Capacity Utilization, Markup Cyclicality, and Inflation Dynamics*

Ignacio González[†] Vasudeva Ramaswamy[‡]

American University American University.

This version: April 2024
First version: May 2023

Abstract

We introduce endogenous capacity utilization into a New Keynesian (NK) model. In our model, firms set capacity under demand uncertainty, utilizing both an effort margin and capacity expansion to meet demand. This mechanism implies that firm-level productivity and desired markups depend on the capacity utilization rate, enabling us to derive three (state-dependent) results that have proven challenging for NK models. First, following an expansionary demand shock, the aggregate markup responds procyclically when desired markups rise enough to overcome the effect of nominal rigidities. Secondly, the labor share can respond countercyclically for reasons that are empirically consistent, namely, when labor productivity, propelled by the utilization of idle capacity, increases more than wages. Finally, inflation typically displays a hump-shaped response, but can also respond sharply during periods of high capacity utilization due to fast-rising markups and reduced productivity effects. We detail the conditions under which these results arise. Using Bayesian IRF matching, we illustrate that our model provides a highly plausible fit to the data. Our results underscore the importance of capacity for the macroeconomic debate surrounding the determinants, dynamics and distributional effects of inflation.

Keywords: Monetary policy, Capacity Utilization, Capacity Constraints, Markups, Labor Share, Inflation, Wages.
JEL codes: E22, E31, E32, E52

*We thank Levent Altinoglu, Justin Bloesch, Chuan Du, Mike Konzcal, Annie McGrew, Ansgar Rannenberg, Luis E. Rojas, Xuguang Simon Sheng, and Joseph E. Stiglitz for constructive comments. We also thank comments from seminar participants at American University, the ASSA annual meeting, the SEA annual meeting, the International Economic Association World Congress, and the panel discussion at the IMF/IPD Conference "Inflation: Origins, Transmissions, Policy Responses and International Implications. We gratefully acknowledge financial support from the Rockefeller Foundation. This paper has circulated previously as "The Cyclical and Distributional Effects of Capacity Utilization". Paper available at SSRN: https://ssrn.com/abstract=4434397

[†]Department of Economics. American University. ignaciog@american.edu

[‡]Department of Economics. American University. ramaswamy@american.edu. Corresponding author.

1 Introduction

The economic recovery after the COVID-19 pandemic has renewed interest in the role of capacity, particularly in the response of inflation to supply and demand shocks. However, standard macroeconomic models seldom model capacity decisions explicitly, focusing instead on variable *capital* utilization costs. The role of *capacity* utilization in business cycles therefore remains understudied despite recent evidence regarding the usual presence of idle resources and its cyclical properties (Boehm and Pandalai-Nayar, 2022). This paper aims to contribute to filling this gap by examining the aggregate effects of capacity utilization in response to monetary policy (MP) shocks. We specifically argue that a model incorporating endogenous capacity and utilization can account for the observed cyclical behavior of the labor share and markups as well as the dynamics and composition of inflation. These aspects have proven challenging to explain within the traditional New Keynesian framework.

In general, the predictions of the New Keynesian (NK) model regarding the effects of MP shocks do not align well with the observed empirical responses of the labor share, markups, and inflation. With respect to the labor share, Cantore et al. (2020) show that, in the data, the labor share ($WL/\mathbf{Y}$) is robustly countercyclical conditional on an MP shock. This occurs because while both wages (W) and labor productivity ($\mathbf{Y}/L$) are procyclical, the latter is more procyclical than the former. The canonical NK model, however, predicts countercyclical productivity and cannot achieve a countercyclical labor share without generating other empirically inconsistent responses such as countercyclical wages. In relation to markups, while measuring them poses notable challenges, studies by Nekarda and Ramey (2020), Stroebel and Vavra (2019), Anderson et al. (2018), and others offer evidence supporting the procyclical nature of markups under demand shocks. Given the close inverse relationship between the markup and the labor share, this is consistent with the countercyclicality of the latter. Nevertheless, in New Keynesian models, the transmission mechanism relies on countercyclical markups. Finally, regarding inflation, a large empirical literature (e.g., Christiano et al. 2005) finds that the response of inflation to an MP shock is either muted or countercyclical initially, but persistently procyclical in subsequent periods. This distinctive hump-shaped response typically cannot be generated in NK models without resorting to counterfactual or unconventional mechanisms such as backward price indexation or alternative forms of (non-rational) expectations (Woodford, 2007; Phaneuf et al., 2018).

The model that we present in this paper can yield procyclical markups, a countercyclical labor share, and hump-shaped inflation under conditions that we find empirically prevalent and without compromising the model's performance on other aggregate variables. We achieve this by incorporating endogenous capacity into an otherwise standard New Keynesian model. In our model, firms face idiosyncratic demand uncertainty, akin to the setup in Fagnart et al. (1999), and plan their productive capacity prior to observing demand. However, in contrast to these authors' approach, firms in our model select all factors, including labor, before the actual demand materializes. This timing implies the presence of a capacity limit—a level of output beyond which the firm cannot produce because its inputs for the period have been predetermined. Thus, when the firm sets capacity, it is optimal for it to keep some excess *precautionary* capacity on hand to service higher-than-average realizations of demand. When the idiosyncratic level of demand actually manifests, output is produced by varying workers' labor effort up to the capacity limit of the firm, in the spirit of the labor hoarding mechanism in Solow (1964), Burnside et al. (1993), and others.[1]

We show that, given this framework, each of the challenges outlined above can be resolved under certain condi-

[1] As Orphanides (1993) shows, embedding labor effort as in the Shapiro and Stiglitz (1984) "shirking model" can be used to rationalize labor hoarding in the presence of idiosyncratic demand shocks, akin to those in our model.

tions. First, following an expansionary MP shock, firms "price in" the increased probability of their being capacity constrained. That is, firms find it optimal to increase prices and eliminate excess expected demand that cannot be serviced because of capacity limits. This manifests in procyclical *desired* markups, though the cyclicality of *realized* markups also depends on the strength of nominal rigidities, which may push them in the opposite direction. Second, until capacity can be expanded sufficiently to meet demand, firms rely on exploiting the intensive margin of their workforce—i.e., requiring higher labor effort from workers. This aligns with the procyclical nature of worker effort observed in the data, and results in higher labor productivity and wages in our model. When wage growth is slower than productivity growth (for example, because of labor market frictions), our model yields a countercyclical labor share for reasons in line with the evidence. Finally, the response of inflation reflects the upward pressures of markups and wages and the downward pressures of procyclical productivity. When the latter pressures dominate early in the cycle, inflation exhibits the characteristic hump shape. However, when capacity utilization is high, procyclical productivity cannot offset the increase in costs and markups, causing inflation to respond on impact. In the subsequent sections, we elaborate on the conditions in which these results occur.

We next turn to the question of whether these conditions are met in the data. To do so, we empirically estimate our model using Bayesian impulse-response function (IRF) matching. Our exercise reveals a very good match between the data and our model. Importantly, when evaluated at the estimated parameters, the model delivers procyclical markups, a countercyclical labor share and a hump-shaped response of inflation that also captures the price-puzzle phenomenon.[2] These results are particularly encouraging given that neither the markup nor the labor share is directly used in the matching exercise and that we generate the inflation response without incorporating backward indexation, nonrational expectations, extended "cost channels" or other elements typically employed for this purpose.

Some important insights emerge from our study. First, the cyclicality of the markup and labor share are state dependent and functions of the degree of slack (excess capacity) in the economy at the time of the MP shock. We show that when capacity utilization in the economy is high, markups are more likely to be procyclical in response to an MP shock, and vice versa. Intuitively, high capacity utilization signals strong demand relative to capacity, which translates in our model into higher pricing power and markups for firms. The capacity utilization rate also emerges as a wedge between the markup and the labor share in our model, breaking the perfectly inverse relationship in the canonical NK model. Indeed, we show that under specific conditions, the markup and the labor share can *both* respond procyclically to a demand shock.

Second, productivity in our model is procyclical and state dependent, with the strength of the response depending on the degree of excess capacity in the economy at the time of the shock. In other words, demand shocks can *induce* higher productivity in the presence of slack—that is, when higher demand meets existing productive capacity, higher output can be achieved without any additional observable inputs. Our model results therefore align with the characterization of cyclical fluctuations in Basu (1996): productivity is procyclical, driven by increased utilization of capital and labor in a production environment exhibiting constant returns to scale. As described above, this occurs because firms rely on the effort margin of their worker base to meet demand; thus, demand shocks induce greater productivity effects when the degree of slack in the economy is higher. In this respect, our approach to the firm's

[2]The price puzzle (Sims, 1992; Eichenbaum, 1992) describes the increase (decrease) in the inflation rate for some periods immediately following an unexpected monetary policy rate increase (decrease), a response that runs counter to the prediction of macroeconomic theory.

labor choice is similar to that of the effort margin in the labor hoarding mechanism in Burnside et al. (1993) and variable labor effort in Galí and Van Rens (2021).[3] There is strong empirical support indicating the importance of the effort margin for labor productivity. For example, Bils and Cho (1994), Sbordone (1996) and Basu and Kimball (1997) provide evidence that variable labor utilization can explain procyclical productivity via the effort margin while matching important business cycle facts related to GNP and the Solow residual. More recently, Lewis and Villa (2023) show using data from the Euro area that worker effort is strongly procyclical and a significant driver of procyclical productivity, whereas Dossche et al. (2023) find that worker effort is crucial for explaining procyclical productivity and employment volatility in OECD countries.

Third, the response of inflation in our model to demand shocks in the economy is state dependent. More specifically, it is characterized by an inertial, hump-shaped response when there is a higher degree of slack in the economy. When there is a high rate of capacity utilization (and therefore low slack) however, inflation responds sharply on impact. These nonlinear effects reflect the convexity of the supply curve due to capacity constraints, which is consistent with the industry-level findings of Boehm and Pandalai-Nayar (2022).[4] We find that in the presence of sufficient slack, expansionary shocks can even cause a *fall* in inflation if the productivity effects outweigh the markup and wage effects, thereby providing an explanation for the so-called price puzzle. Likewise, when capacity utilization is sufficiently high, expansionary shocks can cause *real* wages to fall if the productivity effects are too small, and inflation is driven primarily by markups.[5] We are not aware of any model where the response of inflation is either hump-shaped or immediate depending on state of the economy. To further elucidate these points, we derive a state-dependent Phillips curve and present a decomposition of inflation over the cycle, characterizing the contribution of markups, wages and capacity utilization in the dynamic response of inflation.[6]

Fourth, relative nominal rigidities in the product and labor market interact with capacity utilization to generate important distributional effects. Under conditions of normal capacity utilization, high price rigidities relative to wage rigidities delivers model responses that are observationally equivalent to those of a New Keynesian model—that is, demand shocks cause markups to respond countercyclically, while wages and the labor share are procyclical. When we reverse the relative rigidities, we obtain the empirically consistent result of procyclicality in markups and wages and countercyclicality in the labor share. Under conditions of higher capacity utilization rates, however, markups are more likely to be procyclical, while wages can even be countercyclical (as observed above). In this respect, our results echo the conclusions of Broer et al. (2020), who find a crucial role for wage rigidities in delivering plausible cyclical and distributional responses. Our key contribution in this regard is to highlight the role played by capacity utilization in determining these effects.[7]

[3]The idea that the procyclicality of productivity is related to that of labor effort is an old one, going back at least to Oi (1962), with important contributions by Burnside et al. (1993) and Basu and Fernald (2001). Galí and Van Rens (2021) provide a list of additional contributors.

[4]For the relevance of capacity utilization as an aggregate indicator of inflationary pressures, see Corrado and Mattey (1997).

[5]We find this possibility especially pertinent in the early stages of the post-COVID recovery, which were marked by strong nominal wage growth but declining real wages for most of the worker distribution (Autor et al., 2023).

[6]We believe a decomposition of this type can also contribute toward diagnosing the post-pandemic inflation episode. Whereas the role of supply-side constraints in driving inflation is acknowledged, ambiguity remains regarding the role played by marginal costs vis-à-vis markups.

[7]Note that wage rigidities by themselves cannot generate procyclical markups in a standard NK model. This well-known result emerges from the fact that firms in NK models take the wage as given when determining their optimal price. Thus, regardless of the degree of wage rigidity relative to price rigidity, the markup cannot be procyclical. At the limit, when wages

Our main contribution is to show that a model with capacity utilization can explain the empirical findings related to markups, the labor share, and inflation described above. Explaining the cyclical behavior of these variables is of first-order importance for investigations of the distributional consequences of macroeconomic policy. For example, the cyclicality of the labor share is a direct estimate of inequality in an economy. Typical approaches to achieving a countercyclical labor share—often by adopting significant wage rigidity—primarily result in making wages themselves countercyclical. The desired direction of the response can therefore be achieved only for "the wrong reasons", to borrow a phrase from Cantore et al. (2020). Similarly, the reliance on countercyclical markups in NK models typically results in countercyclical profits (as documented at least since Christiano et al., 1997), which are robustly procyclical in the data. Thus, the presence of countercyclical markups may bias and distort the so-called income-composition channel of MP shocks by mischaracterizing the dynamics of profit income in the economy. Although profit procyclicality can be achieved through the introduction of high wage rigidities (Broer et al., 2020, Bilbiie and Känzig, 2023) or large fixed costs (Lee, 2021), evidence on markup behavior and convex supply curves indicate that procyclical markups may be an important underlying driver for both profits and inflation—particularly in the short-run, when predetermined capacity may impede smooth production expansion.

Other Related Literature Our paper presents an extension of the New Keynesian model that can explain the joint dynamics of markups, the labor share and inflation following a monetary policy shock. To our knowledge, the only other study which shares a similar focus is Qiu and Ríos-Rull (2022), who propose a model where the procyclical search effort of customers within product markets produces procyclical productivity as a response to monetary policy shocks. This differs from our approach, where productivity is procyclical due to the presence of excess capacity which is better utilized during demand-driven expansions. Their model is able to achieve procyclical markups and a countercyclical labor share, although, with respect to inflation, their focus is on achieving a procyclical response. In our model, inflation is also generically procyclical, but we are more concerned with the conditions under which it displays an inertial ("hump-shaped") response and those under which it may respond sharply on impact.

Some other studies share a partial focus with our paper. For example, Phaneuf et al. (2018) present a model that features firm networking and an "extended working capital channel" to align the dynamics of inflation and markups with the evidence. The latter mechanism assumes that the cost of *all* inputs (rather than just wages as in Ravenna and Walsh 2006) is financed through borrowing.[8] Their model generates a hump-shaped response to inflation. However, their results on the cyclicality of the markup rely a different definition of the markup—one that includes the cost of borrowing inputs—from the definition used by Nekarda and Ramey (2020) and others in generating the empirical evidence. Once adjusted for this measurement, the markup in their model is still countercyclical. Similarly, Hyun et al. (2023) propose a model that uses a translog production function that generates procyclical returns to scale. They find a high degree of complementarity between labor and energy, which is also able to deliver procyclical markups and countercyclical labor shares *unconditionally*. This differs from the focus of the present study, which is on markup cyclicality conditional on monetary policy shocks. Finally, in the pursuit of connecting these phenomena, a large post-Keynesian tradition, notably within the Kaleckian/neo-Kaleckian framework, also considers the joint

are rigid but prices are perfectly flexible, the NK model delivers a constant (acyclical) markup. See Nekarda and Ramey (2020) and Cantore et al. (2020) for a detailed discussion.

[8] We follow the literature in assuming a more standard working capital (i.e. only a fraction of the wage bill is borrowed each period). Our estimates are in line with the empirical findings of Galindo Gil (2021).

behavior of capacity utilization and demand fluctuations, albeit typically under the assumption of fixed markups and cost-plus pricing principles (Blecker and Setterfield, 2019). As noted above, in the present study, we rely on capacity utilization and markups which are both endogenous and time-varying.

The importance of capacity constraints is empirically borne out in Boehm and Pandalai-Nayar (2022), who use the Fagnart et al. (1999) framework to show that the degree of capacity utilization is a sufficient statistic for the degree of convexity of the industry's supply curve. Other studies using this model include those of Àlvarez-Lois (2006), who studies the effects of an MP shock in an economy with countercyclical markups, and Kuhn and George (2019), who use it to explain multiple business cycle asymmetries. However, the model presented in Fagnart et al. (1999) suffers from a key limitation, which is that markups can be procyclical following a demand shock only when wages are countercyclical (and vice versa).[9] As a result, the ex ante real interest rate in the model is procyclical for most plausible parameterizations of the model. That is, expansionary MP shocks are associated with a *rise* in the ex ante real interest rate.[10] This is not the case in our model, where the existence of a labor effort margin causes labor productivity to rise following a demand shock, which then allows both wages and markups to respond procyclically. Note, however, that variable labor utilization models (à la Burnside et al., 1993) alone cannot produce procyclical markups, nor can the capacity constraints model of Fagnart et al. (1999) by itself produce a countercyclical labor share for the "right reasons".

