustments, with negative consequences for profitability.

[c]Our results generalize to models of entry with sunk costs where the number of firms acts as a state variable providing propagation and matching profits' dynamics, such as Bilbiie Ghironi, and Melitz (2007, 2012) and Gutierrez, Jones, and Philippon (2021). We nevertheless focus on the free-entry, zero-profits model of entry with a fixed per-period cost for analytical tractability, as in Jaimovich and Floetotto (2008) for flexible prices and Bilbiie (2021) for sticky prices.

[d]Our analysis assumes throughout that the central bank does not act in order to completely "undo" the effect of nominal rigidity (which it can do by changing money supply or interests rates). That is, we derive the implications of supply shocks for a given, suboptimal monetary policy rule—the same suboptimal rule in both models, with and without entry. We then return to the analysis of monetary policy itself.

[e]The response of output with sticky prices itself can be positive (if prices are not entirely fixed, etc.) – but the key point is that, for plausible monetary policy rules, it is always less than one. That is, the output gap (the key summary statistic) is positive in response to negative supply shocks (the response under sticky is smaller than under flexible prices).

[f]This amplification is studied in detail i.a. in Bilbiie, Ghironi, and Melitz (2012) in a model with sunk-cost dynamic entry, and earlier in Devereux, Head and Lapham (1996) and Chatterjee and Cooper (1993) in a "static entry" model. It is also related to the welfare gain of trade and market size in the "new trade theory" with monopolistic competition, e.g. Melitz (2003). Gopinath and Neiman (2014) provide trade-based empirical evidence for the negative efects of

adverse shocks on endogenous productivity.

[g]In this benchmark sticky-price-only case, the responses to shocks under either sticky or flexible prices are identical to the first order; amplification of *negative* shocks still occurs but only through higher-order nonlinear terms driven by the curvature of output in intermediate input variety. We analyze these in an Appendix and the previous working paper version. Such nonlinear terms make the effect larger and are particularly relevant for large shocks like those associated with the COVID-19 crisis, where such nonlinearities are likely to be especially important.

[h]We also show that a model with entry but sticky wages only (flexible prices) suffers from a different issue: since firms can restore profitability by resetting prices and consumers can substitute intertemporally, a negative supply shock implies a future expansion.

Related literature

A large and growing literature emphasizes the role of endogenous entry and variety with *flexible prices* for business cycles, studying macro fluctuations and normative properties i.a. Devereux, Head, and Lapham (1996), Jaimovich (2007), Jaimovich and Floetotto (2008), Colciago and Etro (2010), and Bilbiie et al (2012, 2019). Important recent extensions studied the role of firm heterogeneity and selection, see i.a. Clementi and Palazzo (2015), Lee and Mukoyama (2018), Hamano and Zanetti (2017), and Edmond, Midrigan, and Xu (2020).[3]

Several papers have analyzed these models with nominal rigidities, focusing on the effects and design of monetary policy, i.a. Bilbiie et al (2007, 2014); Bergin and Corsetti (2008); Lewis and Poilly (2014); Bilbiie (2021), and Gutierrez et al (2021). More recently, several subsequent contributions extended this framework to study the interaction of nominal rigidities, endogenous entry, and selection with firm heterogeneity: see Cooke and Damianovic (2021), Colciago and Silvestrini (2022), and Hamano and Zanetti (2022); we also incorporate heterogeneity and the associated selection margin, drawing on these contributions but in an analytically tractable way that may be of independent interest—we show that while selection does effect the magnitude of aggregate effects, our qualitative results remain unchanged.

The standard NK model's failure to produce demand-induced recessions in response to negative supply shocks is the starting point of another recent paper by Guerrieri et al (2020).[4] The authors analyze the impact of the exogenous elimination (exit) of a sector in a 2-sector model. When the goods produced by the two sectors are complements and the intertemporal elasticity of substitution is elastic, this supply disruption induces a demand-induced recession, which the authors call "Keynesian supply shocks". In contrast, we model economy-wide supply disruptions

[3]Other recent developments include Cacciatore and Fiori (2016), Dixon and Savagar (2020), and Michelacci, Paciello, and Pozzi (2019) on unemloyment, endogenous productivity, and demand, respectively. Other earlier pioneering contributions on RBC-like models with entry include Campbell (1998), Chatterjee and Cooper (2014), and Cook (2001).

[4]In isolation, the shortcomings of the standard NK model pertaining to some of the comovements we analyze have also been studied previously. In their seminal contribution, Christiano, Eichenbaum, and Evans (2005) already demonstrated how wage stickiness is necessary, in a fixed-variety model, to replicate the positive response of profits to monetary policy found in the data (see also Christiano, Eichenbaum, and Evans, 1997).

when product-level entry is endogenous and those goods are substitutes. We do not restrict the intertemporal elasticity of substitution so that our mechanism operates when this elasticity is low (inelastic). Thus, our focus is on a different channel that is complementary to the sector-level disruptions examined by Guerrieri et al (2020). Furthermore, importantly—our mechanism generates a negative comovement of inflation and the output gap, whereas in Guerrieri et al "Keynesian" supply shocks are recessionary and also deflationary. Our mechanism makes TFP shocks supply-supply.[5]Finally, our model with sticky wages also reproduces the cyclicality of profits to both supply and demand shocks. See Christiano et al (2005) for the role of wage stickiness in reproducing procyclical profits after demand shocks in a DSGE model and Cantore et al (2011) and Bilbiie and Känzig (2023) for analytical results on wage stickiness and the cyclicality of real wages and profits, respectively; the latter paper also provides VAR evidence for the response of profits to identified monetary policy, TFP, and cost-push (energy price) shocks.

[6] Our model with all three ingredients (endogenous entry, and sticky prices and wages) addresses *all* these comovements jointly.

2 The Entry Multiplier and Aggregate Demand

In this section, we outline the simplest version of our model of endogenous entry with nominal rigidities, focusing on sticky prices. Households maximize the expected present value of utility defined over a consumption good C_t and hours worked L_t, where total consumption is equal to the output of a final-good sector consisting of a CES aggregate of intermediates.

We consider a standard CRRA utility:

$$U\left(C_t, L_t\right) = \frac{C_t^{1-\sigma^{-1}} - 1}{1-\sigma^{-1}} - \frac{L_t^{1+\varphi}}{1+\varphi}.$$

The weight on the disutility of labor is normalized to 1. For now, we model nominal rigidities with fixed prices for intermediates. In our full model, we will model those rigidities more generally

[5]Cesa-Bianchi and Ferrero (2021) quantify empirically the contribution of sectoral shocks to aggregate fluctuations. Other contributions emphasize related supply-side mechanisms, e.g. exit following the revenue fall due to restrictions on a subset of products with rigid operating costs (Auerbach, Gorodnichenko, and Murphy (2021)); inter-sectoral linkages and complementarities (Woodford (2020)); unemployment and endogenous growth (Fornaro and Wolf (2020)), and input-output networks (Baqaee and Farhi (2020)). We abstract from such features to focus on endogenous entry and its interaction with nominal rigidities.

[6]A separate issue is the response of hours worked to supply shocks, see e.g. e.g. Gali (1999), Basu, Fernald and Kimball (2006), Christiano, Eichenbaum, and Vigfusson (2003), Chari, Kehoe, and McGrattan (2008). Cantore et al (2014) show how factor-augmenting shocks and capital-labor substitutability contributes to resolving that debate. A different literature (e.g. Kaplan and Zoch, 2020) studies the the labor *share response to demand* shocks under nominal rigidities.

using sticky prices as well as sticky wages. We discuss the implications for both types of nominal rigidities in detail later on. For now, price rigidity alone is sufficient to develop the main intuition for our aggregate-demand amplification channel.

2.1 Static Equilibrium with Endogenous Entry

At time t, the household consumes C_t, equal to final good production Y_t.[7] The latter is produced using a continuum of intermediate inputs with measure N_t:

$$Y_t = \left(\int_0^{N_t} y_t\left(\omega\right)^{\frac{\theta-1}{\theta}} d\omega \right)^{\frac{\theta}{\theta-1}}, \tag{1}$$

where $\theta > 1$ is the symmetric elasticity of substitution across intermediate goods. Let $p_t\left(\omega\right)$ denote the nominal price of good ω and $P_t = \left(\int_0^{N_t} p_t\left(\omega\right)^{1-\theta} d\omega \right)^{\frac{1}{1-\theta}}$ the price of the final good. The demand for each intermediate ω is then $y_t\left(\omega\right) = \left(p_t\left(\omega\right)/P_t\right)^{-\theta} Y_t$.[8] We assume

$$\theta > \sigma > 1, \tag{2}$$

which ensures that the intermediate varieties ω are Edgeworth-substitutes satisfying $\partial^2 U/\partial y(\omega_1)\partial y(\omega_2) > 0$.[9]

There is a continuum of monopolistically competitive firms, each producing a different intermediate $\omega \in [0, N_t]$. Production requires only one factor, labor, whose productivity is scaled exogenously by a factor A_t. We will pay particular attention to the macroeconomic consequences of substantial negative TFP shocks to A_t (in our simple framework, this is identical to a downward shift in labor supply). Output supplied by firm ω is:

$$y_t(\omega) = A_t l_t(\omega),$$

where $l_t\left(\omega\right)$ is the firm's employment. The constant real marginal cost is then w_t/A_t.

We consider a symmetric equilibrium with N_t firms and drop the ω qualifier. The relative price of intermediates in units of the final good is a key object that captures the aggregate productivity

[7]The two will be distinct after we introduce nominal rigidities.

[8]This specification follows Ethier (1982) and Romer (1987)'s extension of the Spence-Dixit-Stiglitz aggregator. Our results carry through, albeit with some differences in interpretation, to a setup where the CES aggregate is defined over individual varieties of consumption goods instead.

[9]Although $\theta > 1$ uniquely identifies the CES between varieties in consumption C_t, utility $U(C_t, L_t)$ also includes an additive leisure term. This explains the additional restriction on the inter-temporal elasticity σ for substitutability.

benefit of input variety, also known as "increasing returns to specialization":

$$\rho_t \equiv \frac{p_t}{P_t} = N_t^{\frac{1}{\theta-1}}. \tag{3}$$

Variations in the number of goods determine endogenous aggregate productivity, an insight that is at the core of all the expanding-variety endogenous growth literature.

Let μ_t denote the firms' markup (potentially time-varying):

$$\mu_t \equiv \frac{\rho_t}{w_t/A_t}. \tag{4}$$

Firm profit (d for dividends) in period t is:

$$d_t = \frac{p_t}{P_t} y_t - w_t l_t = \left(1 - \frac{1}{\mu_t}\right) \rho_t y_t. \tag{5}$$

We assume free entry subject to an entry cost f_E in units of labor:

$$d_t = w_t f_E.$$

The household's budget constraint is reflected in the aggregate accounting identity equating expenditures (consumption plus the fixed cost "investment" for all firms) with income (labor income and variable profits for all firms). That is:

$$Y_t + N_t w_t f_E = w_t L_t + N_t d_t \iff Y_t = w_t L_t, \tag{6}$$

using free entry. The associated scale of the firm yields the firm-level labor demand:

$$l_t = \frac{1}{\mu_t - 1} f_E.$$

We can then write the aggregate production function as:

$$Y_t = N_t^{\frac{\theta}{\theta-1}} y_t = N_t^{\frac{\theta}{\theta-1}} A_t \frac{1}{\mu_t - 1} f_E. \tag{7}$$

A key equation is *labor demand*, obtained by aggregating l_t across producers:

$$L_t = \frac{\mu_t}{\mu_t - 1} N_t f_E. \tag{8}$$

On the household side, the consumption-leisure choice yields *labor supply*:

$$(L_t)^{\varphi} = (C_t)^{-1/\sigma} w_t, \tag{9}$$

with $C_t = Y_t$.

MM: Combined with the markup rule, this yields:

$$w_t = A_t \left(\frac{\mu_t - 1}{\mu_t} \frac{1}{f_E} L_t \right)^{\frac{1}{\theta - 1}} \frac{1}{\mu_t}.$$

Three important observations are in order: first, endogenous entry implies that the aggregate labor demand is upward sloping. Its slope is the degree of increasing returns. Second, aggregate labor demand shifts as usual with changes in labor productivity, but that effect is amplified here by the increasing returns; and finally, aggregate labor demand shifts with endogenous changes in markups.[a]

[a] As we will show, this effect of markups is also present in sticky-price models with fixed entry. However, in our model with endogenous entry, the markup will also depend on the endogenous response in the number of firms.

2.2 Aggregate Demand and Monetary Policy

An important distinction concerns input versus final-good prices and their corresponding inflation rates. We refer to the former as the producer price p_t and to the latter as the consumer price P_t. Producer-price inflation $1 + \pi_t = p_t / p_{t-1}$ and consumer-price inflation $1 + \pi_t^c = P_t / P_{t-1}$ are related to the growth in the number of intermediates through (3):

$$\frac{1 + \pi_t}{1 + \pi_t^c} = \left(\frac{N_t}{N_{t-1}} \right)^{\frac{1}{\theta - 1}}. \tag{10}$$

This distinction is particularly important with nominal rigidities because these apply at the *individual* firm-level price p_t. The relevant inflation rate for aggregate demand is *consumer* inflation π_t^c, insofar as it determines the ex-ante real interest rate that governs inter-temporal substitution. Indeed, the solution to the household's inter-temporal problem is the standard Euler equation for consumption:[10]

$$(C_t)^{-1/\sigma} = \beta E_t \left(\frac{1 + i_t}{1 + \pi_{t+1}^c} (C_{t+1})^{-1/\sigma} \right), \tag{11}$$

where i_t is the nominal interest rate. The model is closed by specifying the price-setting equation—delivering a Phillips curve for PPI inflation—and monetary policy. The latter is set according to a

[10] The full solution also implies a standard transversality condition.

Taylor rule, responding to expected future producer price inflation π_t with elasticity $\phi \geq 1$:[11]

$$1 + i_t = \beta^{-1} (1 + E_t \pi_{t+1})^{\phi} \exp(-\varepsilon_t), \tag{12}$$

The variable ε_t has two possible interpretations: in one, it is an exogenous (demand) shock—a discretionary decision by the central bank unrelated to current economic conditions—the general-equilibrium effects of which we can trace out. In another interpretation, ε_t is the "intercept" of the Taylor rule, which can be actively managed by the central bank in order to cushion the effects of structural shocks. We shall consider both, being explicit accordingly.

Under a Taylor rule that fixes the ex ante real rate $\phi = 1$, replacing it in the Euler equation (11) and using the definition of CPI inflation (10) and ρ (3) delivers:

$$\rho_t (C_t)^{-1/\sigma} = E_t \left(\rho_{t+1} (C_{t+1})^{-1/\sigma} \right) \exp(-\varepsilon_t) \implies \rho_t (C_t)^{-1/\sigma} = \exp(-\epsilon_t), \text{ where } \epsilon_t \equiv E_t \sum_{j=0}^{\infty} \varepsilon_{t+j}. \tag{13}$$

is the "long" interest rate, measured as the sum of the future changes in interest rates. This is a measure of the monetary stance. We leverage this policy rule for deriving some of our analytical results.

2.3 The Entry Multiplier: Closed-form Solution

In order to distill our core entry multiplier channel, we first consider an extreme form of sticky prices that are indefinitely fixed along with a restriction to the case of **log-consumption utility** ($\sigma = 1$). In the next section, we return to the more general form of preferences along with arbitrary price stickiness. Logarithmic utility in consumption implies that income and substitution effects for labor supplied cancel out. The equilibrium hours worked is then fixed at $L_t = 1$, reflecting the normalization that we used for the disutility of labor in the utility function (see 9).

Under *flexible prices* (denoted with a * superscript), the markup is constant and given by the elasticity of substitution between goods: $\mu_t^* = \theta / (\theta - 1)$. The equilibrium is then fully determined

[11] Results are almost identical when using a Taylor rule that responds to contemporaneous inflation π_t. The advantage of the rule with future inflation is that it nests a policy that fixes the ex ante real rate (when $\phi = 1$). Building on Bilbiie (2008), we use this in the derivation of our analytical results – including those with arbitrary price and wage stickiness.

by the fixed labor supply along with the aggregate production (7) and labor demand (8):

$$N_t^* = \frac{1}{\theta f_E},$$
$$Y_t^* = \frac{\theta - 1}{\theta\left(\theta f_E\right)^{1/(\theta-1)}} A_t.$$

In this simple version with representative firms and log-utility, entry does not respond to TFP A_t under flexible prices (only firm size changes). In the next section, we will show that either selection or more general preferences ($\sigma > 1$) are sufficient to generate a response of N_t^* to A_t: both the extensive and intensive margins respond.

MM: Transition to fixed price where there is a response of entry

---

As we detail below, this contrasts sharply with the expansion/contractions at the intensive margin of firm-size when the extensive margin is exogenously fixed.[12]

Under *fixed prices*, we assume that each firm's price is fixed at $p_t = p^\circ$. The markup μ_t is then endogenous and output is demand determined. With our restriction to log-utility and the fixed-real-rate rule used above, the aggregate demand (13) becomes, using $C_t = Y_t$:

$$M_t \equiv P_t Y_t = \exp \epsilon_t, \text{ where } \epsilon_t \equiv E_t \sum_{j=0}^{\infty} \varepsilon_{t+j}. \tag{14}$$

This is the money demand "quantity equation": $M_t = P_t Y_t$, where M_t also has the interpretation of a money supply controlled by the central bank. Later, we solve the dynamic version of the model with a Phillips curve and Taylor rule that does not entirely neutralize PPI inflation. Writing the price index as a function of the fixed price $P_t = p^\circ N_t^{-\frac{1}{\theta-1}}$ and plugging into the money demand equation yields:

$$Y_t = \frac{M_t}{p^\circ} N_t^{\frac{1}{\theta-1}}.$$

This, along with fixed labor supply $L_t = 1$, aggregate production (7), and labor demand (8) fully

[12]We could also add features in our endogenous entry model that would deliver an intensive margin response along with the extensive margin. But the key distinction is the endogenous extensive margin.This feature of the equilibrium is due to the combination of free entry (no sunk-cost delays) and fixed costs' being denominated in the output of the respective intermediate. Deviating from either of these assumptions would generate some adjustment in the intensive margin too.

determines the fixed-price equilibrium. The number of firms is then:

$$N_t = \frac{1}{f_E}\left(1 - \frac{1}{A_t}\frac{M_t}{p^\circ}\right).$$

MM: With sticky prices, now N_t responds (positively) to A_t

We use hats (ˆ) to denote percent deviations from steady state: $\hat{X}_t = \ln X_t - \ln X$. For any endogenous variable X_t, its steady state level (denoted without time subscript) X is determined assuming steady-state productivity $A = 1$ and the monetary stance M is chosen in the fixed-price models to equalize their steady-state counterpart in the flex-price version (see Appendix D for details).

We use tildes (˜) to denote "gaps" between the equilibrium variables under nominal rigidities and their flexible-price counterparts: $\tilde{X}_t = \hat{X}_t - \hat{X}_t^*$.

Proposition 1. ***The Entry Multiplier.*** *The response of the number of firms N_t to the supply shock A_t is magnified under fixed prices (relative to flexible prices):*

$$\frac{d\tilde{N}_t}{d\hat{A}_t} = \theta - 1 > 0$$

This is a powerful result that operates in models with entry and nominal rigidities. The intuition is very simple and general.

MM: Add footnote that we can extend to 1) $\sigma > 1$, 2) firm heterogeneity, 3) arbitrary stickiness in both prices and wages, and still get entry multiplier $d\tilde{N}_t/d\hat{A}_t > 0$ along with extensive margins response under flex prices $dN_t^*/d\hat{A}_t > 0$.

With sticky prices, the intensive margin cannot adjust in some key dimension and the extensive margin inefficiently bears all the adjustment. For a productivity decrease, the firm would like to increase its price to keep its scale constant, thus selling the same quantity at a higher price, but cannot (sticky prices). This generates a demand shortage and exit, with each remaining firm hiring more workers, producing more, and ending up "too large"; whereas with flexible prices, there would still be exit but each firm would keep its scale constant.[13] Firms are bigger than they would be absent price rigidities, and there are fewer of them. This is a distortion that increases with the

[13]Note that the effect of productivity on entry is symmetric for positive and negative shocks; as we discuss momentarily, this is no longer true for the effect on aggregate output.

demand elasticity θ. In other words, more intensive-margin adjustment would be desirable, and this is relatively more important when inputs are closer substitutes. This last argument is related to the impact on aggregate output, that we will study next.

We note that the equilibrium is determined by two key equations: 1. endogenous entry, implying zero aggregate profits; and 2. individual profit maximization, implying the pricing condition that marginal cost equal marginal revenue.[14] When (say) a negative exogenous productivity shock hits ($dA_t < 0$), there is a ceteris paribus decrease in profits per firm (keeping relative prices ρ_t fixed). *Free entry* implies the number of firms N_t goes down to restore zero profits. Due to increasing returns to specialization, this feeds back into a further—now *endogenous*—fall in aggregate productivity.

To find *how much* equilibrium entry occurs, we need to consider the pricing condition. Notice that marginal revenue is given by ρ_t/μ_t. Keeping the wage fixed, a productivity fall implies an increase in marginal cost. With flexible prices (at given N_t), the markup μ_t is constant and individual prices increase. But the equilibrium response of the *relative* price ρ_t depends on the extent of entry. Each individual firm contracts its labor demand, *and* there is a lower number of firms (one-to-one with the productivity decrease).

With sticky firm prices, marginal cost and revenue are still equalized. But now when firms' profits go down, they are stuck with prices too low, generating an additional incentive for exit. The markup goes down (it was constant under flex prices), which dampens the fall in individual labor demand. The number of firms, however, falls by more, generates exactly the same aggregate labor demand response. Thus, the relative price falls by more under sticky prices to compensate for the fall in the markup and generates the same real marginal cost (and revenue) regardless of whether prices are flexible or sticky. In other words, the final-good price (CPI) P_t falls by more under sticky prices.

The above discussion hints that our mechanism is likely to be more relevant and realistic for negative productivity shocks (rather than for positive shocks), given the relative timing of entry and pricing decisions: For a positive shock, an undesirable model feature is that entry happens before individual firms can adjust their price. This can be addressed by introducing a sunk cost, which is lower than the price adjustment cost (see Bilbiie, Ghironi, and Melitz 2007 for an example of that sunk-cost modeling).

Yet for negative productivity shocks, the same criticism has less bind. If firms are stuck with

[14] A key observation is that the labor market equilibrium is identical under flexible and sticky prices: the real wage and marginal cost change by exactly the same amount.

a price too high and a scale too large, a greater proportion of them fail. In case of a big negative shock, if it were possible to redistribute the fall in individual sales (intensive margin), then more firms would survive. But this is not possible, so disproportionately more firms fail. While price stickiness is probably not the most micro-plausible mechanism for this failure of intensive-margin adjustment, the firms' inability to increase prices enough in a slump certainly seems realistic for large and sudden negative productivity shocks. So we take price stickiness as a metaphor for firms' inability to contract even though a large negative profit results in exit. Furthermore, the difficulty of increasing prices to stabilize individual production is likely to apply to *product* (as opposed to *firm*) level, so the exit emphasized here applies as well to multi-product firms dropping products as it does to the exit of firms.[15]

We have simplified our model as much as possible to focus on the impact of supply shocks. In so doing, we have adopted a formulation of sticky prices that is too simple to adequately analyze the impact of demand (e.g. monetary) shocks in the presence of endogenous entry and exit. We describe the optimal monetary policy in this simplest version of our model in Appendix A.3. It features a "divine coincidence" result that is analogous to models with no entry but similar price rigidities and policy tools (Blanchard and Gali, 2007): the central bank can replicate the efficient flexible-price level of output while at the same time also stabilizing inflation. However, we postpone an analysis of monetary policy until our full model with both prices and wages is developed in Section 3. For given our current model with only sticky prices, a monetary expansion induces a counterfactual prediction of increased exit. The reason is a well-known feature of the standard NK model: markups and profits are countercyclical to demand shocks. As demand goes up, labor demand shifts up, increasing the real wage and real marginal cost and eroding margins; with free entry, this leads to exit. We show in Section 3 below that extending the model to introduce wage rigidity solves this issue and its monetary policy implications—while also providing a quantitatively powerful amplification mechanism.

2.4 The "Standard" New Keynesian Model with Exogenous Entry

It is useful for comparison to review the standard New Keynesian model with exogenous entry: a fixed number of varieties $N_t^{\mathcal{N}} = N^{\circ}$, which we normalize to 1. For reference, we denote throughout variables with a superscript $\mathcal{N}$ (No entry). Labor supply is still given by (9). The markup is similarly defined as $\mu_t^{\mathcal{N}} = A_t/w_t^{\mathcal{N}}$. This also defines labor labor demand, with no aggregate productivity benefit to variety. Furthermore, there is now no distinction between producer and

[15] See Argente et al (2018) for recent evidence on the cyclical relevance of this margin.

consumer prices. Since we normalize the mass of goods to 1, individual and aggregate variables coincide. The production function is $Y_t^{\mathcal{N}} = A_t L_t^{\mathcal{N}}$. Firm profit is $d_t^{\mathcal{N}} = \left[1 - \left(1/\mu_t^{\mathcal{N}}\right)\right] Y_t^{\mathcal{N}}$. Exogenous entry could be the result of a high enough entry cost f_E such that $d_t^{\mathcal{N}} < w_t^{\mathcal{N}} f_E$ for any realization of the TFP or monetary shocks.

