Procyclical Productivity in New Keynesian Models

Zhesheng Qiu,* José-Víctor Ríos-Rull†

Saturday 22nd November, 2025

For the most recent version, please visit here

Abstract

We propose a product-market search friction: households search for merchants before spending, which occupies firms' production capacity. In an otherwise standard New Keynesian environment, this friction creates a link between aggregate demand and TFP. The primitives—search benefits, search costs, and congestion—are disciplined by three micro moments: the average markup, the elasticity of household search to spending, and the elasticity of firm productivity to sales, the last of which is estimated using Compustat data with a Bartik instrument that isolates demand-driven sales variation. These moments yield a closed-form macro elasticity of endogenous productivity to real spending, robust across alternative specifications of the search friction.

This mechanism is externally validated: perishable-capacity sectors exhibit strong comovement among occupancy, profit margins, and sales, whereas storable-capacity sectors do not. Quantitatively, the model outperforms the canonical capital-utilization model in terms of impulse-response matching, reduces the size of required productivity shocks to match business cycle dynamics, rationalizes countercyclical labor share under demand shocks and procyclical consumption under IST shocks, delivers acyclical markups consistent with micro evidence, and implies a flattening Phillips curve over time.

Keywords: Productivity, Profit Margins, Directed Search, Micro Evidence, Macro Implications

JEL Codes: E12, E32, E52

*Department of Economics, Hong Kong University of Science and Technology: zheshqiu@ust.hk.

†Department of Economics, University of Pennsylvania, UCL, CAERP, NBER and CEPR. vr0j@upenn.edu.

We thank Lawrence Christiano, Christiano Cantore, Zhen Huo, Lukas Nord, Federico Di Pace, Jordí Galí, Nils Gottfries, Miguel Leon-Ledesma, Morten Ravn, and Sergio Salgado for constructive comments and discussions. We thank Bowen Fang and Yang Li for excellent research assistance.

1. Introduction

Idle production resources are a defining feature of modern economies. Airlines frequently operate with empty seats, hotels contend with vacant rooms during off-peak seasons, and restaurants often maintain staffing levels that exceed dining demand for large portions of the day. These examples are more than mere anecdotes — they reflect a fundamental economic phenomenon with profound implications. The capacity occupancy rate, defined as the proportion of available production resources actually occupied by customers, fluctuates systematically over the business cycle. Yet its macroeconomic effects remain inadequately understood, especially in service sectors that now dominate advanced economies. This gap in knowledge directly influences three cornerstones of business cycle analysis: total factor productivity (TFP), the cyclical behavior of markups, and inflation dynamics.

The dominant approach to modeling capacity utilization in macroeconomics, exemplified by Greenwood et al. (1988a), treats utilization as a voluntary decision by firms that can adjust capital utilization intensity at increasing marginal cost. This approach, while tractable, fails to capture a crucial aspect of modern economic activity: the involuntary idleness of resources in markets where demand and supply must be actively matched. In service sectors, especially, production is inherently location-specific and time-sensitive. For instance, an airline seat that remains unsold upon departure represents a permanent loss of economic value, rather than a reversible choice. This distinction matters profoundly for how we interpret the aggregate dynamics of productivity, profit, and inflation.

Our paper develops an easy-to-use quantitative framework, introducing directed search in product markets to model these phenomena. We use the term "occupancy" instead of "utilization" to emphasize that resource idleness arises from incomplete matching between households and production locations rather than deliberate scaling back of capital intensity. In our model, households must exert search effort to find and purchase differentiated goods varieties, while firms preinstall labor and capital in production locations that generate output only when matched with a purchasing household. Unmatched locations represent wasted resources and efficiency losses that directly reduce the measured TFP. This mechanism creates a channel through which demand shocks can affect productivity, a comovement that standard New Keynesian models typically have to rely on unrealistically high fixed cost of production to replicate (e.g. Smets and Wouters, 2007).