There exist other approaches to modeling capacity and slack. Hansen and Prescott (2005) consider an economy where firms either operate plants or leave them idle and where capacity constraints bind when all plants are operational. Gilchrist and Williams (2000) incorporate investment irreversibility in a putty-clay environment to generate capacity constraints, which are relaxed as capital of new vintage is installed. A different vein of research Michaillat and Saez (2015, 2022, 2024) motivates slack in both product and labor markets as originating from matching frictions in those markets. Likewise, excess capacity has been studied in the context of insufficient demand from consumers where firms operate in a negligible marginal cost or fixed cost environment (Murphy, 2017; Auerbach et al., 2023). In our model, excess capacity is precautionary; that is, it emerges as the optimal behavior of firms in the presence of idiosyncratic demand uncertainty.

We also contribute to a literature on the behavior of inflation with respect to MP shocks. As discussed above, inflation in our model has a hump-shaped response to MP shocks under normal economic conditions, but can respond sharply on impact when utilization rates are high in the economy. We capture these effects in the state-dependent Phillips curve that we derive in Section 3.4. In addition to Phaneuf et al. (2018) discussed above, approaches to achieving hump-shaped inflation include incorporating departures from rational expectations (Adam, 2005), imperfect information (Mankiw and Reis, 2002) and dynamic externalities (Tsuruga, 2007). In our model, however, we retain purely forward-looking price-setting behavior by agents with rational expectations.

With respect to the state-dependency of the Phillips curve, the resurgence of inflation in the post-COVID recovery has generated renewed interest in this area. For example, Harding et al. (2023) present a state-dependent formulation

[9] We explore the reasons for this analytically in Section 2. See footnote 16 on page 13 for additional discussion.

[10] As is well known, in the 3-equation NK model, the nominal interest rate may rise or fall following an expansionary MP shock. However the ex ante real interest rate always falls. The counterfactual result in the Fagnart et al. (1999) model arises for reasons outlined in Rupert and Šustek (2019). The ex ante real interest rate reflects households' desire to smooth consumption and the feasibility of their doing so. Procyclical markups lead to countercyclical wages, and households respond by borrowing more to smooth consumption. This results in higher ex ante real interest rates.

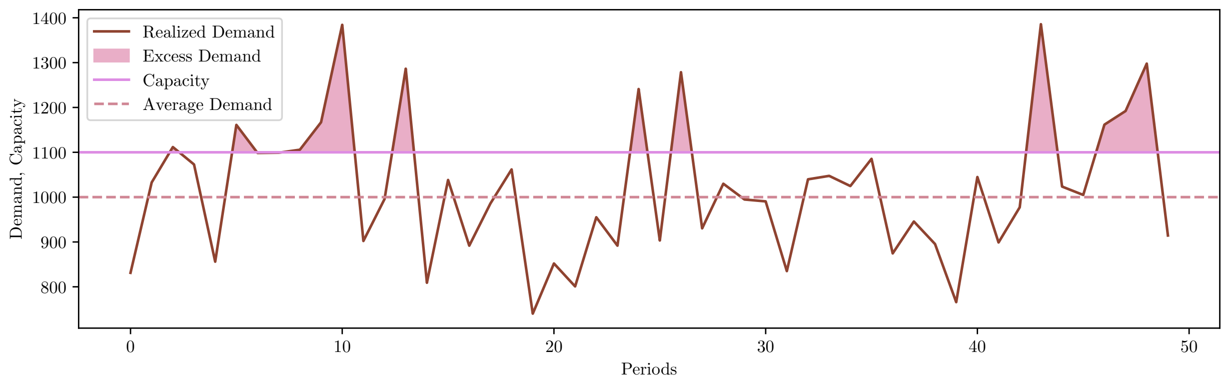


Figure 1: Example realizations of demand relative to capacity in the steady state. Given idiosyncratic demand, a firm's capacity choice implies that it may realize excess demand (relative to installed capacity) in some periods, and hold idle capacity in others.

of the Phillips curve that reflects the quasi-kinked nature of the Kimball aggregator that they assume. Likewise, Benigno and Eggertsson (2023) propose a Phillips curve that incorporates non-linearities reflecting *labor* market tightness. Finally, Comin et al. (2023) examine the effects of COVID era supply chain and capacity constraints on inflation in an model where capacity is exogenously predetermined, and large aggregate shocks cause capacity constraints to occasionally bind across the economy. In our Phillips curve formulation, the relevant state variable is the rate of capacity utilization in the economy, reflecting *product* market tightness, when the shock occurs. Our study offers a general description of inflation behavior—beyond episodes characterized by large shocks—by incorporating a mechanism that also explains the cyclical behavior of productivity, markups, and the labor share. In our approach, idiosyncratic shocks to firms imply binding constraints for some proportion of firms in *every* period, while aggregate demand shocks induce an endogenous expansion of capacity—and therefore capacity utilization rates.

The paper is presented as follows. Section 2 presents the main model and characterizes the equilibrium. Section 3 presents important theoretical and analytical results that emerge from our model. Section 4 outlines our IRF matching estimation procedure and results. Finally, section 5 discusses and concludes.

2 Model

Our economy is composed of a production sector, a household sector, and a monetary authority. The basic structure of the model is familiar: the production sector consists of a single competitive final aggregating firm, and a continuum of intermediate firms that produce individual varieties in a monopolistically competitive market. The household sector is composed of a continuum of individuals who derive utility from consumption and disutility from working. The monetary authority closes the model by setting a nominal interest rate according to a Taylor rule. Aggregate uncertainty is introduced through a shock to this rule.

Capacity and Production In period $t-1$, intermediate firms choose levels of capital and labor which together determine the maximum production capacity of the firm for the next period. In addition to aggregate uncertainty, this capacity is chosen in the presence of idiosyncratic demand uncertainty which is modeled as a serially uncorrelated idiosyncratic shock to the firms' demand function. Firms make the capacity decision balancing the cost of installing

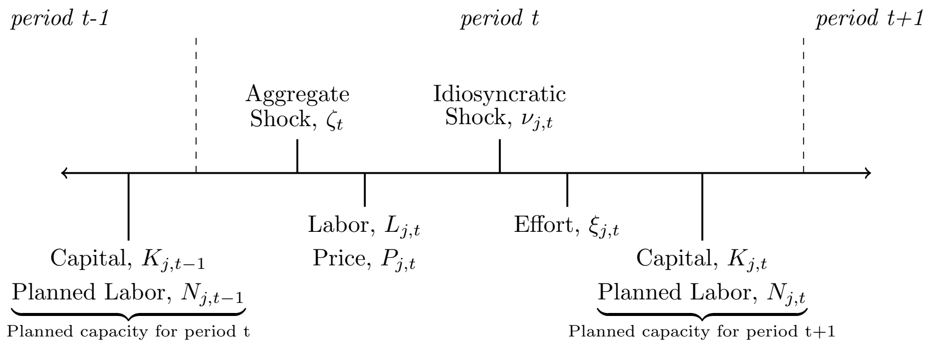


Figure 2: Timing of the model for the production sector.

more capacity with the expected higher revenue that the additional capacity may bring in.

This setup contrasts with the standard approach where only capital is pre-decided, but labor can be varied as necessary after demand is realized such that any level of output can be achieved. In our model, however, the maximum possible level of employment for the firm is determined through its capacity choice in the previous period. Due to its idiosyncratic nature, a firm's demand may exceed this maximum capacity in some periods. This probability of being *capacity constrained* is internalized by the firm as a constraint while setting prices. It is important to note, however, that the realization of the idiosyncratic shock does not affect the capacity setting decision of the firm for the next period. This follows directly from our assumption that the idiosyncratic shock is serially uncorrelated. Figure 1 illustrates graphically the nature of the firm's demand and the implications of its capacity choice in the steady state.

Timing To facilitate aggregation, we adopt the following timing sequence. Prior to the start of each period, each j^{th} firm faces two kinds of demand uncertainty: aggregate uncertainty (monetary policy shock, ζ_t, in this model) and idiosyncratic uncertainty (denoted by $\nu_{j,t}$) as described above. We assume that the two shocks are uncorrelated. In period $t-1$, the firm plans for a level of capacity in period t without knowing the realization of either ζ_t or $\nu_{j,t}$, and based on its expectations which are informed by the (known) moments of both shocks. This planned capacity decision entails choosing a level of capital stock and a level of maximum employment. At the start of the period t, after the aggregate shock ζ_t is observed, the j^{th} firm takes its pricing decision, $P_{j,t}$, and hires workers $L_{j,t}$ at a wage W_t. After the idiosyncratic demand uncertainty is resolved, the firm observes its period demand. The firm then proceeds to extract a level of effort $\xi_{j,t}$ from its workers in order to produce output and meet realized demand. Note that this setup implies that firms adjusts to aggregate shocks through the choice of P and L, while it adjusts to the idiosyncratic shock through movements in ξ.

This timing schema ensures that firms are always ex ante identical (before the shocks) and ex-post heterogeneous in a way that allows aggregation of both quantities as well as economy-wide prices. Figure 2 illustrates the timing sequence for the production sector. The following sections describe each sector in greater detail.

2.1 Final Firm

Our set-up of the final firm follows Fagnart et al. (1999). As in standard New Keynesian frameworks, the final firm aggregates inputs from intermediate firms into a final good that is sold in a perfectly competitive market. These intermediate firms are indexed by $j \in [0, 1]$. The final firm uses a constant-returns-to-scale CES aggregator similar to the one introduced in Dixit and Stiglitz (1977):

$$\mathbf{Y_t} = \left[\int_0^1 (Y_{j,t})^{\frac{\epsilon-1}{\epsilon}} (\nu_{j,t})^{\frac{1}{\epsilon}} \, dj\right]^{\frac{\epsilon}{\epsilon-1}} \tag{2.1}$$

Here, $\epsilon > 1$ represents the elasticity of substitution between the varieties produced by the j firms. As in Fagnart et al. (1999), $\nu_{j,t} \geq 0$ is the realization of the idiosyncratic shock for the j^{th} input producer. The idiosyncratic shock is assumed to be drawn from a serially uncorrelated stochastic i.i.d. process. For the purposes of this paper, we assume that this process is fully characterized by a log-normal distribution with support over $[0, \infty)$. The process has a mean $\mu_\nu = 1$ and variance σ_ν^2 (to be estimated later).

Note that $\nu_{j,t}$ in (2.1) represents the *realized* value of the shock. This is because the final firm operates after all uncertainty has been resolved. Both the actual quantity of output from the j^{th} firm, as well as the quoted prices are known to the final firm. However, the final firm also has to take into account the fact that some firms may be *capacity constrained*, that is to say, they may have experienced a positive shock to their demand large enough that they are constrained by their installed capacity, $\bar{Y}_{j,t}$, which was decided prior to the shock. Thus, the final firm maximizes:

$$\max_{\mathbf{\Pi_t}} \Pi = \mathbf{P_t Y_t} - \int_0^1 P_{j,t} Y_{j,t} \, dj$$

subject to

$$Y_{j,t} \leq \bar{Y}_{j,t}$$

Denoting the relative price of the j^{th} firm by $P_{j,t}/\mathbf{P_t} = \tilde{P}_{j,t}$, the solution to the final firm's problem is given by the following $\forall j \in [0, 1]$:

$$Y_{j,t} = \begin{cases} \tilde{P}_{j,t}^{-\epsilon} \mathbf{Y_t} \nu_{j,t} & \text{if } 0 \leq \nu_{j,t} \leq \bar{\nu}_{j,t} \\ \bar{Y}_{j,t} & \text{otherwise} \end{cases} \tag{2.2}$$

where

$$\bar{\nu}_{j,t} = \frac{\bar{Y}_{j,t}}{\tilde{P}_{j,t}^{-\epsilon} \mathbf{Y_t}} \tag{2.3}$$

Here, $\bar{\nu}_{j,t}$ represents the value of the demand shock at which the j^{th} firm hits its capacity constraint, $\bar{Y}_{j,t}$. As discussed above, the intermediate firms are identical until the realization of the idiosyncratic shock. This implies that they choose the same capacity and price, and have the same shock threshold $\bar{\nu}$. Thus, $\forall j$, $\bar{Y}_{j,t} = \bar{Y}_t$, $\tilde{P}_{j,t} = \tilde{P}_t$ and $\bar{\nu}_{j,t} = \bar{\nu}_t$. Since firms are only differentiated by their shocks each period, the law of large numbers implies that final firm's aggregating function, equation (2.1), can be re-written as

$$\mathbf{Y_t} = \left[(\tilde{P}_t^{-\epsilon} \mathbf{Y_t})^{\frac{\epsilon-1}{\epsilon}} \int_0^{\bar{\nu}_t} \nu_t \, dF(\nu) + (\bar{Y}_t)^{\frac{\epsilon-1}{\epsilon}} \int_{\bar{\nu}_t}^{\infty} \nu_t^{\frac{1}{\epsilon}} \, dF(\nu)\right]^{\frac{\epsilon}{\epsilon-1}} \tag{2.4}$$

where $F(\nu)$ is the distribution function of the idiosyncratic productivity shocks, and the integral has been partitioned in line with (2.2). Given a threshold shock value, $\bar{\nu}_t$, $F(\bar{\nu}_t)$ describes the proportion of firms that are operating with excess capacity (i.e., *demand deficient* firms), while $1 - F(\bar{\nu}_t)$ describes the proportion of firms operating at full capacity, i.e., *capacity constrained* firms.

The set-up of the final firm's problem in Fagnart et al. (1999) allows us to directly obtain closed form representations of aggregate capacity utilization and the relative price as functions only of the threshold shock value $\bar{\nu}$. We restate these formulations here.

Aggregate Capacity Utilization Aggregate capacity utilization is defined as $y_t^* = \frac{\mathbf{Y_t}}{\bar{Y}_t}$. Combining equations (2.3) and (2.4), we have:

$$y_t^* = \frac{\mathbf{Y}_t}{\bar{Y}_t} = \left[\left(\frac{1}{\bar{\nu}_t}\right)^{\frac{\epsilon-1}{\epsilon}} \int_0^{\bar{\nu}_t} \nu_t \, dF(\nu) + \int_{\bar{\nu}_t}^{\infty} \nu_t^{\frac{1}{\epsilon}} \, dF(\nu)\right]^{\frac{\epsilon}{\epsilon-1}} \tag{2.5}$$

with the right-hand side depending entirely on $\bar{\nu}$, the distribution of ν, and ϵ. It can be shown that y^* is strictly decreasing in $\bar{\nu}$ and is bounded by the $[0, 1]$ interval.

Relative Price Further manipulation of equations (2.3) and (2.4), and recalling that $\tilde{P} = P/\mathbf{P}$, yields

$$\tilde{P}_t = \left[\int_0^{\bar{\nu}_t} \nu_t \, dF(\nu) + (\bar{\nu}_t)^{\frac{\epsilon-1}{\epsilon}} \int_{\bar{\nu}_t}^{\infty} \nu_t^{\frac{1}{\epsilon}} \, dF(\nu)\right]^{\frac{1}{\epsilon-1}} = (y_t^* \bar{\nu}_t)^{\frac{1}{\epsilon}} \tag{2.6}$$

The relative price $\tilde{P}$ is strictly increasing in $\bar{\nu}$ and is upper-bounded by 1.

Note that the latter equality also reflects the definition of $\bar{\nu}$, given in (2.3). In any given period t, $\mathbf{P}$ represents the price paid by customers, i.e., households and firms, for the final good which they consume and invest. But P is the nominal price set and received by intermediate firms for their variety. In a standard New Keynesian setup with Rotemberg pricing, the final price is the same as the price set by intermediate firms and $\tilde{P} = 1$. This equivalence result of $P = \mathbf{P}$ is only achieved in our model when capacity constraints are completely absent, i.e., as $\bar{\nu} \to \infty$.

2.2 Intermediate Firms

There exists a continuum of intermediate firms on the unit interval that produce differentiated goods. As described above, the production process involves setting a capacity and hiring workers before observing demand, and varying the effort of workers to meet demand after it is observed. The decisions of the firm are made under different information sets. We describe the production process and these decisions in detail below.

Production In period $t-1$, firms determine capacity for period t by choosing a level of capital, K, and the maximum planned number of workers who can operate on that capital, denoted by N, according to a CES capacity function:

$$\bar{Y}_{j,t} = \left(\alpha_K K_{j,t-1}^{\psi} + \alpha_N N_{j,t-1}^{\psi}\right)^{\frac{1}{\psi}} \tag{2.7}$$

where $\bar{Y}_{j,t}$ is the j^{th} firm's level of capacity planned in period $t-1$ for period t, ψ is the substitution parameter capturing the degree to which K and N can be substituted for each other in setting capacity, while α_K and α_N are distribution parameters.[11] Note that the choices of K and N in $t-1$ are made under both aggregate and idiosyncratic uncertainty; that is, neither ζ_t nor $\nu_{i,t}$ are realized, although the moments of ν are known to the firm. Since capacity planned for period t is pre-determined in period $t-1$ through the choice of factor levels, we can rewrite planned capacity as:

$$\bar{Y}_t = \left[\alpha_K \left(\frac{K_{t-1}}{N_{t-1}}\right)^{\psi} + \alpha_N\right]^{\frac{1}{\psi}} N_{t-1} = A_t N_{t-1} \tag{2.8}$$

[11] If $\sigma \in [0, \infty)$ is the capital-labor elasticity of substitution, $\psi = \frac{\sigma-1}{\sigma}$, which implies that $\psi \in (-\infty, 1]$. For values of $\sigma < 1.0$ ($\psi < 0$), N and K are considered complements. See section 4.3 for further discussion on the role of σ in the dynamics of the markup.

where A_t is the productivity of each worker implied by the capital-labor ratio chosen in equation (2.8). Notice that we have dropped the j subscript as all firms are *ex ante* identical and have the same information set when they choose their capacity, leading to identical choices of capacity, capital levels and planned employment levels.