Under *flexible prices*, optimal pricing implies a constant markup rule $\mu_t^{*\mathcal{N}} = \frac{\theta}{\theta-1}$. This implies that the wage is $w_t^{*\mathcal{N}} = \frac{\theta-1}{\theta} A_t$, with hours and consumption given by $L_t^{*\mathcal{N}} = \left(\frac{\theta-1}{\theta}\right)^{\frac{1}{1+\varphi}}$ and $Y_t^{*\mathcal{N}} = A_t L_t^{*\mathcal{N}}$. Hours are constant in this equilibrium, even though labor supply is endogenous, because income and substitution effects cancel out (a consequence of logarithmic utility in consumption). With *fixed prices* we now have $Y_t^{\mathcal{N}} = A_t L_t^{\mathcal{N}} = \frac{M_t}{P^\circ}$, or equivalently $L_t^{\mathcal{N}} = \frac{M_t}{P^\circ} \frac{1}{A_t}$. Hours go up when productivity goes down in order to keep consumption constant at the demand-determined level.[16]

MM: When prices are flexible, output is supply-determined and the demand side only determines nominal variables (the price level). When prices are fixed, output is demand-determined and the supply side determines hours worked (to produce the demand-determined output level).

2.5 Aggregate-Demand Amplification Through Entry

When does this entry multiplier of the supply-side productivity shock lead to aggregate demand amplification—a higher response of aggregate output (and consumption) relative to the flexible price equilibrium?

MM: Transition? Moved description of NK higher up...

We plot final output Y_t as a function of the shock A_t for the two equilibria in the left panel of Figure 1. Since output is demand-determined under flexible prices, it is the upward sloping line with slope L°. Under sticky prices, it is the horizontal line $Y_t^{\mathcal{N}} = \frac{M_t}{P^\circ}$. We choose the domain of A_t such that there is no rationing. That is, the equilibrium level of output is equal to demand and the adjustment is borne by hours worked. Those hours increase to compensate for the productivity-driven shortfall.[17]

[16] Assuming no rationing when hours $L_t^{\mathcal{N}}$ are below the total time endowment $L_{\max}$. This effect on hours is due to income effects. Note that wages and profits are: $w_t^{\mathcal{N}} = A_t^{-\varphi} \left(\frac{M_t}{P^\circ}\right)^{1+\varphi}$ and $D_t^{\mathcal{N}} = (M_t/P^\circ)\left(1 - A_t^{-(1+\varphi)} (M_t/P^\circ)^{1+\varphi}\right)$. Wages are countercyclical and profits pro-cyclical conditional on supply shocks. In particular, wages go up and profits down in response to a bad shock. Agents work more because of the extra income effect of profits relative to the free-entry ($Y_t = w_t L_t$) case, whereby income and substitution effects cancel out. See Gali (1999) and a large literature thereafter discussing the evidence for this co-movement.

[17] With a negative enough shock, demand can exceed supply (what can be produced), so the equilibrium amount produced and consumed would be represented by a kinked line (where to the left, the upward-sloping part would be supply-determined). The kink point itself is determined by labor supply elasticity: for instance with inelastic labor, any small negative TFP shock would lead to rationing. In particular, there is rationing as soon as $\frac{M_t}{P^\circ}\frac{1}{A_t} > L^\circ_{\max}$ (the

MM: The main point to emphasize is that efficiency of flex (with subsidy) holds in both standard NK and Endogenous entry (envelope property). BUT consumption-leisure choice is distorted in standard NK – so no envelope for consumption. The envelope still holds for consumption with endogenous entry because entry "eliminates" the consumption-leisure distortion. The only distortion is entry. Conditional on entry, the fixed-price equilibrium is efficient. Further note: The optimal subsidy to employment plays no role for the shape of the curves (not sure how much to emphasize this more subtle point).

The main takeaway is that, in response to a bad supply shock (lower A_t), output goes down proportionally under flexible prices. But under sticky prices, it either stays unchanged (if labor is elastic enough) or at most falls by as much as under flexible prices: in other words, there is never a demand shortage in response to a negative supply shock, and there can even be excess demand. The "output gap" is positive in response to supply disruptions. This is a well-known property of the standard no-entry sticky-price model restated here as a benchmark.

FB: Need to change the slopes for the red-dashed line to 1. Also see comment for Figure 3 regarding axis tick marks (and then the issue regarding the slope of 1 becomes irrelevant... And we don't need to discuss what value we choose for θ...

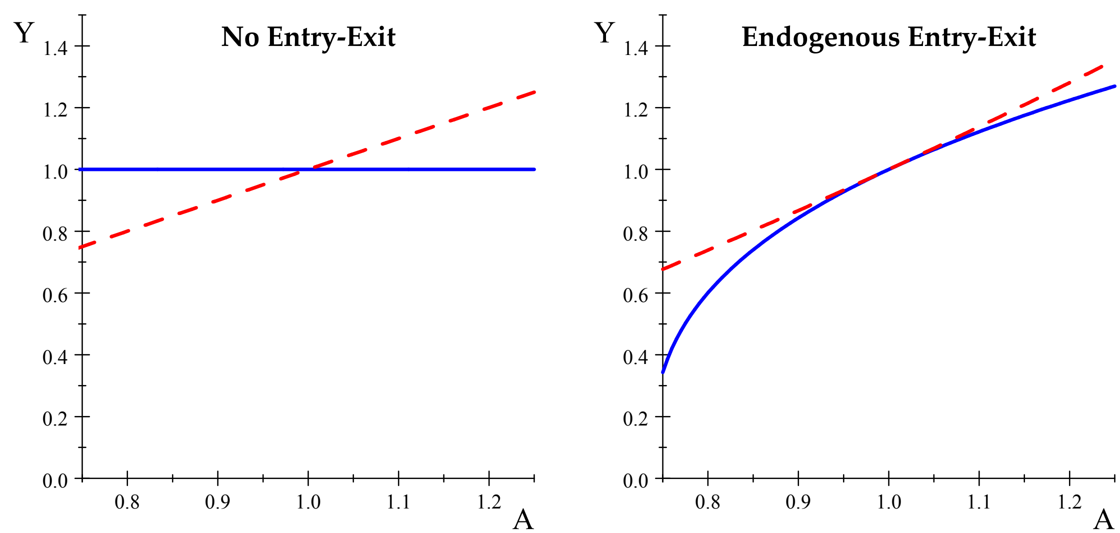


Figure 1: CES. Y^* (flex. price) red dash, Y (fixed price) solid blue

Consider now the role of endogenous entry and exit. In the right panel of Figure 1, we plot the flex- (red dash) and fixed (blue solid) price equilibria of our endogenous-entry model. The economies are again calibrated so that the steady-state equilibria ($A_t = 1$) coincide, and also coincide with the steady-state of the no-entry model (see Appendix A.1 for details). The only remaining

total time endowment), which calibrating real money balances to equate the two equilibria at $A_t = 1$ (in the absence of shocks) delivers $A_t < \frac{M^\circ}{P^\circ} \frac{1}{L^\circ_{\max}} = \left(\frac{\theta-1}{\theta}\right)^{\frac{1}{1+\varphi}} \frac{1}{L^\circ_{\max}}$.

free parameter is θ, which we set to 3.8, which is a conservative value in line with estimates from the trade literature.

The first order elasticities are identical under flexible and sticky prices. This is because under the CES aggregator, the market equilibrium is Pareto optimal: as in Dixit and Stiglitz, the number of input varieties is efficient. By an envelope argument, first-order deviations from that allocation are negligible (a consequence of the neutrality proposition in Bilbiie (2021)). As the figure makes clear, output under sticky prices is always lower than under flexible prices, but only through second-order effects in this simplest version. In particular, output falls by more in response to a bad supply shock when prices are sticky.[18] That is, the output gap is negative in response to supply disruptions—but only through second-order effects. Moreover, the same second-order effects engender welfare costs: we show in Appendix A.3 that even in this simple, first-order-neutral model there are second-order welfare costs of (squared) fluctuations in inflation and in the entry multiplier—the gap between the number of firms under sticky and flexible prices. First-order effects on welfare and the output gap occur in realistic extensions of the model, such as the one we study next—which features sticky wages in addition to prices, as well as heterogeneity and selection—and a more general utility function. As we show in the Online Appendix D.1, first-order effects can alternatively occur through external returns to variety.

In is worth emphasizing that in our economy with endogenous entry the "envelope" result holds for both output (consumption) and welfare: that is, the allocation of hours worked is efficient by virtue of the efficiency of entry combined with the $C_t = w_t L_t$ aggregate accounting—the wage is both the MRS and the MRT. This efficiency breaks when wage adjustment is no longer frictionless, which generates frictions in the allocation of labor, creating first-order effects. The key, as we shall see, is that the output effects of these inefficiencies in our economy with entry go in the exact opposite direction of an economy without entry.

3 The "Full" Heterogeneous-Firm NK Model with Endogenous Entry and Nominal Rigidities

We now develop our full model. On the "macro" side, we incorporate arbitrary levels of price and wage stickiness. And on the "micro" side, we introduce firm heterogeneity and selection. We show how such a model featuring the entry-multiplier described in the previous section replicates a se-

[18]Note that a similar "envelope" result holds in the no-entry economy, but for *welfare* only: the allocations of consumption and hours are both distorted in the sticky equilibrium. In the entry model instead, they are both efficient.

ries of business-cycle co-movements. The addition of sticky wages generates first-order effects for the aggregate-demand amplification of supply shocks and positive co-movements between monetary policy and entry. Yet, we also show how sticky wages *on their own* cannot generate the key aggregate-demand amplification that arises with sticky prices. This is why we started with a simpler model with just that single nominal rigidity in order to highlight the key intuition underlying that mechanism. Nominal rigidities take the form of (Rotemberg) quadratic adjustment costs for both prices and wages, delivering standard Phillips curves for both prices and wages.[19] The full model derivations are relegated to the Appendix A.[20]

Along with the more realistic modeling of price and wage rigidities, our full model also incorporates firm heterogeneity and selection. Although those forces play an important role in shaping the aggregate returns for the economy, we show that they do not interact with the entry multiplier channel that we previously highlighted. Or put differently, selection is "orthogonal" to the differences between the flexible and sticky price equilibria. We incorporate firm heterogeneity in both productivity and quality.

FB: We need a bit more description of these papers for the related literature. There are tons of paper with firm heterogeneity and dynamics under flex prices – so there should be another relevant feature to highlight the Clementi-Pallazo paper? And I think we need a sharper description/differentiation for the other papers with nominal rigidities? Is the key distinction that the literature has not modeled price stickiness along with firm heterogeneity?

Clementi and Pallazzo (2016) study the amplifying role of endogenous entry and exit in a model with heterogeneous firms in the Hopenhayn tradition (with perfect competition and decreasing returns), and with flexible prices. Closer to our framework, Hamano and Zanetti (2017, 2022) incorporate heterogeneity into Bilbiie et al (2012) and study ...[21] We draw on these contributions but introduce heterogeneity in *quality*, which allows us to model *price* stickiness in a tractable way.

[19]We model wage stickiness in a standard way with a quadratic cost parameterization for nominal wages adjustments by a " labor union", which bundles the differentiated labor types of a unit mass of households, giving rise to a standard nonlinear " wage Phillips" curve (we follow the classic references on wage rigidity, Erceg et al, 2000; and Schmidt-Grohé and Uribe, 2006). In the limit as the adjustment cost increases, we recover the case of fixed nominal wages, which we can fully solve analytically, without linearization.

[20]In an online appendix we also derive a model extension with demand externalities that generate additional aggregate-demand amplification.

[21]See also Hamano and Pappada (2023) for an open-economy application, and Hamano (2025) for a study of aggregate and idiosyncratic volatility.

3.1 Heterogeneous Firms

MM: Mention that $C_t = Y_t$ in flex price? More generally, carry over GE equations from section 2.

Firms have heterogeneous productivity (physical output per unit labor) and quality (quality-adjusted output per unit of physical output). We assume that those productivity and quality differences across firms do not change over time. All firms therefore face the same productivity shifter A_t over time that we modeled in the previous section with representative firms. Across firms, we normalize output units to equal output per unit labor. This normalization delivers the same reference price p_t for all firms. We then capture both productivity and quality differences terms across firms as an index $q > 0$, which we just call quality (though this could also reflect productivity differences). Our main restriction is that this index q does not vary over time: only aggregate productivity A_t fluctuates.[22]

Demand across varieties is still CES with elasticity θ, with quality-adjusted prices p_t/q and quality-adjusted output $qy_t(q)$ for a firm with quality q:

$$qy_t\left(q\right) = \left(\frac{p_t/q}{P_t}\right)^{-\theta} Y_t,$$

where final good production Y_t is the CES aggregate of intermediate good production $y_t(q)$ as we previously defined in (1). The price index P_t now aggregates over the quality-adjusted prices: $P_t^{1-\theta} = \int_0^\infty \left(p_t/q\right)^{1-\theta} dG_S\left(q\right)$, where $G_S(q)$ is the distribution of quality for producing (surviving) firms. And similarly, output Y_t now aggregates over quality-adjusted units: $Y_t^{\frac{\theta-1}{\theta}} = \int_\omega \left(qy_t\left(q\right)\right)^{\frac{\theta-1}{\theta}} dG_S\left(q\right)$. A firm's production function for quality-adjusted units is:

$$qy_t(q) = A_t q\left[l_t(q) - f_O\right],$$

where $l_t(q)$ is firm employment and f_O captures a per-period overhead production cost in units of labor.[23] This clarifies how $A_t q$ captures heterogenous productivity with a common time-varying trend A_t.

The definitions for the relative price ρ_t, markup μ_t, and (real) profit per firm d_t remain unchanged (cf. 3, 4, 5). Using the quality-adjusted forms for demand and production above, firm

[22]Our framework also easily incorporates aggregate quality changes over time. This affects the inflation metrics – depending on whether the prices are adjusted for quality or not; but does not change any of the implications of our model in terms of the output gap and entry-multiplier responses to productivity shocks and the impact of monetary supply responses.

[23]With CES demand, an overhead production cost is required to generate selection into production for the heterogeneous producers. We will also introduce an entry cost later in a similar way to the model in Section 2.

profit can now be written:

$$d_t\left(q\right)=\left(1-\frac{1}{\mu_t}\right)\mu_t^{1-\theta}\left(\frac{w_t}{qA_t}\right)^{1-\theta}Y_t-\frac{w_t}{A_t}f_O.$$

The overhead cost induces selection: Only firms with quality $q > q_{0t}$ will produce in period t, where the quality cutoff is determined by the zero profit cutoff (ZPC) condition $d_t(q_{0t}) = 0$. Also note that variable profit $d_t(q) + w_t f_O/A_t$ is proportional to $q^{\theta-1}$ across firms with different quality q.

As we assumed in Section 2, firms must pay an entry cost paid in labor units that does not change with aggregate productivity A_t. We now denote this cost f_E (in this version with heterogenous firms, there are both overhead costs and entry costs). We assume that this cost is paid "ex ante" upon entry, before firms learn their idiosyncratic quality q. Following Hopenhayn (1992) and Melitz (2003), we model this "learning" as a draw from an exogenous distribution $G(q)$. Firms with high quality then survive and select into production; and pay the overhead cost f_O "ex post". The average profit among those firms is:

$$\bar{d}_t=\left[1-G\left(q_{0t}\right)\right]\int_{q_{0t}}^{\infty}d_t(q)\;dG\left(q\right).$$

As shown by Melitz (2003), this economy with heterogeneous firms is isomorphic to one with representative firms that earn profit $\bar{d}_t$ with an "average" quality level $\bar{q}_t$ defined by:

$$\bar{q}_t^{\theta-1}=\left[1-G\left(q_{0t}\right)\right]\int_{q_{0t}}^{\infty}q^{\theta-1}dG\left(q\right). \tag{15}$$

We assume that firms observe the realization of aggregate productivity A_t prior to entry. This leads to N_{Et} entrants and $N_t = \left[1 - G\left(q_{0t}\right)\right] N_{Et}$ producing/surviving firms such that the average profit among all entrants is equalized with the entry cost:

$$\left[1-G\left(q_{0t}\right)\right]\bar{d}_t=w_t f_E,$$

where average profit can be written:

$$\bar{d}_t=\left(1-\frac{1}{\mu_t}\right)\frac{Y_t}{S_t}-\frac{w_t}{A_t}f_O.$$

Note that we recover our representative firm model from the previous section by setting the overhead cost to $f_O = 0$. In this version, there is still firm heterogeneity, but that heterogeneity is immaterial as there is no selection when there is no overhead production cost: All entering firms with quality over the whole support of G produce in every period ($N_t = N_{Et}$).

MM: Introduce $\tilde{k} = k/(\theta-1) \geq 1$ and make sure there are no more κs (now k)

FB: Should we use ϑ for $k/(\theta-1)$? We can't use $\tilde{k}$...

To obtain analytical expressions, we follow the literature (e.g. Helpman, Melitz and Yeaple; Chaney) and employ a by now standard Pareto distribution for quality, with:

$$G(q) = 1 - \left(\frac{q_{\min}}{q}\right)^k,$$

which will also imply a Pareto distribution with shape $\frac{k}{\theta-1}$ for firm size (revenue). Then (15) implies that the average quality is always proportional to the cutoff quality:

$$\bar{q}_t = \left(\frac{k}{k-(\theta-1)}\right)^{\frac{1}{\theta-1}} q_{0t}$$

So for the price we have e.g.

$$\frac{p_t/\bar{q}_t}{P_t} = S_t^{\frac{1}{\theta-1}} \rightarrow \frac{p_t}{P_t} = \left(\frac{k}{k-(\theta-1)}\right)^{\frac{1}{\theta-1}} q_{0t} S_t^{\frac{1}{\theta-1}}$$

Recall zero-profit

$$\left(1-\frac{1}{\mu_t}\right)\mu_t^{1-\theta}\left(\frac{w_t}{q_{0t}A_t}\right)^{1-\theta} Y_t = \frac{w_t}{A_t} f_O$$

For any producing firm with $q \geq q_{0t}$, we can write:

$$\frac{d(q) + \frac{w_t}{A_t} f_O}{d(q_{0t}) + \frac{w_t}{A_t} f_O} = \left(\frac{q}{q_{0t}}\right)^{\theta-1}$$

which implies:

$$\frac{\bar{d} + \frac{w_t}{A_t} f_O}{d^* + \frac{w_t}{A_t} f_O} = \left(\frac{\tilde{q}}{q^*}\right)^{\theta-1} = \frac{k}{k-(\theta-1)}$$

Using that by definition the cutoff-quality firm makes zero profits $d_t(q_{0t}) = 0$, average profits become:

$$\bar{d} + \frac{w}{A} f_O = \frac{k}{k-(\theta-1)} \frac{w}{A} f_O \rightarrow$$
$$\bar{d} = \frac{\theta-1}{k-(\theta-1)} \frac{w}{A} f_O$$

Then, for the cutoff we have, since average profits are still

$$\bar{d}_t = \frac{\theta-1}{k-(\theta-1)} \frac{w_t}{A_t} f_O$$

$$\bar{d}_t + \frac{w_t}{A_t} f_O = \left(1 - \frac{1}{\mu_t}\right) \frac{Y_t}{S_t} = \frac{k}{k-(\theta-1)} \frac{w}{A} f_O \rightarrow \frac{w}{A} f_O = \left(1 - \frac{1}{\mu_t}\right) \frac{Y_t}{S_t} \frac{k-(\theta-1)}{k}$$

So

$$\bar{d}_t = \frac{\theta-1}{k} \left(1 - \frac{1}{\mu_t}\right) \frac{Y_t}{S_t}$$

The free-entry condition and the ZCP imply:

$$f_E = \left(\frac{q_{\min}}{q_{0t}}\right)^k \frac{\theta-1}{k-(\theta-1)} \frac{f_O}{A_t}$$

We confine attention to equilibria whereby this condition always holds with equality, in other words the cutoff level is larger than $q_{\min}$. In steady state, $q_{0t} > q_{\min}$ implies a restriction

$$\left(\frac{q_{0t}}{q_{\min}}\right)^k = \frac{\theta-1}{k-(\theta-1)} \frac{f_O}{A_t f_E} > 1 \rightarrow \frac{k}{\theta-1} < 1 + \frac{f_O}{A_t f_E}.$$

This imposes an upper bound on the shape parameter of the size distribution; we return to this when discussing the role of heterogeneity for shaping aggregate fluctuations.

The rest of the model is identical to the previous, homogeneous-firms model—and is orthogonal to the introduction of firm heterogeneity, by virtue of the assumption that heterogeneity pertains to quality, and individual prices are not a function of the individual state. Thus, in particular, heterogeneity does not interact with price stickiness, which renders the model particularly tractable.

MM: This subsection will end by deriving: Pricing, Variety, Profits, Free Entry, Agg Acct, Survivors, Exit Cutoff. Describe how we add the additional dynamic/sticky price equations. Then show them in loglinear form (with non-linear versions in appendix). Contrast with equations in standard NK version. Note: notation changes still need to be carried through here.

Flexible prices and wages

MM: All derivations hold for flex, fixed, and sticky prices.

Under flexible prices and wages, we have the desired markup $\mu_t^* = \theta/(\theta-1)$, labor supply equation $w_t^* = (L_t^*)^{\varphi}(Y_t^*)^{1/\sigma}/\mu_t^*$ and production function $Y_t^* = w_t^* L_t^*$. This is equivalent to a subsidy μ_t^* to employment labor to restore the efficient labor/leisure choice. The closed-form solution for the "natural", efficient equilibrium obtained under fully flexible prices and wages in our model with heterogeneity and selection is emphasized in the following proposition.

Proposition 2. *The effect of TFP shocks on equilibrium output (under flexible prices and wages) is on the one hand amplified by entry thorugh the variety effect (the elasticity is proportional to $\frac{\theta}{\theta-1}$). On the other, it is mitigated through selection and heterogeneity; the elasticity is proportional to $-1/k$, which is the elasticity of the zero-profit cutoff to TFP. Therefore, as k decreases (selection increases) from infinity to its lower bound $k=\theta-1$ the output elasticity (aggregate return to scale) goes from $\frac{\theta}{\theta-1}$ down to 1.*

The proposition follows immediately from direct inspection of the closed-form solution of the model, given by:

$$Y_t^* = (A_t)^{\left(\frac{\theta}{\theta-1}-\frac{1}{k}\right)(1+\varphi)\Phi}\left(\frac{1}{f_O\theta}\right)^{\left(\varphi+\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}\Phi-1}\left(\frac{\theta-1}{\theta}\right)^{1-\frac{1}{\sigma}\frac{\theta}{\theta-1}\Phi} F^{1+\left(1-\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}\Phi}; \quad (16)$$
$$q_{0t}^* = (A_t)^{-\frac{1}{k}} F,$$

where $F \equiv \left(\frac{\theta-1}{k-(\theta-1)}\frac{f_O}{f_E}\right)^{\frac{1}{k}} q_{\min}$ and $\Phi \equiv 1/\left(1+\varphi-\frac{\theta}{\theta-1}\left(1-\frac{1}{\sigma}\right)\right)$.

The solution clearly illustrates the two key forces, which can be most clearly seen under inelastic labor ($\varphi \to \infty$): product variety leads to an amplified effect of TFP shocks through endogenous productivity improvement via the first term in the exponent, $\frac{\theta}{\theta-1} > 1$. On the other hand, heterogeneity and selection lead to a dampening of business cycles: an increase in aggregate TFP reduces the zero-profit cutoff (the more so, the larger is $\frac{1}{k}$, i.e. the more concentrated/skewed the distribution): as less "productive" firms enter, selection leads to a decrease in endogenous aggregate productivity. As the Pareto shape decreases towards its lower feasible bound $\theta-1$,the two effects on aggregate productivity eventually perfectly offset each other. Note that we will then show that this impact of selection of RTS is orthogonal to entry multiplier.

To further clarify intuition and relate back to the log-linearized NK model, it is instructive to consider the solution in log-form (including for the number of firms):

$$\hat{N}_t^* = \frac{(\sigma-1)\left(\theta\tilde{k}-1\right)}{\tilde{k}\left(\theta\left(1+\sigma\varphi\right)-\sigma\left(1+\varphi\right)\right)}\hat{A}_t; \tag{17}$$

$$\hat{Y}_t^* = \frac{\sigma\left(1+\varphi\right)\left(\theta\tilde{k}-1\right)}{\tilde{k}\left[\theta\left(1+\sigma\varphi\right)-\sigma\left(1+\varphi\right)\right]}\hat{A}_t;$$

$$\hat{q}_{0t}^* = -\frac{1}{\tilde{k}\left(\theta-1\right)}\hat{A}_t$$

$$\hat{S}_t^* = \hat{N}_t^* - k\hat{q}_{0t}^* = \left(\frac{(\sigma-1)\left(\theta-\frac{1}{\tilde{k}}\right)}{\theta\left(1+\sigma\varphi\right)-\sigma\left(1+\varphi\right)}+\frac{1}{\theta-1}\right)\hat{A}_t$$

where the last variable is the number of surviving, producing firms.[24] The parameter restriction $\sigma > 1$ (low income effects on labor) is needed in order to have both entry and hours worked increase in response to TFP shocks, an uncontroversial business-cycle property and a property shared with the RBC model.

One notable implication is that under either log utility $\sigma = 1$ or inelastic labor, the number of firms is invariant to productivity—even though the number of surviving, producing firms still fluctuates via the cutoff. This invariance is due to the shock not affecting the entry cost and is reminiscent of results in the trade literature (see Arkolakis et al and Melitz and Redding).

3.2 New Keynesian Block: Price and Wage Setting and Monetary Policy

MM: Emphasize that all derivations from Sec 3.1 (before flex price derivation) also hold with sticky prices. In this case, profits are "gross" of price adjustment costs. Only other difference is specification of the markup. And emphasize that the common markup across q with sticky prices comes from the fact that all firms face the same cost of price changes (in percentage terms) – and this leads all firms to pick the same markup.

In the case of arbitrary nominal rigidities, we assume that price and wage adjustment are both subject to Rotemberg adjustment costs. The Phillips curve (which determines the firms' markup) and the wage Phillips curve are obtained as follows (since this is entirely standard we only describe it here but outline the two curves along with all the equilibrium conditions of the full model for completion in Table 2 in Appendix A.4).