Our paper makes three primary contributions. First, we develop a new theory to measure the effect of capacity occupancy rate on business cycles. Second, we provide new empirical evidence on how firm profit margins comove with sales across various levels, bolstering our proposed mechanisms. Third, we develop a new quantitative toolkit for medium-scale New Keynesian models.

A Theory to Measure the Effect of Capacity Occupancy The notion of involuntary production capacity idleness has already been well-established in the literature. Michaillat and Saez (2015) develop a search theory to formalize this concept, while Bai et al. (2025) show quantitatively that this mechanism can account for a substantial proportion of the business cycle fluctuations under a certain latent demand shock. What remains unexplored, however, is the precise role that such idleness plays in the transmission and amplification of observable economic shocks within an otherwise standard New Keynesian framework.

To address this issue, we propose a novel directed search model for the product market that captures the transmission of demand shocks through two standard channels of search frictions: the household's incentive to search more intensively as spending increases, and the congestion effect that dampens the effectiveness of such additional effort. While the presence of the two channels has been well-recognized in the literature (e.g. Huo and Ríos-Rull, 2020), our main contribution lies in proposing two empirically measurable micro-elasticities to quantify each of the two channels, particularly the second one.

While the first channel can be directly captured by the elasticity of search activities to expenditure, quantifying the second channel hinges on a key property of directed search: as overall market congestion intensifies, a rise in demand for an individual firm translates into a smaller increase in customer arrivals and, accordingly, a weaker improvement in firm-level TFP. We therefore propose using the elasticity of firm-level TFP to demand-driven sales to measure the strength of the second channel, offering a novel approach to quantifying the second channel that has no previous observations.

The key theoretical contribution is summarized as a robust mapping from the two observable micro elasticities, along with the steady-state markup, to the macro elasticity of occupancy-driven TFP with respect to the real aggregate expenditures. We demonstrate that this mapping is robust to alternative assumptions regarding the wealth effects in search decisions, partial waste of idle inputs, and limited exposure to search frictions. This robustness indicates that the mapping itself is a robust sufficient statistic for the aggregate effects of search frictions, largely independent of modeling details.

Empirical Support from the movement of Firm Profit Margins We also provide a full package of empirical evidence to measure and validate the proposed mechanisms in our theory. While the evidence for the elasticity of search activities to expenditure and that for the average (steady-state) firm markups can be borrowed in the literature (Dolfen et al., 2023; Michelacci et al., 2022; De Loecker et al., 2020), the elasticity of firm-level TFP to demand-driven sales requires a new measure: we choose the quarterly sale-to-COGS ratio to proxy TFP, controlling for the 6-digit level fixed effects and price rigidities to rule out the potential confounding impacts from input and output prices.The demand variations are isolated from the heterogeneous exposures of firm sales to seasonality, via a Bartik method.

To validate our model mechanisms, we examine the comovement between the sales-to-COGS ratio and sales across multiple untargeted dimensions. First, our theory predicts that the search-driven TFP fluctuations arise from the pre-installation of productive resources, which are wasted if left unused. This mechanism is particularly relevant in service sectors, as opposed to goods-producing sectors (including wholesale and retail trade). Consistent with this, our firm-level evidence from Compustat shows that the positive comovement between the sales-to-COGS ratio and sales is significantly stronger across nearly all 2-digit service sectors, while being close to zero in non-service sectors. Second, our model implies that in sectors characterized by pre-installed inputs, the sector-level sales-to-COGS ratio strongly co-moves with occupancy rate. Transportation and accommodation are two prototypical examples of such sectors.

A Quantitative Toolkit for Medium-Scale New Keynesian Models Our theory, supported by new micro-evidence, can be elevated into a quantitatively useful framework for medium-scale New Keynesian models, providing a refined toolkit for understanding business cycle fluctuations. The theoretical structure is intentionally designed to integrate naturally within a New Keynesian environment. It introduces endogenous efficiency and markup wedges through search frictions and significantly improves the model's ability to match empirical impulse responses following the tradition of Christiano et al. (2016), while also offering a coherent explanation for several key empirical patterns in business cycles.