At the start of period t, after observing any aggregate shocks, the firm hires a level of labor L_t subject to

$$L_t \leq N_{t-1} \tag{2.9}$$

The inequality above indicates that the firm cannot choose to hire more employees in period t than was planned for in period $t-1$.[12]

Note that L and N are related but distinct labor variables. N_{t-1} refers to the maximum *planned* number of workers for period t, so we refer to $A_t N_{t-1}$ as *planned* capacity. L_t refers to the number of workers actually hired in period t; $A_t L_t$ therefore refers to the *actual* productive capacity of the firm in period t. The inequality constraint in (2.9) allows for the possibility that firms choose a lower actual productive capacity in period t compared to what was planned in $t-1$. For simplicity, we assume that (2.9) always binds so that

$$\bar{Y}_t = A_t N_{t-1} = A_t L_t$$

for all t. We maintain both L and N in the exposition below to underscore that the labor market clears and wages are determined after the aggregate shock is observed, in keeping with the standard timing.

Once both the aggregate and idiosyncratic shocks have been realized and demand is known, the firm undertakes production. The production function is linear in effective labor and is given by:

$$Y_{j,t} = A_t \xi_{j,t} L_t \tag{2.10}$$

where $\xi_{j,t} \in [0, 1]$ is the level of effort extracted from hired labor L_t by the j^{th} firm.[13] Here, we re-introduce the j subscripts since firms are *ex-post* heterogeneous due to different realizations of demand. Thus, the idiosyncratic demand that materializes is met through adjustments in labor effort, ξ. When the firm produces at capacity, i.e., $Y_{j,t} = \bar{Y}_t$, then $\xi_{j,t} = 1$ and output equals actual capacity, $A_t L_t$. Thus, the *ex-post* marginal productivity of labor in the period of production is given by $A_t \xi_{j,t}$ and is dependent on the level of effort required from labor.

Firm Optimization We relate the optimal decisions of the firm in reverse order below. After all uncertainty (aggregate and idiosyncratic) is resolved, the firm observes its demand, $Y_{j,t}$. The firm determines the level of effort required to meet $Y_{j,t}$. This effort decision follows directly from (2.10):

$$\xi_{j,t} = \frac{Y_{j,t}}{A_t L_t} \tag{2.11}$$

Apart from $\xi_{j,t}$, the firm makes all its decisions before the idiosyncratic shock is realized and demand $Y_{j,t}$ is exactly known. Firms therefore use the possible outcomes described in (2.2) to form probability weighted expectations about their future demand as follows:

$$\mathbb{E}_\nu\{Y_t\} = (\tilde{P}_t)^{-\epsilon} \mathbf{Y_t} \int_0^{\bar{\nu}_t} \nu \, dF(\nu) + \bar{Y}_t \int_{\bar{\nu}_t}^{\infty} dF(\nu) \tag{2.12}$$

[12] The inequality constraint in (2.9) additionally highlights the Leontief nature of labor and installed capital in the short run. This production structure, where the elasticity of substitution between labor and capital differs over time-horizons, has significant empirical support. See Koh and Santaeulàlia-Llopis (2022) and Chirinko and Mallick (2017) for recent evidence.

[13] In Burnside et al. (1993), the value of ξ belongs in the $[0, \infty)$ interval.

Note that although the choice of $\xi_{j,t}$ is static after the idiosyncratic shock, the effort mechanism is dynamic over time. More specifically, the firm forms expectations over the level of effort to be extracted from workers in the future. Using equations (2.11) and (2.12) above, the firms' expectations on labor effort, denoted by $\bar{\xi}_t$, can be defined as:

$$\bar{\xi}_t = \mathbb{E}_\nu\{\xi_{j,t}\} = \frac{\mathbb{E}_\nu\{Y_t\}}{A_t L_t} \tag{2.13}$$

We can show that $\bar{\xi}$ is a strictly decreasing function of $\bar{\nu}$:

$$\bar{\xi}_t = \frac{1}{\bar{\nu}_t}\int_0^{\bar{\nu}_t} \nu \, dF(\nu) + \int_{\bar{\nu}_t}^{\infty} dF(\nu) \tag{2.14}$$

Value Function The value of the firm is given by the following:

$$V(K_{t-1}, N_{t-1}, \tilde{P}_{t-1}, L_{t-1}) =$$
$$\max_{\tilde{P}_t, L_t, K_t, N_t} \tilde{P}_t \cdot \mathbb{E}_\nu\{Y_t\} - W_t L_t - I_t - \Phi^P(P_t, P_{t-1}) - \Phi^H(H_t, L_{t-1}) + \mathbb{E}_t\{\rho_{t,t+1} V'(K_t, N_t, \tilde{P}_t, L_t)\} \tag{2.15}$$

subject to:

$$K_t = (1-\delta)K_{t-1} + \Phi^I(I_t, I_{t-1}) \tag{2.16}$$

$$L_t = (1-\varrho)L_{t-1} + H_t \tag{2.17}$$

and equations (2.8), (2.9) as an equality, (2.12) and (2.13), where term $\rho_{t,t+1}$ represents the stochastic discount factor.[14]

Here, (2.16) and (2.17) are the capital and labor laws of motion. Term I is investment while H is new hiring to replace the exogenous per-period rate of separations given by ϱ. Expressions $\Phi^P(\tilde{P}_t, \tilde{P}_{t-1})$, $\Phi^I(I, I_{t-1})$ and $\Phi^H(H_t, L_{t-1})$ represent price adjustment, investment adjustment, and labor adjustment costs respectively.[15] These take the following functional forms:

$$\Phi^P(\tilde{P}_t, \tilde{P}_{t-1}) = \frac{\phi^P}{2}\left(\Pi \frac{\tilde{P}_t}{\tilde{P}_{t-1}} - 1\right)^2 \mathbf{Y}_t \tag{2.18}$$

$$\Phi^I(I_t, I_{t-1}) = \left[1 - \frac{\phi^K}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2\right] I_t \tag{2.19}$$

$$\Phi^H(H, L_{t-1}) = \frac{\phi^H}{2}\left(\frac{H}{\tilde{L}_{t-1}} - \varrho\right)^2 \mathbf{Y}_t \tag{2.20}$$

Price Decision Maximizing the above with respect to the relative price $\tilde{P}$ yields the following optimality condition:

$$\tilde{P}_t = \frac{\epsilon\Gamma(\bar{\nu}_t)}{\epsilon\Gamma(\bar{\nu}_t) - 1}\left[\frac{W_t}{A_t\bar{\xi}_t} - \frac{\phi^P}{\epsilon\Gamma(\bar{\nu}_t)\mathbb{E}_\nu\{Y_t\}}(\Upsilon_t - \mathbb{E}_t\{\rho_{t,t+1}\Upsilon_{t+1}\})\right] \tag{2.21}$$

[14] To simplify the notation, the shocks have been excluded from the state vector.

[15] We include labor adjustment costs to differentiate between the notion of adjusting *capacity* versus adjusting capital outlays. In Section 4, we estimate the hiring adjustment cost parameter ϕ^H and find it to be very small relative to investment adjustment costs.

where

$$\Gamma(\bar{\nu_t}) = \frac{(\tilde{P}_t)^{-\epsilon}\mathbf{Y}_t}{\mathbb{E}_\nu\{Y_t\}} \int_0^{\bar{\nu_t}} \nu \, dF(\nu) \tag{2.22}$$

$$\Upsilon_t = \left(\Pi_t \frac{\tilde{P}_t}{\tilde{P}_{t-1}} - 1\right)\left(\Pi_t \frac{\tilde{P}_t}{\tilde{P}_{t-1}}\right)\mathbf{Y}_t \tag{2.23}$$

As in Fagnart et al. (1999), the term $\Gamma(\bar{\nu}_t)$ plays a key role in the dynamics of our model, and can be understood as the proportion of expected output that is produced by firms unconstrained by their capacity. In other words, this is the weighted probability of the firm holding excess capacity in the period. Indeed, combining equations (2.12) and (2.2), it can be shown that this this term is a strictly increasing function only of $\bar{\nu}$:

$$\Gamma(\bar{\nu_t}) = \frac{\int_0^{\bar{\nu}_t} \nu \, dF(\nu)}{\int_0^{\bar{\nu}_t} \nu \, dF(\nu) + \bar{\nu}_t \int_{\bar{\nu}_t}^{\infty} \, dF(\nu)} \tag{2.24}$$

However, the pricing policy in (2.21) differs from Fagnart et al. (1999) in a crucial way. Specifically, the interaction of $\Gamma(\bar{\nu})$ and $\bar{\xi}$ plays a crucial role in the dynamics and cyclicality of the markup and wages. To see this, assume that $\phi^P = 0$, i.e., a model with no price rigidities. In this case, the firm's optimal price choice is simply given by:

$$\tilde{P}_t = \underbrace{\frac{\epsilon\Gamma(\bar{\nu_t})}{\epsilon\Gamma(\bar{\nu_t}) - 1}}_{\text{markup}} \underbrace{\left(\frac{W_t}{A_t\bar{\xi}_t}\right)}_{\text{marginal cost}}$$

Equation (2.6) shows that $\tilde{P}$ is strictly increasing in $\bar{\nu}$. That is, following an expansionary demand shock, the relative price falls. Since $\Gamma(\bar{\nu_t})$ is strictly increasing in $\bar{\nu}$, expansionary shocks lead to a decline in $\Gamma(\bar{\nu_t})$ and an increase in the markup. Since A is predetermined when choosing $\tilde{P}$, it is clear procyclical wages require $\bar{\xi}$ to rise faster than the markup. Thus, in our model, achieving procyclical aggregate markups and aggregate wages simultaneously requires a strong productivity response indicated by $\bar{\xi}$. We explore this point further in Section 3.[16]

As is well known, the steady state markup in the New Keynesian is given by

$$\mu^{NK} = \frac{\epsilon}{\epsilon - 1} \tag{2.25}$$

The standard New Keynesian case is therefore simply a special case of (2.25) where $\bar{\nu}_t \to \infty$ and $\Gamma(\bar{\nu_t}) = 1$, i.e., when firms are *never* constrained by their capacity and *all* the output is produced by unconstrained firms.

Capacity Decisions The firm's capacity decisions consist of choosing a level of capital stock K and a maximum level of labor N. These decisions are made under full uncertainty (i.e., the realizations of both ν and ζ are unknown).

The firm once again uses the value function in (2.15) and the associated constraints. Maximization then yields the following optimality condition for capital stock K:

$$\mathbb{E}_t\left\{\rho_{t,t+1}\,\alpha_K\left(\frac{\bar{Y}_{t+1}}{K_t}\right)^{1-\psi}\left(\underbrace{\left(\tilde{P}_{t+1} - \frac{W_{t+1}}{A_{t+1}\bar{\xi}_{t+1}}\right)\int_{\bar{\nu}_{t+1}}^{\infty} dF(\nu)}_{\mathscr{A}} + \underbrace{\frac{W_{t+1}}{A_{t+1}\bar{\xi}_{t+1}}\left(\frac{\mathbb{E}_\nu\{Y_{t+1}\}}{\bar{Y}_{t+1}}\right)}_{\mathscr{B}}\right)\right\} \tag{2.26}$$
$$= Q_t - \mathbb{E}_t\{\rho_{t,t+1}(1-\delta)Q_{t+1}\}$$

[16]The above exposition also highlights our contribution to this class of models. In Fagnart et al. (1999) and other studies using their model, the marginal cost is given by W/A. Without a mechanism for increasing productivity, the model is incapable of producing procyclical markups without also producing countercyclical wages.

where Q is the marginal price of capital given by:

$$Q_t \left(1 - \phi^K \frac{I_t}{I_{t-1}} \left(\frac{I_t}{I_{t-1}} - 1\right) - \Phi^I(\cdot)\right) + \phi^K \mathbb{E}_t \left\{\rho_{t,t+1} Q_{t+1} \left(\frac{I_{t+1}}{I_t}\right)^2 \left(\frac{I_{t+1}}{I_t} - 1\right)\right\} = 1 \tag{2.27}$$

Equation (2.26) above is intuitively understood as follows. The right hand side represents the discounted cost of installing an additional unit of capital. The left hand side represents the discounted value of two elements, normalized by the marginal product of capital. The first element, indicated by $\mathscr{A}$, is the expected profits (expected price minus the expected marginal cost) generated from a marginal unit of capital, adjusted for the probability of being capacity constrained—i.e., *not* being able to operate that marginal unit of capital. This, therefore, represents the expected opportunity cost involved in maintaining a certain level of capacity. The second element, indicated by $\mathscr{B}$ represents the expected production costs at a certain level of capacity, conditioned on the probability of *operating* that capacity. The firm, therefore, chooses its level of capital such that the expected operating and opportunity costs are balanced by the costs of installation.

The maximum labor level N, in turn, is determined by the following optimality condition:

$$\mathbb{E}_t \left\{\rho_{t,t+1} \left[\underbrace{\alpha_N \left(\frac{\bar{Y}_{t+1}}{N_t}\right)^{1-\psi} \left(\tilde{P}_{t+1} - \frac{W_{t+1}}{A_{t+1}\bar{\xi}_{t+1}}\right) \int_{\bar{\nu}_{t+1}}^{\infty} dF(\nu)}_{\mathscr{A}} + \underbrace{\frac{W_{t+1}}{A_{t+1}\bar{\xi}_{t+1}} \left(\alpha_N \left(\frac{\bar{Y}_{t+1}}{N_t}\right)^{1-\psi} \int_0^{\bar{\nu}_{t+1}} \nu \, dF(\nu) - \frac{\mathbb{E}_\nu\{Y_{t+1}\}}{N_t}\right)}_{\mathscr{B}}\right]\right\} = \mathbb{E}_t \left\{\rho_{t,t+1} \Xi_{t+1}\right\} \tag{2.28}$$

where Ξ_t is given by:

$$\Xi_t = \phi^H \left(\frac{H_t}{L_{t-1}} - \varrho\right) \frac{Y_t}{L_{t-1}}$$

The intuitive explanation for (2.28) closely resembles (2.26). $\mathscr{A}$ represents the opportunity cost of maintaining a certain level of employment, normalized by the marginal product of a planned employee. $\mathscr{B}$ represents the expected cost of operation at a level of employment, similarly normalized. The right hand side equals the discounted labor adjustment costs associated with achieving a level of employment N_t. Thus, the firm chooses a production capacity with a level of maximum labor such that the expected wage bill, at the margin, is exactly justified by expected opportunity and operating costs of maintaining that capacity.

2.3 Households

The household sector features a continuum of households indexed by the unit interval. Households supply labor at a wage W_t to the production sector. To introduce nominal wage rigidities, we assume the existence of a labor agency that costlessly aggregates the different labor types into a homogeneous labor unit to be sold in a perfectly competitive market to the intermediate firms.

Aggregate Labor Supply The labor aggregating agency uses a CES aggregator of the form:

$$L_t = \left(\int_0^1 (L_{i,t})^{\frac{\epsilon^W - 1}{\epsilon^W}} \, di\right)^{\frac{\epsilon^W}{\epsilon^W - 1}}$$

where L_t is the aggregate level of labor demanded, $L_{i,t}$ is the level of labor supplied by the i^{th} household, $\epsilon^W > 0$ is the degree of substitutability between various labor types. Assuming labor of the i^{th} household is supplied at the real wage rate of $W_{i,t}$, optimization yields the labor demand for each household's labor:

$$L_{i,t} = \left(\frac{W_{i,t}}{W_t}\right)^{-\epsilon^W} L_t \tag{2.29}$$

where W_t is the aggregate wage index given by:

$$W_t = \left[\int_0^1 W_{i,t}^{1-\epsilon^W}\, di\right]^{\frac{1}{1-\epsilon^W}} \tag{2.30}$$

The Problem of Households The objective function of the household i is given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t\, U(C_i, L_i)$$

where $U(C_i, L_i)$ is the expected period utility of the household, defined by:

$$U(C_i, L_i) = \mathbb{E}_\nu \left[\log C_{i,t} - \omega \frac{(\xi_{i,t} L_{i,t})^{1+\varphi}}{1+\varphi}\right] = \log C_{i,t} - \omega \frac{L_{i,t}^{1+\varphi}}{1+\varphi}\, \mathbb{E}_\nu[\xi_{i,t}^{1+\varphi}] \tag{2.31}$$

Here, $C_{i,t}$ is the period consumption of the i^{th} household, and $\xi_{i,t} L_{i,t}$ is the effective labor supplied by the household. This latter concept captures both the number of hours as well as the actual effort expended each hour in the period. For simplicity, we assume that effort and hours enter the utility function symmetrically. This treatment of effective labor as a function of effort and hours is similar to the specification in Burnside et al. (1993).

At the start of period t, after observing the aggregate shock and before observing the idiosyncratic shock, household members contract to supply labor. But since the firm's demand has not yet manifested, households may have to expend different effort levels $\xi_{i,t}$ depending on the demand received by the firm. Households therefore choose a level of labor $L_{i,t}$ based on their *expected* level of effort.