For price setting, we closely follow Bilbiie et al (2007, 2014) and assume that firms pay an adjustment cost for resetting their price relative to last period's price (or relative to the average producer price in the market, for those producers who enter afresh). By close parallel with the

[24]This this is not a linearization–it is the non-linear solution in log-form.

aforementioned contributions. maximization of profits net of this price adjustment cost leads to the nonlinear Phillips curve for price inflation outlined in Table 2 in the Appendix A.4.

The "wage stickiness" part is also standard—wage setting is done by an union bundling the differentiated labor inputs of households, setting the nominal wage subject to adjustment costs—and its details are unaffected by the introduction of firm entry. For the sake of space, we do not report all derivations of this block but refer to Erceg et al (2000) and Schmitt-Grohé and Uribe (2006); the nonlinear Phillips curve is included for completion in the same Appendix table referenced above and determines the dynamics of wage inflation, which is defined by:

$$1 + \pi_{w,t} = \frac{w_t}{w_{t-1}} \left(1 + \pi_t^C\right).$$

Both Phillips curves are included in loglinearized form in Table 1: as anticipated, they are exactly identical to those of the standard NK model with no entry. Their slopes ψ and ψ_w respectively are inverse functions of the respective adjustment-cost parameter (a lower ψ means more stickiness, with full rigidity at $\psi = 0$ and full flexibility for $\psi \to \infty$).[25] Finally, monetary policy is set according to the same Taylor rule (12) from the previous section. The full nonlinear model is summarized for completion by Table 2.

MM: Consider moving previous footnote to text?

Log-linearizing around the steady state with $A = 1$,$\varepsilon = 0$,$\pi = \pi_w = 0$ and $1 + i = \beta^{-1}$ delivers the linear equilibrium conditions outlined in Table 1. We use again hats (ˆ) to denote percent deviations from steady state: $\hat{X}_t = \ln X_t - \ln X$ for any endogenous variable, with the exception of rates which are already in percentage points and are expressed as deviations from their steady-state value (since inflation rates are zero in steady state, that coincides with their actual value—hence, there are no hats on the π variables).

The first 8 equations in the Table are exactly the same as in the standard, no-entry NK model, modulo the right modifications: in particular, equation 1 would have $\hat{\rho}_t = 0$ (no benefit of variety), implying simply that the markup is the inverse of real marginal cost; in equation 6, CPI inflation

[25] Specifically, $\psi = (1 + s_w)(\theta - 1)/\kappa$ and $\psi_w = (1 + s_w)(\theta_w - 1)/\kappa_w$ where the κs are the adjustment cost parameters and s_w is the labor subsidy. See appendix A.4 for the full derivation of the non-linear model with the functional forms for the cost of price adjustments and the optimal subsidy. The optimal subsidy offsets the markup distortions for both labor and intermediate goods. This is just a recalibration of the ψs; it will be relevant for our welfare analysis in section 4. In the log-linearized version, the ψs (slopes of the Philips curves) are a sufficient statistic for the impact of nominal rigidities – because the inflation costs are second order.

Table 1: Loglinearized Model, Summary

Sec 2 Equations	
1. Pricing	$\hat{\rho}_t = \hat{\mu}_t + \hat{w}_t - \hat{A}_t$
2. Agg. acct.	$\hat{Y}_t = \hat{w}_t + \hat{L}_t$
3. Profits	$\hat{d}_t + \hat{S}_t = \hat{Y}_t + (\theta - 1)\,\hat{\mu}_t$
Dynamic:	
4. Price PC	$\pi_t = \beta E_t \pi_{t+1} - \psi \hat{\mu}_t$
5. Wage PC	$\pi_{w,t} = \beta E_t \pi_{w,t+1} + \psi_w \left(\sigma^{-1}\hat{Y}_t + \varphi \hat{L}_t - \hat{w}_t\right)$
6. Euler bonds	$\hat{Y}_t = E_t \hat{Y}_{t+1} - \sigma \left(\hat{i}_t - E_t \pi^C_{t+1}\right)$
7. Taylor rule	$\hat{i}_t = \phi E_t \pi_{t+1} - \varepsilon_t$
8. Wage inflation	$\hat{w}_t = \hat{w}_{t-1} + \pi_{w,t} - \pi^C_t$
9. CPI inflation	$\pi_t = \pi^C_t + \hat{\rho}_t - \hat{\rho}_{t-1}$
Entry-specific equations:	
10. Free entry	$\hat{w}_t + \hat{N}_t = \hat{d}_t + \hat{S}_t$
11. Variety effect	$\hat{\rho}_t = \hat{q}_{0t} + \frac{1}{\theta-1}\hat{S}_t$
12. Survivors	$\hat{S}_t = \hat{N}_t - k\hat{q}_{0t}$
13. Exit Cutoff	$(\theta - 1)\,\hat{q}_{0t} + \hat{Y}_t - \theta \hat{\rho}_t = -\theta \hat{\mu}_t$

Notes: The table displays all the loglinearized equilibrium conditions, around a zero-inflation steady state with an optimal subsidy.

would this be the same as producer inflation ($\pi_t = \pi^C_t$). In equation 2, aggregate accounting and the CRS production function would simply imply $\hat{Y}_t = \hat{A}_t + \hat{L}_t$.[26] Equation 3 would simply be the definition of profits (with no change in entry,$\hat{S}_t = 0$), which are irrelevant for the equilibrium outcome.

Our model has a reduced, 5-equation representation that is isomorphic to the standard NK model (with both nominal rigidities)—with endogenous entry acting as a wedge in the model's key equations.

An NK, "Gap" Representation

The model can be further reduced and rewritten in the NK tradition, in "gaps" of the sticky-equilibrium variables from the flexible-equilibrium one, namely for any variable x, we have $\tilde{X}_t \equiv \hat{X}_t - \hat{X}^*_t$.

A key object, which acts as a composite reduced-form shock for the dynamics of the model, is r-star: the "natural" real interest rate occurring under flexible prices and wages. This is obtained by replacing the equilibrium values under flexible prices (17) into the Euler equation for bonds,

[26]Or equivalently from the income side: $\hat{Y}_t = \left(1 - \theta^{-1}\right)\left(\hat{w}_t + \hat{L}_t\right) + \theta^{-1}\hat{d}_t$, with $\hat{d}_t = \hat{Y}_t + (\theta - 1)\,\hat{\mu}_t$.

using the definition of CPI inflation (4 and 10 in Table 1). In other words, this is the "real" with respect to PPI inflation π_t interest rate:

$$\hat{r}_t^* = (1 - \rho_a)\,\Gamma \hat{A}_t \tag{18}$$

$$\text{where } \Gamma \equiv -\frac{(\sigma - 1)\,\varphi + \tilde{k}\,(\theta - \sigma)\,(1 + \varphi)}{\tilde{k}\,(\theta\,(1 + \sigma\varphi) - \sigma\,(1 + \varphi))} < 0$$

MM: Consider eliminating $\tilde{k}$ and use:

$$\Gamma = -\frac{(\sigma - 1)\varphi(\theta - 1) + k(\theta - \sigma)(1 + \varphi)}{k\,[(\theta - \sigma) + \sigma\varphi(\theta - 1)]}$$

denotes the elasticity of r-star to a one-time TFP shock and ρ_a is the persistence of the shock, $E_t\hat{A}_{t+1} = \rho_a\hat{A}_t$. As in the standard NK model, a fall in TFP engenders an increase in the natural interest rate.[27] Note that in the no-entry model, we similarly have (under constant returns and around an efficient steady state). $\Gamma^{\mathcal{N}} = -\frac{1+\varphi}{1+\sigma\varphi} < 0$. Moreover, in both models the elasticity is larger than > -1 under the exact same condition: $\sigma > 1$ (and $\varphi > 0$). Therefore, the dynamics of r-star are qualitatively similar to the NK model, and are only affected quantitatively by entry, and—importantly—heterogeneity and selection; the last point is worth stressing because the dynamics of r-star are the only channel through which heterogeneity affects aggregate dynamics in our model; as we shall see, the dynamics of the cutoff are exactly the same under the flexible and sticky equilibria.

The system then becomes strikingly similar to the standard NK model, with some important differences.

$$\pi_t = \beta E_t \pi_{t+1} + \frac{\psi}{\theta}\tilde{Y}_t - \psi\frac{1}{\theta - 1}\tilde{N}_t \tag{19}$$

$$\pi_{w,t} = \beta E_t \pi_{w,t+1} + \frac{\psi_w}{\theta}\left(\frac{\theta}{\sigma} - 1 + (\theta - 1)\,\varphi\right)\tilde{Y}_t \tag{20}$$

$$\hat{i}_t = \phi E_t \pi_{t+1} - \varepsilon_t \tag{21}$$

$$\tilde{Y}_t = E_t\tilde{Y}_{t+1} - \sigma\left(\hat{i}_t - E_t\pi_{t+1} - \hat{r}_t^*\right) - \sigma\frac{1}{\theta - 1}\left(E_t\tilde{N}_{t+1} - \tilde{N}_t\right) \tag{22}$$

$$\frac{1}{\theta}\tilde{Y}_t = \frac{1}{\theta}\tilde{Y}_{t-1} + \pi_{w,t} - \pi_t - \left(\hat{A}_t - \hat{A}_{t-1}\right) + \frac{1}{\theta - 1}\tilde{N}_t - \frac{1}{\theta - 1}\tilde{N}_{t-1} \tag{23}$$

[27] This holds under the sufficient condition $\theta > \sigma > 1$, which will also be a sufficient condition for our key comovement; thus, our mechanism does not rely on overturning the response of r-star, but on endogenous forces overturning the reaction to r-star as a reduced-form shock.

Like the standard NK model with both sources of nominal rigidity, this is a 5-equation model.[28] The main differences are that the price PC and the IS equation feature new endogenous terms/wedges, related to endogenous entry; the same applies, naturally, to the definition of real wage.

The TFP shocks generates cyclical fluctuations in the familiar way: it impacts "r-star" $\hat{r}_t^*$, and also appears in the wage growth equation. However, its equilibrium effects are radically different, because of the presence of the endogenous number of varieties in both the aggregate demand and supply sides.

For instance, as in the standard NK model, r-star goes up with negative TFP. This normally leads to a positive output gap (and it would do so here too, with fixed entry) by inter-temporal substitution in (22); but—as we will show momentarily—the last term is endogenous and goes down, such that there is a recession (output falls below potential) instead. At the same time, on the supply side, the fall in number of firms acts as a shifter of the Phillips curve (19) at given consumption, so it is inflationary; this is akin to a fall in endogenous productivity.

We first illustrate the main comovement properties in a parameterized, quantitative version of the model. Then, we derive the equilibrium in closed form in a special case and substantiate these insights while showing their generality beyond the specific parameterization.

Selection and the Entry Multiplier

An important property should be emphasized, which is a corrolary of the derivation of our reduced model in gap form.

Proposition 3. *Selection is orthogonal to nominal rigidity: the zero-profit cutoff follows the same dynamics under sticky and flexible prices and wages, so the gap is zero $\tilde{q}_{0t} = 0$. Consequently, the gaps of the number of producing and entering firms are identical $\tilde{S}_t = \tilde{N}_t$, which implies very similar aggregate dynamics with and without selection.*

The proposition follows directly as a corrolary of the derivation of the gap system (19)-(23), and can also be directly shown by using equations 1, 3, and 10-13 in Table 1, which delivers $\hat{q}_{0t} = \hat{q}_{0t}^* = -\frac{1}{k}\hat{A}_t$. This "irrelevance" result, however, does not mean that heterogeneity itself is irrelevant for aggregate dynamics. Indeed, heterogeneity does shape the dynamics of macroeconomic aggregates in the "gap" form used in (19)-(23)—its effect is just encapsulated in the effect on r-star (through the composite parameter Γ introduced in (18)), which acts as a reduced-form shock and is thus a "sufficient statistic" for the effect of heterogeneity on aggregate dynamics.

[28]Equation 23 uses the flex-equilibrium version of wage growth, CPI inflation, and pricing (6, 10, and 1 in Table 2) to replace: $\pi_{w,t}^* - \pi_t^* = \hat{w}_t^* - \hat{\rho}_t^* - (\hat{w}_{t-1}^* - \hat{\rho}_{t-1}^*) = \hat{A}_t - \hat{A}_{t-1}$.

MM: This will be a new result that we will emphasize: That selection does not magnify/dampen the impact of the entry-multiplier for the output gap and welfare (even though selection *does* affect the entry multiplier). AND point out that this result was not *obvious* ex-ante!!!

3.3 Cyclical Implications: *TFP Shocks become true "supply shocks"*

We quantitatively illustrate how the addition of sticky wages generates the first-order amplification effects for TFP shocks, how this interacts with heterogeneity and selection, and what the model implies for demand shocks and for how a policymaker uses aggregate-demand leverage (monetary policy) to respond to the supply-driven recession.

FB: Re-write paragraph. 1) Describe where calibration comes from. For Philips curve, we start with ψs equal to 0.01 (similar to literature, our results our robust to changes in this). And then in a footnote, mention that this value of ψ and the value of the optimal subsidy s_w implies a value of κ. In other words, we do not separately calibrate the κs.

Consider a baseline parameterization with values that are commonly used in the New Keynesian literature. The elasticity of substitution between goods is $\theta = 3.8$, as in e.g. our previous work (Bilbiie et al, 2007, 2012); we assume the same elasticity of substitution for labor types ($\theta_w = 3.8$; this parameter is largely inconsequential). The Pareto shape parameter is set in line with the trade literature to $k = 4$, see e.g. Melitz and Redding (2015); this implies a shape parameter for the size distribution of firms of $k/(\theta - 1) = 1.43$. The CRRA coefficient is $\sigma^{-1} = 0.5$, implying relatively low income effects on labor, and labor supply elasticity is $\varphi = 1$. For the nominal rigidity paramters, we adopt empirically standard Phillips-curve slopes consistent with medium-scale DSGE estimation and micro evidence on nominal rigidities. The price and wage Phillips curve slopes are set to $\psi = \psi_w = 0.01$ per quarter, corresponding (in a "Calvo-equivalent" reinterpretation) to average durations of roughly three to five quarters (our results are robust to variations around this magnitude, including by considering relatively stickier wages than prices). These values lie squarely within the consensus range documented in Smets and Wouters (2007), Justiniano, Primiceri, and Tambalotti (2010, 2011, 2013), Del Negro, Giannoni, and Schorfheide (2015), and the micro-price evidence summarized by Klenow and Malin (2011). The Taylor rule respon response is $\phi = 1.5$. Finally, we assume an optimal subsidy eliminating the steady-state goods and labor market inefficiency:[29]

$$1 + s_w = \frac{\theta}{\theta - 1}\frac{\theta_w}{\theta_w - 1}.$$

[29]Note that the values of the ψs and the optimal subsidy s_w imply a values of the adjustment-cost parameters κs. In other words, we do not separately calibrate the κs.

Figure 2 illustrates the central positive implication of our framework; we plot the dynamic equilibrium responses of key variables to a 1% fall in productivity with persistence $\rho_a = 0.9$) under flexible prices (red dash) and under sticky prices and wages (solid blue). The output gap plots the difference between the two respective output values. Following a persistent but transitory negative TFP shock, output falls by more under sticky prices and wages than under flexible prices, generating a negative output gap. In our baseline calibration (Panel 2b), the 1% decline in productivity reduces the mass of firms by about 4 percent and produces a negative output gap of roughly 2 percent. These large magnitudes reflect the first-order effects of endogenous entry emphasized above. Lower productivity reduces profits and expected profitability, discouraging entry and shrinking productive capacity. Relative to the flexible-price equilibrium, the economy therefore experiences an aggregate-demand recession in response to a supply disturbance.

This implication sharply contrasts with the standard New Keynesian model without entry. As shown in the upper panel 2b, when product variety is fixed, a negative TFP shock generates a positive output gap under nominal rigidities, because flexible-price output contracts more than actual output. With endogenous entry, the opposite occurs: hours and output under sticky prices fall by more than in the flexible equilibrium. Interpreted through the lens of unemployment—defined, following Galí (2011), as the wedge between sticky-wage and flexible-wage hours—a TFP-driven recession in our model is associated with an increase in inefficient unemployment. This implication is consistent with a growing body of empirical evidence. VAR studies identifying transitory TFP shocks using utilization-adjusted productivity, news-shock decomposition, or sign restrictions find that adverse productivity shocks raise inflation while unemployment increases or remains flat and capacity utilization rises, indicating a contraction relative to the productive frontier rather than an expansion (Basu, Fernald, and Kimball 2006; Barsky and Sims 2011; Fernald and Wang 2016). Semi-structural state-space estimates of output gaps and DSGE-based measures of natural output deliver the same message: productivity improvements are associated with rising output gaps, while negative productivity shocks reduce them.

In this sense, endogenous entry "transforms" TFP shocks into truly supply shocks, in that they generate the negative comovement between inflation and slack typically associated, in the standard NK model, with ad hoc cost-push or markup shocks.[30] Indeed, for all variables common across models, the impulse responses to TFP shocks with entry closely resemble those generated by markup shocks in a no-entry NK model—with one crucial exception. In our framework, markups

[30]Endogenous entry thus provides a microfoundation for endogenous cost-push shocks; in section 3.2, we show formally how entry appears as an endogenous cost-push, markup-shock wedge in the Phillips curve.

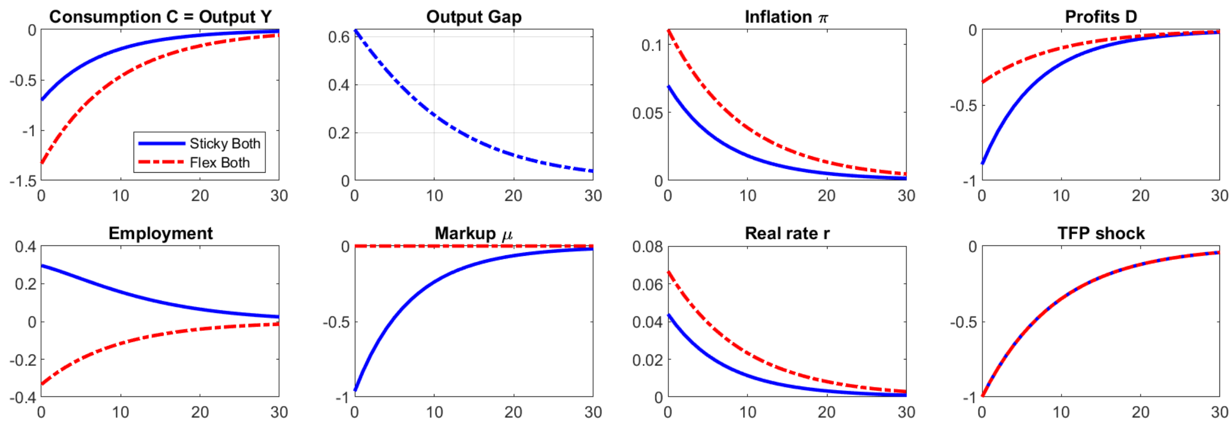
and profits fall following adverse productivity shocks, whereas markup shocks raise markups and imply countercyclical profits in the standard model. This distinction is empirically central. Profits are strongly procyclical in the data, both unconditionally and conditional on supply disturbances. Aggregate profits rise following positive TFP shocks and fall following negative ones, and firm-level evidence shows that productivity improvements increase profits, survival, and entry, while productivity declines trigger exit (Basu et al. 2006; Fernald 2014; Syverson 2011; De Loecker and Eeckhout 2017; Gourio and Rudanko 2014; Bilbiie and Känzig 2023). By contrast, cost-push episodes such as large oil price shocks are stagflationary but are not associated with increases in aggregate profits; if anything, profits tend to fall as higher costs and lower quantities dominate any increase in markups (Barsky and Kilian 2004; Blanchard and Galí 2007; Bilbiie and Känzig 2023).

Finally, while our model predicts—like the standard NK framework—that inflation rises following a negative TFP shock, inflation is lower than under flexible prices. Relative to the efficient allocation, there is therefore a deflationary gap associated with the increase in slack. This explains why, despite inflationary pressure, the appropriate policy response in our framework is expansionary: easing monetary policy closes the negative output gap by counteracting the endogenous contraction in productive capacity. We return to this point in the policy discussion below.

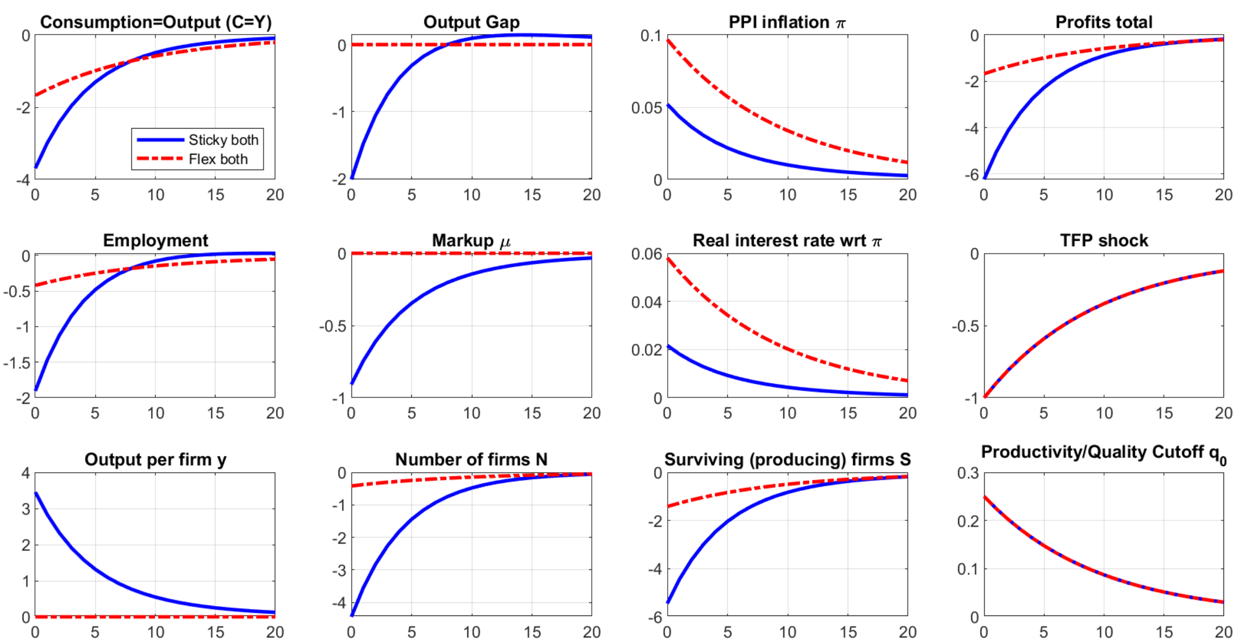
FB Figures: get rid of black lines? (we emphasize the "orthogonality" result analytically)

The role of price versus wage stickiness: Although wage stickiness plays an important role in amplifying the quantitative effects of the entry multiplier, it is not sufficient on its own to generate aggregate-demand amplification of supply disruptions. In Appendix A.5, we analyze a version of the model with sticky wages but flexible prices. In that environment, firms can restore profitability by adjusting prices, which prevents a persistent contraction in entry. Following a negative TFP shock, hours increase to smooth nominal wages, the economy expands relative to the flexible-wage allocation, and inflation turns persistently negative. There is no sustained negative output gap. Price stickiness is therefore essential: it generates the persistent profitability decline that endogenously triggers exit and the associated persistent demand amplification.

Demand shocks and entry: Wage stickiness is also crucial for the model's behavior in response to demand shocks. Figure 6 in Appendix A.7 reports the impulse responses to a temporary monetary expansion. With sticky prices but flexible wages, a demand shock lowers markups and profits, inducing firm exit—mirroring the standard NK result that profits are countercyclical to demand shocks. When both prices and wages are sticky, higher demand instead raises profitability and entry, increasing aggregate activity. Entry allows real wages to rise through a lower price index rather



(a) No Entry



(b) Endogenous Entry

Figure 2: Effects of a 1% productivity fall: Flexible (red dash) vs Sticky (solid blue) prices & wages. Dash black in b: representative-firm, sticky

than higher nominal wages, dampening the fall in markups and restoring profits. We relegate the detailed discussion to the Appendix; below, we derive a novel analytical condition for the relative degree of wage versus price stickiness required to obtain procyclical entry.

Bottom line: Our framework delivers a combination that is both empirically plausible and difficult to obtain in existing models: adverse productivity shocks generate stagflation—output gaps fall and inflation rises—while profits and entry are procyclical. Neither standard New Keynesian TFP shocks nor cost-push shocks can replicate these joint dynamics. By endogenizing entry and productive capacity, our model reconciles stagflationary supply shocks with procyclical profits, in line with the empirical evidence.

MM: Add some of this back to main text?
Interpretation and policy implications
Our main finding is particularly important in this context because endogenous entry overturns the standard New Keynesian prediction for the output gap. In our framework, adverse TFP shocks reduce profits and expected profitability, leading to a contraction in entry and in the mass of active producers. The resulting endogenous destruction of productive capacity implies that actual output falls **more** than flexible-price output, generating a negative output gap even though inflation rises. As a result, TFP-driven inflation is not a sign of overheating, and deflation following favorable productivity shocks can coexist with a positive output gap. Sticky prices combined with strong entry responses can therefore generate what might be described as "good deflation."
This logic has immediate implications for monetary policy. Central banks should not mechanically contract in response to inflation associated with productivity disturbances. The appropriate response depends on the source of inflation or disinflation—whether it reflects demand pressures or endogenous movements in productive capacity through entry. Our mechanism is related to the "Keynesian supply shock" framework of Guerrieri, Lorenzoni, Straub, and Werning, but with a crucial difference. While both mechanisms generate negative output gaps following adverse productivity shocks, the inflation response has the opposite sign. In Guerrieri et al., adverse supply shocks are deflationary at the aggregate level and thus observationally similar to demand contractions. In our model, adverse TFP shocks are inflationary because they trigger an endogenous contraction in the number of producers. The shock is therefore "supply–supply," not "supply–demand," a distinction with first-order consequences for policy design.