The improved model performance is demonstrated by comparing specifications with and without search frictions. This comparison is especially telling when the search-based model is estimated without the canonical capital utilization mechanism, while the alternative model retains capital utilization with an additional free parameter. Even under such a demanding comparison, the model incorporating search frictions still outperforms its counterpart in both Bayesian and frequentist metrics. The enhancement stems primarily from a closer fit to the impulse responses of labor productivity and labor share following monetary policy shocks, as well as improved alignment of GDP component and consumption responses to investment-specific technology shocks.

The estimated model yields three main implications. First, it reveals a new transmission channel for monetary policy shocks, with endogenous TFP contributing to approximately 10% of output response. Second, the model generates procyclical markups in response to monetary policy shocks based on the metric of Nekarda and Ramey (2020), and acyclical markups in that of Burstein et al. (2025). Third, it also helps account for the flattening of the Phillips curve documented in Del Negro et al. (2020).

Related Literature. Our paper also broadly contributes to three strands of literature. We tackle the classical question of how to model capacity utilization following the agenda of Bai et al. (2025). Other ways of modeling include the canonical costly choice of capital utilization in Greenwood et al. (1988a), labor utilization in Basu et al. (2006), and supply side constraint in Boehm and Pandalai-Nayar (2022). Our approach emphasizes both the mechanism of endogenous TFP generated from capacity utilization, and the way to measure it using micro-evidence.

In terms of theory, we contribute to the literature of product market search in macroeconomics. In particular, our approach is in line with Dolfen et al. (2023); Michelacci et al. (2022); Li (2021), which emphasize the extensive margin of expenditure. A complementary literature including, but not limited to Burdett and Judd (1983); Kaplan and Menzio (2016); Nord (2023), emphasizes the price comparison motive of product market search.

We furthermore contribute to the unsettled debate of markup cyclicality. The view of countercyclical markups, represented by Basu and House (2016); Bils et al. (2018), emphasizes the convexity of labor costs, while the view of procyclical markups, represented by Nekarda and Ramey (2020), emphasizes the difference between conditional and unconditional measures. While we do not aim to terminate this debate, our results suggest that markups in New Keynesian models can be more procyclical with search frictions. The markup cyclicality patterns in our model and data are mostly consistent with the recent findings of Burstein et al. (2025), and helps address the labor share puzzle in Cantore et al. (2021).

Layout of the Paper. The remainder of the paper proceeds as follows. Section 2 develops our static search model of occupancy rates, formalizing the core mechanism through which household search effort affects firm-level productivity and deriving analytical results for the cyclical behavior of TFP, markups, and inflation, as well as the mapping from observable empirical moments to the macroeconomic effects of search frictions. Section 3 provides empirical evidence to both measure the magnitude of the search frictions, and to validate the associated mechanisms. Section 4 extends the framework to a full-fledged medium-scale New Keynesian model suitable for estimation. Section 5 presents our estimation strategy and compares the performance of our model with conventional alternatives. Section 6 discusses the macroeconomic implications of our estimated model, focusing on the transmission of monetary policy, markup cyclicality, and the flattening Phillips curve. Section 7 examines the robustness of our results to alternative specifications, including the wealth effects in search decisions, partial waste of idle inputs, and limited exposure to search frictions. Section 8 concludes.

Appendix materials provide additional technical details, including proofs of key lemmas and propositions, description of the empirical analysis, and further robustness checks.

2. A Static Search Model of Occupancy Rate

Our analysis starts with the development of a static and transparent version of our model that captures our core mechanism. In the model, households search for the differentiated goods varieties to purchase. Upon making a purchase, they occupy the production locations of the corresponding producer. Crucially, labor is preinstalled in each of the locations, implying that the unutilized labor is wasted and an efficiency loss.[1] Section 2.1 formalizes the core mechanism of production location occupancy with a static directed search model. Section 2.2 discusses the aggregate implications of this model for the cyclicality of TFP, markups, and inflation conditional on demand shocks. Section 2.3 connects the endogenous movement of TFP driven by occupancy rate to the observable individual behaviors of the households and the firms.