The individual household members maximize these preferences subject to the budget constraint, which is standard:

$$C_{i,t} + B_{i,t} + V_t S_{i,t} \;\leq\; W_{i,t} L_{i,t} - \underbrace{\frac{\phi^W}{2}\left(\Pi_t \frac{W_{i,t}}{W_{i,t-1}} - 1\right)^2 W_t L_t}_{\text{Rotemberg wage adjustment costs}} + R_t B_{i,t-1} + S_{i,t-1}(V_t + D_t) \tag{2.32}$$

where $B_{i,t}$ are the real bond holdings of the household, R_t the risk-free interest rate, $S_{i,t}$ are the stock holdings representing ownership of the intermediate firms, V_t is the price associated with the stock, and D_t is the dividends issued by the intermediate firms. The Rotemberg wage adjustment costs represent the household's lost income associated with changing the nominal wage across periods. The parameter ϕ^W governs the degree of nominal wage stickiness.

Since the idiosyncratic uncertainty of the firm only affects the effort levels of the household, using the definition of $\mathbb{E}_\nu[\xi_{i,t}]$ from (2.13), we get:

$$\mathbb{E}_\nu\left[\xi_{i,t}^{1+\varphi}\right] = \underbrace{\left[\left(\frac{1}{\nu}\right)^{1+\varphi} \int_0^{\bar{\nu}} \nu^{1+\varphi}\, dF(\nu) + \int_{\bar{\nu}}^{\infty} dF(\nu)\right]}_{\text{Expected disutility from effort}}$$

Thus, after dropping the i-indexation, the expected utility function can be re-written as:

$$\mathbb{E}_\nu U(C, L) = \log C_t + \omega \frac{L_t^{1+\varphi}}{1+\varphi} \left[\left(\frac{1}{\nu}\right)^{1+\varphi} \int_0^{\bar{\nu}} \nu^{1+\varphi}\, dF(\nu) + \int_{\bar{\nu}}^{\infty} dF(\nu)\right] \tag{2.33}$$

Denoting the Lagrangian multiplier on the household's budget constraint by λ_t^H, the solutions to the household's problem are presented below:

W.r.t. Consumption

$$\lambda_t^H = \frac{1}{C_t} \tag{2.34}$$

W.r.t. Stocks

$$\mathbb{E}_t \left\{ \rho_{t,t+1} \left(\frac{V_{t+1} + D_{t+1}}{V_t} \right) \right\} = 1 \tag{2.35}$$

W.r.t. Bonds

$$\mathbb{E}_t \left\{ \rho_{t,t+1} \frac{R_t}{\Pi_{t+1}} \right\} = 1 \tag{2.36}$$

W.r.t. Wages

$$W_t(1+\chi_t) - \mathbb{E}_t \left\{ \rho_{t,t+1} W_{t+1} \frac{L_{t+1}}{L_t} \chi_{t+1} \right\} = \underbrace{\frac{\epsilon^W}{\epsilon^W - 1}}_{\text{Wage markup}} \underbrace{\frac{\omega L_t^{\varphi}}{\lambda_t^H} \left[\left(\frac{1}{\bar{\nu}_t} \right)^{1+\varphi} \int_0^{\bar{\nu}} \nu_t^{1+\varphi} + \int_{\bar{\nu}}^{\infty} dF(\nu) \right]}_{\text{Expected MRS}} \tag{2.37}$$

where the auxiliary variable χ_t is defined as

$$\chi_t \equiv \frac{\phi^W}{\epsilon^W - 1} \Pi_t \frac{W_t}{W_{t-1}} \left(\Pi_t \frac{W_t}{W_{t-1}} - 1 \right) \tag{2.38}$$

Note that the first order condition with respect to wages contains a term that captures households' uncertainty related to effort. Since households form expectations over the level of effort they have to expend, they evaluate their *expected* marginal rate of substitution (MRS). The choice of optimal wage (adjusted for expected wage growth and income loss from wage adjustment costs) then equals a markup over this *expected* MRS.

2.4 Central Bank

Monetary policy is managed by a Central Bank that targets both output and inflation following a Taylor rule. The Taylor rule is given by

$$\frac{R_t}{\bar{R}} = \left(\frac{R_t}{R_{t-1}} \right)^{\rho_S} \left(\left(\frac{\Pi_t}{\bar{\Pi}} \right)^{\rho_\pi} \left(\frac{Y_t}{Y_{SS}} \right)^{\rho_Y} \right)^{1-\rho_S} e^{\zeta} \tag{2.39}$$

where $\bar{R}$, $\bar{\Pi}$ and Y_{SS} are the target steady state (gross) nominal interest rate, (gross) inflation rate and output level respectively; ρ^π and ρ^Y are feedback coefficients on inflation, and output respectively, ρ^R is an interest-rate smoothing coefficient and ζ is the nominal interest rate shock that follows an AR(1) process in logs

$$\ln \zeta_t = \rho_R \ln \zeta_{t-1} + \epsilon^r$$

where $\epsilon^r \sim N(0, \sigma_r^2)$ is a random shock.

2.5 Equilibrium

Finally, the model is closed with a resource constraint that equates output in the economy to the sum of consumption, investment and price adjustment costs.

$$\mathbf{Y}_t = C_t + I_t + \Phi^P(P_t, P_{t-1}) \tag{2.40}$$

The dynamic equilibrium of this economy can be summarized as a vector of prices, quantities and proportions such that the optimality conditions outlined above are satisfied, and markets clear. Specifically, these include the price vector $\{\mathbf{P}_t, P_t, W_t, R_t\}$, the quantity vector $\{\mathbf{Y}_t, Y_t, C_t, L_t, K_t, N_t\}$, and the proportion of firms with excess capacities, $\{\Gamma(\bar{\nu}_t)\}$.

3 Analytical Results

In this section, we illustrate some of the properties of the model focusing on the implications for the markup, labor share and inflation in the model. We derive log-linearized approximations of the nonlinear model equations, and analytically establish the role of key parameters in determining the cyclicality of the markup, labor share and wages. We then derive the Phillips curve implied by our model, and outline the conditions necessary for the characteristic hump-shaped response.

3.1 Markup Cyclicality

In this section, we establish the conditions for the cyclicality of the markup. Unlike in the canonical NK model, the cyclicality of the markup depends on the parameterization and state of the economy. We start by presenting a log-linearized expression for the cyclicality of the markup.

Proposition 1. The response of the markup is given by:

$$\hat{\mu}_t = \underbrace{\frac{-\hat{\Gamma}(\bar{\nu}_t)}{\epsilon\Gamma(\bar{\nu}) - 1}}_{\textcircled{1}} - \frac{\phi^P}{\Psi^1}\left\{\underbrace{(\hat{\pi}_t^p - \beta\mathbb{E}_t\hat{\pi}_{t+1}^p) - (\hat{\pi}_t^w - \beta\mathbb{E}_t\hat{\pi}_{t+1}^w)}_{\textcircled{2}} - \underbrace{\frac{\epsilon^W - 1}{\phi^W}(\hat{mrs}_t - \hat{w}_t)}_{\textcircled{3}}\right\} \tag{3.1}$$

where $\Psi^1 = \mathbb{E}\{Y\}\tilde{P}(\epsilon\Gamma(\bar{\nu}) - 1)$ is a constant, all hatted variables are deviations from the steady state, and variable names without a time subscript are steady state values. $\hat{\pi}^p$ represents the inflation rate in the relative-price $\tilde{P}$; $\hat{\pi}^w$ is the real wage inflation rate; and $\hat{mrs}$ is the change in the household's marginal rate of substitution seen in (2.37).

Proof. See Appendix A.1. □

From Proposition 1, we see that the markup is composed of 3 elements. The first, indicated by (1), represents the impact of desired markups due to capacity constraints on the firm's pricing decision. The second term, indicated by (2), is the relative inflationary trajectories of the intermediate firm price $\tilde{P}$ and the real wage, W. The final term, indicated by (3), is a measure of the rigidities in the labor market. While the variables are determined in general equilibrium, we focus here on the role of specific parameters.

This expression of the markup allows us to explore the role of nominal and real rigidities in the model.[17] We start with the role of nominal price rigidities in the model, denoted by ϕ^P. The sign of $\partial\hat{\mu}_t/\partial\phi^P$ is ambiguous, but note that setting $\phi^P = 0$ makes the markup equal to $-\hat{\Gamma}(\bar{\nu}_t)$. As defined above, $\hat{\Gamma}(\bar{\nu}_t)$ is the percentage change in the proportion of output from firms with idle capacity—this is strictly countercyclical (see equation 2.24). Thus, in the absence of price rigidities, the markup is always procyclical. More generally, lower price rigidities increase the likelihood of procyclical markups because they allow firm prices to reflect capacity considerations.

To explore the role of nominal wage rigidities, note that

$$\frac{\partial\hat{\mu}_t}{\phi^W} = \frac{\phi^P}{(\phi^W)^2}\left(\frac{\epsilon^W - 1}{\Psi^1}\right)(\hat{mrs}_t - \hat{w}_t)$$

which depends on the sign of $\hat{mrs}_t - \hat{w}_t$, that is, the difference between the marginal rate of substitution of the household and the wage rate. This difference is always procyclical after a demand shock when $\phi^W > 0$, since wages cannot update fast enough to match the marginal rate of substitution of the household. Thus, we have the condition $\frac{\partial\hat{\mu}_t}{\phi^W} > 0$, which implies that higher wage rigidities correspond to a higher increase in markups.

At the limit, as $\phi^W \to \infty$, the term ③ disappears, and the direction of the markup depends only on the interplay of nominal intermediate price inflation and real capacity constraints:

$$\hat{\mu}_t = \frac{-\hat{\Gamma}(\bar{\nu}_t)}{\epsilon\Gamma(\bar{\nu}) - 1} - \frac{\phi^P}{\Psi^1}\underbrace{\left\{(\hat{\pi}^p_t - \beta\mathbb{E}_t\hat{\pi}^p_{t+1}) + (\hat{\pi}_t - \beta\mathbb{E}_t\hat{\pi}_{t+1})\right\}}_{\text{Nominal intd. price inflation}} \tag{3.2}$$

Next, we explore the role of real rigidities in the model. On the production side, the key real rigidities are the capacity constraints. Relaxing this in our model is equivalent to setting $\bar{\nu}_t \to \infty$, i.e., firms are never constrained by their capacity. From (2.24) and (2.6), we know that this means that $\tilde{P}_t = \Gamma(\bar{\nu}_t) = 1\ \forall t$, implying $\Psi^1 = \epsilon - 1$. With $\tilde{P}_t$ and $\Gamma(\bar{\nu}_t)$ converging to constants, $\hat{\pi}^p_t = \hat{\Gamma}(\nu_t) = 0$, reducing (3.2) to:

$$\hat{\mu}_t = -\frac{\phi^P}{(\epsilon - 1)}\left[(\hat{\pi}_t - \beta\mathbb{E}_t\hat{\pi}_{t+1})\right]$$

which is, of course, identical to the log-linearized expression for the Phillips curve in the textbook New Keynesian model. In other words, in the absence of capacity constraints, the markup in our model is observationally identical to the New Keynesian model.

Finally, the only real rigidity in the labor market in our model is the monopoly power of workers selling differentiated labor. Specifically, lower ϵ^W is associated with greater monoply power of the households in the labor market. Notice again that $\frac{\partial\hat{\mu}_t}{\epsilon^W} < 0$, implying that higher household labor market power in the labor market implies higher price-cost markup in the product market. This highlights the pass-through of labor market markups into the product market.

Together, the results emerging from Proposition 1 indicate that higher capacity utilization rates at the time of the aggregate shock elicit higher desired markups from firms. Greater labor market imperfections—both nominal and real—allow firms to realize these higher markups, driving aggregate markups to be procyclical. In contrast, product market, *nominal* imperfections work in the opposite direction, making it harder for firms to realize the higher markups.

[17] Note that $\epsilon\Gamma(\bar{\nu}) > 1$ at the steady state, so that the coefficient on the RHS of equation (3.1) is positive. For $\epsilon > 1$ and $\alpha > 0$, this is guaranteed by the steady state relationship $\Gamma(\bar{\nu}) = \frac{1}{\alpha(\epsilon-1)+1}$. See Section 4.2 for more discussion on this.

3.2 Labor Share Cyclicality

Next, we establish a relationship between the markup and the labor share. Unlike in the textbook New Keynesian models, there is no perfectly inverse relationship between the markup and the labor share. Instead, the output loss due to capacity constraints forms a wedge between these two variables. Depending on the degree of loss associated with capacity constraints, the labor share and markup can be either pro- or countercyclical, and they can also both be procyclical. We formalize this finding in the following proposition, and then investigate the conditions under which empirically consistent cyclicality is obtained.

Proposition 2. *The relationship between the labor share and the markup can be given by:*

$$\hat{s}_t = \frac{\Omega}{1-\Omega}\hat{\Omega}(y_t^*) - \hat{\mu}_t \tag{3.3}$$

where all hatted variables again represent log-deviations from the steady state, s_t is the labor share, μ_t is the markup, and $\Omega(y^*) \equiv 1 - \dfrac{P\mathbb{E}_\nu\{Y\}}{\mathbf{PY}}$ is the output loss due to capacity constraints. It can be shown that $\Omega(y^*) \in [0,1]$ and is strictly increasing in the utilization rate y^*.

Proof. See Appendix A.2 for derivation. □

The wedge $\Omega(y^*)$ has an intuitive interpretation. For any period t, the sum of all intermediate input is given by

$$\int_0^1 Y_{j,t} = \mathbb{E}_\nu\{Y_t\}$$

which derives from the ergodicity of the IID idiosyncratic shock, ν. Recall that $\mathbb{E}_\nu\{Y\} > \mathbf{Y}$ and $P < \mathbf{P}$ as long as firms face any capacity constraints (i.e., $\bar{\nu}_t < \infty$). Thus $\Omega(y^*) > 0 \implies \frac{P}{\mathbf{P}}\mathbb{E}_\nu\{Y\} < \mathbf{Y}$. This inequality presents a relationship between inputs (i.e., the output of intermediate firms) and aggregate output. In other words, capacity constraints imply that the real value of inputs is less than the real value of final output. This "distortion" reflects output that is neither captured by the labor share nor by the markup, and $\Omega(y^*)$ is a measure of that output. Note that since y^* is strictly procyclical, $\Omega(y^*)$ strictly increases following an expansionary demand shock—that is, expansionary shocks cause constraints to bind tighter, and the output value loss to be increased. Whereas in the textbook New Keynesian economy the labor share falls by the same amount that markups rise, in our economy with capacity constraints, the loss to labor share is greater than the rise in markups.

Equation (3.3) implies that as the capacity constraints bind more (y^* is higher), the wedge between the labor share and the inverse-markup is increased. When no constraints bind (i.e., $y^* \to 0 \implies \hat{\Omega}(y^*) = 0$), the labor share is exactly equal to the reciprocal of the markup, reflecting the textbook New Keynesian case.[18]

From this expression, we glean two key insights into the joint behavior of the markup and labor share. First, the labor share is always procyclical if the markup is countercyclical. Second, if the markup is procyclical, the cyclical response of labor share depends on relative responses of the output loss and the markup. In other words, the evidence-consistent cyclicality of the markup does not automatically imply the evidence-consistent cyclicality of the labor share.

[18] The perfectly reciprocal case of labor share and the markup emerges in NK models that feature Cobb-Douglas production functions or linear production functions. More complex production technologies and features can introduce a wedge even in the NK model. See Nekarda and Ramey (2020) for additional discussion.

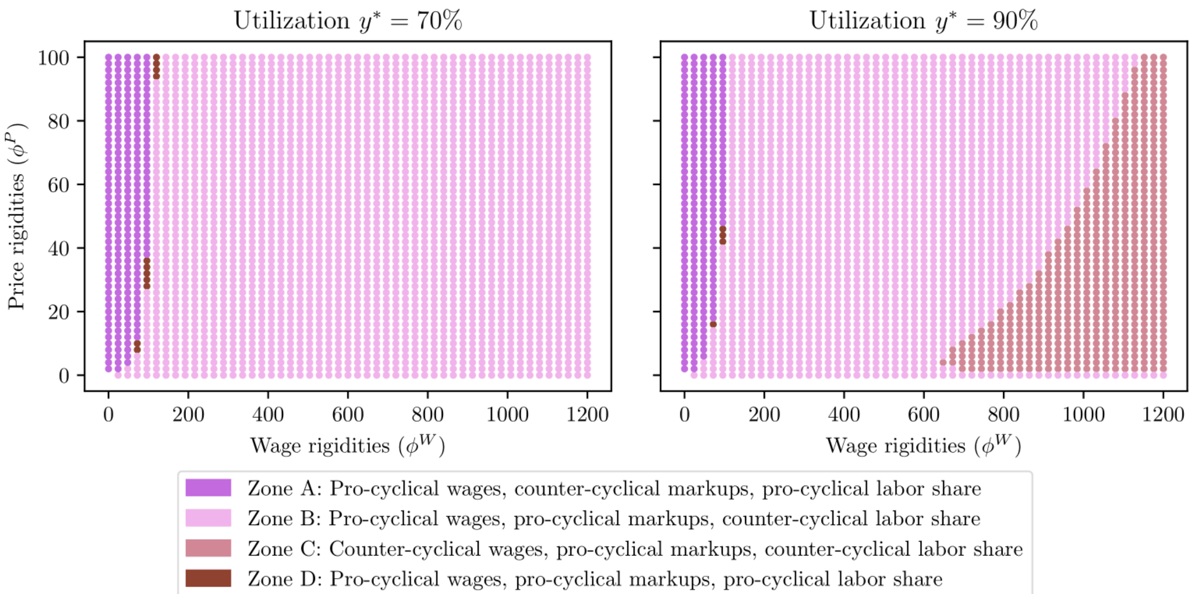


Figure 3: Distributional regimes: cyclicality of real wages, markups, and labor share depending on price rigidities, wage rigidities and capacity utilization rates.