3.4 Closed-form Analytical Solution

FB: Maybe a couple of sentences here regarding why these further assumptions are not very restrictive (maybe with some references to where they have been used elsewhere)? State that those assumptions do not substantially change the quantitative impulse responses.

To obtain a closed-form solution, we make only two further simplifying assumptions—made without losing generality, as we illustrate by simulations of the quantitative model without these special assumptions we just saw. The assumptions are that 1. both Phillips curves are static ($\beta = 0$

in 19 and 20 only); and 2. the monetary policy rule perfectly neutralizes changes in expected inflation and thus keep the ex-ante real rate fixed, $\phi = 1$. Under just these two assumptions, the full solution of the model is described in Proposition 4. These assumptions are not very restrictive at all: they have been employed in other contexts to derive analytical results shown to be robust to forward-looking Phillips curves and standard Taylor rules (e.g., Bilbiie, 2025; Bilbiie and Känzig, 2023); and most importantly for our purpose, they do not substantially change the quantitative impulse responses (we provide a comparison of the analytically restricted calibration with the baseline one in Appendix A.6).

Proposition 4. *The output gap and entry multiplier response to a negative TFP shock is negative so long as prices are sticky "enough".*

To see this note that the output gap and entry multiplier responses are given by:

$$\tilde{Y}_t = \sigma \frac{1}{\theta - 1} \tilde{N}_t + \sigma \left(\epsilon_t + \Gamma \hat{A}_t \right) \tag{24}$$

$$\tilde{N}_t = \Theta \tilde{N}_{t-1} + \sigma \frac{\theta - 1}{\theta - \sigma} \Theta \left[\left(1 + \psi - \Psi_w\right) \epsilon_t - \epsilon_{t-1} \right] + \frac{\theta - 1}{\theta - \sigma} \Theta \left[\left(\theta + \left(1 + \psi - \Psi_w\right) \sigma \Gamma \right) \hat{A}_t - \left(\theta + \sigma \Gamma \right) \hat{A}_{t-1} \right]. \tag{25}$$

We denote $\Psi_w = \psi_w \left(\frac{\theta}{\sigma} - 1 + (\theta - 1)\varphi \right)$ as the slope of the wage Philips curve and $\Theta = \left(1 + \psi + \Psi_w \frac{\sigma}{\theta - \sigma} \right)^{-1}$. Θ is an index of overall price and wage stickiness and governs endogenous persistence. When both are flexible, $\Theta = 0$; and it increases with stickiness reaching its upper-bound of 1 under fixed price and wage. The impact elasticities are then:

$$\frac{\partial \tilde{Y}_t}{\partial \hat{A}_t} = \frac{\sigma \theta}{\theta - \sigma} \Theta \left(1 + (1 + \psi) \Gamma \right)$$

$$\frac{\partial \tilde{N}_t}{\partial \hat{A}_t} = \frac{\theta - 1}{\theta - \sigma} \Theta \left(\left(1 + \psi - \Psi_w\right) \sigma \Gamma + \theta \right).$$

Thus, $\partial \tilde{Y}_t / \partial \hat{A}_t > 0$ so long as $\psi < -\left(1 + \Gamma\right)/\Gamma$. The latter implies a positive threshold for price stickiness (recall that lower ψ is inversely related to price stickiness)[31] *In addition, $\partial \tilde{N}_t / \partial \hat{A}_t > 0$ so long as $\left(1 + \psi - \Psi_w\right) \sigma \Gamma + \theta > 0$. This will be satisfied under the previous condition for price stickiness.*

Note that the output gap is zero under flexible wages, for any degree of price-stickiness: $\Theta = 0$ in that case. This is a generalization of the envelope result in Section 2.

MM: Tie this to the key distortion associated with sticky wages: the consumption-leisure choice is distorted.

[31] This requirement and ensuing result for the output-gap response also extend to the full dynamic solution beyond the impact elasticity (recall that $\theta > \sigma > 1$).

[32]

The response of entry to monetary policy. As we emphasized before in this sticky-price only version monetary expansions are contractionary on entry—the addition of sticky wages fixes this. We can solve for the threshold value of wage stickiness beyond which the response of entry to a monetary easing becomes positive.

Proposition 5. *The response of entry to an interest rate cut under arbitrary stickiness is positive when wages are sticky "enough":*

$$\frac{\partial \hat{N}_t}{\partial \varepsilon_t} = \frac{\theta - 1}{\frac{\theta}{\sigma}\left(1 - \frac{\Psi_w}{1+\psi}\right)^{-1} - 1} > 0 \qquad \textit{when} \qquad \Psi_w < 1 + \psi.$$

The intuition for the first part is that when wages are flexible, the increase in labor demand translates directly into an increase in the nominal wage—with the ensuing negative consequences for firm profitability and exit. However, this counterfactual prediction is overturned once we add sticky enough wages. Then, the increase in labor demand is accommodated by increased entry. This provides workers with real consumption gains at a given nominal wage—and thus substitutes for the sharp increase in nominal wages, which is no longer feasible. In other words, when both nominal wages and individual prices are sticky, the consumption good price index P_t must fall to induce a rise in the real wage—and this occurs via entry, dampening the fall in markups. Recall than when wages are sticky, entry is inefficiently low. The positive monetary shock can then boost real output by offsetting this inefficiency.

MM: Then, so long as the conditions for proposition 3 and 4 hold:

Proposition 6. *A monetary expansion closes the output gap:*

$$d\epsilon_t|_{d\tilde{Y}_t=0} = \Gamma d\left(-\hat{A}_t\right) - \frac{1}{\theta - 1} d\tilde{N}_t = \left(\Gamma + \frac{1}{1+\psi}\right) d\left(-\hat{A}_t\right) > 0$$

MM: The idea here is to find the $d\epsilon_t$ such that $d\tilde{Y}_t = 0$. And then $d\epsilon_t$ measures the associated monetary policy response.

The intuition for the second result is that the first term $\Gamma d\left(-\hat{A}_t\right)$ does indeed do the same job as in the standard NK model—on its own, it would imply a contraction (increase in interest rate)

[32] In this case, the entry multiplier response is $\tilde{N}_t = -(\theta - 1)\left(\epsilon_t + \Gamma \hat{A}_t\right)$. This generalizes the entry multiplier that we derived in Section 2. It further incorporates the role of selection through $\tilde{k}$ (which only affects r-star) and general labor and inter-temporal elasticities through Γ.

in order to track r-star, which is going up with the adverse TFP shock. But the key is the second term $-\frac{1}{\theta-1}d\tilde{N}_t$: because of the reduction in entry and thus the fall in endogenous productivity, the demand-relevant r-star (with respect to inflation in the price index of the final good, π^c) is going down, and that requires a cut in interest rates—the opposite of the standard NK model. This intuition is sharpest in the extreme case of fully-rigid prices and wages covered below.

It cannot be overemphasized how these implications are the diametrical opposite of the standard NK model's—and are thus entirely due to the endogenous-entry margin, rather than the presence of sticky prices and wages. In that model, of which we include a short recap (and a novel analytical illustration) in Appendix 3.2:

1. the output gap goes up with negative TFP

2. it can be closed by tracking r-star, which entails a monetary contraction. (the discussion of this exact implication is to our knowledge new in the context of the model with both sticky prices and wages).

3. this tracking r-star policy is equivalent to completely stabilizing the "average" composite inflation measure.

We have just shown how 1 and 2 are overturned in our model, and we now turn to 3.

MM: Contrast with standard NK:

$$d\epsilon_t^{\mathcal{N}}\big|_{d\tilde{Y}_t^{\mathcal{N}}=0} = \Gamma^{\mathcal{N}} d\left(-\hat{A}_t\right) < 0$$

Divine Coincidence

In the standard NK model with sticky prices and wages, Woodford (2003) and Gali (2015) show that a modified version of divine coincidence holds: the central bank can close the output gap by stabilizing an appropriately aggregated measure of price and wage inflation, namely:

$$\bar{\pi}_t \equiv \frac{\psi_w}{\psi_w+\psi}\pi_t + \frac{\psi}{\psi_w+\psi}\pi_{w,t}.$$

This is no longer true in our model due to the entry multiplier. Writing the aggregate Phillips curve for this inflation measure by aggregating (19) and (20) we obtain:

$$\begin{aligned}\bar{\pi}_t &= \beta E_t\bar{\pi}_{t+1} + \bar{\psi}\left(\sigma^{-1}+\varphi-(1+\varphi)\frac{1}{\theta}\right)\tilde{Y}_t - \bar{\psi}\hat{\mu}_t \\ &= \beta E_t\bar{\pi}_{t+1} + \bar{\psi}\left(\sigma^{-1}+\varphi\frac{\theta-1}{\theta}\right)\tilde{Y}_t - \bar{\psi}\frac{1}{\theta-1}\tilde{N}_t,\end{aligned}$$

where we define $\bar{\psi} \equiv \frac{\psi_w \psi}{\psi_w + \psi}$. This shows how stabilizing that measure of average inflation still induces an output gap that will be proportional to the entry multiplier $\tilde{N}_t$. In our model, monetary policy should target wage inflation: $\pi_{w,t} = 0$ closes the output gap (by definition, it mimics the sticky-price-only equilibrium whereby the envelope result holds). However, closing the output gap is no longer optimal (for welfare). We describe this in further detail in section 4.

TFP or endogenous cost-push?

A large empirical literature—especially based on estimated small- and medium-scale New Keynesian DSGE models—uses *cost-push* (*markup*) shocks as the primary driver of inflation fluctuations that are orthogonal to real activity, precisely because they generate the desired *negative comovement* between inflation and measures of slack. In canonical estimated models such as Smets and Wouters (2007), Justiniano et al (2012), and Del Negro et al (2015), markup shocks enter the Phillips curve as exogenous disturbances that move inflation independently of marginal cost, and are typically identified as the dominant source of high-frequency inflation variation; in particular, they can raise inflation while output and the output gap decline, thus replicating stagflationary episodes. Quantitatively, these models often attribute a substantial fraction of short-run inflation volatility to such shocks, even though other disturbances (e.g. demand) may matter more at lower frequencies. More broadly, both DSGE-based decompositions and VAR approaches that isolate "cost-push" innovations (e.g. oil or input cost shocks) find a similar pattern: inflation increases while activity contracts, producing a negative comovement between inflation and slack consistent with the New Keynesian notion of a Phillips-curve shift (see e.g. Bilbiie and Känzig, 2023 for recent evidence). This empirical strategy—treating markup shocks as reduced-form wedges in the Phillips curve—has therefore become the standard way to reconcile observed inflation dynamics with relatively flat Phillips curves, although it remains somewhat controversial precisely because it relies on large, largely exogenous residuals to account for inflation movements. For if those supply shocks are instead negative TFP shocks, then standard NK model predicts a *positive* output gap. The literature has thus reconciled this by assuming that those supply shocks were not TFP but cost-push shocks. But then why are profits—and, with them, entry too—not increasing during those episodes? The uncontroversial fall in entry and profits during such recessionary episodes is indirect evidence supporting the mechanism of our model. We discuss normative predictions, including welfare improving or maximizing monetary policy response to such recessions, in the next section.

FB: This paragraph discusses empirical evidence on negative "supply shocks" that are associated with inflation and negative output gaps (such as 70s oil shocks and COVID). If those supply shocks are negative TFP shocks, then standard NK model would predict a *positive* output gap. The literature has reconciled this by assuming that those supply shocks were not TFP but cost-push shocks. But then why are profits not increasing during those episodes? Evidence of our model: negative impact of entry during those episodes (also connected to negative profits). Mention that we will discuss normative predictions (welfare improving/maximizing monetary policy response) in the next section.

MM: the part below used to be together with the discussion after the PC, just before this previous sub(sub)section. Now having written this long-ish paragraph above, I (Florin) am not sure what the right sequencing is anymore. Perhaps slightly rewritten it does flow well here and gives a good transition to the Welfare section. That would be my favorite option and happy to either do this or follow you.

In our model, closing the output gap does not restore efficiency due a combination of the inflationary costs and distortions associated with entry. This inefficiency result for the output gap is reminiscent of a "cost-push" shock, traditionally motivated as an exogenous shock to the markup. In our model, the markup also varies endogenously with the endogenous entry response. It induces a similar response for the output gap as an exogenous "cost-push" shock.[33]

However, in the standard NK model with exogenous entry, the optimal monetary response to a stagflationary "cost-push" shock (resulting from an exogenous increase in the markup) is *tightening*. This worsens the output gap but raises welfare by reducing inflation. In the next section, we show how this response of monetary policy is reversed in our model with endogenous entry. Closing the output gap with monetary *easing* in response to the stagflationary TFP shock will improve welfare.

4 Welfare and Policy Implications

In this section we analyze the welfare implications of our model combining TFP and monetary policy responses. This requires the full derivation for the time path of the nonlinear equilibrium response under both price and wage rigidities. Due to the complexities associated with this problem, we focus on the case of fixed price and wages and derive results for the direction of the welfare responses. Thus, we impose $\psi = \psi_w = 0$. As we described in Section 2, the resulting Euler equation under the fixed-real rate is (13), $\rho_t Y_t^{-\frac{1}{\sigma}} = \exp(-\epsilon_t)$. Solving for output, cutoff, and surviving

[33] As an (endogenous) time-varying real distortion, this has a similar flavor to the implications of real wage rigidities studied by Blanchard and Gali (2010).

firms yields:

$$Y_t = F\left(\frac{\theta A_t - (\theta-1)}{\theta f_O}\right)^{\frac{\sigma}{\theta-\sigma}} (A_t)^{-\frac{1}{k}\frac{\sigma(\theta-1)}{\theta-\sigma}} \exp(\frac{\sigma\theta}{\theta-\sigma}\epsilon_t) \tag{26}$$

$$L_t = \frac{\theta}{\theta-1} Y_t^{1-\frac{1}{\sigma}} \exp(\epsilon_t) \tag{27}$$

$$q_{0t} = (A_t)^{-\frac{1}{k}} F$$

$$S_t = Y_t^{1-\frac{1}{\sigma}} \frac{k-(\theta-1)}{k} \frac{\theta A_t - (\theta-1)}{\theta f_O} \exp(\epsilon_t).$$

MM: F is defined in section 3 in the 16 system of equations.

The response of output is similar to the one derived under representative firm, with an important extra term having to do with selection that acts as a dampening force; however, this effect of heterogeneity and selection on aggregates closely follows the one obtained under flexible prices, as can be seen from the expression of the cutoff. In other words, there is no meaningful interaction between selection and nominal rigidity. We view this as an important benchmark, not as a realistic description of how these two features operate in the real world: the study of features that break this separation is independent in its own right and should be facilitated by this benchmark.

Welfare implications of alternative policies

In Figure 3 we plot, first, the nonlinear solution of the model as a function of the TFP shock under the benchmark calibration—for flexible (red dash) and fixed (solid blue), and under two alternative policies: an optimal monetary policy (blue empty dots) derived above, and a gap-closing monetary expansion (red dot-dash).

In the left panel, we have our main "positive" message: there is a positive output gap in response to positive TFP shocks, and a negative one to negative shocks; the effect is large ("first order"), unlike the model with sticky prices only. A monetary easing leads to an expansion, and a properly-designed monetary rule can—by systematic *easing*—close the output gap.

In the right panel, we have our main normative results: We plot nonlinear (lifetime) utility using the full model solution. Since the equilibrium is efficient under flexible prices and wages (by virtue of the constant labor subsidy), utility in that equilibrium is an upper envelope of utility under nominal rigidity: the welfare gap (between flex and fixed) is positive for both negative and positive shocks. A monetary easing (blue dots) does lead to a welfare improvement when it responds to a negative shock, because, being expansionary, it mitigates the negative output gap. (It would lead to a welfare deterioration to positive shocks, because it generates an inefficient expansion: output is already too high in the sticky equilibrium, the output gap is positive, so increasing it further moves the economy further from the efficient frontier). Note that this is the opposite conclusion relative to the standard NK model, where a monetary easing would be good (and indeed required) for positive shocks. and a contraction for negative shocks.

A state-contingent monetary policy that closes the output gap (red dot-dash) does lead to a welfare improvement throughout; this is because it de facto automatizes a monetary contraction in response to positive TFP shocks. Again, this is completely reversed relative to the NK model. Furthermore, this has a useful interpretation in terms of r-star. In the NK model, a policy that closes the output gap would track r-star, and thus be contractionary (because r-star is increasing with TFP shocks). In our economy, tracking r-star is indeed the opposite of a welfare-improving policy and would uniformly deteriorate welfare.

MM: Contrast to Figure 1 and remind reader that Y curve for standard NK has "opposite" output gap, in the sense of the curve starting *above* flex (negative TFP shock) and *below* (positive TFP shock)

FB: Should we get rid of the gap-closing curve? We derive analytically what it will look like in both panels (LHS is obvious overlap with flex-price; RHS will have same first-order properties as welfare-maximizing policy). Some more minor formatting: I propose getting rid of the axis ticks marks and just show one level at $A = 1$? And should we label with t subscript? These last comments also apply to Figure 1.

Finally, we study the effect of monetary policy on welfare in the fully-rigidity case. The main result is the following Proposition.

Proposition 7. *A monetary policy expansion improves welfare when starting from a situation with below-trend TFP and a negative output gap, as long as the condition* (2) *holds; conversely, a monetary contraction improves welfare with above-trend TFP and a positive output gap.*

This is neither trivial nor immediate: even though a monetary expansion goes towards closing the output gap (which is "good"), it also leads to higher hours worked (which is "bad")—so there is a tradeoff. In our model, as long as the condition (2) holds, that tradeoff is resolved in favor of the monetary expansion in a TFP-driven recession. We prove this formally in the Appendix C by replacing the equilibrium expressions derived in (26) in the utility function and doing comparative statics with respect to a monetary expansion in a low-TFP state. A monetary contraction is instead optimal in an (above-trend) TFP-driven expansion: that reduces the output gap but also reduces

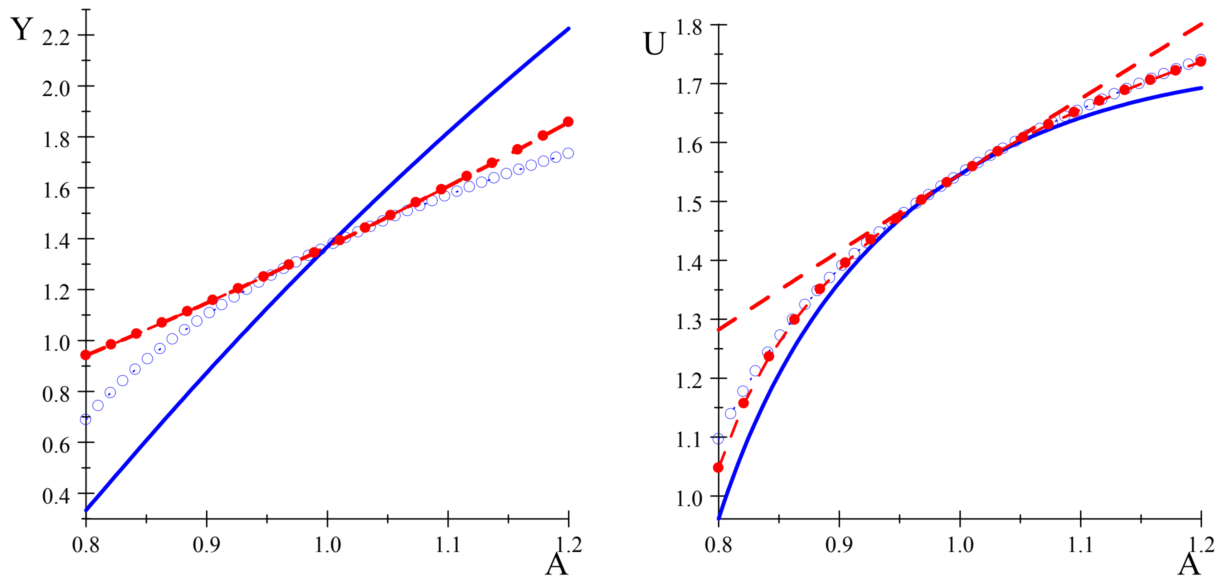


Figure 3: Y^* (flex. prices) red dash, Y (sticky prices) solid blue, Adding monetary expansion (blue dots) and gap-closing MP (red dots)

hours worked. This property overturns a key implication of the standard NK model, where a monetary expansion is welfare-reducing in response to a low-TFP state, and increases welfare in a high-TFP state. As we recall from the left panel of Figure 1 and the discussion therein, a low TFP realization triggers an expansion in that model, which is obtained by working more hours. Since an "envelope" result applies to that model too for welfare, the effect on utility is unambiguously negative; but since a monetary expansion increases output even further (with a further increase in hours), it deteriorates welfare—this is why a contraction is in fact needed in the no-entry NK model. We prove this formally in the Appendix C too.

Finally, we can find the welfare-maximizing monetary policy (a simple, sttaic version of a Ramsey problem) by maximizing utility subject to the equilibrium values derived under fixed prices and wages in (26). This delivers our final proposition.

Proposition 8. *Optimal monetary policy is countercyclical, requiring an expansion in a TFP-driven recession.*

$$\begin{aligned}\exp(\epsilon_t^{opt}) &= \left(\frac{\theta}{\theta-1}\right)^{-\frac{\varphi}{1+\varphi}} Y_t^{-\left(1-\frac{1}{\sigma}\right)\frac{\varphi}{1+\varphi}} \\ &= \left(\left(\frac{\theta}{\theta-1}\right)^{\sigma\frac{\theta-1}{\theta-\sigma}} \left(\theta A_t-(\theta-1)\right)^{\frac{\sigma-1}{\theta-\sigma}} \left(A_t\right)^{-\frac{1}{k}(\theta-1)\frac{\sigma-1}{\theta-\sigma}}\right)^{-\frac{\varphi}{1+\varphi\sigma\frac{\theta-1}{\theta-\sigma}}}\end{aligned}$$

Furthermore, to first order, this is identical to the monetary easing required to close the output gap ϵ_t^{opt}. Thus, under the optimal policy, the output gap is stabilized to a linear approximation.

It can be seen by direct inspection that (under our maintained assumption $\sigma > 1$) optimal policy requires easing "automatically" whenever there is a recession (the first line) and also in reduced form, whenever there is a negative TFP shock (the second line). This and the proof of the second statement can be seen most clearly in linearized form, taking logs around the same steady state as above:

$$
\begin{aligned}
d\epsilon_t^{opt} &= -\left(1 - \frac{1}{\sigma}\right) \frac{\varphi}{1+\varphi} d\hat{Y}_t \\
&= -\frac{(\sigma - 1)}{(\theta - \sigma)\,\varphi^{-1} + \sigma\,(\theta - 1)} \left(\theta - \frac{\theta - 1}{k}\right) d\hat{A}_t \\
&= (\Gamma + 1)\left(-d\hat{A}_t\right)
\end{aligned}
$$

An important observation is that the value of the optimal monetary expansion thus derived $d\epsilon_t^{opt}$ is identical to the gap-closing value $d\epsilon_t|_{d\tilde{Y}_t=0}$ we derived in Proposition 6, Note, however, that this equivalence holds *only to first order*. The gap-closing policy is still inefficient du to higher-order terms, for two reasons. First, the inefficient change in the number of varieties (which already appears in the simplest version of our model and we characterized by a second-order approximation to the welfare function); indeed, the gap in the number of firms (aka the entry multiplier) under the policy closing the output gap $d\epsilon_t|_{d\tilde{Y}_t=0}$ is $\hat{N}_t - \hat{N}_t^* = (\theta - 1)\,\hat{A}_t$. And second, when prices are not fixed, there is also inflation, which we also characterized in our second-order approximation as a welfare cost. Both of these welfare costs feature analytically—even when the output gap is closed—in the second-order approximation to the welfare function we derived in Appendix A.3.

The blue dot curves in Figure 3 plot output and welfare under this optimal policy. It is important to understand that while the output gap is closed to first order under the optimal policy $\tilde{Y}_t^{opt} = 0$, it is still negative to higher order and the policy is still "second best"—there are welfare costs associated with the change in the number of varieties, which under both (optimal and gap-closing) policies vary to first order $\hat{N}_t\left(\epsilon_t^{opt}\right) - \hat{N}_t^* = (\theta - 1)\,\hat{A}_t$, creating second-order welfare costs as we established analytically in AppendixA.3 even in the simplest sticky-price-only model (the equilibrium of which is in fact replicated by the optimal policy). This is illustrated in our Figure 3: welfare under the optimal policy is identical to first-order to the gap-closing value, and output under the optimal policy is also first-order identical to the flexible-equilibrium value (just as under fixed-prices only).

FB: Add some description regarding contrast of our result with fixed prices relative to standard NK with arbitrary stickiness **with monetary policy that eliminates inflation costs**

The above results apply for fixed prices and wages. In standard NK with both sticky prices and wages where "modified divine coincidence" holds, a policy that stabilizes the average measure of inflation derived above is exactly equivalent to a policy tracking r-star in that model. In our model, all these policies are very different. We already discussed how targeting average $\bar{\pi}_t = 0$ leaves an entry-related endogenous cost-push-like term breaking divine coincidence. As to tracking r-star, this is in fact not only inefficient (and not equivalent to $\bar{\pi}_t = 0$, either)—but it does the exact **opposite** relative to RANK (where it is equivalent to $\bar{\pi}_t = 0$ and closing the output gap). Therefore, it worsens welfare relative to the baseline (without policy) equilibrium: it induces a contraction (r-star still goes up in our model) when in fact the economy needs an expansion. This is one of the sharpest policy lessons of our model: simply tracking r-star can dramatically backfire in an economy with endogenous entry.