2.1. A Directed Search Model Formalizing the Core Mechanism

We model the occupancy of firms' production locations by households using a simple one-period directed search model, in which each market is indexed by a pair of price and tightness $\{p, q\}$. The representative household conducts searches in all markets, with an aggregate search effort of $D(p, q)$ in market $\{p, q\}$. There exists a unit continuum of firms, each producing a differentiated goods variety as well as operating a unit continuum of production locations. All locations of a firm must be in the same market, and each of them is matched with at most a unit measure of households.[2] In the market with a measure $J(p, q)$ of varieties, firms, and production locations, the total number of matches is described by a constant returns to scale (CRS) matching function $\psi(J, D)$, where tightness is the ratio $q = \frac{D(p,q)}{J(p,q)}$, and the numbers of matches per firm and per unit of search are $\psi^f(q) \equiv \frac{\psi[J(p,q),D(p,q)]}{J(p,q)}$ and $\psi^h(q) \equiv \frac{\psi[J(p,q),D(p,q)]}{D(p,q)}$. ψ^f also represents the fraction of occupied production locations, where the preinstalled inputs can be converted to output, resembling the firm's production efficiency, while $\psi^h(q)$ resembles the search efficiency.

2.1.1. Household Problem for a Given Set of Available Markets

The representative household exhibits a love-for-variety preference, but accessing these varieties requires search effort. To optimize this trade-off, the household purchases only a subset of the available varieties. Specifically, the household chooses search effort d in each market $\{p, q\} \in \Phi$, where Φ denotes the set of available markets, and we denote the market-specific search effort as $d(p, q)$. Given search efficiency $\psi^h(q)$, the household finds a number $d(p, q)\,\psi^h(q)$ of goods varieties and the corresponding firms, and purchases a quantity $c(p, q)$ of each goods variety they find.

The love-for-variety preference is modeled as a utility function $U(c^A, d^A)$, with c^A aggregating the market-specific quantity of varieties $\{c(p, q)\}$ weighted by the corresponding numbers of varieties found

[1] A typical example is an airline company, where flight seats function as production locations, while aircraft, pilots, and cabin crew are preinstalled inputs. These inputs must be preinstalled in the short run, but are not fixed costs, as they can be adjusted over longer horizons.

[2] See Appendix A.14 for further discussion on this setup.

$\{d(p,q)\,\psi^h(q)\}$, and d^A aggregating the market-specific search effort $\{d(p,q)\}$. The aggregators are

$$c^A \equiv \left(\int_\Phi d(p,q)\,\psi^h(q)\,c(p,q)^{\frac{1}{\rho}}\,dp\,dq\right)^\rho, \quad \text{with} \quad \rho > 1 \tag{1}$$

$$d^A \equiv \int_\Phi d(p,q)\,dp\,dq. \tag{2}$$

The household searching in a set of markets Φ with total expenditure e chooses the market specific search effort $d(p,q)$ and the quantity of each variety to purchase $c(p,q)$ to solve the following problem:

$$v(e,\Phi) = \max_{c^A, d^A, \{c(p,q), d(p,q)\}\geq 0} U(c^A, d^A), \tag{3}$$

$$\text{s.t.} \qquad e = \int_\Phi d(p,q)\,\psi^h(q)\,p\,c(p,q)\,dp\,dq, \tag{4}$$

and the definition of aggregates (1) and (2).

Lemma 1. *The solution to problem (1)-(4) features $d(p,q) > 0$ only in the markets $\{p,q\}$ that solves*

$$\min_{\{p,q\}\in\Phi} \left\{\psi^h(q)^{-(\rho-1)}\,p\right\},$$

and the household is indifferent between the markets where $d(p,q) > 0$.