One additional special case relates to when there are no price rigidities, (i.e., $\phi^P = 0$). In this case, the markup is also only a strictly increasing function of $\bar{\nu}$ (see equation 2.24), making the labor share a strictly decreasing function only of $\bar{\nu}$. In this case, all expansionary shocks cause the labor share to decline.

3.3 Distributional Regimes

Based on Propositions 1 and 2, we are in a position to graphically represent the implications of capacity utilization, through its interactions with wage and price rigidities, for the behavior of markups, wages and labor share. We demonstrate this through simulation, where we vary the price and wage rigidity parameters, ϕ^P and ϕ^W while keeping the rest of the parameters the same as in Table 2. We then plot the cyclicality of the labor share and markup in the first period following the shock. We perform this simulation for two different states: when the economy is at a low utilization rate when the shock occurs, and when the utilization rate is high. The results are exhibited in Figure 3.

The result provides a key insight into the role of capacity utilization, price rigidities and labor market rigidities in determining the response of economic variables. Zone A is characterized by a dominance of price rigidities over labor market rigidities. Here, wages are procyclical, but the labor share and the markup exhibit counterfactual behavior. Specifically, the labor share is procyclical and the markup countercyclical. Note that with respect to these variables, the zone is observationally equivalent to a standard textbook New Keynesian model. The second case, Zone B is the zone that aligns with the empirical evidence reviewed above. Wages, the labor share and markups all display the correct cyclical properties. Generally speaking, this zone is characterized by a dominance of labor market rigidities over price rigidities, as is generally the real world case.

Zone C is a zone where the labor share and the markup display the correct cyclicality, but wages are countercyclical.

We find this case particularly intriguing, as it pertains to a case where the rate of inflation outstrips the rate of wage growth, such that real wages fall. The evidence in Autor et al. (2023), where high inflation following the COVID pandemic caused real wages to fall for a large section of the wage distribution, could represent a scenario where the economy may have found itself in Zone C.[19] Note additionally that the probability of Zone C is more likely when capacity constraints bind more aggressively, as can be seen in the right panel. Generally speaking, when capacity utilization is high, a higher degree of nominal price rigidity can still deliver procyclical markups than when utilization rates are low.[20]

A striking aspect of Figure 3 is the dominance of the Zone B for the presented values of ϕ^P and ϕ^W, which encompass typical values in the literature. Recall that the features of Zone B—procyclical wages, procyclical markups and countercyclical labor share—cannot be generated in a standard New Keynesian model. However, Figure 3, which is generated using parameter values estimated from the data (see Table 2), suggests that these are the features that typically prevail under normal economic conditions. The typical responses seen in New Keynesian models (Zone A) constitute a relatively small portion of the possible outcome space. A second interesting element, though less visually apparent, is that this New Keynesian zone diminishes when utilization rates are higher. This once again highlights the importance of capacity utilization as a driver of markup and labor share cyclicality, and reinforces the points highlighted in Propositions 1 and 2.

3.4 Hump-shaped Inflation Response

In this section, we establish the conditions for the a hump-shaped response to inflation. Intuitively, an expansionary demand shock has two opposing effects on inflation. The first effect is upward pressure on inflation: as in the standard NK model, the expansionary shock causes an increase in real wages and inflationary expectations, which directly raises inflation. Additionally, in our model with capacity constraints, firms increase their desired markup in response to their increased probability of hitting their production capacity, and this, too, pushes prices upwards. The second effect is downward pressure on inflation: as demand increases, firms first meet their output goals by relying on the intensive margin of labor (extracting more effort from workers), which, as we discussed above, raises the productivity of labor. These productivity effects of demand shocks are strongest early in the cycle, and can mitigate inflationary pressures substantially, particularly when capacity utilization rates are low. As long as the productivity effect dominates the wage and markup effects, inflation will remain subdued. Indeed, for appropriate parameterizations, inflation can even *fall* in the immediate aftermath of an expansionary demand shock, providing an explanation for the so-called "price puzzle". In the section below, we provide analytical results demonstrating these effects.

Proposition 3. *The Phillips curve based on our model can be written as follows:*

$$\hat{\pi}_t = \Psi \hat{mc}_t + \beta \mathbb{E}\hat{\pi}_{t+1} + G(\mathbb{E}\hat{y}^*_{t+1}, \hat{y}^*_t, \hat{y}^*_{t-1}) \tag{3.4}$$

where $\Psi = \dfrac{\Psi^1}{\phi^p}$ *is a constant, and the function* $G(\cdot)$ *represents the effect of capacity utilization rate,* y^**, on the rate of inflation.*

[19] To be precise, the evidence in Autor et al. (2023) shows that real wages grew for low-wage workers but declined for the median worker in the period following the COVID pandemic.

[20] We find an additional edge case of a Zone D, which is characterized by a narrow zone where wages and the markup display appropriate behavior, but the labor-share is procyclical.

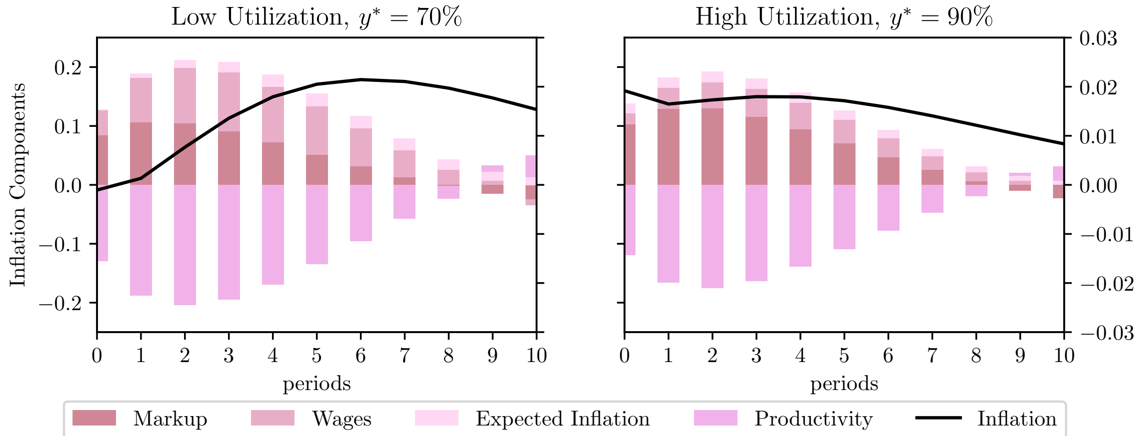


Figure 4: The decomposition of inflation (*scale on the right axis*) into the component driving parts (*scale on the left axis*) over the cycle. Note that when capacity utilization is low (left panel), the productivity effects are large and dominate the upward pressures (markups and wages), resulting in a distinct "price-puzzle" effect. When utilization is higher (right panel), markups are the main source of upward pressures, which dominate the mitigating effects of productivity. Inflation increases on impact.

Proof. See Appendix A.3 for derivation. □

Proposition 3 tells us that the realized inflation rate depends on the response of marginal costs, expected inflation and on current and expected capacity utilization. Crucially, notice that the value of $G(\cdot)$ depends on y^*_{t-1}, which is a state variable. Thus, the Phillips curve in our model is state-dependent. In other words, following an expansionary demand shock, the magnitude and sign of the inflation response depends on the utilization rate in the economy when the shock occurs. This has implications for whether the disinflationary productivity effects or the inflationary wage and markup effects dominate. Intuitively, when y^*_{t-1} is low, the productivity effects dominate because firms are able to satisfy the additional demand with the available excess capacity more easily. Additionally, the probability of firms hitting their capacity constraint rises, but less so. In contrast, when utilization is high at the time of the shock, firms have little headroom to expand output to meet demand. Thus, the adjustment occurs through prices, which rise faster as reflected in higher markups. This mechanism aligns with the arguments in Boehm and Pandalai-Nayar (2022) regarding the reasons for the convexity of the supply curve.

To further drive intuition, we restate the Phillips curve to explicate the role of markups, wages, productivity and expectations. We use the log-linearized definition of the markup, $\hat{\mu}_t = \hat{\tilde{p}}_t - \hat{mc}_t$ and marginal cost, $\hat{mc}_t = \hat{w}_t - \hat{a}_t - \hat{\tilde{\xi}}_t$ to rewrite equation 3.4 as

$$\hat{\pi}_t = \tilde{\Psi}\hat{w}_t + \mu_t + \beta\mathbb{E}\hat{\pi}_{t+1} + H(\hat{a}_t, \mathbb{E}\hat{y}^*_{t+1}, \hat{y}^*_t, \hat{y}^*_{t-1}) \tag{3.5}$$

Here, $\tilde{\Psi}$ is a constant, and $H(\cdot)$ is a function that captures the productivity effects in the model. This function is also state-dependent whose value depends on the marginal productivity of labor determined through the combination of K and N seen in equation (2.8), and the utilization rate in the economy.

Notice that a representation similar to equation (3.5), where inflation is decomposed into the contribution of wages and markups, is not possible in a textbook New Keynesian model. This is because in a textbook model with Rotemberg price rigidities, the price chosen by the intermediate firms (P in our model) is the same in equilibrium as the aggregate price level ($\mathbf{P}$ in our model). Since $P/\mathbf{P} = 1$, the markup in the standard model is the same as the inverse of the marginal cost, which is a function of wages. Given such a set-up, the standard linearized Phillips Curve can only show the relationship between the change in price (inflation) and the marginal cost, or—equivalently—the negative markup.

In contrast, our model distinguishes between the intermediate good price and the aggregate price level. The relationship is given in equation (2.6). This allows us to separate the changes in the aggregate price level—reflected in inflation—from changes to the markup and wages, which reflect the intermediate good's price.

In Figure 4, we provide a decomposition of inflation based on the components in the Phillips curve as stated in equation (3.5). We provide the same results for two different parameterizations of capacity utilization rates, while keeping the rest of the economy at the baseline parameterization.[21] At a 70% utilization rate, we see that inflation falls on impact, generating the "price puzzle". This is due to the fact that although markups and wages are both rising, the productivity effects of utilizing capacity more effectively, captured in $H(\cdot)$, dominate. This mitigating effect drives inflation down. At a higher 90% rate of capacity utilization, we see that inflation is driven primarily by the sharp rise in markups. Real wage inflation plays a relatively small role; indeed, later in the cycle, real wages even fall. The hump-shaped response of inflation is still visible, however, as productivity effects are larger earlier in the cycle.

It can be shown that when capacity utilization tends to zero such that firms are completely unconstrained by capacity considerations, the function $G(\cdot) \to 0$, and we recover the standard New Keynesian Phillips curve[22]

$$\hat{\pi}_t = \Psi \hat{mc}_t + \beta \mathbb{E} \hat{\pi}_{t+1}$$

Despite the functional similarity between the standard New Keynesian Phillips curve and our formulation in equation (3.4), the two curves contain important differences in the specification of the marginal cost. In the canonical NK framework, the marginal cost is increasing in output. This emerges from the CES (usually Cobb-Douglas) specification of the production function which features diminishing marginal returns to individual factors.[23] Our model, however, features a distinction between the marginal cost of *production* and the marginal cost of *capacity*. Whereas the marginal cost of increasing capacity is always increasing for a firm, the marginal cost of increasing *production* is decreasing in output until the capacity limit.

We illustrate this theoretical mechanism in Figure 5. Assuming the wage to be given in the period, the marginal cost of production falls until the level of installed capacity (Cap_1 and Cap_2 in the figure). This is because firms utilize their installed capacity more effectively, manifesting as higher productivity of their workers. At capacity, the marginal cost reaches its minimum. Output cannot be raised beyond capacity—the marginal cost of output is theoretically infinite beyond the installed level of capacity. Raising *capacity*, however, invites diminishing returns to the individual factors of capacity creation, K and N. This is reflected in the upward sloping marginal cost of capacity.

[21] The parameterization is based on the results from our Bayesian IRF matching exercise, which are provided in Table 2.

[22] See, for example, equation (24) in Galí and Gertler (1999). Alternative, and equivalent, representations of the Phillips curve focus on the (log deviations of) output gap $\tilde{y}_t \equiv y^f - y$ or the markup directly. See equation (22) and (17), and Chapter 3, footnote 4 in Galí (2015) for a discussion of the equivalence.

[23] In medium-scale extensions of the NK model, variable capital (not capacity) utilization, the "working capital channel" (a la Ravenna and Walsh, 2006) and other modifications attempt to ameliorate the increase in the marginal cost.

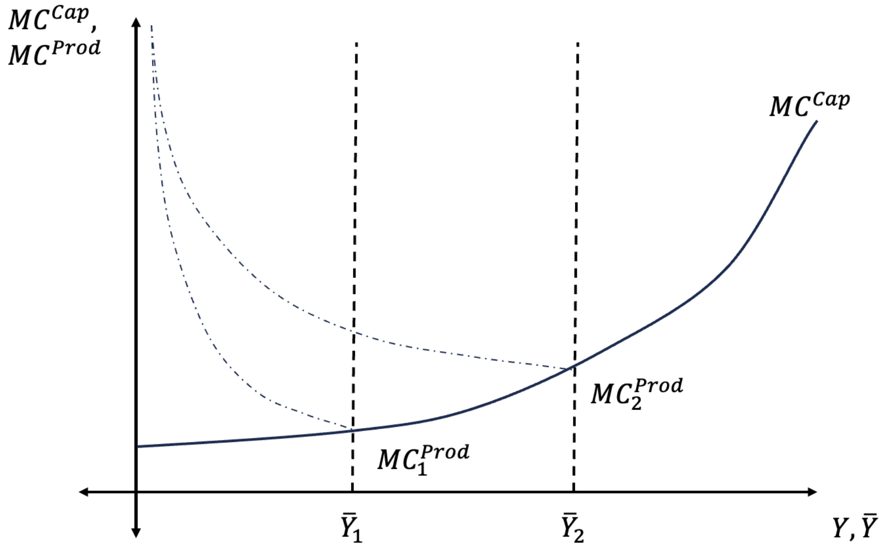


Figure 5: Marginal cost of production versus marginal cost of capacity. The x-axis plots output (Y) and capacity ($\bar{Y}$). The y-axis plots marginal costs. The marginal cost of raising output, MC^{Prod}, is downward sloping, given some level of capacity. The marginal cost of raising capacity, MC^{Cap}, is upward sloping.

4 Estimation and Quantitative Analysis

We now proceed to identify parameter values for which the model produces typical aggregate responses found in the empirical literature. We find these parameters using Bayesian impulse-response function (IRF) matching, following the empirical methodology and the VAR results presented in Christiano et al. (2010) ("CTW" hereafter). Under this approach, our parameters of interest are identified such that the distance between the empirical impulse responses and our DSGE model is minimized. We center our empirical approach around CTW because the VAR results presented in this study have been used in several other papers (Christiano et al., 2016, 2021 etc.). Other studies have used the models presented in CTW as baseline results against which new results are compared (Cantore et al., 2020; Qiu and Ríos-Rull, 2022 etc.). Staying close to CTW allows us to immediately compare our results and contribution across models and studies.

4.1 Approach

The empirical impulse responses are obtained from the estimation of a two-lag VAR using seasonally adjusted quarterly data over the period 1951Q1 and 2008Q4. The estimation strategy we employ focuses on the IRFs of 8 of the 14 variables included in the CTW's VAR exercise.[24] We stack the contemporaneous and 14 lagged values of each of these IRFs in a 120×1 vector. Following CTW, we excise the contemporaneous responses of variables from the matching vector for all variables except the Federal Funds Rate, since the VAR estimation strategy requires these to be zero in the empirical IRFs. Thus, our matching vector, denoted by $\hat{\Psi}$, has 113 elements.