Furthermore, even if in response to a cost-push (markup) shock in the standard NK model a monetary easing will indeed close the output gap, it will in fact generically worsen welfare because of its inflaitonary effect (in other words, the optimal policy in the standard NK model in response to a stagflationary cost-push shock is a monetary contraction, trading off its inflation-reduction benefit against a worsening of the output gap). Our main result is that even though the (TFP) supply shock is now stagflationary, the welfare-maximizing monetary response is reversed: it amounts to *easing* instead of contracting.

That a monetary easing in response to a negative TFP shock *does not create extra inflation* is a subtle point that requires some elaboration. Prima facie, this makes it sounds like a downward-sloping Phillips curve. However, this is not about the PC slope, but rather about an endogenous, entry-driven shift of the curve when demand shifts. A negative TFP reduces both productivity and profits and implies fewer entrants, fewer varieties, and lower markups. Inflation rises due to these variety and markup effects. But, crucially, when the central bank eases, demand and profitability rise, which boosts entry; variety expands, markups increase, and consequently inflation declines again. The mechanism behind "easing curtails inflation" is there: not a slope reversal, but an endogenous "supply restoration" channel. In other words, it is not that the Phillips curve itself flips but that policy-induced entry shifts the effective supply curve endogenously.

FB: Emphasize our main result: Even though easing will close the output gap in that case, this *worsens* welfare. So it still gets the prediction for the welfare maximizing monetary response *reversed*?

5 Conclusion

The following table summarizes the qualitative predictions of our framework relative to the standard New Keynesian model (both with and without cost-push shocks): TFP shocks become observationally supply shocks without counterfactual profit dynamics, and with intuitive monetary-policy implications. This comparison highlights the paper's central contribution: endogenous entry allows productivity shocks to generate stagflationary dynamics and procyclical profits, a combination that existing models struggle to reconcile. And it implies that monetary easing is the right policy response to adverse (stagflation-inducing) TFP shocks.

Need some text to introduce this as our main result

Table: Summary of Model Predictions

Model/Shock	$d\pi$	$d\,(gap)$	Optimal MP(stabilization)
Standard NK			
TFP shock	↑	↑	Tighten: both $\pi \downarrow$ & close gap
Cost-push	↑	↓	**Tighten**, trade-off $\pi \downarrow$ & worsen gap
Entry NK			
TFP shock	↑	↓	**Ease**, No trade-off: $\pi \downarrow$ & close gap

This paper studies the macroeconomic and policy implications of productivity shocks in a New Keynesian economy with endogenous firm entry, selection, and nominal rigidities. Our central finding is that allowing for endogenous adjustment along the extensive margin fundamentally alters the nature of TFP shocks. Once entry is taken seriously, productivity shocks no longer behave like disguised demand shocks—as they do in the standard New Keynesian model—but instead acquire the defining features of true supply shocks.

The core mechanism is an entry multiplier. Under nominal rigidities, productivity shocks induce large and persistent movements in firm entry and exit, far exceeding those obtained under flexible prices. In a benchmark environment with flexible wages, this amplification operates entirely through the extensive margin and leaves the output gap unaffected to first order. Despite large reallocations in the number of active firms, actual output coincides with its flexible-price counterpart. This neutrality result provides a sharp analytical benchmark: nominal rigidities dramatically magnify entry responses, even when allocative efficiency is preserved.

Introducing sticky wages breaks this neutrality. In the full model, adverse productivity shocks reduce profits and expected profitability; with sticky prices, firms cannot immediately restore margins, triggering exit and a contraction in product variety. Sticky wages prevent hours from adjusting efficiently, so output falls by more than in the flexible-price-and-wage allocation. As a result, negative TFP shocks generate a negative output gap while remaining inflationary. This comovement—higher inflation and increased slack—matches the defining characteristics of stagflationary supply disturbances and stands in stark contrast to the standard New Keynesian prediction.

A key implication of this mechanism is that TFP shocks endogenously generate a cost-push wedge in the Phillips curve. Entry and exit affect marginal costs through endogenous movements in product variety and markups, making productivity shocks observationally similar to markup shocks in representative-firm models. Crucially, however, the implications for profits and entry are very different. In our framework, profits and entry are procyclical in response to productivity shocks, consistent with the empirical evidence. This resolves a long-standing tension in New Keynesian models, where cost-push shocks can generate stagflation only at the cost of countercyclical profits.

These results have first-order implications for monetary policy. When wages are sufficiently sticky relative to prices, a monetary expansion raises profits, stimulates entry, and expands productive capacity. In this region, expansionary policy can close the negative output gap induced by an adverse productivity shock and improve welfare. This prescription is the opposite of the standard New Keynesian one, which calls for contractionary policy in response to inflationary productivity shocks or stagflationary markup shocks. In our framework, tightening in response to TFP-driven inflation would exacerbate the endogenous contraction in entry and deepen the recession.

We characterize analytically the degree of wage stickiness required for entry to respond positively to monetary expansions. When this condition holds, monetary policy operates primarily through the extensive margin: demand expansions increase profitability, induce entry, and raise output by expanding the economy's productive base. This channel is absent in representative-firm models and highlights the importance of firm dynamics for policy transmission.

While our quantitative results are obtained in a model with firm heterogeneity and endogenous selection, these features affect magnitudes rather than mechanisms. The model remains mathematically isomorphic to the standard New Keynesian framework, with entry and selection appearing as tractable wedges in the Phillips curve and resource constraints. This structure allows us to deliver closed-form analytical results while preserving the familiar logic and transparency of New Keynesian analysis.

More broadly, our results suggest that firm dynamics are not a second-order feature for monetary policy analysis. By endogenizing the creation and destruction of productive capacity, entry transforms the interpretation of inflation, slack, and the appropriate policy response to supply disturbances. Productivity-driven disinflation need not signal weakness, and productivity-driven inflation need not signal overheating. Understanding which margin of adjustment is operative—intensive or extensive—is therefore essential for both positive and normative analysis.

XX–OLD VERSION–XX
The responses of entry to adverse supply shocks such as supply disruptions are amplified by firms' inability to increase their prices, which leads to additional losses for individual firms. This in turn amplifies the response of entry relative to a flexible-price benchmark. We call this simple mechanism *the entry multiplier*, and we show that it operates in a wide range of models with endogenous entry and nominal rigidities. This "supply-side" amplification further induces an aggregate-demand recession; that is a fall in output under nominal rigidities that is larger than the fall in its flexible-price counterpart: a negative output gap.
We show that the only "necessary ingredients" for this aggregate demand amplification are endogenous entry and sticky good prices. However, the addition of sticky wages is necessary for two reasons: for this amplification channel to have "large" (first-order) effects, and for the model's suitability to realistically consider responses to (and deisgn of) monetary policy. Indeed, we analyze how an expansionary monetary policy can be used to dampen—and even eliminate—the negative output gap induced by supply disruptions.
In terms of our demand-side (utility) assumptions, we find that those key amplification channels and the associated monetary policy responses only rely on the empirically accepted benchmark that the elasticity of substitution (typically in the range between 4 and 8) is higher than the intertemporal elasticity of substitution for the consumption aggregate (typically below 2).
The table below compares our model's implications to those of the standard New Keynesian (NK) benchmark. We contrast (i) positive implications (how TFP and cost–push shocks affect inflation and the output gap) and (ii) normative implications (the appropriate monetary policy response). Our central message is that the model delivers a set of comovements that most closely align with (1) the economic notion of a supply shock and (2) the intuitive stabilization policy for such shocks.
A negative TFP shock in our model under a standard Taylor rule generates stagflation: inflation increases, the output gap turns negative. This matches the classic notion of an adverse supply shock—whereas in standard NK the same shock leads to a *positive* output gap, thus leading to the introduction of ad-hoc cost-push or "markup" shocks to obtain that negative comovement.
TFP shocks in our model deliver the same negative comovement that cost-push shocks do in the NK model. However, unlike in the NK model with cost-push shocks:
i. Profits are **procyclical**, not countercyclical (in NK, markup/cost-push shocks raise profits in recessions).
ii. The appropriate stabilization policy is to **ease**, not tighten.
Monetary easing raises the output gap without increasing inflation, so its tradeoff is improved.

By contrast, in the standard NK model—whether the shock is TFP or cost-push—the prescription is to tighten, even though output contracts.
These contrasts are summarized in the Table, where by "standard NK" we mean a model with both price and wage rigidity.

Model/Shock	$d\pi$	$d\,(gap)$	Divine coincidence	Optimal MP(stabilization)
Standard NK				
TFP shock	↑	↑	Modified version	Tighten: both π & gap ↓
Cost-push	↑	↓	Breaks	Tighten, trade-off: π ↓*worsen* gap
Entry NK				
TFP shock	↑	↓	Breaks (sign reversed)	**Ease**: gap ↑ but π ↓

Why easing lowers inflation in our model It may appear that monetary easing reduces inflation because the Phillips curve " flips" and becomes downward-sloping. This is not the mechanism. The slope is unchanged. Instead, easing triggers an endogenous supply-restoration mechanism: A negative TFP shock lowers productivity and expected profits. This depresses entry and reduces product variety, raising markups and inflation. When the central bank eases, demand and profitability increase. This stimulates **entry**, expands product variety, and reduces markups. Inflation falls because *supply capacity endogenously improves*. Thus the economy does *not* move along a downward-sloping Phillips curve; the Phillips curve itself *shifts* due to endogenous product-market dynamics. Monetary policy restores the supply side.
The conclusion of our analysis is that New Keynesian models should be updated to include endogenous entry—in addition to both sources of nominal rigidities, sticky prices and wages—in order to generate reasonable macroeconomic fluctuations and thus serve as a guide to analyze and design stabilization policies.

References

Aghion, P., and P. Howitt., 1992, "A Model of Growth through Creative Destruction." Econometrica, 60(2), 323–351

Alexopoulos, M. 2011, " Read All About It! What Happens Following a Technology Shock?" American Economic Review, 101(4), 1144–1179.

Argente, D., M. Lee, and S. Moreira, 2018, "Innovation and product reallocation in the great recession," Journal of Monetary Economics

Auerbach, A., Y. Gorodnichenko and D. Murphy, 2021, "Inequality, Fiscal Policy and COVID19 Restrictions in a Demand-Determined Economy," European Economic Review

Baqaee, D. and E. Farhi, 2020, " Supply and Demand in Disaggregated Keynesian Economies with an Application to the Covid-19 Crisis," American Economic Review

Barsky, R. and L. Kilian 2004: "Oil and the Macroeconomy Since the 1970s." Journal of Economic Perspectives.

Barsky, R. and E. Sims 2011: "News Shocks and Business Cycles." Journal of Monetary Economics.

Basu, S., J. G. Fernald, and M. S. Kimball, 2006. "Are Technology Improvements Contractionary?," American Economic Review, 96(5), 1418-1448

Bénassy, J.P., 1996, "Taste for variety and optimum production patterns in monopolistic competition," Economics Letters, vol. 52, pp 41-47.

Bergin, P. and G. Corsetti, 2008, “The extensive margin and monetary policy,” Journal of Monetary Economics, 55(7), pp. 1222–1237

Bilbiie F. 2008. “Limited Asset Market Participation, Monetary Policy, and (Inverted) Aggregate Demand Logic”. Journal of Economic Theory 140, 162-196.

Bilbiie, F. 2020, “The New Keynesian Cross,” Journal of Monetary Economics

Bilbiie, F. 2021, “Monetary Neutrality With Sticky Prices and Free Entry,” Review of Economics and Statistics

Bilbiie, F. 2025, “Monetary Policy and Heterogeneity: An Analytical Framework,” Review of Economic Studies

Bilbiie, F. O., F. Ghironi, and M. Melitz, 2007, “ Monetary Policy and Business Cycles with Endogenous Entry and Product Variety,” NBER Macroeconomics Annual, vol. 22, pp. 299-353

Bilbiie, F. O., F. Ghironi, and M. Melitz, 2012, “Endogenous Entry, Product Variety and Business Cycles,” Journal of Political Economy,120(2), 304-345

Bilbiie, F. O., F. Ghironi, and M. Melitz, 2019, “Monopoly Power and Endogenous Product Variety: Distortions and Remedies” American Economic Journal: Macroeconomics.

Bilbiie, F. O., I. Fujiwara, and F. Ghironi, 2014, “Optimal Monetary Policy with Endogenous Product Variety,” Journal of Monetary Economics, 2014, vol. 64, pp. 1-20.

Bilbiie, F. O. and D. Känzig 2023: "Greed? Profits, Inflation, and Aggregate Demand." NBER Working Paper 31618

Blanchard, O. and J. Galí, 2007, “Real Wage Rigidities and the New Keynesian Model”, Journal of Money, Credit, and Banking

Blanchard, O., and J. Galí. 2010. "Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment." American Economic Journal: Macroeconomics 2 (2): 1–30.

Broer, T., N. J. Hansen, P. Krusell, and E. Oberg, 2020 “The New Keynesian transmission mechanism: a heterogeneous agent perspective,” Review of Economic Studies

Cacciatore, M. and G. Fiori, 2016, “The Macroeconomic Effects of Goods and Labor Markets Deregulation,” Review of Economic Dynamics

Campbell, J., 1998, “Entry, Exit, Embodied Technology, and Business Cycles.” Review of Economic Dynamics, 1(2), 371-408

Cantore, C., M. León-Ledesma, P. McAdam, A. Willman, 2014, “Shocking Stuff: Technology, Hours, And Factor Substitution,” Journal of the European Economic Association, 12(1), 108-128

Cantore, C., F. Ferroni, and M. León-Ledesma, 2021, "The Missing Link: Monetary Policy and the Labor Share ." Journal of the European Economic Association, 19(3), 1592-1620.

Cesa-Bianchi, A. and A. Ferrero, 2021, “The Transmission of Keynesian Supply Shocks,” Mimeo.

Chaney, T., 2008 "Distorted Gravity: The Intensive and Extensive Margins of International Trade." American Economic Review, 98(4), 1707-1721

Chari, V.V., P. Kehoe, and E. McGrattan, 2008. “Are structural VARs with long-run restrictions useful in developing business cycle theory?,” Journal of Monetary Economics, 55(8), 1337-1352

Chatterjee, S., and R. Cooper, 2014, “ Entry and Exit, Product Variety and the Business Cycle,” Economic Inquiry

Christiano, L., M. Eichenbaum, and C. Evans, 1997, “Sticky Price and Limited Participation Models of Money: A Comparison”, European Economic Review

Christiano, L., M. Eichenbaum, and C. Evans, 2005 “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy”, Journal of Political Economy, 113:1, 1-45

Christiano, L., M. Eichenbaum, and R. Vigfusson, 2003, “ What Happens after a Technology Shock?” Mimeo

Clementi, G. L., and B. Palazzo. 2016. “Entry, Exit, Firm Dynamics, and Aggregate Fluctuations.” American Economic Journal: Macroeconomics, 8 (3): 1-41

Colciago, A. and F. Etro, 2010 “Endogenous Market Structures and the Business Cycle” The Economic Journal

Colciago, A. and R. Silvestrini, 2022, “Monetary Policy, Productivity, and Market Concentration,” European Economic Review

Cook, D., 2001, “ Time to Enter and Business Cycles,” Journal of Economic Dynamics and Control 25 (8): 1241-1261

Cooke, D. and T. Damjanovic, 2020, “Monetary Policy and Welfare with Heterogeneous Firms and Endogenous Entry,” mimeo

De Loecker, J., and J. Eeckhout, 2017 "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper No. 23687.

Devereux, M. B., A. C. Head, and B. J. Lapham, 1996, “ Aggregate Fluctuations with Increasing Returns to Specialization and Scale,” Journal of Economic Dynamics and Control 20: 627-656.

Dixit, A. K., and J. E. Stiglitz, 1977, “ Monopolistic Competition and Optimum Product Diversity,” American Economic Review 67: 297-308.

Dixon, H. and A. Savagar, 2020, “Firm Entry, Excess Capacity and Endogenous Productivity,” European Economic Review

Edmond, C., V. Midrigan, and D. Xu, 2019, “How costly are markups?,” Journal of Political Economy

Erceg, C. J., D. W. Henderson, and A. T. Levin, 2000, “ Optimal monetary policy with staggered wage and price contracts”. Journal of Monetary Economics 46 (2), 281-313.

Ethier, W., 1982, “National and international returns to scale in the modern theory of international trade,” American Economic Review, 72, 389-405.

Etro, F. and L. Rossi, 2015, “New-Keynesian Phillips curve with Bertrand competition and endogenous entry,” Journal of Economic Dynamics and Control

Fernald, John G., 2014, "Productivity and Potential Output Before, During, and After the Great Recession." NBER Macroeconomics Annual (29), 1-51

Fernald, John G., and J. Christina Wang, 2016, "Why Has the Cyclicality of Productivity Changed? What Does It Mean?" Annual Review of Economics, 8, 465-496

Fornaro, L. and M. Wolf, 2020, “The Scars of Supply Shocks,” Mimeo

Galí, J. 1999, “ Technology, Employment and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” American Economic Review 89/1, 249–7

Galí, J., 2011, " The Return of the Wage Phillips Curve", Journal of the European Economic Association, 9(3), 436–461.

Galí, J. 2015, “Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications”, Princeton University Press, Second Edition

Ghironi, F. and M. J. Melitz., 2005, “International Trade And Macroeconomic Dynamics With Heterogeneous Firms,” Quarterly Journal of Economics, 120(3), 865-915

Gopinath, G., and B. Neiman. 2014. “Trade Adjustment and Productivity in Large Crises.” American Economic Review, 104 (3): 793-831

Gourio, F., and L. Rudanko., 2014, "Customer Capital." The Review of Economic Studies, 81(3), 1102-1136

Guerrieri, V., G. Lorenzoni, I. Werning, and L. Straub, 2020 “Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages,” American Economic Review

Gutierrez, G., C. Jones and T. Philippon, 2021, “Entry Costs and Aggregate Dynamics,” mimeo

Hamano M. and F. Zanetti, 2017, “Endogenous Product Turnover and Macroeconomic Dynamic,” Review of Economic Dynamics

Hamano M. and F. Zanetti, 2022, “Monetary Policy, Firm Heterogeneity, and Product Variety,” European Economic Review

Hamano, M., 2025, "Heterogeneity in Tastes, Productivities, and Macroeconomic Volatility." Mimeo

Hamano, M., and F. Pappadà., 2023 "Exchange Rate Policy and Firm Heterogeneity." IMF Economic Review, 71(3), 759-790

Helpman, E., M. J. Melitz, and S. R. Yeaple. "Export Versus FDI with Heterogeneous Firms." American Economic Review, 94(1), 300-316

Hopenhayn, H. A., 1992, "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." Econometrica, 60(5), 1127-1150

Jaimovich, N., 2007, “Firm Dynamics and Markup Variations: Implications for Multiple Equilibria and Endogenous Economic Fluctuations,” Journal of Economic Theory, 137:1, 300-325.

Jaimovich, N., and M. Floetotto, 2008, Firm Dynamics, Markup Variations and the Business Cycle, Journal of Monetary Economics, 55:7, 1238-1252.

Justiniano, A., G. E. Primiceri, and A. Tambalotti, 2010. "Investment Shocks and Business Cycles." Journal of Monetary Economics, 57(2), 132–145.

Justiniano, A., G. E. Primiceri, and A. Tambalotti, 2011. "Investment Shocks and the Relative Price of Investment." Review of Economic Dynamics, 14(1), 101–121.

Justiniano, A., G. E. Primiceri, and A. Tambalotti, 2013. "Is There a Trade-Off between Inflation and Output Stabilization?" American Economic Journal: Macroeconomics, 5(2), 1–31

Kaplan, G. and P. Zoch, 2020, “Markups, Labor Market Inequality and the Nature of Work,” Mimeo

Kim, J., 2004. “What determines aggregate returns to scale?,” Journal of Economic Dynamics and Control, 28(8), 1577-1594

Klenow, P. and B. Malin, 2011, "Microeconomic Evidence on Price-Setting", the Handbook of Monetary Economics 3A, B. Friedman and M. Woodford ed.: Elsevier, 231-284.

Krugman, P., 1980, “Scale Economies, Product Differentiation, and the Pattern of Trade,” American Economic Review, 70:5, 950-59

Lee, Y. and T. Mukoyama, 2018. “A model of entry, exit, and plant-level dynamics over the business cycle,” Journal of Economic Dynamics and Control, 96(C), 1-25.

Lewis, V., 2009, “ Business Cycle Evidence on Firm Entry.” Macroeconomic Dynamics

Lewis, V. and C. Poilly, 2012, “Firm Entry, Markups and the Monetary Transmission Mechanism,” Journal of Monetary Economics 59(7), 670-685.

Melitz, M., 2003, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” Econometrica 71(6), pp. 1695-1725

Melitz, M. J., and S. J. Redding, 2015, "New Trade Models, New Welfare Implications." American Economic Review, 105(3), 1105–1146

Michelacci,C., L. Paciello, and A. Pozzi, 2019, “The extensive margin of aggregate demand,” Review of Economic Studies

Mussa, M., and S. Rosen, 1978, "Monopoly and Product Quality." Journal of Economic Theory, 18(2), 301-317

Romer, P., 1986, Increasing Returns and Long-Run Growth, Journal of Political Economy, Vol. 94, No. 5. pp. 1002-1037

Romer, P., 1987, Growth Based on Increasing Returns Due to Specialization, American Economic Review, Vol. 77, No. 2, pp. 56-62

Schmitt-Grohé, S. and M. Uribe, 2006, “ Optimal fiscal and monetary policy in a medium-scale macroeconomic model”. NBER Macroeconomics Annual 2005, vol. 20, 383-462.

Sedlacek, P. and V, Sterk, 2017, “The growth potential of startups over the business cycle,” American Economic Review.

Sedlacek, P., 2020, “Lost generations of firms and aggregate labor market dynamics,” Journal of Monetary Economics

Syverson, C., 2011, "What Determines Productivity?" Journal of Economic Literature, 49(2), 326-365

Woodford, M., 2003, Interest and prices: foundations of a theory of monetary policy, Princeton University Press

Woodford, M., 2020 “Effective Demand Failures and the Limits of Monetary Stabilization Policy,” American Economic Review.

Appendix

A Derivations and Complete General Model Outline

In this Appendix, we first provide some key derivations for the simplest benchmark model and then outline the full general model with both sticky prices and wages.

A.1 Calibration equalizing steady states across models

In Figure 1, we choose the fixed cost f in order to make models consistent in the steady state, when the shock is absent $A = 1$; i.e. we pick f that equalizes Y^* to $Y^{*\mathcal{N}}$, $\theta f = \bar{L}\left(\frac{\theta-1}{\theta}\right)^{\theta-1}$. Then, we choose money supply M to equalize the SP equilibrium Y with FE to this same $Y^* = Y^{*\mathcal{N}}$. This requires, using f:

$$Y = \frac{\theta-1}{\theta}\left(\frac{\bar{L}}{f} - \frac{M_t}{f\bar{p}}\right)^{\frac{1}{\theta-1}} = \bar{L} \rightarrow \frac{M}{\bar{p}} = \bar{L}\left(1 - \frac{1}{\theta}\right)$$

With no entry N (left panel) we plot $Y_t^{*\mathcal{N}} = A_t\bar{L}$ and $Y_t^{NS} = \frac{M}{P}$, while for the free entry (E) model we plot, for the two cases (replacing $f = \frac{\bar{L}}{\theta}\left(\frac{\theta-1}{\theta}\right)^{\theta-1}$ and $\frac{M}{\bar{p}} = \bar{L}\left(1-\frac{1}{\theta}\right)$):

$$Y_t = \bar{L}\left(\theta A_t - \theta + 1\right)^{\frac{1}{\theta-1}} \text{ and } Y_t^* = A_t^{\frac{\theta}{\theta-1}}\bar{L}.$$

A.2 Second-order solution in the simplest model

To understand what drives the key result that completely overturns the propagation of supply shocks in the no-entry New Keynesian model, we compare the two free-entry equilibria using a second-order approximation around the point $Y = Y^*$. This leads to our Proposition 9. Let temporarily small letters denote log deviations, and small letters *with a tilde* denote deviations as a share of the steady-state value, i.e. $\tilde{x}_t \equiv (X_t - X)/X$. Of course, when focusing on a first order approximation this distinction is immaterial, since to first order $\tilde{x}_t \simeq x_t$; but to second order the two are different ($\tilde{x}_t \simeq x_t + \frac{1}{2}x_t^2$).

Proposition 9. *To second order, output under flexible and sticky prices is, respectively:*

$$\tilde{y}_t^* \simeq \frac{\theta}{\theta-1}a_t + \frac{1}{2}\frac{\theta}{(\theta-1)^2}a_t^2,$$
$$\tilde{y}_t \simeq \frac{\theta}{\theta-1}a_t + \frac{1}{2}\frac{\theta^2(2-\theta)}{(\theta-1)^2}a_t^2. \tag{28}$$

Therefore, the "output gap" is:

$$\tilde{y}_t - \tilde{y}_t^* \simeq -\frac{1}{2}\theta a_t^2. \tag{29}$$

But the output gap response is always negative, due to the second-order effect. There is also an asymmetry: output increases by less in response to positive shocks, but falls by more in response to negative shocks. For large negative shocks in particular, the response under sticky prices can be substantially larger.