Proof. See Appendix A.1. □

Lemma 1 characterizes the tradeoff between search efficiency $\psi^h(q)$ and the price of goods varieties p for the household. The tighter market (higher q and hence lower ψ^h) must compensate the household with a lower price to stay attractive. Given total effort and spending, the household is indifferent between the market where it can find more good varieties but purchase a smaller quantity of each, and one with opposite features. This tradeoff does not depend on the total amount of spending e.

Now we specialize our environment to GHH preference à la Greenwood et al. (1988b) in consumption and search effort: $U(c^A, d^A) = c^A - \zeta\frac{(d^A)^{1+\nu}}{1+\nu}$. This preference generates the following relation between search and consumption aggregates

$$(d^A)^{1+\nu} = \frac{\rho-1}{\zeta}\cdot c^A, \tag{5}$$

which helps us derive the following Lemma 2 to characterize the solution to problem (1)-(4).

Lemma 2. *The solution to problem (1)-(4) satisfies*

$$v(e,\Phi) \propto c^A(e,\Phi) \propto d^A(e,\Phi)^{1+\nu} \propto \left(\frac{e}{\min_{\{p,q\}\in\Phi}\left\{\psi^h(q)^{-(\rho-1)}\,p\right\}}\right)^{\frac{1+\nu}{(1+\nu)-(\rho-1)}}. \tag{6}$$

Proof. See Appendix A.2. □

Lemma 2 characterizes the indirect utility (and the consumption and search decisions) solved from problem (1)-(4). When $(1+\nu)-(\rho-1)>0$, the utility of the household is higher if the total spending e is higher, the search efficiency $\psi^h(q)$ is higher, or the price p is lower.

Now, we can characterize the household's reaction to prices under a given utility level with the help of Lemma 2. Specifically, given the household's total spending e and utility level $\overline{v}$, a market with price p and $d(p,q)>0$ must have a corresponding tightness $\widetilde{q}$. If $q>\widetilde{q}$, the household stays away, leading to $d(p,q)=0$. If $q<\widetilde{q}$, staying within this single market yields a higher utility $v>\overline{v}$. Therefore, the reaction of $\widetilde{q}$ to p can be described by a function $\widetilde{q}(e,\overline{v},p)$, which characterizes the indifference curve of the household conditional on spending e at utility level $\overline{v}$. Along this difference curve, the household optimally chooses the quantity of each variety to purchase $\widetilde{c}(e,\overline{v},p)$. The two functions $\widetilde{q}(e,\overline{v},p)$ and $\widetilde{c}(e,\overline{v},p)$ are taken as given by the firms and characterized in the following Lemma 3.

Lemma 3. *The two functions $\widetilde{q}(e,\overline{v},p)$ and $\widetilde{c}(e,\overline{v},p)$ described above satisfy*

$$\psi^h\Big[\widetilde{q}(e,\overline{v},p)\Big] \propto \overline{v}^{\frac{(1+\nu)-(\rho-1)}{(1+\nu)(\rho-1)}}\left(\frac{p}{e}\right)^{\frac{1}{\rho-1}}, \tag{7}$$

$$\widetilde{c}(e,\overline{v},p) \propto \overline{v}^{-\frac{1}{\rho-1}}\left(\frac{p}{e}\right)^{-\frac{\rho}{\rho-1}}. \tag{8}$$

Proof. See Appendix A.3. □

2.1.2. Firm Problem Conditional on the Utility Delivered to the Household

A firm that produces a goods variety transforms y units of labor into y units of output. The firm solves the price-setting problem, taking as given the nominal wage rate W, the nominal spending e, the utility $\overline{v}$ delivered to the household, and the functions $\{\widetilde{q}(e,\overline{v},p),\widetilde{c}(e,\overline{v},p)\}$ characterized in Lemma 3:

$$\Omega(e,W,\overline{v},p^-) = \max_p\left\{\left(p\,\psi^f\Big[\widetilde{q}(e,\overline{v},p)\Big]-W\right)\widetilde{c}(e,\overline{v},p)-\chi\left(\frac{p}{p^-}\right)e\right\}, \tag{9}$$

where p^- is the reference price, and $\chi(\cdot)$ is the convex function of price adjustment costs that satisfies $\chi(1)=0$, $\chi'(1)=0$, and $\chi''(1)\geq 0$.