The estimation procedure follows CTW closely, which we summarize briefly here. Let the parameters of the model

[24] We drop the variables which do not have a corresponding object in our model. These are the variables for vacancies, job finding rate, job separation rate, unemployment rate, and relative price of investment.

be denoted by the vector θ, and the associated model impulse responses by $\Psi(\theta)$. If the true values of the model parameters is given by θ_0, then the values in $\hat{\Psi}$ correspond to estimates of the values reflected in $\Psi(\theta_0)$. When the number of observations T is large, asymptotic sampling theory implies

$$\sqrt{T}(\hat{\Psi} - \Psi(\theta_0)) \overset{a}{\sim} N(0, W(\theta_0, \zeta_0))$$

Here, ζ_0 denotes the true values of the parameters of the shocks of the model. In our case, $\Psi(\theta_0)$ is independent of ζ_0 since we solve our model using a first-order perturbation solution method, but, as Christiano et al. (2016) note, the sampling distribution of $\hat{\Psi}$ is still a function of ζ_0. The asymptotic distribution of $\hat{\Psi}$ is given by

$$\hat{\Psi} \overset{a}{\sim} N(\Psi(\theta_0), V) \tag{4.1}$$

$$\text{where } \; V = W(\theta_0, \zeta_0)/T \tag{4.2}$$

Following CTW's method, $\hat{\Psi}$ is treated as "data". We seek to identify parameters such that $\Psi(\theta)$ is close to $\hat{\Psi}$. This is done by specifying priors for θ and computing the posterior distribution for θ using Bayes rule. Since (4.1) indicates our "data" is normally distributed, the likelihood of $\hat{\Psi}$ given θ can be written as:

$$f(\hat{\Psi}|\theta, V) = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\hat{\Psi} - \Psi(\theta))' V^{-1}(\hat{\Psi} - \Psi(\theta))\right\}$$

We seek to obtain the value of θ that maximizes the above likelihood. This would represent a maximum likelihood estimator of θ, but as CTW show, it is only an *approximate* one because (1) the central limit theorem underlying (4.1) holds only when $T \to \infty$, (2) because the estimate of V holds only when $T \to \infty$, and (3) the IRFs are based on a linearized model solution, which represents a further approximation.[25]

4.2 IRF Matching Considerations

We begin by highlighting three considerations related to parameter estimation.

First, in order to further align our model with CTW, we extend our model to include some additional features present in their model. These features include a "working capital channel" reflecting the need of firms to finance their wage bill. Correspondingly, the wage bill facing the firm is given by

$$\bar{W}_t = (1 - \iota + \iota R_t) W_t$$

where $\iota \in [0, 1]$ is the share of the wage bill that is borrowed by the firm. CTW assume that $\iota = 1$, implying that the entire wage bill is borrowed. This assumption of "full working capital" channel is standard in the literature. However, empirical evidence suggests that this somewhat lower, at about 0.76 with large variation across industries (Galindo Gil, 2021). We test the full channel assumption, and leave ι to be estimated by the model.

An additional extension is the introduction of long-run growth such that the economy is on a balanced growth path. We include this to align our model with the assumptions adopted in the VAR estimation methodology in CTW. However, unlike CTW who have two sources of growth in their model, we can only accommodate labour-augmenting technological growth. This is due to our choice of a CES capacity function. As Jones (2005) and others

[25] We follow CTW in using a diagonal matrix as a consistent estimator of the matrix V. In practice, this amounts to selecting a θ such that the model implied IRFs lie within the confidence tunnel around the point estimates in $\hat{\Psi}$.

demonstrate, a balanced growth path requires either a Cobb–Douglas technology or technical change that is purely labor-augmenting.

A second matter for consideration relates to certain parameters and steady-state variables whose values are jointly determined due to cross-restrictions implied by the model equations. The first such set relates to the normalization of the CES capacity function following the "re-parameterization" procedure outlined in Cantore et al. (2015). Specifically, the distributional parameters α_K and α_N are given by:

$$\alpha_K = \alpha \left(\frac{\bar{Y}}{K}\right)^{\psi} \qquad\qquad \alpha_N = (1-\alpha)\left(\frac{\bar{Y}}{N}\right)^{\psi}$$

where $\bar{Y}$, K and N are steady-state values of capacity, capital and planned labor respectively. Therefore, our estimation procedure focuses on specifying α, capital's share of *capacity*.

Next, the steady state rate of capacity utilization y^*_{ss} and the steady state labor share s_{ss} have important implications for parameter estimation due to the state-dependent nature of the model. Both y^* and s depend only on $\bar{\nu}$ in the steady-state. In turn, $\bar{\nu}$ in the steady-state is fully determined by α, ϵ and the distribution function $F(\nu)$. The latter is assumed to be a unit-mean lognormal process, whose variance σ_ν is the sole parameter to be specified. As in Fagnart et al. (1999), we evaluate (2.28) at the steady state and combine it with (2.24) to obtain the following relationship:

$$\Gamma(\bar{\nu})_{ss} = \frac{\int_0^{\bar{\nu}_{ss}} \nu \, dF(\nu)}{\int_0^{\bar{\nu}_{ss}} \nu \, dF(\nu) + \bar{\nu}_{ss} \int_{\bar{\nu}_{ss}}^{\infty} \, dF(\nu)} = \frac{1}{\alpha(\epsilon-1)+1}$$

Specifying values for y^*_{ss} and s_{ss} therefore fixes the values of the parameters σ_ν and α, and the steady-state value of $\bar{\nu}$. However, while this may ensure that the parameters are consistent with y^*_{ss} and s_{ss}, the resulting model *dynamics* may not be. This issue emerges directly from the fact that the dynamics of our model are state-dependent. That is, the variables respond to shocks differently depending on whether y^*_{ss} is low or high at the time of the shock. To align the *dynamics* of the model with the values chosen at the steady state, we allow for the utilization rate experienced by the firms to differ from the aggregate utilization rate indicated by our chosen y^*_{ss}. We do this by extending equation (2.5) as follows:

$$y^*_t = \frac{\mathbf{Y}_t}{\bar{Y}_t} = \left[\left(\frac{1}{\bar{\nu}_t}\right)^{\frac{\epsilon-1}{\epsilon}} \int_0^{\bar{\nu}_t} \nu_t \, dF(\nu) + \int_{\bar{\nu}_t}^{\infty} \nu_t^{\frac{1}{\epsilon}} \, dF(\nu)\right]^{\frac{\epsilon}{\epsilon-1}} + \tilde{m}$$

where $\tilde{m}$ is a fixed parameter. Rewriting this compactly gives us:

$$\mathbf{Y}_t = \mathbf{Y}(\bar{Y}_t, F(\nu)) + \tilde{m}$$

In other words, we assume that aggregate output is a function of capacity installed and the distribution of idiosyncratic shocks $F(\nu)$, plus an exogenous quantity of output not determined by the model's capacity dynamics, $\tilde{m}$.[26] The parameter $\tilde{m}$ operates in a fashion similar to the fixed cost parameter in standard models. That is, it dampens the effect of the utilization rate on inflation and other dynamics. We leave $\tilde{m}$ to be estimated by the data.

One final matter for consideration is the empirical response of capacity utilization in CTW's study. The VAR exercise in CTW uses capacity utilization data for the manufacturing industry only, since this data is not collected or published at a national aggregate level. This poses a problem for our IRF matching exercise because while other

[26] For intuition, if $\mathbf{Y}_t$ represents GDP, $\tilde{m}$ can be thought of as imports or government production whose hiring and pricing dynamics do not depend on domestic capacity constraints directly.

Table 1: Non-estimated Parameters

Parameter		Description	Parameter		Description
β	0.993	discount factor	G^{ss}	0.12	govt. consumption to GDP
δ	0.025	depreciation rate	ϱ	0.05	exogenous separation rate
ϵ^W	6.0	elast. of sub. between labor varieties	ω	1.0	weight on disutility of labor
y^*_{ss}	0.80	steady-state capacity utilization	s_{ss}	0.62	steady-state labor share
γ	1.0059	gross balanced growth rate	$\bar{\pi}$	1.0083	gross inflation rate

variables (such as output, employment etc.) are matched to economy-wide aggregate responses, utilization will be matched to the manufacturing industry's response only.

Given our emphasis on the importance of capacity utilization for aggregate dynamics, we are keen to demonstrate the ability of the model to match the VAR dynamics. We therefore additionally estimate an expanded model where the manufacturing industry is modeled separately from the rest of the economy. To do so, we split the production sector into a manufacturing industry (denoted with an M super-script) and a non-manufacturing industry (denoted with an S super-script). Each industry is structurally identical to the production sector described in Section 2.2, with an industry-specific aggregating firm as in Section 2.1. Thus

$$\mathbf{Y}^i_t = \left[\int_0^1 \left(Y^i_{t,j} \right)^{\frac{\epsilon^i - 1}{\epsilon^i}} \nu_{t,j}^{\frac{1}{\epsilon^i}} dj \right]^{\frac{\epsilon^i}{\epsilon^i - 1}}$$

where the final firm's maximization yields

$$Y^i_j = \left(\frac{P^i_{t,j}}{\mathbf{P}^i_t} \right)^{-\epsilon^i} Y^i_t$$

Here, P^i_j and $\mathbf{P}^i$ refer to the j^{th} intermediate firm's price and industry price for the i^{th} industry for $i \in M, S$. The output of the two industries, Y^M and Y^S, are then aggregated by a final firm that aggregates the output of the two firms into a final good, Y^F, that is consumed by households and purchased for investment by intermediate firms.

$$Y^F_t = \left(\alpha_M (Y^M_t)^{\psi^F} + \alpha_S (Y^S_t)^{\psi^F} \right)^{\frac{1}{\psi^F}}$$

The enhancement described here allows us to preserve a unified household sector with a single wage facing all firms. This helps retain intuition gained from our explorations in section 3 by deviating minimally from the single-industry model described earlier. The full set of equilibrium equations for this extension, along with additional discussion related to estimating it, is presented in Appendix B.

4.3 Estimation Results

As is standard in the estimation literature, a subset of the parameters of the model are set *a priori*. These parameters are given in Table 1. The values for the balanced path growth rate and the steady state gross inflation rate are taken from CTW. The steady-state utilization rate, y^*_{ss}, reflects the approximate average for the manufacturing industry

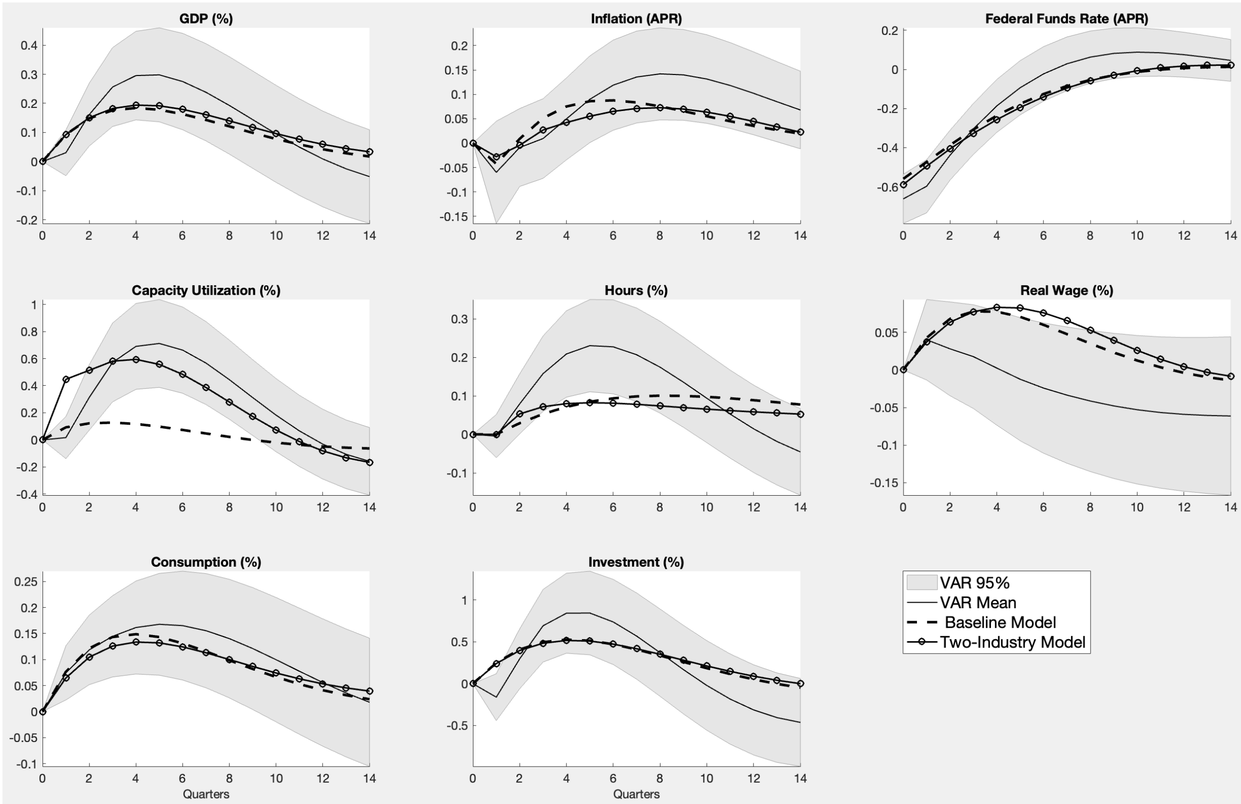


Figure 6: Impulse Responses from Bayesian IRF Matching.

only. The steady-state labor share, s_{ss} is the average aggregate labor share as in the data. The rest of the parameters are standard in the literature.

The estimated parameter values for the baseline model is presented in Table 2, along with details related to the priors posited for each parameter.[27] We highlight a few interesting points. First, the nominal price rigidity values are smaller than typical estimates in the NK literature. To illustrate, based on a standard NK Phillips curve, the Rotemberg adjustment parameter value of 9.2 corresponds to prices changing on average about every 1.7 quarters, while a more generally accepted value is about 4 quarters.[28] This low value of ϕ^P should not however be interpreted as the measure of price rigidity as in standard NK models. Recall that capacity constraints in our model impose nonlinear effects on firms' price setting. That is, firms are less inclined to raise prices when excess capacity is high and vice-versa. Once the effect of this *real* price rigidity is accounted for, the remaining nominal rigidity is accordingly small. Thus, observed prices in our model will exhibit more rigidity than the low value for ϕ^P suggests.

Second, the value for ι (the extent of the working capital channel) we obtain is substantially smaller than 1.0 as widely assumed in the literature. A high value for ι mitigates the increase in marginal costs facing the firm in the immediate aftermath of an expansionary monetary policy shock, and helps generate an inertial, hump-shaped

[27] The top panel of Table 2 displays the parameters which were directly estimated, while the bottom panel displays the parameters whose values are themselves functions of the steady-state values and estimated parameters. These latter values are evaluated at the posterior mean of the estimated parameters.

[28] Ascari and Rossi (2012) provide conditions establishing the equality of Calvo and Rotemberg parameters in NK Phillips curves, under specific assumptions.

inflation response (Christiano et al., 2005; Phaneuf et al., 2018). While studies like Barth and Ramey (2001), Ravenna and Walsh (2006) and others provide evidence for the presence of some degree of a working capital cost channel, the empirical evidence suggests that the assumption of a "full working capital channel" that transmits fully and immediately to the private sector is not justified (Galindo Gil, 2021; Ippolito et al., 2017). Additionally, inflation exhibits an inertial response even following demand shocks where the interest rate does not explicitly decline. For example, Jørgensen and Ravn (2022) find evidence for a "fiscal price puzzle", whereby inflation *falls* following an expansionary *fiscal* shock. Thus, finding an ι that is less than 1.0 implies that our model generates a mechanism for inertial inflation and price -puzzle that is not tied to the mechanical effect of a working capital channel alone. As discussed in Section 3.4 above, the driving feature in our model is the procyclical productivity effect due to the increased use of installed capacity.

Third, the elasticity of substitution between capital and labor in the firm's capacity function, σ, is substantially lower than 1.0, implying that labor and capital are strongly complementary, rather than substitutes. This is in line with a large literature that estimates σ and finds it to be less than 1.0. For example, Chirinko and Mallick (2017) finds a σ under putty-clay assumptions of about 0.19, which is close to our own estimate of 0.13.[29] The low substitutability between labor and capital plays a key role in the dynamics of markups in our model, because it makes it harder for firms to add capacity following expansionary demand shocks. This leads capacity to remain constrained for longer, generating more persistently elevated markups in our model.[30]

The remaining model parameter estimates are largely in line with the data and broader literature, although the Rotemberg wage rigidity parameter is somewhat higher than typical values for this parameter. This partly reflects the fact that we have focused our innovations on the production sector while retaining as standard an NK set-up as feasible.

We present the model's impulse responses based on the results of the IRF-matching exercise in Figure 6. Alongside our baseline results, we present the results of our two-industry extension also, to demonstrate the ability of the model to match the response of manufacturing capacity utilization. The VAR mean is indicated by the solid lines, while the grey shaded area represents the 95% confidence bands. The dashed lines are the impulse responses of the baseline model. The lines with circle markers correspond to the two-industry model.