Dissecting the Mechanism. The key to understanding these second-order (concavity) effects lies in the equilibrium dependence of aggregate output to the number of intermediate inputs $Y(N)$. As we already noted, N itself is a linear function of A. In other words, N is (linearly) amplified through our entry multiplier. Y is then amplified further through second-order effects. In particular, consider the "aggregate production function":

$$Y_t = N_t^{\frac{\theta}{\theta-1}}\left(\frac{A_t\bar{L}}{N_t} - f\right). \tag{30}$$

A second-order approximation around the steady-state equilibrium yields:

$$\begin{aligned}\tilde{y}_t &\simeq \frac{\bar{L} - \theta f N}{(\theta-1)\left(\bar{L} - fN\right)} n_t + \frac{1}{2}\frac{(2-\theta)\bar{L} - \theta f N}{(\theta-1)^2\left(\bar{L} - fN\right)} n_t^2 \\ &= \frac{\theta}{\theta-1}\frac{1 - \frac{N}{N^*}}{\theta - \frac{N}{N^*}} n_t - \frac{1}{2}\frac{\theta}{(\theta-1)^2}\frac{\theta + \frac{N}{N^*} - 2}{\theta - \frac{N}{N^*}} n_t^2 \\ &= -\frac{1}{2}\frac{\theta}{(\theta-1)^2} n_t^2.\end{aligned} \tag{31}$$

The first linear term drops out given that $N = N^* = \bar{L}/\theta f$ (The *steady-state* value is the same as in the efficient flexible-price equilibrium). Thus, the output gap is always zero to a first order. Intuitively, entry implies an adjustment mechanism such that if demand is too high and therefore profits too low (due to sticky prices), some firms exit. This reduces aggregate output through the variety effect, and since variety provision is efficient with CES preferences and flexible prices, output is the same as in the flex-price level to a first order (although the number of varieties is inefficiently small); in other words, the individual per-firm labor demand shifts, but the number of firms moves so as to offset the effect on aggregate labor demand to first order. This is a more general case of the local neutrality result in response to monetary shocks first emphasized in Bilbiie (2021).[34]

The amplification of exit (lower N) to a negative productivity shock (lower A) thus generates

[34] We show in Appendix D.2 that this first-order irrelevance holds generally for arbitrary price stickiness.

amplification for the fall of real output Y. This effect, which operates through the concavity of consumption in the number of intermediates, is decreasing with the benefit of variety $(\theta-1)^{-1}$.[35] It is thus determined by the same parameter governing the amplification of entry itself. Naturally, when the benefit of variety (the degree of increasing returns to specialization) vanishes there is no curvature of output in the number of varieties. The degree of increasing returns to specialization is crucial for the balance between the extensive and intensive margin adjustment that becomes distorted under sticky prices. More intensive-margin adjustment would be desirable but is unfeasible, and this distortion is *less* important when goods are closer substitutes: θ larger, less returns to scale (less benefit of variety), less distortion. To summarize, θ determines both entry amplification and the concavity distortion, but has opposite effects on these two forces.

Overall, the net effect of θ is to amplify the difference between flexible and sticky-price allocations: the positive effect through the entry multiplier is proportional to θ^2, while the negative effect through (31) is proportional to θ^{-1}, (i.e. $\theta(\theta-1)^{-2}$). We disentangle these two forces subsequently, using preferences that break this link between the degree of returns to scale and the elasticity of substitution.

A.3 Stabilization Policy Implications and Second-order approximation to utility

As our discussion at page 15 anticipated, the simplest, stripped-down version of our model is unsuitable for studying monetary policy or demand shocks; we postpone a detailed discussion of this to section (3). Here, we nevertheless prove a "divine coincidence" result analogous to fixed-entry economies (Blanchard and Gali, 2007): the central bank can replicate the efficient flexible-price level of output while at the same time *also* stabilizing inflation (in a version where prices are not fixed but arbitrarily sticky).[36]

This can be seen directly by replacing the free-entry condition written for arbitrarily sticky prices, as a function of the markup (8) into the aggregate accounting equation (7), obtaining (assuming log utility in consumption without loss of generality, so that hours are constant at $\bar{L}$):

$$Y_t = \frac{1}{\mu_t}\left[\frac{1}{f}\left(1-\frac{1}{\mu_t}\right)\right]^{\frac{1}{\theta-1}}\left(A_t\bar{L}\right)^{\frac{\theta}{\theta-1}}. \tag{32}$$

[35] In Appendix A.3, we provide an alternative interpretation (for an arbitrary degree of price stickiness). We take a second-order approximation to household utility (Woodford, 2003, Chapter 6) delivering a loss function in squared inflation and the gap of the number of firms from its flex-price level (equation (33) therein); replacing the latter equilibrium expressions we obtain the equivalent of Proposition 9 above.

[36] We are grateful to an anonymous referee who suggested both emphasizing this property and the connection with the second-order welfare approximation.

It follows directly that stabilizing the markup at its flexible-price level $\mu^* = \theta / (\theta - 1)$ and thus eliminating inflation in individual prices, delivers the flexible-price level of real activity Y_t^* in Table 1, or vice versa: there is no conflict between the two objectives, i.e. "divine coincidence".

This can be further illustrated by taking a second-order approximation to household utility, following e.g. Woodford (2003, Chapter 6). The derivation, described in the Appendix A.3 (for the benchmark case of log utility in consumption that isolates our channel) delivers the quadratic loss function:

$$\mathcal{L}_t^E \simeq -\frac{1}{2} \left[\kappa \pi_t^2 + \frac{\theta}{(\theta - 1)^2} (n_t - n_t^*)^2 \right], \tag{33}$$

capturing the costs of (squared) individual-prices inflation (κ is the Rotemberg adjustment-cost coefficient) and the gap of the number of firms relative to the flexible-price level. The latter is a sufficient statistic for the welfare loss, due to the local neutrality result emphasized above: output is equal to the first order to flexible-price output, but it is different to the second order because of the extensive-margin concavity effects discussed above. Replacing the equilibrium expressions of n_t (under fixed prices, so for $\pi_t = 0$) and n_t^*, we obtain exactly the second-order approximation of the output gap in Proposition 9 above.

This welfare criterion can be used to assess the implications of different, suboptimal policy rules. For example, the suboptimal rule of fixing the money supply (or, with fixed prices, the real interest rate) "costs" $L_t^E = -\frac{1}{2} \theta a_t^2$ in the endogenous-entry model; while in the fixed-entry model (where the loss function is readily derived in e.g. Woodford, 2003; or Gali, 2015), it is $L_t^N = -\frac{1}{2} (1 + \varphi) a_t^2$, where φ is the inverse labor elasticity . This illustrates that the models are not directly comparable (they are not nested in one another); the welfare costs are determined by different features of the economy—the benefit of variety and elasticity of substitution, in the former case; and labor elasticity, in the latter. Furthermore, while the fixed-monetary-policy rule is suboptimal in response to negative productivity shocks and thus costly in both economies, its underlying implications are radically different: a positive output gap under fixed entry, but a negative output gap in our free-entry economy. This is the key takeaway of our benchmark model.[37]

[37] Another possibility is that the policymaker stabilizes output at the flex-price level of the no-entry economy. This is evidently costly in the free-entry economy, for there is a first-order, linear term distortion too. In particular, the gap between the flex-price output of the fixed-entry economy and the efficient free-entry equilibrium, approximated to second order, is:

$$y_t^{*\mathcal{N}} - y_t^* \simeq -\frac{1}{\theta - 1} a_t - \frac{1}{2} \frac{\theta}{(\theta - 1)^2} a_t^2.$$

Second-order Approximation

Note that we approximate around the steady state of the flex-price equilibrium (which is the same as for the SP equilibrium) with $N^* = \frac{A\bar{L}}{f\theta}$ and $C^* = (N^*)^{\frac{1}{\theta-1}} \left(A\bar{L} - N^* f\right) = (N^*)^{\frac{1}{\theta-1}} \frac{\theta-1}{\theta} A\bar{L}$. A second-order approximation to utility around this steady state (which, by virtue of free entry, is efficient) delivers:

$$\hat{U}_t \equiv U(C_t, L_t) - U(C, L) \simeq U_C C \frac{C_t - C}{C} + U_L L \frac{L_t - L}{L} + \frac{1}{2} U_{CC} C^2 \left(\frac{C_t - C}{C}\right)^2 + U_{LL} L^2 \left(\frac{L_t - L}{L}\right)^2$$
$$= U_C C \left[c_t + \frac{1 - \sigma^{-1}}{2} c_t^2\right] + U_L L \left[l_t + \frac{1 + \varphi}{2} l_t^2\right] + t.i.p + O\left(\| \zeta \|^3\right),$$

where small letters denote log-deviations from steady state $c_t \equiv \log \frac{C_t}{C}$ and we used

$$\frac{C_t - C}{C} \simeq c_t + \frac{1}{2} c_t^2,$$

and same for L_t. Finally, $t.i.p$ are terms independent of policy and $O\left(\| \zeta \|^3\right)$ groups all terms of order 3 or higher.

Next, note that we focus here on the simple case of logarithmic utility in consumption $\sigma = 1$, implying that hours worked are always fixed in equilibrium (regardless of price stickiness and regardless of monetary policy). Therefore, the second term in the approximation drops out (this allows us to focus on the entry channel that is of the essence here). Furthermore, the term in squared consumption deviations c_t^2 also drops out since $\sigma = 1$, so we are left with the linear term—that we nevertheless need to approximate to *second* order. Taking a second-order approximation of the aggregate production function/resource constraint:

$$C_t = \left(1 - \frac{\kappa}{2} \pi_t^2\right) N_t^{\frac{1}{\theta-1}} \left(A_t L_t - N_t f\right)$$

around the steady-state using that hours are always constant in equilibrium and denoting by $\delta_t = -\ln\left(1 - \frac{\psi}{2}\pi_t^2\right)$ the inflation welfare cost (which we then approximate to second order below):

$$C_t \simeq C - N^{\frac{1}{\theta-1}} (L - Nf)\, \delta_t + \left(\left(\frac{1}{\theta - 1}\right) N^{\frac{1}{\theta-1}-1} \left(\bar{L} - Nf\right) - N^{\frac{1}{\theta-1}} f\right) (N_t - N)$$
$$+ N^{\frac{1}{\theta-1}} L (A_t - A) + \frac{1}{2} \begin{pmatrix} \frac{1}{\theta-1}\left(\frac{1}{\theta-1} - 1\right) N^{\frac{1}{\theta-1}-2}\left(\bar{L} - Nf\right) \\ -\frac{1}{\theta-1} N^{\frac{1}{\theta-1}-1} f - \frac{1}{\theta-1} N^{\frac{1}{\theta-1}-1} f \end{pmatrix} (N_t - N)^2$$
$$+ \frac{1}{\theta - 1} N^{\frac{1}{\theta-1}-1} \bar{L} (N_t - N)(A_t - A)$$

and writing with percentage deviations (recall $A = 1$ by normalization):

$$\frac{C_t - C}{C} = -\delta_t + \frac{1}{\theta - 1} \frac{N^{\frac{\theta}{\theta-1}} \left(\frac{\bar{L}}{N} - \theta f\right)}{C} \frac{N_t - N}{N} + \frac{N^{\frac{1}{\theta-1}} L}{C} \frac{A_t - A}{A}$$
$$+ \frac{1}{2} \frac{1}{\theta - 1} \frac{N^{\frac{\theta}{\theta-1}} \left(\left(\frac{1}{\theta-1} - 1\right)\left(\frac{\bar{L}}{N} - f\right) - 2f\right)}{C} \left(\frac{N_t - N}{N}\right)^2 + \frac{1}{\theta - 1} \frac{N^{\frac{1}{\theta-1}} \bar{L}}{C} \frac{N_t - N}{N} \frac{A_t - A}{A}$$

Use the steady-state, replacing $N = \frac{\bar{L}}{f\theta}$ and noticing that the 1st order term disappears:

$$\frac{C_t - C}{C} = -\delta_t + \frac{\theta}{\theta - 1} \frac{A_t - A}{A} - \frac{1}{2} \frac{\theta}{(\theta - 1)^2} \left(\frac{N_t - N}{N}\right)^2 + \frac{\theta}{(\theta - 1)^2} \frac{N_t - N}{N} \frac{A_t - A}{A}$$

Finally, using the second-order approximation of the cost of inflation $\delta_t \simeq \frac{1}{2}\kappa\pi_t^2$ and the expression of the flexible-price number of firms $n_t^* = a_t$ and ignoring terms independent of policy and of order higher than 2, the loss function is proportional to

$$-\frac{1}{2}\left(\kappa\pi_t^2 + \frac{\theta}{(\theta - 1)^2}(n_t - n_t^*)^2\right).$$

A.4 Nonlinear general model with both price and wage stickiness

Firms face nominal rigidity in the form of a quadratic cost of adjusting prices over time (Rotemberg, 1982). Specifically, the real cost (in units of the composite basket) of output-price inflation volatility around a steady-state level of inflation equal to 0 facing firm q is:

Each firm faces:

$$pac_t(q) = \frac{\kappa}{2}\left(\frac{p_t(q)}{p_{t-1}(q)} - 1\right)^2 \rho(q) y_t^D(q), \tag{34}$$

where $\rho_t(q) \equiv p_t(q)/P_t$ denotes the real (relative) price of firm q's output. This expression is interpreted as the amount of marketing materials that the firm must purchase when implementing a price change. We assume that this basket has the same composition as the consumption basket. The cost of adjusting prices is proportional to the real revenue from output sales, $(p_t(q)/P_t)\, y_t^D(q)$, where $y_t^D(q)$ is firm q's output demand.

Firms face demand for their output from consumers and firms themselves when they change prices. In each period, there is a mass S_t of firms producing and setting prices in the economy. [38]

[38] When a new firm sets the price for the first time, we appeal to symmetry across firms and interpret the $t-1$ price in the expression of the price adjustment cost for that firm as the notional price that the firm would have set at time $t-1$ if it had been producing in that period. See Bilbiie et al (2007) for a justification of this assumption—which is consistent with the original Rotemberg (1982) setup—and a relaxation. Note that symmetry of the equilibrium will imply $p_{t-1}(q) = p_{t-1} \forall q$.

The total demand for the output of firm q is thus

$$qy_t(q) = \left(\frac{p_t/q}{P_t}\right)^{-\theta} Y_t. \tag{35}$$

where $Y_t. = C_t + PAC_t = C_t + S_t pac_t\,(q)$

Then, the objective function of a firm choosing its price is equal to the firm q's real profit in period t (distributed to households as dividend) *net* of the price adjustment cost; this can be written as

$$d_t^R(q) = \rho_t\,(q)\,y_t^D\,(q) - w_t l_t\,(q) - \frac{\kappa}{2}\left(\frac{p_t(q)}{p_{t-1}(q)} - 1\right)^2 \rho_t\,(q)\,y_t^D\,(q)\,.$$

At time t, firm q chooses $l_t\,(q)$ and $p_t\,(q)$ to maximize $d_t^R\,(q)$ subject to $y_t\,(q) = y_t^D\,(q)$, taking w_t, $P\,_t$, C_t, PAC_t, and A_t as given. Letting $\lambda_t\,(q)$ denote the Lagrange multiplier on the constraint $y_t\,(q) = y_t^D\,(q)$, the first-order condition with respect to $l_t\,(q)$ yields:

$$\lambda_t\,(q) = \frac{w_t}{qA_t}.$$

The shadow value of an extra unit of output is simply the firm's marginal cost, common across all firms in the economy.

Maximizing the PDV of net profits (discounted wth the SDF, i.e. the marginal rate of intertemporal substitution of the household $\Lambda_{t,t+1}$) we obtain, imposing symmetry of the equilibrium $p_t(q) = p_t$ and denoting $\pi_t = \frac{p_t}{p_{t-1}} - 1$, after some substitutions the Phillips curve corresponding to the pricing decision is:

$$(1+\pi_t)\,\pi_t = \beta E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}} \frac{Y_{t+1}}{Y_t}\frac{S_t}{S_{t+1}}\,(1+\pi_{t+1})\,\pi_{t+1}\right] + \frac{\theta}{\kappa}\left[\frac{1}{\mu_t} - (1+s_w)\,\frac{\theta-1}{\theta}\left(1-\frac{\kappa}{2}\pi_t^2\right)\right]$$

where $1 + s_w$ is a labor subsidy that will be set by a fiscal authority to eliminate both distortions in the goods and labor market. Notice that by virtue of the assumption of heterogeneity in product quality, price stickiness does not interact with selection. Note that the total GDP of the economy, inclusive of the adjustment cost, is $Y_t = \left(1 - \frac{\kappa}{2}\pi_t^2\right)^{-1} C_t$ and we used this in rewriting the Phillips curve.

The "wage stickiness" part is standard—wage setting is done by an union bundling the differentiated labor inputs of households, setting the nominal wage subject to adjustment costs—and its details are unaffected by the introduction of firm entry. For the sake of space, we do not report all derivations of this block but refer to Erceg et al (2000) and Schmitt-Grohé and Uribe (2006). We first provide the nonlinear equations (used to derive the loglinearized equations in text).The disutility of labor can now be written as $\frac{1}{1+\varphi}\left(\frac{L_t}{1-\frac{\kappa_w}{2}\pi_{w,t}^2}\right)^{1+\varphi}$ where we already replaced labor supply using labor market clearing $L_t/\left(1-\frac{\kappa_w}{2}\pi_{w,t}^2\right)$ where L_t is total labor demand and the denominator is related to the labor cost of adjusting nominal wages paid by the union. Denote the wage inflation rate by:

$$1+\pi_{w,t}=\frac{w_t}{w_{t-1}}\left(1+\pi_t\right)$$

The optimality condition for each union setting wages for a differentiated labor type subject to a downward sloping labor demand with elasticity θ_w, Rotemberg adjustment costs paid in labor units κ_w and a labor subsidy s_w is:

$$\begin{aligned}\frac{\pi_{w,t}\left(\pi_{w,t}+1\right)}{1-\frac{\kappa_w}{2}\pi_{w,t}^2}&=\beta\mathrm{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}}\frac{L_{t+1}}{L_t}\frac{\pi_{w,t+1}\left(\pi_{w,t+1}+1\right)^2}{1-\frac{\kappa_w}{2}\pi_{w,t+1}^2}\right]\\&+\frac{\theta_w-1}{\kappa_w}\left[\frac{\theta_w}{\theta_w-1}\frac{1}{w_tC_t^{-\frac{1}{\sigma}}}\left(\frac{L_t}{1-\frac{\kappa_w}{2}\pi_{w,t}^2}\right)^{\varphi}-\left(1+s_w\right)\right].\end{aligned}$$

The full model is described by Table A1. In the simulations, we set $\theta_w=\theta$, $\kappa_w=\kappa$ and an optimal subsidy eliminating the steady-state inefficiency stemming from both market power sources $1+s_w=\frac{\theta}{\theta-1}\frac{\theta_w}{\theta_w-1}$.

Note that the price adjustment cost is second order and it does not appear in the log-linearized version around a zero-inflation steady state.

Finally, aggregate accounting, free entry and the profits definition imply that the "aggregate production function" takes the form:

$$Y_t=A_t\rho_t\left(L_t-N_tf_E-S_t\frac{f_O}{A_t}\right);$$

in other words, labor used in production equals total labor minus labor used to pay for the entry fost (for all entrants) and labor used to pay for the fixed overhead cost in efficiency units (for all producting firms)—i.e., the labor market clears by Walras' law.

A.5 Sticky-wage only model

Consider the model with sticky wages but flexible prices, i.e. same as our benchmark but with $\kappa=0$. We plot the impulse responses of this case for the endogenous-entry model in Figure 4. As discussed in text, the recession is now only one-period lived and turns into an expansion from next

Table 2: Benchmark Model, Summary

1. Pricing	$\rho_t = \mu_t \frac{w_t}{A_t}$
2. Price PC	$(1+\pi_t)\,\pi_t = \beta E_t \left[\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}} \frac{Y_{t+1}}{Y_t} \frac{S_t}{S_{t+1}} (1+\pi_{t+1})\,\pi_{t+1} \right] + \frac{\theta}{\kappa} \left[\frac{1}{\mu_t} - (1+s_w)\,\frac{\theta-1}{\theta} \left(1 - \frac{\kappa}{2}\pi_t^2\right) \right]$
3. Wage PC	$\frac{\pi_{w,t}(\pi_{w,t}+1)}{1-\frac{\kappa_w}{2}\pi_{w,t}^2} = \beta \mathrm{E}_t \left[\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}} \frac{L_{t+1}}{L_t} \frac{\pi_{w,t+1}(\pi_{w,t+1}+1)^2}{1-\frac{\kappa_w}{2}\pi_{w,t+1}^2} \right] + \frac{\theta_w - 1}{\kappa_w} \left[\frac{\theta_w}{\theta_w - 1} \frac{1}{w_t C_t^{-\frac{1}{\sigma}}} \left(\frac{L_t}{1-\frac{\kappa_w}{2}\pi_{w,t}^2} \right)^{\varphi} - (1+s_w) \right]$
4. Euler bonds	$C_t^{-\frac{1}{\sigma}} = \beta E_t \left(\frac{1+i_t}{1+\pi_{t+1}^C} C_{t+1}^{-\frac{1}{\sigma}} \right)$
5. Taylor rule	$1 + i_t = \beta^{-1} (1 + E_t \pi_{t+1})^{\phi} \exp(-\varepsilon_t)$
6. Wage inflation	$1 + \pi_{w,t} = \frac{w_t}{w_{t-1}} (1+\pi_t)$
7. Agg. acct.	$C_t + \frac{\kappa}{2}\pi_t^2 Y_t + N_t w_t f_E = w_t L_t + S_t \bar{d}_t$
8. Profits	$\bar{d}_t = \left(1 - \frac{1}{\mu_t}\right) \frac{Y_t}{S_t} - \frac{w_t}{A_t} f_O$
9. Free entry	$w_t f_E = (1 - G(q_{0t}))\, \bar{d}_t = \left(\frac{q_{\min}}{q_{0t}}\right)^k \bar{d}_t$
10. CPI inflation	$\frac{1+\pi_t}{1+\pi_t^C} = \frac{\rho_t}{\rho_{t-1}}$
11. Variety effect	$\rho_t = \left(\frac{k}{k-(\theta-1)}\right)^{\frac{1}{\theta-1}} q_{0t} S_t^{\frac{1}{\theta-1}}$
12. Survivors	$S_t = (1 - G(q_{0t}))\, N_t = \left(\frac{q_{\min}}{q_{0t}}\right)^k N_t$
13. Exit Cutoff	$q_{0t} = \left(\frac{f_O}{\mu_t - 1} \frac{1}{Y_t}\right)^{\frac{1}{\theta-1}} \rho_t^{\frac{\theta}{\theta-1}}$
14. Output	$C_t = \left(1 - \frac{\kappa}{2}\pi_t^2\right) Y_t$

Notes: The table displays all the nonlinear equilibrium conditions.

period onwards, as firms cut prices to restore profitability and households increase hours worked.

This can be described analytically using the loglinearized model under fixed wages $\psi_w = 0$ and flexible prices $\kappa = 0$, implying $\mu_t = 0$ and $\pi_t^w = 0$ respectively. Imposing these in Table A2 and replacing the wage in the wage inflation definition we obtain:

$$\pi_t^C + \frac{1}{\theta - 1} (n_t - n_{t-1}) = \hat{A}_{t-1} - \hat{A}_t,$$

which determines equilibrium inflation in individual producer prices:

$$\pi_t = \hat{A}_{t-1} - \hat{A}_t.$$

Replacing in the Euler equation (noting aggregate accounting implies $y_t = \frac{\theta}{\theta-1} n_t$) yields:

$$n_t - E_t n_{t+1} = \sigma \frac{\theta - 1}{\theta - \sigma} \left(\hat{A}_t - E_t \hat{A}_{t+1} \right) - \sigma \frac{\theta - 1}{\theta - \sigma} \phi \left(\hat{A}_{t-1} - \hat{A}_t \right) + \sigma \frac{\theta - 1}{\theta - \sigma} \varepsilon_t,$$

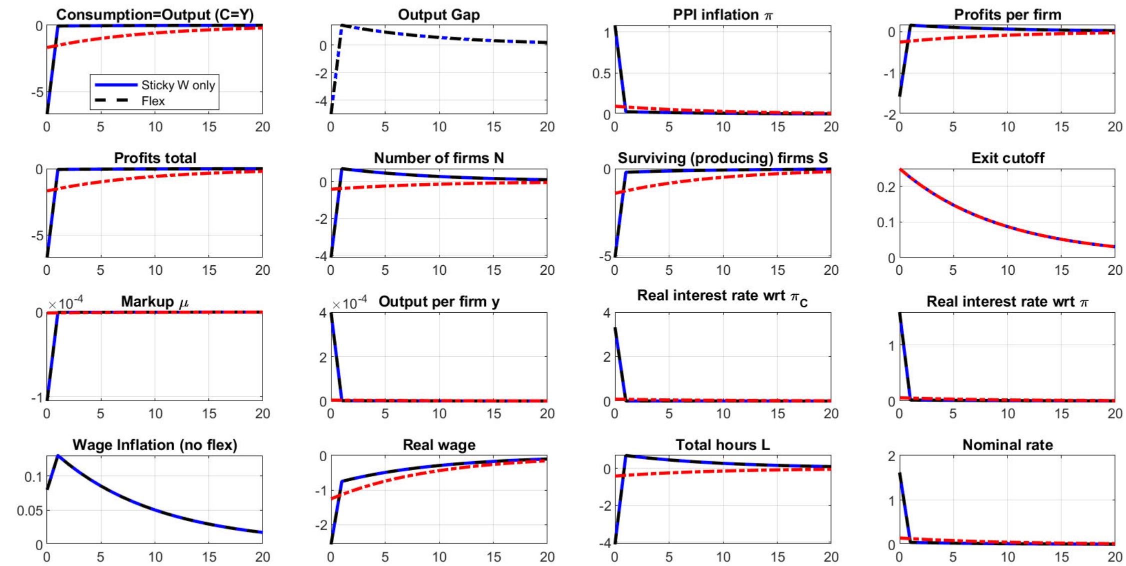


Figure 4: Effects of a 1% productivity fall: Flexible (red dash) vs Sticky (solid blue) *wages*. Flex prices.

which delivers the closed-form solution

$$n_t^{ES_w} = \sigma\frac{\theta-1}{\theta-\sigma}\hat{A}_t - \sigma\frac{\theta-1}{\theta-\sigma}\phi\hat{A}_{t-1} + \sigma\frac{\theta-1}{\theta-\sigma}\sum_{i=0}^{\infty}\varepsilon_{t+j}.$$

This illustrates clearly the reversal dynamics visible in Figure A2, whereby current TFP shocks are negatively correlated to future entry (and aggregate activity).