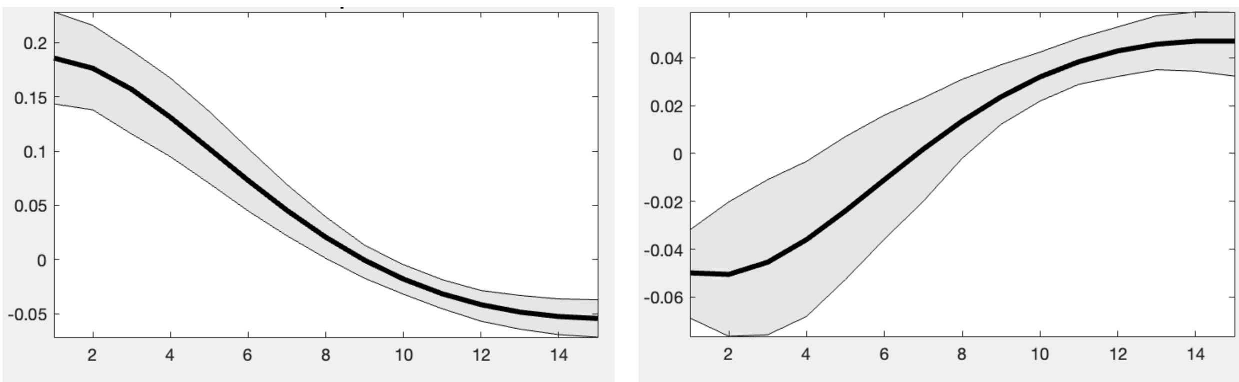
Overall, the model fits the empirical responses relatively well. In most cases, the model IRFs are situated inside the 95% confidence bands. Additionally, the shape of the model's responses closely match the behavior of the VAR mean. This is particularly true for the IRFs for GDP, inflation, the Federal Funds rate, the real wage, consumption and investment.

As noted above, the IRF for inflation captures the price puzzle phenomenon very closely. This is particularly interesting because it does not rely on standard features like the "full working capital channel" or backward-indexation of inflation.

However, the impulse response of the capacity utilization rate in the baseline model does not match the data very well. As discussed above, this is due to the fact that the empirical IRF corresponds to utilization in the manufacturing industry only. The extended two-industry model, however, replicates the response of capacity utilization very well without sacrificing the responses of the other variables. We find this particularly encouraging because the two-industry

[29]Chirinko and Mallick (2017) report a benchmark "long-run" σ of ≈ 0.41, which is still well below unity, in line with our finding. See Knoblach and Stöckl (2020) for a comprehensive recent survey.

[30]See Dolado et al. (2021) for additional discussion on the demand amplification role of capital-labor complementarity.



(a) Model-implied markup

(b) Model-implied labor share

Figure 7: Impulse responses of model implied markup and labor share. The gray shaded areas provide highest posterior density intervals (5% and 95%) based on 1000 draws from the posterior distribution.

extension we design deviates minimally from the model presented in Section 2.

In Figure 7, we present the Bayesian IRFs for the markup and labor share based on the baseline estimated parameters. Both variables exhibit the correct cyclicality conditional on an expansionary monetary policy shock, which we find to be particularly encouraging since neither variable was included in the matching exercise.

5 Concluding Remarks

In this paper, we present a model where firms' choice of capacity and its utilization are incorporated within a New Keynesian framework. Firms face idiosyncratic demand uncertainty and respond by holding additional precautionary capacity. Following an expansionary monetary policy shock, increased demand is met both through higher utilization of existing capacity and through capacity expansion. To achieve higher utilization, firms rely on an "effort margin" of their workers, aligning with recent evidence on effort cyclicality. In addition, firms respond to the decline in excess capacity by raising their markups, which has a positive impact on prices. Meanwhile, capacity expansion occurs through increased investment and hiring, which mitigates the effect of demand on prices over time.

We suggest that our mechanism helps reconcile the dynamics of labor share, markups, and inflation in New Keynesian models with recent empirical evidence. Specifically, in our model, the labor share responds countercyclically to a monetary policy shock, markups typically display a procyclical response, and inflation manifests either as a hump-shaped pattern or an immediate response, contingent upon the state of the macroeconomy. We offer analytical results outlining the conditions under which these outcomes occur. We then estimate the model parameters using Bayesian IRF matching and show that these conditions hold in the data.

Our results emphasize the importance of capacity utilization for the macroeconomic debate surrounding inflation, inequality, and productivity, potentially carrying significant policy implications. We believe that at least two policy implications are of particular importance. Firstly, our analysis reveals an additional trade-off for central bankers to consider when formulating policy. We show that markups can exert a significant influence on inflation, particularly during periods of heightened capacity utilization. In such scenarios, while raising interest rates may dampen inflation by discouraging demand, it could also slow down capacity expansion, thereby prolonging higher markups and inflation.

Secondly, since our model indicates that the responses of markups, wages, and labor share—variables associated with the cyclical behavior of inequality—are state-dependent, we can utilize this framework to design taxes and transfers over the economic cycle. In this regard, a version of the model incorporating heterogeneous agents could offer valuable insights for fiscal policy design. In forthcoming and future research, we aim to delve into these topics.

Table 2: Estimated (top panel) and model-implied (bottom panel) parameters for baseline model.

		Prior			Posterior		
		Mean	Dist.	St. Dev	Mean	90% HPD interval	
σ	K-N elasticity of sub.	0.1	Gamm	0.08	0.1348	0.0293	0.237
ϕ^K	Invest. adj. costs	8	Gamm	2	7.8078	5.6085	9.9569
ϕ^P	Rotemberg price adj.	8	Gamm	2	9.2362	6.029	12.3906
ϕ^H	Labor adj. costs	0.01	Gamm	0.005	0.005	0.0017	0.0081
ι	Working capital pct.	0.7	Beta	0.1	0.6124	0.4319	0.7896
ϵ	Intd. varieties elast. of sub.	8	Gamm	1	8.4294	6.7772	10.0735
$\tilde{m}$	Non-capacity output pct.	0.2	Beta	0.15	0.1591	0.1248	0.192
ϕ^W	Rotemberg wage adj.	800	Gamm	100	912.4364	749.4163	1075.8
φ	Inverse frisch elast.	1.5	Gamm	1	1.9075	0.7464	3.0413
h	Habits in consumption	0.8	Beta	0.1	0.8149	0.7883	0.8401
ρ_S	Taylor rule smoothing	0.86	Beta	0.1	0.856	0.8238	0.8864
ρ_π	Taylor rule inflation	1.8	Gamm	0.25	1.744	1.3449	2.1474
ρ_Y	Taylor rule output	0.02	Norm	0.05	0.0493	0.001	0.097
Model-implied parameters: Evaluated at posterior mean of estimated parameters							
σ_ν	Var. of idiosyncratic shock	0.82					
α	K share of capacity	0.13					
α_K	K dist. param.	1506.9					
α_N	L dist. param.	0.18					
ψ	K-L sub. param.	-13.4					

References

Adam, K. (2005). Learning to forecast and cyclical behavior of output and inflation. *Macroeconomic Dynamics*, 9(1):1–27.

Àlvarez-Lois, P. P. (2006). Endogenous capacity utilization and macroeconomic persistence. *Journal of Monetary Economics*, 53(8):2213–2237.

Anderson, E., Rebelo, S., and Wong, A. (2018). Markups across space and time. Technical report, National Bureau of Economic Research.

Ascari, G. and Rossi, L. (2012). Trend inflation and firms price-setting: Rotemberg versus calvo. *The Economic Journal*, 122(563):1115–1141.

Auerbach, A. J., Gorodnichenko, Y., and Murphy, D. (2023). Macroeconomic frameworks: Reconciling evidence and model predictions from demand shocks. *American Economic Journal: Macroeconomics*.

Autor, D., Dube, A., and McGrew, A. (2023). The unexpected compression: Competition at work in the low wage labor market. Technical report, National Bureau of Economic Research.

Barth, M. J. and Ramey, V. A. (2001). The cost channel of monetary transmission. *NBER Macroeconomics Annual*, 16:199–240.

Basu, S. (1996). Procyclical productivity: Increasing returns or cyclical utilization? *The Quarterly Journal of Economics*, 111(3):719–751.

Basu, S. and Fernald, J. (2001). Why is productivity procyclical? why do we care? In *New developments in productivity analysis*, pages 225–302. University of Chicago Press.

Basu, S. and Kimball, M. S. (1997). Cyclical productivity with unobserved input variation. Working Paper 5915, National Bureau of Economic Research. DOI: 10.3386/w5915.

Benigno, P. and Eggertsson, G. B. (2023). It's baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve. Technical report, National Bureau of Economic Research.

Bilbiie, F. O. and Känzig, D. R. (2023). Greed? Profits, Inflation, and Aggregate Demand. Working Paper 31618, National Bureau of Economic Research. DOI: 10.3386/w31618.

Bils, M. and Cho, J.-O. (1994). Cyclical factor utilization. *Journal of Monetary Economics*, 33(2):319–354.

Blanchard, O. and Galí, J. (2010). Labor markets and monetary policy: A new keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2(2):1–30.

Blecker, R. A. and Setterfield, M. (2019). *Heterodox Macroeconomics: Models of Demand, Distribution and Growth*. Edward Elgar Publishing.

Boehm, C. E. and Pandalai-Nayar, N. (2022). Convex supply curves. *American Economic Review*, 112(12):3941–69.

Broer, T., Harbo Hansen, N.-J., Krusell, P., and Öberg, E. (2020). The New Keynesian Transmission Mechanism: A Heterogeneous-Agent Perspective. *The Review of Economic Studies*, 87(1):77–101.

Burnside, C., Eichenbaum, M., and Rebelo, S. (1993). Labor hoarding and the business cycle. *Journal of Political Economy*, 101(2):245–273.

Cantore, C., Ferroni, F., and León-Ledesma, M. (2020). The Missing Link: Monetary Policy and The Labor Share. *Journal of the European Economic Association*, 19(3):1592–1620. jvaa034.

Cantore, C., Levine, P., Pearlman, J., and Yang, B. (2015). Ces technology and business cycle fluctuations. *Journal of Economic Dynamics and Control*, 61:133–151.

Chirinko, R. S. and Mallick, D. (2017). The substitution elasticity, factor shares, and the low-frequency panel model. *American Economic Journal: Macroeconomics*, 9(4):225–253.

Christiano, L., Eichenbaum, M., and Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1997). Sticky price and limited participation models of money: A comparison. *European Economic Review*, 41(6):1201–1249.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84(4):1523–1569.

Christiano, L. J., Trabandt, M., and Walentin, K. (2010). *DSGE Models for Monetary Policy Analysis*, volume 3, page 285–367. Elsevier.

Christiano, L. J., Trabandt, M., and Walentin, K. (2021). Involuntary unemployment and the business cycle. *Review of Economic Dynamics*, 39:26–54.

Comin, D. A., Johnson, R. C., and Jones, C. J. (2023). Supply chain constraints and inflation. Working Paper 31179, National Bureau of Economic Research. DOI: 10.3386/w31179.

Corrado, C. and Mattey, J. (1997). Capacity utilization. *Journal of Economic Perspectives*, 11(1):151–167.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American economic review*, 67(3):297–308.

Dolado, J. J., Motyovszki, G., and Pappa, E. (2021). Monetary policy and inequality under labor market frictions and capital-skill complementarity. *American Economic Journal: Macroeconomics*, 13(2):292–332.

Dossche, M., Gazzani, A., and Lewis, V. (2023). Labor adjustment and productivity in the oecd. *Review of Economic Dynamics*, 47:111–130.

Eichenbaum, M. (1992). 'comments on 'interpreting the time series facts: The effects of monetary policy'by christopher sims,". *European Economic Review*, 36(5):1001–1011.

European Central Bank Monthly Bulletin (2014). Output, demand and the labour market. Technical report, European Central Bank.

Fagnart, J.-F., Licandro, O., and Portier, F. (1999). Firm heterogeneity, capacity utilization, and the business cycle. *Review of Economic Dynamics*, 2(2):433–455.

Galesi, A. and Rachedi, O. (2018). Services deepening and the transmission of monetary policy. *Journal of the European Economic Association*, 17(4):1261–1293.

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 edition.

Galí, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222.

Galí, J. and Van Rens, T. (2021). The vanishing procyclicality of labour productivity. *The Economic Journal*, 131(633):302–326.

Galindo Gil, H. (2021). Is the working capital channel of the monetary policy quantitatively relevant? a structural estimation approach. Working paper.

Gilchrist, S. and Williams, J. C. (2000). Putty-clay and investment: A business cycle analysis. *Journal of Political Economy*, 108(5):928–960.

Hansen, G. D. and Prescott, E. C. (2005). Capacity constraints, asymmetries, and the business cycle. *Review of Economic Dynamics*, 8(4):850–865.

Harding, M., Lindé, J., and Trabandt, M. (2023). Understanding post-covid inflation dynamics. *Journal of Monetary Economics*, 140:S101–S118.

Hyun, J., Kim, R., and Lee, B. (2023). Business cycles with cyclical returns to scale. *International Economic Review*.

Ippolito, F., Ozdagli, A. K., and Perez-Orive, A. (2017). The transmission of monetary policy through bank lending: The floating rate channel. *Finance and Economics Discussion Series*, 2017(026):n/a.

Jones, C. I. (2005). The shape of production functions and the direction of technical change. *The Quarterly Journal of Economics*, 120(2):517–549.

Jørgensen, P. L. and Ravn, S. H. (2022). The inflation response to government spending shocks: A fiscal price puzzle? *European Economic Review*, 141:103982.

Knoblach, M. and Stöckl, F. (2020). What determines the elasticity of substitution between capital and labor? a literature review. *Journal of Economic Surveys*, 34(4):847–875.

Koh, D. and Santaeulàlia-Llopis, R. (2022). Countercyclical elasticity of substitution. Technical report, Center for Economic and Policy Research.

Kuhn, F. and George, C. (2019). Business cycle implications of capacity constraints under demand shocks. *Review of Economic Dynamics*, 32:94–121.

Lee, D. (2021). Quantitative easing and inequality.

Lewis, V. and Villa, S. (2023). Labor Productivity, Effort and the Euro Area Business Cycle. Discussion Paper 18389, Centre for Economic Policy Research.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Merz, M. (1995). Search in the labor market and the real business cycle. *Journal of Monetary Economics*, 36(2):269–300.

Michaillat, P. and Saez, E. (2015). Aggregate demand, idle time, and unemployment. *The Quarterly Journal of Economics*, 130(2):507–569.

Michaillat, P. and Saez, E. (2022). An economical business-cycle model. *Oxford Economic Papers*, 74(2):382–411.

Michaillat, P. and Saez, E. (2024). Moen meets rotemberg: An earthly model of the divine coincidence. *arXiv preprint arXiv:2401.12475*.

Murphy, D. (2017). Excess capacity in a fixed-cost economy. *European Economic Review*, 91:245–260.

Nekarda, C. J. and Ramey, V. A. (2020). The cyclical behavior of the price-cost markup. *Journal of Money, Credit, and Banking*, 52(S2):319–53.

Oi, W. Y. (1962). Labor as a Quasi-Fixed Factor. *Journal of Political Economy*, 70(6):538–555.

Orphanides, A. (1993). Labor hoarding when unemployment is a worker discipline device. *The Scandinavian Journal of Economics*, 95(1):111–118.

Phaneuf, L., Sims, E., and Victor, J. G. (2018). Inflation, output and markup dynamics with purely forward-looking wage and price setters. *European Economic Review*, 105:115–134.

Qiu, Z. and Ríos-Rull, J.-V. (2022). Procyclical productivity in new keynesian models. Technical report, National Bureau of Economic Research.

Ravenna, F. and Walsh, C. E. (2006). Optimal monetary policy with the cost channel. *Journal of Monetary Economics*, 53(2):199–216.

Rupert, P. and Šustek, R. (2019). On the mechanics of new-keynesian models. *Journal of Monetary Economics*, 102:53–69.

Sbordone, A. M. (1996). Cyclical productivity in a model of labor hoarding. *Journal of Monetary Economics*, 38:331–361.

Shapiro, C. and Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American economic review*, 74(3):433–444.

Sims, C. A. (1992). Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy. *European Economic Review*, 36(5):975–1000.

Solow, R. M. (1964). On the short-run relation of employment and output. Draft of Presidential Address, Econometric Society. Unpublished.

Stroebel, J. and Vavra, J. (2019). House prices, local demand, and retail prices. *Journal of Political Economy*, 127(3):1391–1436.

Tsuruga, T. (2007). The hump-shaped behavior of inflation and a dynamic externality. *European Economic Review*, 51(5):1107–1125.

Woodford, M. (2007). Interpreting inflation persistence: Comments on the conference on" quantitative evidence on price determination". *Journal of Money, Credit and Banking*, 39:203–210.