A.6 Baseline vs Analytical Calibration

In Figure 5 we repeat the experiment of Figure 1, comparing the baseline calibration $\beta = .99$ with the static-PCs calibration used in the closed-form analytical solution $\beta = 0$. As the figure makes clear, the responses are barely distinguishable for all real variables; differences naturally appear for inflation (and hence the real rate), since such a sharp change in β essentially amounts to a recalibration of the long-run Phillips curve's slope.

A.7 Demand shocks: with vs without entry and sticky wages

In Figure 6, we report the impulse responses to a 1% interest rate cut for the no-entry NK (panel a) and our entry (panel) models, respectivly—without sticky wages (only sticky prices) in dashed green, and with sticky both prices and wages in solid blue.

In our model with entry (panel b) but sticky prices only, we see how the demand shock induces

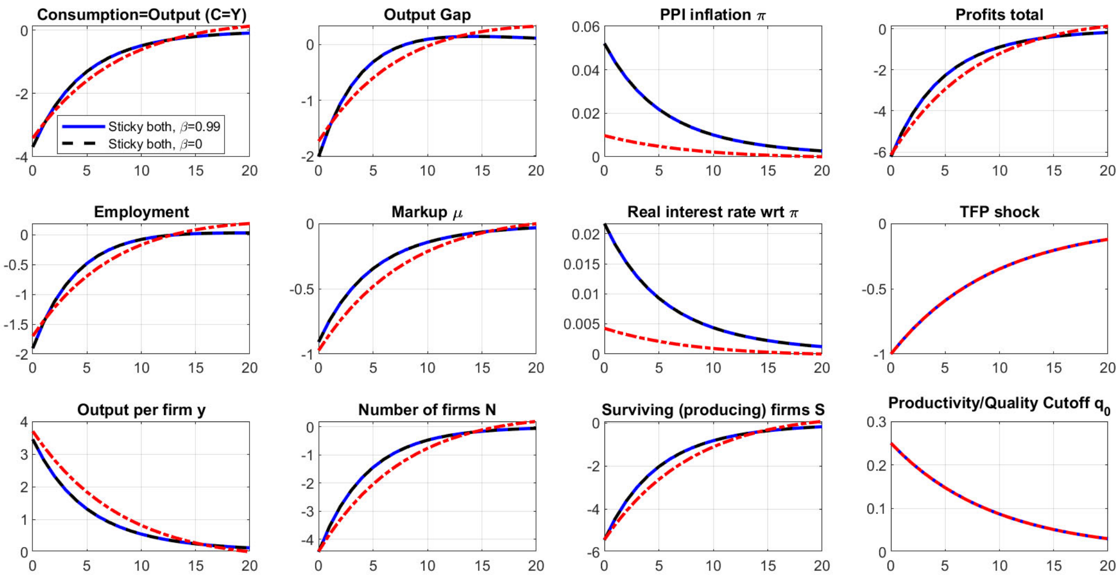


Figure 5: Effects of a 1% productivity fall: $\beta = 0$ (red dash) vs $\beta = 0.99$ (solid blue).

exit. As we previously discussed (and shown analytically), this counterfactual prediction is driven by the sharp drop in the markup, and hence profits (per firm).

[39] In the no-entry NK model (upper panel a), the countercyclicality of markups in response to demand shocks implies that profits are countercyclical too under flexible wages. This is a well-known issue of sticky-price (only) New Keynesian models, for which wage stickiness is an equally well-known solution making profits procyclical; we see in panel a that the markup and profits are countercyclical in the sticky-price-only case, but the markup gets dampened and profits become procyclical under wage stickiness.

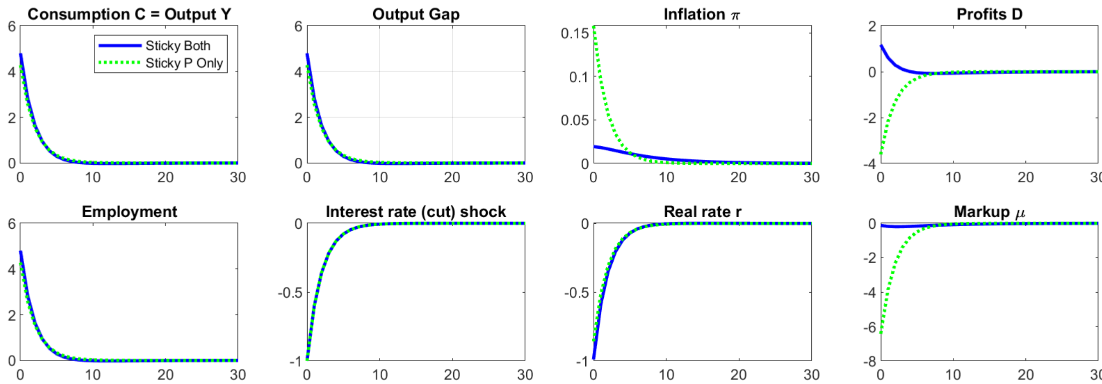
B No-Entry NK Model Summary Recap

Main points, under DRS: $y_t = a_t + \alpha l_t$, fixed real rate.

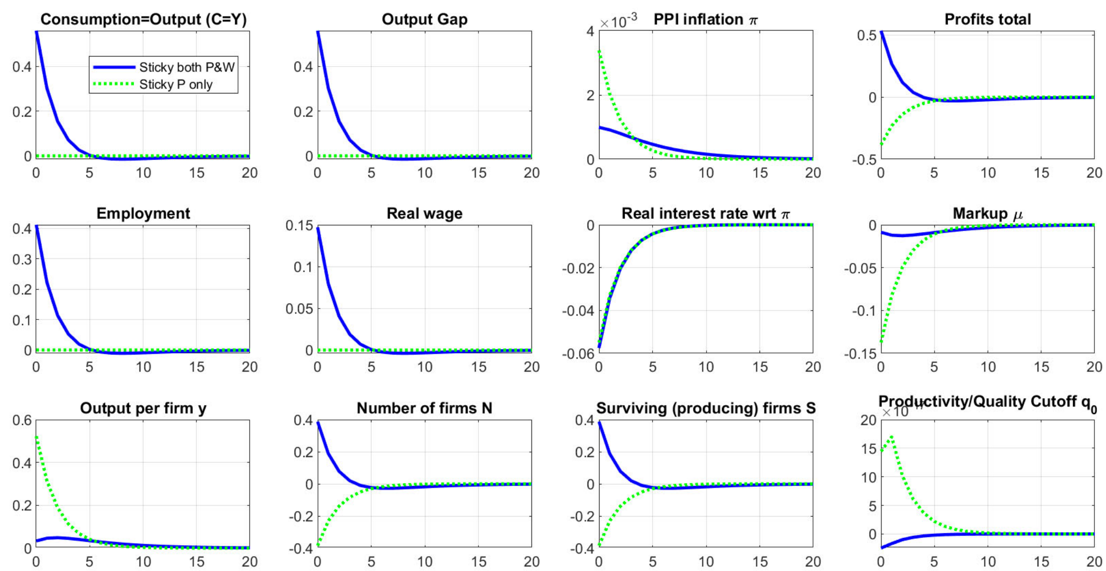
1. output gap goes up with negative TFP
2. it can be closed by tracking r-star, which entails a monetary contraction.
3. this tracking r-star policy is equivalent to completely stabilizing the "average" composite inflation measure.

Natural rate is $\mathcal{N}$

[39]The cylical properties of profits have nontrivial (and sometimes perverse) aggregate-demand implications throgh distributional mechanisms in models with heterogeneous agents, see e.g. Bilbiie (2008, 2020). Wage stickiness helps alleviate some of those issues, as discussed recently e.g. by Broer et al (2020).



(a) No Entry



(b) Endogenous Entry

Notes: Panel (a): No-Entry NK. Panel (b) Endogenous-Entry NK.

Figure 6: Effects of a 1% rate cut: Sticky P only (green dots) vs Sticky P & W (solid blue)

Table 3: Loglinearized NK Model, in Gaps

Common equations:	
1. Pricing	$.-\hat{\mu}_t = \tilde{w}_t + \frac{\alpha}{1-\alpha}\tilde{Y}_t$
2. Price PC	$\pi_t = \beta E_t \pi_{t+1} + \psi \left(\tilde{w}_t + \frac{\alpha}{1-\alpha}\tilde{Y}_t \right)$
3. Wage PC	$\pi_{w,t} = \beta E_t \pi_{w,t+1} + \psi_w \left(\left(\sigma^{-1} + \frac{\varphi}{1-\alpha} \right) \tilde{Y}_t - \tilde{w}_t \right)$
4. Euler bonds	$\tilde{Y}_t = E_t \tilde{Y}_{t+1} - \sigma \left(\hat{i}_t - E_t \pi_{t+1} - \hat{r}^*_t \right)$
5. Taylor rule	$\hat{i}_t = \phi E_t \pi_{t+1} - \varepsilon_t$
6. Wage inflation	$\tilde{w}_t = \tilde{w}_{t-1} + \pi_{w,t} - \pi_t - \left(w^*_t - w^*_{t-1} \right)$

Notes: The table displays all the loglinearized equilibrium conditions of the no-entry NK model, in "gap" notation, around a zero-inflation steady state with an optimal subsidy.

$$r_t^{*\mathcal{N}} = (1 - \rho_a)\, r_a^{*\mathcal{N}} \hat{A}_t$$

where

$$r_a^{*\mathcal{N}} = -\sigma^{-1} \frac{1 + \frac{\varphi+\alpha}{1-\alpha}}{\sigma^{-1} + \frac{\varphi+\alpha}{1-\alpha}}$$

so it co-moves negatively to TFP.

Write in gaps (ignore superscript $\mathcal{N}$ when obvious to lighten notation)

Take difference assume $\phi = 1$

$$\pi_{w,t} - \pi_t = \beta \left(E_t \pi_{w,t+1} - E_t \pi_{t+1} \right) + \psi_w \left(\left(\sigma^{-1} + \frac{\varphi}{1-\alpha} \right) \tilde{Y}_t - \hat{w}_t \right) - \psi \left(\hat{w}_t + \frac{\alpha}{1-\alpha} \tilde{Y}_t \right)$$

Demand side fully determines output gap in this version of the standard NK model:

$$\tilde{Y}_t^{\mathcal{N}} = E_t \tilde{Y}_{t+1}^{\mathcal{N}} + \sigma \left(\varepsilon_t + r_t^{*\mathcal{N}} \right) \rightarrow \tilde{Y}_t^{\mathcal{N}} = \sigma \left(\epsilon_t + r_a^{*\mathcal{N}} \hat{A}_t \right)$$

PC+ wage

$$\tilde{w}_t - \tilde{w}_{t-1} + \left(w^*_t - w^*_{t-1} \right) = \beta \left(E_t \tilde{w}_{t+1} - \tilde{w}_t + \left(w^*_{t+1} - w^*_t \right) \right) + \left(\psi_w \left(\sigma^{-1} + \frac{\varphi}{1-\alpha} \right) - \psi \frac{\alpha}{1-\alpha} \right) \tilde{Y}_t - \left(\psi_w + \psi \right) \tilde{w}_t$$

$$\beta E_t \tilde{w}_{t+1} - \left(1 + \beta + \psi_w + \psi \right) \tilde{w}_t + \tilde{w}_{t-1} = \left(\psi \frac{\alpha}{1-\alpha} - \psi_w \left(\sigma^{-1} + \frac{\varphi}{1-\alpha} \right) \right) \tilde{Y}_t$$
$$+ \left(w^*_t - w^*_{t-1} \right) - \beta \left(E_t w^*_{t+1} - w^*_t \right)$$

Use static PC

$$\left(1+\psi_w+\psi\right)\tilde{w}_t=\tilde{w}_{t-1}+\left(\psi_w\left(\sigma^{-1}+\frac{\varphi}{1-\alpha}\right)-\psi\frac{\alpha}{1-\alpha}\right)\tilde{Y}_t-\left(w_t^*-w_{t-1}^*\right)$$

This determines the wage gap, but output gap already given by demand side, given r-star.

TFP down, r-star goes up, output gal goes up. "Cure" this by contracting.

$$d\tilde{Y}_t^{\mathcal{N}}=0\rightarrow d\epsilon_t^{*\mathcal{N}}=r_a^{*\mathcal{N}}d\left(-\hat{A}_t\right)$$

, increase interest rates (negative ϵ) in response to a negative TFP shock because it engenders an output-gap expansion.

Finally, note that this "tracking r-star" policy is exactly equivalent to a policy stabilizing the average inflation measure defined in text, since in this model modified divine coincidence holds (closing the output gap stabilizes average inflation). It can be easily shown by aggregating the two PCs (and it is a textbook result known from Woodford 2003 and Gali 2015) that in this model:

$$\bar{\pi}_t=\beta E_t\bar{\pi}_{t+1}+\bar{\psi}\left(\sigma^{-1}+\frac{\varphi+\alpha}{1-\alpha}\right)\tilde{Y}_t.$$

As we saw, all these policy are very different in our endogenous-entry model.

C Proof that monetary easing increases utility in low-TFP state

Suppose we start from a realization of TFP under steady state, $\bar{A}<1$ which leads to a recession under fixed prices and wages.

Under the calibration used in the welfare figure, the solution for output (consumption) derived in Proposition **??** becomes, denoting $M_t=\exp(\epsilon_t)$ for ease of interpretation:

$$\begin{aligned}
Y_t&=\left(\theta\bar{A}-(\theta-1)\right)^{\frac{1}{\frac{\theta}{\sigma}-1}}\bar{A}^{-\frac{1}{k}\frac{\theta-1}{\frac{\theta}{\sigma}-1}}\left(\frac{\theta}{\theta-1}M_t\right)^{\frac{\theta}{\frac{\theta}{\sigma}-1}}\\
L_t&=M_t\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}=\left(\theta\bar{A}-(\theta-1)\right)^{\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{\theta-1}{k}\frac{\sigma-1}{\theta-\sigma}}\left(\frac{\theta}{\theta-1}M_t\right)^{\sigma\frac{\theta-1}{\theta-\sigma}}
\end{aligned}$$

where the second equation is the solution for hours worked.

The change in utility in response to a monetary expansion will have two terms, one from utility of consumption the second from disutility of labor. The first term of $dU\left(.\right)$is:

$$
\begin{aligned}
Y_t^{-\frac{1}{\sigma}}\frac{dY_t}{dM_t} &= Y_t^{-\frac{1}{\sigma}}\left(\theta\bar{A}-(\theta-1)\right)^{\frac{1}{\frac{\theta}{\sigma}-1}}\bar{A}^{-\frac{1}{k}\frac{\theta-1}{\frac{\theta}{\sigma}-1}}\left(\frac{\theta}{\theta-1}\right)^{\sigma\frac{\theta}{\theta-\sigma}}\sigma\frac{\theta}{\theta-\sigma}\left(M_t\right)^{\sigma\frac{\theta}{\theta-\sigma}-1} \\
&= Y_t^{1-\frac{1}{\sigma}}\sigma\frac{\theta}{\theta-\sigma}\frac{1}{M_t}
\end{aligned}
$$

The second term is

$$
\begin{aligned}
-L_t^{\varphi}\frac{dL_t}{dM_t} &= -L_t^{\varphi}\left(\theta\bar{A}-(\theta-1)\right)^{\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{\theta-1}{k}\frac{\sigma-1}{\theta-\sigma}}\left(\frac{\theta}{\theta-1}\right)^{\sigma\frac{\theta-1}{\theta-\sigma}}\sigma\frac{\theta-1}{\theta-\sigma}\left(M_t\right)^{\sigma\frac{\theta-1}{\theta-\sigma}-1} \\
&= -L_t^{1+\varphi}\sigma\frac{\theta-1}{\theta-\sigma}\frac{1}{M_t}
\end{aligned}
$$

This needs to be evaluated at steady-state M:

$$
M=\left(\frac{\theta}{\theta-1}\right)^{\frac{\frac{\theta}{\sigma}-1}{\theta-1}\frac{1}{1+\varphi-\left(1-\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}}-1}
$$

Collecting terms

$$
\begin{aligned}
\frac{dU_t}{dM_t} &= Y_t^{1-\frac{1}{\sigma}}\sigma\frac{\theta}{\theta-\sigma}\frac{1}{M_t}-L_t^{1+\varphi}\sigma\frac{\theta-1}{\theta-\sigma}\frac{1}{M_t} \\
&= \sigma\frac{\theta-1}{\theta-\sigma}\frac{1}{M_t}\left(\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}-L_t^{1+\varphi}\right)
\end{aligned}
$$

Evaluate sign of paranthetic term, show it is larger than 0 (derivative of utility is then positive under the parameter condition (2)):

$$
\begin{aligned}
\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}-L_t^{1+\varphi} &= \frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}-\left(M_t\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}\right)^{1+\varphi} \\
&= \frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}\left(1-\left(M_t\right)^{1+\varphi}\left(\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}\right)^{\varphi}\right)
\end{aligned}
$$

Show

$$
\left(M_t\right)^{1+\varphi}\left(\frac{\theta}{\theta-1}Y_t^{1-\frac{1}{\sigma}}\right)^{\varphi}<1
$$

Use

$$
Y_t^{1-\frac{1}{\sigma}}=\left(\theta\bar{A}-(\theta-1)\right)^{\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}}\left(\frac{\theta}{\theta-1}M_t\right)^{\frac{\theta(\sigma-1)}{\theta-\sigma}}
$$

$$
\left(M_t\right)^{1+\varphi}\left(\frac{\theta}{\theta-1}\left(\theta\bar{A}-(\theta-1)\right)^{\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}}\left(\frac{\theta}{\theta-1}M_t\right)^{\frac{\theta(\sigma-1)}{\theta-\sigma}}\right)^{\varphi}<1
$$

$$\left(M_t\right)^{1+\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}}\left(\frac{\theta}{\theta-1}\right)^{\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}}\left(\theta\bar{A}-(\theta-1)\right)^{\varphi\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}\varphi}<1$$

$$M^{SS}=\left(\frac{\theta}{\theta-1}\right)^{\frac{\frac{\theta}{\sigma}-1}{\theta-1}\frac{1}{1+\varphi-\left(1-\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}}-1}$$

$$\left(\left(\frac{\theta}{\theta-1}\right)^{\frac{\frac{\theta}{\sigma}-1}{\theta-1}\frac{1+\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}}{1+\varphi-\left(1-\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}}-1-\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}+\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}}\right)\left(\theta\bar{A}-(\theta-1)\right)^{\varphi\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}\varphi}<1$$

$$\left(\left(\frac{\theta}{\theta-1}\right)^{\frac{\frac{\theta}{\sigma}-1}{\theta-1}\frac{1+\varphi\frac{(\theta-1)\sigma}{\theta-\sigma}}{1+\varphi-\left(1-\frac{1}{\sigma}\right)\frac{\theta}{\theta-1}}-1}\right)\left(\theta\bar{A}-(\theta-1)\right)^{\varphi\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}\varphi}<1$$

Simplifying, the term in the first bracket is in fact 1 (the exponent is 0), so it becomes:

$$\left(\theta\bar{A}-(\theta-1)\right)^{\varphi\frac{\sigma-1}{\theta-\sigma}}\bar{A}^{-\frac{1}{k}\frac{(\theta-1)(\sigma-1)}{\theta-\sigma}\varphi}<1$$

$$\left(\left(\theta\bar{A}-(\theta-1)\right)\bar{A}^{-\frac{\theta-1}{k}}\right)^{\varphi\frac{\sigma-1}{\theta-\sigma}}<1$$

As long as the power is positive (and, obviously, so is the term in brackets), it is therefore sufficient to show that $\left(\theta\bar{A}-(\theta-1)\right)\bar{A}^{-\frac{\theta-1}{k}}<1$. Since $\theta>1$, $k>\theta-1$ and, crucially, $\bar{A}<1$ it can be shown by standard calculus that this inequality always holds. (Define the function $J\left(\bar{A}\right)=\bar{A}^{\frac{\theta-1}{k}}-\left(\theta\bar{A}-(\theta-1)\right)$ and show $J\left(.\right)>0$ over the whole domain. Notice that $J\left(1\right)=0$ and $J'\left(1\right)=\frac{\theta-1}{k}-\theta<0$, so J is positive to the left of zero; and finally, J is concave: $J''\left(1\right)=\frac{\theta-1}{k}\left(\frac{\theta-1}{k}-1\right)<0$.)

The exact same derivation—with opposite sign—shows that a monetary contraction improves welfare with $\bar{A}>1$, i.e. in a TFP-driven expansion.

In the No-Entry NK model, the opposite holds. We can prove this using the expressions from Table **??** (for the log-utility case, but easily generalizable), reproducing them here $Y_t^{\mathcal{N}}=\frac{M_t}{P^\circ}$ and $L_t^{\mathcal{N}}=\frac{1}{A_t}\frac{M_t}{P^\circ}$ and evaluating utility starting from steady-state money balances $\frac{M}{P^\circ}=1$; The change in utility starting at $\bar{A}$ is

$$\frac{dU_t^{\mathcal{N}}}{dM_t}=Y_t^{-\frac{1}{\sigma}}\frac{1}{P^\circ}-L_t^{\varphi}\frac{1}{P^\circ}\frac{1}{A_t}=\left(\frac{M_t}{P^\circ}\right)^{-\frac{1}{\sigma}}\frac{1}{P^\circ}-\left(\frac{M_t}{P^\circ}\frac{1}{\bar{A}}\right)^{\varphi}\frac{1}{P^\circ}\frac{1}{\bar{A}}=1-\left(\frac{1}{\bar{A}}\right)^{1+\varphi}<0\text{ when }\bar{A}<1$$

Conversely, utility increases with money when starting in a high-TFP (recession-inducing, in the no-entry model) state.

Online Appendix

D Derivations for External Returns to Variety: General CES Model

This Appendix outlines the derivations for the benchmark nonlinear model pertaining to stabilization (monetary) policy and to the role of external returns to variety.

Table A1. Sticky-price Model Summary

1 Euler equation	$Y_t^{-\frac{1}{\sigma}} = \beta E_t \left(\frac{1+i_t}{1+\pi_{t+1}^C} Y_{t+1}^{-\frac{1}{\sigma}} \right)$
2 Labor supply	$L_t^{\varphi} = w_t Y_t^{-\frac{1}{\sigma}}$
3 CPI inflation	$\frac{1+\pi_t}{1+\pi_t^C} = \left(\frac{N_t}{N_{t-1}} \right)^{\lambda + \frac{1}{\theta-1}}$
4 Aggregate accounting	$Y_t = w_t L_t$
5. Markup (Phillips curve)	$\mu_t \left(1 - \frac{\kappa}{2}\pi_t^2\right) = \frac{\theta}{(\theta-1)+\kappa\left\{(1+\pi_t)\pi_t - \beta E_t\left[\left(\frac{Y_{t+1}}{Y_t}\right)^{1-\frac{1}{\sigma}} \frac{N_t}{N_{t+1}} \frac{(1+\pi_{t+1})\pi_{t+1}}{1-\frac{\kappa}{2}\pi_{t+1}^2}\right]\right\}}$
6 Policy	$1 + i_t = \beta^{-1} (1+\pi_t)^{\phi} \exp(-\varepsilon_t)$
7 Pricing	$N_t^{\lambda + \frac{1}{\theta-1}} = \mu_t \frac{w_t}{A_t}$
8 Free Entry	$A_t L_t \left[1 - \frac{1}{\mu_t \left(1 - \frac{\kappa}{2}\pi_t^2\right)} \right] = f N_t$

In the sticky-wage case the disutility of labor can be written as $\frac{1}{1+\varphi}\left(\frac{L_t}{1-\frac{\kappa_w}{2}\pi_{w,t}^2}\right)^{1+\varphi}$ where we already replaced labor supply using labor market clearing $L_t/\left(1-\frac{\kappa_w}{2}\pi_{w,t}^2\right)$ where L_t is total labor demand and the denominator is related to the labor cost of adjusting nominal wages paid by the union. Denote the wage inflation rate by:

$$1 + \pi_{w,t} = \frac{w_t}{w_{t-1}} (1 + \pi_t)$$

The optimality condition for each union setting wages for a differentiated labor type subject to a downward sloping labor demand with elasticity θ_w, Rotemberg adjustment costs paid in labor units κ_w and a labor subsidy s_w is:

$$\frac{\pi_{w,t}(\pi_{w,t}+1)}{1-\frac{\kappa_w}{2}\pi_{w,t}^2} = \beta \mathrm{E}_t \left[\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\sigma}} \frac{L_{t+1}}{L_t} \frac{\pi_{w,t+1}(\pi_{w,t+1}+1)^2}{1-\frac{\kappa_w}{2}\pi_{w,t+1}^2} \right] + \frac{\theta_w - 1}{\kappa_w} \left[\frac{\theta_w}{\theta_w - 1} \frac{1}{w_t C_t^{-\frac{1}{\sigma}}} \left(\frac{L_t}{1-\frac{\kappa_w}{2}\pi_{w,t}^2} \right)^{\varphi} - (1+s_w) \right].$$

The full model is described by Table A1, where eq. 2 (labor supply) is replaced by the wage Phillips curve, and adding the wage inflation definition. In the simulations, we set $\theta_w = \theta$, $\kappa_w = \kappa$

and an optimal subsidy eliminating the steady-state labor market inefficiency $s_w = (\theta_w - 1)^{-1}$ (the goods market is efficient by free entry).