A Proofs

A.1 Proof to Proposition 1

Log-linearizing the firm's first-order condition for the price, given in (2.21), and wage choice, given in equation (2.37), yield the following equations after some manipulations:

$$\hat{\mu}_t = \frac{\hat{\Gamma}(\bar{\nu}_t)}{1-\epsilon\Gamma(\bar{\nu})} - \frac{\phi^P}{L(\epsilon\Gamma(\bar{\nu})-1)}\left(\{\hat{\pi}^p_t - \beta\mathbb{E}\hat{\pi}^p_{t+1}\} + \{\hat{\pi}_t - \beta\mathbb{E}\hat{\pi}_{t+1}\}\right)$$

and

$$\hat{mrs}_t - \hat{w}_t = \frac{\phi^W}{\epsilon^W - 1}\left(\{\hat{\pi}^w_t - \beta\mathbb{E}_t\hat{\pi}^w_{t+1}\} + \{\hat{\pi}_t - \beta\mathbb{E}_t\hat{\pi}_{t+1}\}\right)$$

where all hatted variables are deviations from the steady state value, and terms without a time subscript are steady state values. We define $\hat{\pi}^p$ as the relative-price inflation rate, $\hat{\pi}^w$ as the real wage inflation rate, and $\hat{mrs}$ is the change in the household's marginal rate of substitution. Additionally, we have used the definitions of the markup $\mu_t = \frac{\tilde{P}_t}{MC_t}$ and marginal cost $MC_t = \frac{W_t}{\bar{\xi}_t A_t}$ in their log-linearized form to obtain:

$$\hat{\mu}_t = \hat{\tilde{p}}_t - \hat{w}_t + \hat{a}_t + \hat{\bar{\xi}}_t$$

Equating the expressions for the trajectory of inflation, $\hat{\pi}_t - \beta\mathbb{E}_t\hat{\pi}_{t+1}$, we find the following expression for the markup:

$$\hat{\mu}_t = \underbrace{\frac{-\hat{\Gamma}(\bar{\nu}_t)}{\epsilon\Gamma(\bar{\nu})-1}}_{①} - \frac{\phi^P}{\Psi^1}\left\{\underbrace{(\hat{\pi}^p_t - \beta\mathbb{E}_t\hat{\pi}^p_{t+1}) - (\hat{\pi}^w_t - \beta\mathbb{E}_t\hat{\pi}^w_{t+1})}_{②} - \underbrace{\frac{\epsilon^W-1}{\phi^W}(\hat{mrs}_t - \hat{w}_t)}_{③}\right\} \tag{A.1}$$

where $\Psi^1 = \mathbb{E}\{Y\}\tilde{P}(\epsilon\Gamma(\bar{\nu})-1)$ is a constant.

$$\tag{A.2}$$

A.2 Proof to Proposition 2

The ex-post (i.e., after all shocks have been realized) real profits of the j^{th} firm in period t is given by:

$$\Pi_{j,t} = \tilde{P}_t Y_{j,t} - W_t L_t - I_t - \Phi^P(\cdot) - \Phi^H(\cdot)$$

subject to

$$Y_{j,t} = \xi_{j,t} A_t L_t$$

$\Phi^P(\cdot) = \Phi(\cdot) = 0$ since all hiring and price-setting related costs are already incurred. Denoting the lagrangian multiplier on the production function constraint as λ_t, the following definition of the marginal cost emerges from the profit maximization problem of the firm along the labor margin.

$$W_t = \lambda_t \frac{\partial Y_{j,t}}{\partial L_t}$$

Since λ_t is the shadow price of raising production by a unit, it is also the marginal cost of the firm. Rearranging the above expression, multiplying both sides by the real price of the firm $\tilde{P}$, and multiplying and dividing the RHS by $L/\mathbf{Y}$, we get:

$$\frac{\tilde{P}_t}{\lambda_t} = \tilde{P}_t \frac{\mathbf{Y}_t}{W_t L_t} \left(\frac{\partial Y_{j,t}}{\partial L_t} \frac{L_t}{\mathbf{Y}_t} \right)$$

Recall that the marginal product of labor is just $\xi_{j,t} A_t$ in our model (see equation 2.10). Define the labor share $s \equiv W_t L_t / \mathbf{Y}_t$. Thus the firm's realized markup, μ, is given by:

$$\mu_t = \frac{1}{s_t} \left(\tilde{P}_t \frac{Y_{j,t}}{\mathbf{Y}_t} \right)$$

Integrating this over all j and rearranging, we get

$$s_t = \frac{1}{\mu_t} \left(\tilde{P}_t \frac{\mathbb{E}_\nu\{Y_t\}}{\mathbf{Y}} \right) = \frac{1 - \Omega(\bar{\nu}_t)}{\mu_t} \tag{A.3}$$

where we have defined $\Omega \equiv 1 - \dfrac{P\mathbb{E}_\nu\{Y\}}{\mathbf{PY}}$. We can additionally show, from Proposition 2 and (2.12), that $\Omega(\cdot)$ is only dependent on $\bar{\nu}$:

$$1 - \frac{P\mathbb{E}_\nu\{Y_t\}}{\mathbf{P}_t \mathbf{Y}_t} = 1 - \frac{\int_0^{\bar{\nu}} \nu \, dF(\nu) + \bar{\nu} \int_{\bar{\nu}}^{\infty} \, dF(\nu)}{\int_0^{\bar{\nu}} \nu \, dF(\nu) + \bar{\nu}^{\frac{\epsilon-1}{\epsilon}} \int_{\bar{\nu}}^{\infty} \, \nu^{\frac{1}{\epsilon}} dF(\nu)}$$

Log-linearizing equation (A.3) yields:

$$\hat{s}_t = \frac{\Omega}{1-\Omega} \hat{\Omega}(\bar{\nu}_t) - \hat{\mu}_t \tag{A.4}$$

A.3 Proof to Proposition 3

Here, we derive the Phillips curve implied by our model. Consider the log-linearized version of (2.21):

$$\hat{\tilde{p}}_t = \hat{w}_t - \hat{mpl}_t + \frac{\hat{\Gamma}(\bar{\nu}_t)}{1 - \epsilon\Gamma(\bar{\nu})} + \frac{1}{\Psi^2} \left(\{\hat{\pi}_t^p - \beta \mathbb{E} \hat{\pi}_{t+1}^p\} + \{\hat{\pi}_t - \beta \mathbb{E} \hat{\pi}_{t+1}\} \right)$$

where $\Psi^2 = \dfrac{\mathbb{E}_\nu\{Y\}\tilde{P}(\epsilon\Gamma(\bar{\nu}) - 1)}{\phi^P}$ and $\hat{mpl}_t = \hat{a}_t + \hat{\bar{\xi}}_t$ is the log deviations of marginal productivity of labor from its steady state. Isolating $\hat{\pi}_t$ on the RHS:

$$\hat{\pi}_t = \Psi^2 \left(\hat{w}_t - \hat{mpl}_t - \hat{\tilde{p}}_t \right) + \Psi^2 \frac{\hat{\Gamma}(\bar{\nu}_t)}{(1 - \epsilon\Gamma(\bar{\nu}))} + \left\{ \hat{\pi}_t^p - \beta \mathbb{E} \hat{\pi}_{t+1}^p \right\} + \beta \mathbb{E} \hat{\pi}_{t+1} \tag{A.5}$$

Recalling that $\hat{w}_t - \hat{mpl}_t = \hat{mc}_t$ is the marginal cost, and noting that $\hat{\tilde{p}}_t = \frac{1}{\epsilon}(\hat{\bar{\nu}}_t + \hat{y}_t^*)$, we have:

$$\hat{\pi}_t = \Psi^2 \hat{mc}_t + \beta \mathbb{E} \hat{\pi}_{t+1} + \underbrace{\Psi^2 \frac{\hat{\Gamma}(\bar{\nu}_t)}{(1 - \epsilon\Gamma(\bar{\nu}))} + \frac{1}{\epsilon}(\hat{\bar{\nu}}_{t-1} + \hat{y}_{t-1}^*) - \frac{(\Psi^2 + 1 + \beta)}{\epsilon} \left(\hat{\bar{\nu}}_t + \hat{y}_t^* \right) + \frac{\beta}{\epsilon} \mathbb{E}(\hat{\bar{\nu}}_{t+1} + \hat{y}_{t+1}^*)}_{G(\mathbb{E}\hat{y}_{t+1}^*, \hat{y}_t^*, \hat{y}_{t-1}^*)}$$

Note that the variables $\hat{\Gamma}(\bar{\nu}_t)$ and $\hat{y}^*$ are both purely functions of $\bar{\nu}$ (see equation 2.6 and equation 2.3). Since there exists a one-to-one correspondence between $\bar{\nu}$ and the utilization rate y^* (see equation 2.5 and associated discussion), we can re-write the Phillips curve derived as follows:

$$\hat{\pi}_t = \Psi^2 \hat{mc}_t + \beta \mathbb{E} \hat{\pi}_{t+1} + G(\mathbb{E}\hat{y}_{t+1}^*, \hat{y}_t^*, \hat{y}_{t-1}^*)$$

B Two-Industry Model

Here we briefly describe the two industry model introduced in Section 4. As outlined above, we endeavor to design the two industry model such that it deviates minimally from the single industry model, so that the intuition developed in the model above can be applied in a straightforward manner to the extension. In this spirit, we ignore many modeling elements which may be important and relevant, such as the non-homothetic nature of demand across manufactured and non-manufactured goods, the phenomenon of "services deepening" (Galesi and Rachedi, 2018), the role of imports in manufacturing etc. For brevity, we keep the discussion here to deparures from the single-industry model.

B.1 Production Sector

As discussed earlier, we split the production sector into a manufacturing industry (denoted with an M super-script) and a non-manufacturing industry (denoted with an S super-script). Each industry is structurally identical to the production sector described in Section 2.2, with an industry-specific aggregating firm as in Section 2.1. Thus

$$\mathbf{Y}_t^i = \left[\int_0^1 \left(Y_{t,j}^i\right)^{\frac{\epsilon^i - 1}{\epsilon^i}} \nu_{t,j}^{\frac{1}{\epsilon^i}} dj\right]^{\frac{\epsilon^i}{\epsilon^i - 1}}$$

where the final firm's maximization yields

$$Y_j^i = \left(\frac{P_{t,j}^i}{\mathbf{P}_t^i}\right)^{-\epsilon^i} Y_t^i$$

Here, P_j^i and $\mathbf{P}^i$ refer to the j^{th} intermediate firm's price and industry price for the i^{th} industry for $i \in M, S$. The output of the two industries, Y^M and Y^S, are then aggregated by a final firm that aggregates the output of the two firms into a final good, Y^F, that is consumed by households and purchased for investment by intermediate firms.

$$Y_t^F = \left(\alpha_M (Y_t^M)^{\psi^F} + \alpha_S (Y_t^S)^{\psi^F}\right)^{\frac{1}{\psi^F}}$$

B.2 Household Sector

The problem of the household sector remains the same as in the single industry model. For reasons we will expand upon shortly, we model the household along the lines of Merz (1995), such that there is full consumption risk sharing between households. This set-up is standard in many search-and-matching models of unemployment (Blanchard and Galí, 2010). While this is obviously a heroic assumption, it allows us to maintain a single representative household with a single wage across sectors.

As before the household chooses its labor supply and wage prior to demand manifesting. An additional wrinkle, however, is that there are two sources of uncertainty in the household's decision to supply labor. First, household's are uncertain which sector they will be assigned to. Second, household's are uncertain regarding how much effort they will be required to expend. Thus, they form expectations over both the labor demand from the secotrs, as well as the effort within each sector. The first order condition with respect to labor therefore takes the form:

$$W_t(1+\chi_t) - \mathbb{E}_t\left\{\rho_{t,t+1} W_{t+1} \frac{L_{t+1}}{L_t} \chi_{t+1}\right\} =$$

$$\frac{\epsilon^W}{\epsilon^W - 1} \frac{\omega L_t^{\varphi}}{\lambda_t^H} \sum_{i \in M,S} \frac{L^i}{L} \left[\left(\frac{1}{\bar{\nu}_t^i}\right)^{1+\varphi} \int_0^{\bar{\nu}^i} \nu_t^{1+\varphi} + \int_{\bar{\nu}^i}^{\infty} dF(\nu^i)\right]$$

Table 3: Non-estimated Parameters, Two Industry Model

Parameter		Description	Parameter		Description
β	0.993	discount factor	G^{ss}	0.12	govt. consumption to GDP
δ	0.025	depreciation rate	ϱ	0.05	exogenous separation rate
ϵ^{W}	6.0	elast. of sub. between labor varieties	ω	1.0	weight on disutility of labor
γ	1.0059	gross balanced growth rate	$\bar{\pi}$	1.0083	gross inflation rate
s_{ss}	0.62	steady-state labor share			
${y^*}^{M}_{ss}$	0.80	manuf. steady-state cap. utilization	${y^*}^{S}_{ss}$	0.80	non-manuf. steady-state cap. util.
${s^*}^{M}$	0.68	manuf. steady-state labor share	$Y^M/\mathbf{Y}$	0.205	manuf. steady state % of GDP
ϵ^S	8.0	elast. of sub. non-manuf varieties	${\phi^P}^S$	9.0	non-manuf. Rotemberg price adj.
${\phi^K}^S$	8.0	non-manuf. invest. adj.	σ^S	0.30	non-manuf. K-L elast. of sub.

where L^i is the labor demanded by the $i^t h$ industry. When uncertainty is resolved, workers are directed either toward the manufacturing or non-manufacturing industry. This exposes the possibility that households might have differential earnings ex-post depending on the sector they get assigned to. Our assumption of full risk sharing within a joint "family household" as in Merz (1995) allows us to ignore the implications of this.

The remainder of the optimality conditions remain unchanged.

B.3 Estimation

Estimation is once again performed using Bayesian IRF matching, as described in Section 4. As before, we set some parameters in advance. In addition to the parameters set in Table 1, we also set the parameters associated with the non-manufacturing sector. Thus, the parameters estimated are those related to the household sector, the manufacturing industry and the final firm which aggregates the manufacturing and non-manufacturing goods. We do this to aid parameter identification; estimating parameters related to both industries and the final aggregating firm produces a posterior that is badly behaved. Our non-estimated parameters are given in Table 3.

As discussed above, the state-dependency of our model implies that the choice of steady-state is important. We choose a manufacturing steady-state capacity utilization rate that is consistent with the data. For the non-manufacturing industry, this object is not measured. However, there is some evidence that the *services* industry has a slightly higher rate of utilization, although the definition of utilization is different (European Central Bank Monthly Bulletin, 2014). We choose to match the utilization rate in the non-manufacturing sector to the manufacturing sector. The labor share of the manufacturing industry gross value added is matched to the 1951-2008 average. Likewise, the share of manufacturing in GDP reflects the average for the sample period, although this value has a strong downward trend in the sample. The remainder of the non-manufacturing industry parameters are kept close to the values obtained from the estimation of the single industry model. The values of the parameters for the manufacturing sector are therefore assumed to account for all the adjustment required to match the aggregate dynamics.

Table 4: Estimated (top panel) and model-implied (bottom panel) parameters for baseline model.

		Prior			Posterior		
		Mean	Dist.	St. Dev	Mean	90% HPD interval	
ϕ^W	Rotemberg wage adj.	800	Gamm	100	864.3508	708.2055	1026.3627
φ	Inverse frisch elast.	1.5	Gamm	1	1.6158	0.4996	2.6832
h	Habits in consumption	0.8	Beta	0.1	0.8375	0.812	0.8627
ρ_S	Taylor rule smoothing	0.86	Beta	0.1	0.8407	0.809	0.8731
ρ_π	Taylor rule inflation	1.8	Gamm	0.25	1.8906	1.4953	2.2674
ρ_y	Taylor rule output	0.02	Gamm	0.015	0.0146	0.0003	0.0295
ϵ^F	Manuf./non-manuf. elast. sub.	4	Gamm	1	3.8546	2.5587	5.1533
ι^M	Manuf. working capital	0.4	Beta	0.2	0.1671	0.0086	0.327
ϵ^M	Manuf. varieties elast	8	Gamm	1	7.9942	6.3906	9.5698
ϕ^P	Manuf. Rotemberg price adj.	4	Gamm	1	6.0205	4.2117	7.818
ϕ^K	Manuf. investment adj.	7	Gamm	1	7.5202	5.9445	9.0819
ϕ^N	Manuf. labor adj.	0.001	Gamm	0.0001	0.0004	0.0003	0.0006
σ	Manuf. K-L substitution	0.12	Gamm	0.1	0.1043	0.0622	0.1439
$\tilde{m}^M$	Manuf. supply curve shifter	0.1	Beta	0.02	0.1119	0.0907	0.1332
Model-implied parameters: Evaluated at posterior mean of estimated parameters							
σ_ν^M	Manuf. idio. shock var.	0.72	σ_ν^S	Non-manuf. idio. shock var.			0.91
α^M	K share of capacity, manuf.	0.11	α^S	K share of capacity, non-manuf.			0.25
α_K^M	K dist. param., manuf.	1.21E-05	α_K^S	K dist. param., non-manuf.			3.86
α_N^M	L dist. param., manuf.	1.07E-07	α_N^S	L dist. param., non-manuf.			0.24
ψ^M	K-L substitution, manuf.	-8.58	ψ^S	K-L substitution, non-manuf.			-2.33
s^{*S}	Non-manuf. labor share	0.80					

The results of the estimation procedure are given in Table 4. The results for the household parameters remain close to the single industry model. We find that the elasticity of substitution between capital and labor in the manufacturing sector is substantially lower than in the single aggregated industry, which aligns with intuition, given that manufacturing is a much more structured production process. Likewise, the investment adjustment costs are higher than in the single industry model, reflecting greater difficulty in capacity expansion. Incidentally, price rigidities are lower than in the rest-of-the-economy case. This is in line with the findings in Galesi and Rachedi (2018) and other studies which show that the prices in the services sector demonstrate higher degree of stickiness.

The impulse response functions for the model are provided in Figure 6 in the main body of the essay, so we do not repeat them here. Given that this two-industry extension is designed only to showcase the ability of the broader model to match the empirical dynamics, we find that the results are surprisingly well aligned with empirical findings and intuition related to the manufacturing sector. A fuller model which explicitly seeks to model the specificities of the manufacturing sector may uncover additional insights into the role of manufacturing in aggregate outcomes.