D.1 External Returns to Variety: General CES

Solving the benchmark model with the general CES aggregator with external returns introduced in text under flexible and fixed prices, respectively, delivers:

$$Y_t^* = \left(\frac{1}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} (A_t\bar{L})^{\lambda+\frac{\theta}{\theta-1}} \frac{\theta-1}{\theta}$$

$$Y_t = \frac{M_t}{\bar{p}}\left(\frac{A_t\bar{L}}{f} - \frac{M_t}{f\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}}.$$

Consider a steady state equilibrium with $\frac{M}{f\bar{p}} = \frac{\theta-1}{\theta f}A\bar{L} \to \frac{A\bar{L}}{f} - \frac{M}{f\bar{p}} = \frac{A\bar{L}}{\theta f}$

$$Y^* = \left(\frac{1}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} (A\bar{L})^{\lambda+\frac{\theta}{\theta-1}} \frac{\theta-1}{\theta}$$

$$Y = \frac{M}{\bar{p}}\left(\frac{A\bar{L}}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} = \frac{\theta-1}{\theta}A\bar{L}\left(\frac{A\bar{L}}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} = Y^*$$

Taking a Taylor approximation around $Y^* = \left(\frac{1}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} (A\bar{L})^{\lambda+\frac{\theta}{\theta-1}} \frac{\theta-1}{\theta}$

$$Y_t^* - Y^* = \left(\lambda+\frac{\theta}{\theta-1}\right)\frac{\theta-1}{\theta}\left(\frac{1}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} (A\bar{L})^{\lambda+\frac{\theta}{\theta-1}}\left(\frac{A_t-A}{A}\right)$$

$$+\frac{1}{2}\left(\lambda+\frac{1}{\theta-1}\right)\left(\lambda+\frac{\theta}{\theta-1}\right)\frac{\theta-1}{\theta}\left(\frac{1}{\theta f}\right)^{\lambda+\frac{1}{\theta-1}} (A\bar{L})^{\lambda+\frac{\theta}{\theta-1}}\left(\frac{A_t-A}{A}\right)^2 \to$$

$$\frac{Y_t^*-Y^*}{Y^*} = \left(\lambda+\frac{\theta}{\theta-1}\right)\left(\frac{A_t-A}{A}\right) + \frac{1}{2}\left(\lambda+\frac{1}{\theta-1}\right)\left(\lambda+\frac{\theta}{\theta-1}\right)\left(\frac{A_t-A}{A}\right)^2$$

$$Y_t - Y^* = \left(\lambda+\frac{1}{\theta-1}\right)\frac{A\bar{L}}{f}\frac{M}{\bar{p}}\left(\frac{A\bar{L}}{f}-\frac{M}{f\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}-1}\left(\frac{A_t-A}{A}\right)$$

$$+\frac{1}{2}\left(\lambda+\frac{1}{\theta-1}\right)\left(\lambda+\frac{1}{\theta-1}-1\right)\left(\frac{A\bar{L}}{f}\right)^2\frac{M}{\bar{p}}\left(\frac{A\bar{L}}{f}-\frac{M}{f\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}-2}\left(\frac{A_t-A}{A}\right)^2$$

Recall $Y = \frac{M}{\bar{p}}\left(\frac{A\bar{L}}{f} - \frac{M_t}{f\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}}$ and $\frac{M}{f\bar{p}} = \frac{\theta-1}{\theta f}A\bar{L} \to \frac{A\bar{L}}{f} - \frac{M}{f\bar{p}} = \frac{A\bar{L}}{\theta f}$

$$\frac{Y_t-Y^*}{Y^*} = \left(\lambda+\frac{1}{\theta-1}\right)\theta\left(\frac{A_t-A}{A}\right) + \frac{1}{2}\left(\lambda+\frac{1}{\theta-1}\right)\left(\lambda+\frac{1}{\theta-1}-1\right)\theta^2\left(\frac{A_t-A}{A}\right)^2$$

The sticky response is larger to first-order iff $\lambda > 0$, as discussed in text. Here, we focus on the second-order difference. The sticky response is larger second-order iff $\lambda > \frac{\theta}{\theta+1}$. (But now even with *negative* externality there can be over-reaction to negative shocks driven by higher-order effects. If the shock is negative enough, the higher-order term eventually kicks in.)

In the figure, we plot the case $\lambda = 0.2$ for the two respective cases: blue solid for sticky prices, red dash for flexible prices. We use again the normalization with f that equalizes Y^* to $Y^{*\mathcal{N}}$, $\bar{L}^{\frac{\theta-1}{\theta}} = (\theta f)^{\lambda+\frac{1}{\theta-1}}$. Then, we choose money supply M to equalize the SP equilibrium Y with FE to this same $Y^* = Y^{*\mathcal{N}}$. This requires, using f:

$$Y = \bar{L} = \frac{M}{\bar{p}}\left(\bar{L} - \frac{M}{\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}} \frac{\theta^{\lambda+\frac{1}{\theta-1}}}{\bar{L}^{\lambda+\frac{1}{\theta-1}}(\theta-1)} \to 1 = \frac{M}{\bar{p}\bar{L}}\left(1 - \frac{M}{\bar{L}\bar{p}}\right)^{\lambda+\frac{1}{\theta-1}} \frac{\theta^{\lambda+\frac{1}{\theta-1}}}{\frac{\theta-1}{\theta}},$$

again delivering $\frac{M}{\bar{p}} = \bar{L}\left(1 - \frac{1}{\theta}\right)$. Replacing these, we thus plot

$$Y_t^* = A_t^{\lambda+\frac{\theta}{\theta-1}}\bar{L} \text{ and } Y_t = \bar{L}\left(\theta A_t - (\theta-1)\right)^{\lambda+\frac{1}{\theta-1}}.$$

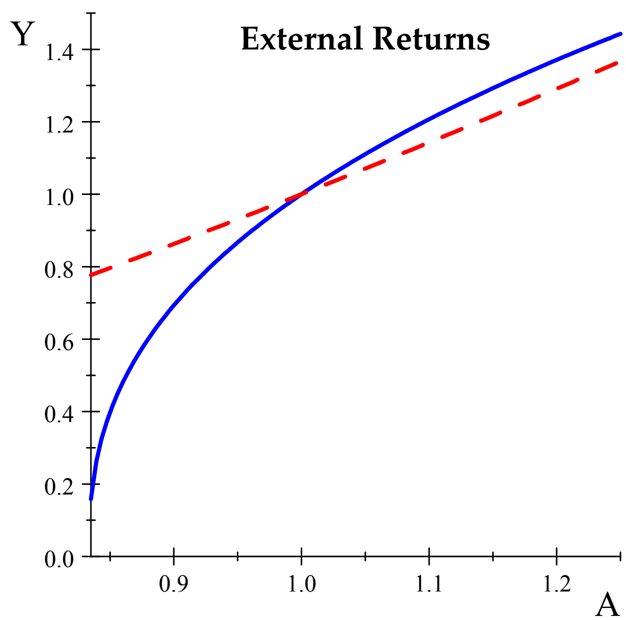


Figure 7: Y^* (flex. prices) red dash, Y (sticky prices) solid blue. External-returns $\lambda = 0.2$

Consumption (output) to second order is:

$$c_t^* = \left(\lambda + \frac{\theta}{\theta-1}\right)a_t + \frac{1}{2}\left(\lambda + \frac{1}{\theta-1}\right)\left(\lambda + \frac{\theta}{\theta-1}\right)a_t^2$$
$$c_t = \left(\lambda + \frac{1}{\theta-1}\right)\theta a_t + \frac{1}{2}\left(\lambda + \frac{1}{\theta-1}\right)\left(\lambda + \frac{1}{\theta-1} - 1\right)\theta^2 a_t^2$$

The effect of supply shocks is thus:

$$\frac{dc_t^*}{da_t} = \left(\lambda + \frac{\theta}{\theta - 1}\right) + \left(\lambda + \frac{1}{\theta - 1}\right)\left(\lambda + \frac{\theta}{\theta - 1}\right) da_t$$

$$\frac{dc_t}{da_t} = \left(\lambda + \frac{1}{\theta - 1}\right)\theta + \left(\lambda + \frac{1}{\theta - 1}\right)\left(\lambda + \frac{1}{\theta - 1} - 1\right)\theta^2 da_t$$

$$\rightarrow \frac{d\left(c_t - c_t^*\right)}{da_t} = \lambda\left(\theta - 1\right) + \left(\lambda + \frac{1}{\theta - 1}\right)\left[\lambda\left(\theta^2 - 1\right) + \theta - \theta^2\right] da_t$$

The effect on the output gap is larger than zero (falls more to negative shocks) if:

$$\lambda\left(\theta - 1\right) + \left(\lambda + \frac{1}{\theta - 1}\right)\left[\lambda\left(\theta^2 - 1\right) + \theta - \theta^2\right] da_t > 0$$

Even with negative externalities $\lambda < 0$ this still holds for a negative enough shock, i.e.:

$$\left(\lambda + \frac{1}{\theta - 1}\right)\left[\lambda\left(\theta^2 - 1\right) + \theta - \theta^2\right] da_t > -\lambda\left(\theta - 1\right) > 0,$$

we need $\lambda < \frac{\theta}{\theta+1}$, which is always satisfied when $\lambda < 0$, so the condition is:

$$d\left(-a_t\right) > \frac{\lambda}{\left(\lambda + \frac{1}{\theta-1}\right)\left[\lambda\left(\theta + 1\right) - \theta\right]}.$$

For a calibration with the overall benefit of variety $(\lambda + \frac{1}{\theta-1})$ equal to half the markup, $\lambda\left(\theta - 1\right) = -.5$ (a property of the translog preferences used in Bilbiie et at 2012 to match the cyclicality of markups and profits), the threshold is 0.215.

Insights from approximating $Y\left(N\right)$

Taking a second-order approximation of the aggregate production function:

$$Y_t = N_t^{\lambda + \frac{1}{\theta-1}}\left(A_t\bar{L} - N_t f\right)$$

around the steady-state of the flex-price equilibrium (same as for SP equilibrium) with $N^* = \frac{A\bar{L}}{f\theta}, Y^* = (N^*)^{\lambda+\frac{1}{\theta-1}}\left(A\bar{L} - N^* f\right) = (N^*)^{\lambda+\frac{1}{\theta-1}}\frac{\theta-1}{\theta}A\bar{L}$

$$Y_t \simeq Y + \left(\left(\lambda + \frac{1}{\theta-1}\right)N^{\lambda+\frac{1}{\theta-1}-1}\left(A\bar{L} - Nf\right) - N^{\lambda+\frac{1}{\theta-1}}f\right)(N_t - N)$$
$$+\frac{1}{2}\left(\begin{array}{c}\left(\lambda+\frac{1}{\theta-1}\right)\left(\lambda+\frac{1}{\theta-1}-1\right)N^{\lambda+\frac{1}{\theta-1}-2}\left(A\bar{L}-Nf\right)\\ -\left(\lambda+\frac{1}{\theta-1}\right)N^{\lambda+\frac{1}{\theta-1}-1}f-\left(\lambda+\frac{1}{\theta-1}\right)N^{\lambda+\frac{1}{\theta-1}-1}f\end{array}\right)(N_t-N)^2$$

and writing with percentage deviations:

$$\tilde{y}_t \simeq \lambda n_t + \frac{1}{2}\left(\lambda + \frac{1}{\theta-1}\right)\left(\lambda - \frac{\theta}{\theta-1}\right)n_t^2.$$

D.2 Loglinearized general-CES NK model

This Appendix presents the loglinearized NK model with *arbitrary* benefit of input variety and first-order welfare effects, directly comparable with the plain-vanilla textbook version of the no-entry NK model. In Table OA1, we outline the key equilibrium responses of the loglinearized model, around a steady state with no supply shock mirroring the same structure as for Table 1, but for the loglinearized model. Letting a small letter denote the log-deviation from the respective steady-state, the loglinearized Euler equation (11) and Taylor rule are, respectively:

$$c_t = E_t c_{t+1} - \left(i_t - E_t\pi^C_{t+1}\right);\text{ and} \tag{36}$$

$$i_t = \phi\pi_t. \tag{37}$$

Under endogenous entry and flexible prices, the solution is readily obtained by noticing that $\mu_t^* = \frac{\theta}{\theta-1}$. By virtue of logarithmic utility in consumption, hours worked stay constant (income and substitution effects on labor cancel out). The real wage responds to labor productivity with elasticity $\lambda + \frac{\theta}{\theta-1}$; the effect is amplified relative to the no-entry model by the standard variety effect that acts like a form of increasing returns, making output and consumption also move with the shock in the same manner. The number of firms changes proportionally to the shock: a decrease in productivity triggers exit because it induces losses. The lower left quadrant of Table OA1 outlines the full solution of the endogenous entry, flexible-price model. Other than substantiating the above, notice that the natural interest rate responds with the same sign as under no entry but with a larger elasticity, driven by the increasing returns. Since the natural rate increases with bad shocks, there is inflation in producer prices. And since there is exit, there is even higher inflation in consumer prices through the benefit of input variety. These inflation dynamics are nevertheless *still irrelevant* for the real allocation since prices are flexible.

Matters are different with sticky prices. Hours worked are again fixed in equilibrium, because income and substitution effects of the real wage cancel out (log utility in consumption), and in addition there are no extra income effects due to profits, which are zero by virtue of free entry. This can be seen by combining equations (9) and $C_t = w_t L_t$, and recalling the discussion after the latter, which implies that in loglinearized terms we have $w_t = c_t$.

Combining the loglinearized Euler equation (36) with the loglinearized (10) relating CPI, PPI inflation and variety growth:

$$\pi_t = \pi_t^C + \left(\lambda + \frac{1}{\theta - 1}\right)(n_t - n_{t-1}), \tag{38}$$

and imposing fixed producer prices $\pi_t = 0$, we obtain:

$$c_t = E_t c_{t+1} - \left(\lambda + \frac{1}{\theta - 1}\right)(E_t n_{t+1} - n_t). \tag{39}$$

Loglinearization of the pricing rule (4) combined with the relative price (3) delivers:

$$w_t - a_t = \left(\lambda + \frac{1}{\theta - 1}\right) n_t - \mu_t, \tag{40}$$

while the free-entry condition (8) is:

$$n_t = a_t + l_t + (\theta - 1)\mu_t. \tag{41}$$

Combining the last two while imposing that hours are constant in this equilibrium $l_t = 0$ and replacing in $w_t = c_t$, we obtain:

$$c_t = \frac{\theta}{\theta - 1} a_t + \lambda n_t$$

Together with the Euler equation under fixed prices (39), this delivers

$$n_t = \theta a_t; \; c_t = \left(\lambda + \frac{1}{\theta - 1}\right)\theta a_t,$$

and the rest of the solution reported in the lower right quadrant of Table OA1. Direct comparison with the solution under flexible prices delivers our condition for (first-order) negative output gap following a negative supply shock, $\lambda > 0$.

To help intuition, consider again the *labor market equilibrium*. With free entry and flexible prices, there is a larger recession than with no entry ($*\mathcal{N}$) because of the variety effect which generates aggregate returns to scale: aggregate LD is upward sloping (with slope $\lambda + \frac{1}{\theta-1}$) and shifts by $\lambda + \frac{\theta}{\theta-1}$ with supply disturbances. Individual labor demand is as before, but now an increase in marginal cost and fall in markup triggers *exit* (product destruction); since prices can be freely set, the amount of product destruction is dictated by the benefit of variety. This is represented with blue dashes in Figure 4.

Consider next sticky (fixed) prices ES. Since prices cannot increase now, the *markup goes down*. The crucial questions is: does LD shift up, or down? This depends on the benefit of input variety $\lambda + \frac{1}{\theta-1}$ versus the net markup $\frac{1}{\theta-1}$. When external returns are positive $\lambda > 0$, the benefit of input variety is higher and LD shifts further down: instead of a fall in profits, as under no entry), there is now exit. As a result, LS shifts further right due to the further negative income effect and, as we will see, consistent with intertemporal substitution. In other words, there is a *negative output gap*: consumption and income fall more than under flexible prices.

A complementary intuition starts from recalling that since prices cannot increase, the *markup goes down*. When the benefit of variety is higher than the markup, labor demand shifts further down: instead of a fall in profits (as under no entry), there is now exit. The loglinear approximation of aggregate labor demand is:

$$w_t = \left(\lambda + \frac{\theta}{\theta - 1}\right) a_t + \left(\lambda + \frac{1}{\theta - 1}\right) l_t + \lambda (\theta - 1) \mu_t;$$

when the markup falls and real marginal cost increases there is a shift downwards in labor demand when $\lambda > 0$: demand forces dominate, labor demand plunges, and this demand shortage is met by dropping products. As a result, labor supply shifts further right due to the further negative income effect and, as we discuss in the dynamic model, consistent with intertemporal substitution.

Neutrality without external effects

The first-order irrelevance (of price stickiness) under CES, without external effects $\lambda = 0$, applies for arbitrary price stickiness and can be seen most clearly by inspecting the loglinearized markup rule (the combination of (3) and (4)) and the free entry condition (8), respectively:[40]

$$w_t - a_t = \frac{1}{\theta - 1} n_t - \mu_t;$$
$$n_t = a_t + l_t + (\theta - 1)\, \mu_t.$$

Combining the two delivers a loglinearized version of aggregate labor demand:

$$w_t = \frac{\theta}{\theta - 1} a_t + \frac{1}{\theta - 1} l_t. \tag{42}$$

This illustrates that, as stated in text in the discussion of equation (31), to a first-order approximation, any endogenous changes in markups and in the extensive margin perfectly offset each other when it comes to the aggregate labor-demand effects of productivity shocks (they drop out from the aggregate labor demand equation, 42). (In contrast, in the fixed-entry model, productivity changes engender endogenous changes in markups that shift the aggregate labor demand.)

Aggregate Demand and Variety: Intertemporal Interpretation

A key element of the model is the aggregate Euler equation governing aggregate demand (36), which written in gaps from the flexible-price equilibrium is:

$$c_t - c_t^* = E_t c_{t+1} - E_t c_{t+1}^* - \left(i_t - E_t\left(\pi_{t+1}^C\right) - r_t^*\right)$$

where $r_t^* = \left(\lambda + \frac{\theta}{\theta-1}\right)(E_t a_{t+1} - a_t)$ is the natural interest rate. In this Euler equation, the relevant real rate is defined relative to CPI inflation. Spelling out CPI inflation using (38) we have:

$$c_t - c_t^* = E_t c_{t+1} - E_t c_{t+1}^* - \left[i_t - E_t \pi_{t+1} + \left(\lambda + \frac{1}{\theta - 1}\right)(E_t n_{t+1} - n_t) - r_t^*\right], \tag{43}$$

which generalizes the aggregate-Euler IS curve with entry derived in Bilbiie, Ghironi, and Melitz (2007, equation 12).

With entry, even when producer prices are fixed (or the real rate defined with respect to PPI inflation $i_t - E_t \pi_{t+1}$ is fixed), the output gap is no longer proportional to the natural interest rate,

[40] These are equations 7 and 8 in Tables A1 and A2, which outline the full set of equilibrium conditions, nonlinear and loglinearized, for the most general version of the model which nests this as a special case.

as in a no-entry model. Indeed, the output gap then falls with bad supply shocks $\frac{\partial(c_t - c_t^*)}{\partial(-a_t)} < 0$ if:

$$\left(\lambda + \frac{1}{\theta - 1}\right) \frac{\partial \left(E_t n_{t+1} - n_t\right)}{\partial \left(-a_t\right)} > \frac{\partial r_t^*}{\partial \left(-a_t\right)},$$

that is if the increase in "expected inflation" that is purely due to the variety effect exceeds the increase in the natural rate. Replacing the responses of n_t and r_t^* we recover $\lambda > 0$.

Thus, with $i_t - E_t \pi_{t+1}$ fixed, the real rate that is relevant for aggregate demand—i.e. real relative to CPI inflation—goes up since there is exit today, thus triggering intertemporal substitution towards the future. The labor supply then shifts right because of intertemporal substitution. This is a general mechanism that translates to our setup where producer prices are arbitrarily sticky, not fixed, outlined next. The AD representation (43) also suggests a possible way out of a supply-driven, exit-amplified crisis: subsidize entry or sales temporarily so as to break the exit loop and generate future expected CPI inflation, and a boost in aggregate demand today by intertemporal substitution. This policy works even when interest rates are constrained against the lower bound.

The 3-Equation NK model with Free Entry

Like the textbook NK model (Woodford 2003, Gali 2015) our model can be summarized by an Aggregate Demand (IS curve) and Aggregate Supply (Phillips curve), with arbitrary degree of price stickiness. The former is given by (43), where we replace the number of firms using aggregate accounting $c_t = \frac{\theta}{\theta-1} a_t + \lambda n_t$ to obtain, after substitutions and using the flex-price equilibrium $c_t^* = \left(\lambda + \frac{\theta}{\theta-1}\right) a_t$ and $r_t^* = E_t c_{t+1}^* - c_t^* = \left(\lambda + \frac{\theta}{\theta-1}\right)\left(E_t a_{t+1} - a_t\right)$:

$$c_t - c_t^* = E_t \left(c_{t+1} - c_{t+1}^*\right) + \lambda \left(\theta - 1\right) \left(i_t - E_t \pi_{t+1} - \frac{1}{\lambda + \frac{\theta}{\theta-1}} r_t^*\right) \tag{44}$$

or in levels (instead of gaps):

$$c_t = E_t c_{t+1} - \left(\lambda + \frac{1}{\theta - 1}\right) \theta \left(E_t a_{t+1} - a_t\right) + \lambda \left(\theta - 1\right) \left(i_t - E_t \pi_{t+1}\right) \tag{45}$$

Aggregate Supply: Starting from the Phillips curve for PPI inflation obtained when it is costly for individual producers to change their prices (Bilbiie, Ghironi and Melitz, 2007):

$$\pi_t = \beta E_t \pi_{t+1} - \psi \mu_t, \tag{46}$$

where $\psi = \left(\theta - 1\right)/\kappa$ and κ is the Rotemberg adjustment-cost coefficient ranging from 0 (flexible prices) to infinity (fixed prices). The loglinearized free entry condition, using that hours worked are fixed in equilibrium, implies that $\mu_t = (\theta - 1)^{-1}(n_t - a_t)$ and using aggregate accounting $c_t = \frac{\theta}{\theta-1}a_t + \lambda n_t$ to replace the number of goods we obtain:

$$\mu_t = \frac{1}{\lambda(\theta-1)}\left[c_t - \left(\lambda + \frac{\theta}{\theta-1}\right)a_t\right].$$

Replacing in the pricing equation and using $c_t^* = \left(\lambda + \frac{\theta}{\theta-1}\right)a_t$ we obtain:

$$\pi_t = \beta E_t \pi_{t+1} - \psi \frac{1}{\lambda(\theta-1)}(c_t - c_t^*) \tag{47}$$

Equations (44) and (47), together with a standard Taylor rule

$$i_t = \phi \pi_t - \varepsilon_t, \tag{48}$$

constitute a full description of the model. Notice that when prices are flexible, the equilibrium is fully determined by the supply side AS (47), $c_t = c_t^*$. While when prices are completely rigid, it is determined exclusively by the demand side, AD (44) or (45) $c_t = \left(\lambda + \frac{1}{\theta-1}\right)\theta a_t$. In between these two extremes, we need to solve the model.

To do so, first notice that the requirement for equilibrium determinacy in the entry model is exactly the same as in the no-entry model: the Taylor principle $\phi > 1$. To prove this, replace (47) and (48) into (44) to eliminate the output gap and interest rate, obtaining:

$$\pi_t - \beta E_t\pi_{t+1} = E_t\pi_{t+1} - \beta E_t \pi_{t+2} - \psi\left(\phi\pi_t - E_t\pi_{t+1} - \frac{1}{\lambda + \frac{\theta}{\theta-1}}r_t^*\right), \tag{49}$$

Solving under AR1 shock with persistence ρ_a, $E_t a_{t+1} = \rho_a a_t$ and letting $\tilde{\psi} \equiv \psi/(1-\beta\rho_a)$:

$$\pi_t = -\tilde{\psi}\frac{1-\rho_a}{1-\rho_a + (\phi - \rho_a)\tilde{\psi}}a_t;\; c_t - c_t^* = \lambda(\theta-1)\frac{1-\rho_a}{1-\rho_a+(\phi-\rho_a)\tilde{\psi}}a_t$$

The result generalizes the previous one, derived with fixed prices: when the condition making demand, variety forces dominate supply, entry forces holds ($\lambda > 0$), a bad supply shock causes a *negative output gap* and *PPI inflation*. Whereas in the opposite case, it causes a *positive output gap* that is still accompanied by PPI inflation. As a side note, this points to the possibility of deriving an implicit empirical test, based on macro comovements, of the mysterious micro parameter λ. Since there is exit regardless of whether $\lambda \gtreqless 0$, CPI inflation is also going up.

An important point, which is related to determinacy results staying unchanged relative to the no-entry model, is that the crossing of the threshold $\lambda = 0$ triggers a swiveling of *both* AD and AS: in the $\lambda > 0$ region, AD slopes upwards and AS slopes downwards. A shift upwards of AD (as happens when the natural interest rate goes up, in response to an adverse supply shock) moves us leftward along the downward sloping AS, thus triggering a fall in output gap and inflation. Whereas for $\lambda < 0$ AS and AD have regular slopes and a shock shifting AD up causes an increase in the output gap and inflation, moving along an upward sloping AS curve.[41]

[41] In the CES-DS case, AD is vertical and price stickiness is irrelevant, the neutrality result in Bilbiie (2021).