When a firm raises its prices, three countervailing effects erode profits. First, the function $\widetilde{c}(e, \overline{v}, p)$ captures the reduction in demand within each customer match. Second, $\widetilde{q}(e, \overline{v}, p)$ captures the decline in the number of matches of the firm. Third, price adjustment costs further reduce profits. The optimal pricing decision is characterized by Lemma 4.

Lemma 4. *The solution of the price-setting problem (9) $p(e, W, \overline{v}, p^-)$ satisfies*

$$\chi'\left(\frac{p}{p^-}\right)\frac{p}{p^-} = \frac{1}{\rho - 1}\left\{\frac{\rho W}{p} - \frac{\psi^f\left[\widetilde{q}(e, \overline{v}, p)\right]}{1 - \mathcal{E}\left[\widetilde{q}(e, \overline{v}, p)\right]}\right\}\frac{p\,\widetilde{c}(e, \overline{v}, p)}{e}. \tag{10}$$

where $\mathcal{E}(q) \equiv \frac{d \ln \psi^f(q)}{d \ln q}$ is the elasticity of the matching function per firm $\psi^f(q)$.

Proof. See Appendix A.4. □

Equation (10) in Lemma 4 characterizes the optimal pricing decision $p(e, W, \overline{v}, p^-)$, which is driven by the marginal cost of labor W as in models with no search frictions, the endogenous productivity ψ^f as in models with random search à la Huo and Ríos-Rull (2020), and the endogenous component $1 - \mathcal{E}$ that is unique in our directed model.

2.1.3. Equilibrium Given Nominal Spending and Wage Rate

Given exogenous nominal spending e and wage rate W, as well as a reference price P^-, the equilibrium features a single market $\{P^*(e, W, P^-), Q^*(e, W, P^-)\}$ satisfying the following consistency

$$\begin{aligned} P^* &= p(e, W, \overline{v}, P^-), \\ Q^* &= d^A(e, \{P^*, Q^*\}), \\ \overline{v} &= v(e, \{P^*, Q^*\}), \end{aligned}$$

between the equilibrium objects $\{P^*, Q^*, \overline{v}\}$ and the corresponding solutions to firm's and household's problems. As a result, we have Proposition 1 below to characterize the equilibrium market.

Proposition 1. *The equilibrium market $\{P^*(e, W, P^-), Q^*(e, W, P^-)\}$ satisfies*

$$\chi'\left(\frac{P^*}{P^-}\right)\frac{P^*}{P^-} = \frac{\rho}{\rho - 1}\left\{\frac{W}{P^*\psi^f(Q^*)} - \frac{1}{\rho\left[1 - \mathcal{E}(Q^*)\right]}\right\}, \tag{11}$$

$$\frac{e}{P^*} = \frac{\zeta}{\rho - 1}\frac{(Q^*)^{1+\nu}}{\psi^f(Q^*)^{\rho - 1}}. \tag{12}$$

Proof. See Appendix A.5. □

Within proposition 1, condition (11) is derived from equation (10) in Lemma 4, and condition (12) is from equation (5) and budget (4). Conditions (11) and (12) characterize how price P^* and tightness Q^* respond to the changes in nominal spending e and nominal wage rate W, which allows us to discuss the cyclicality of TFP, markups, and inflation in Section 2.2.

2.2. Aggregate Implications of TFP, Markup, and Inflation Cyclicality

Now we explore the aggregate implications of the model for the cyclicality of TFP, markup, and inflation conditional on demand shocks. Our analysis is built on the equilibrium conditions (11) and (12) provided in Proposition 1. To further the analysis, we log-linearize these conditions, which requires us to choose a reference point (e, W, P^-). A natural reference point has a given exogenous pair (e, W), and satisfies $P^- = P^*(e, W, P^-)$ when prices are fully flexible, i.e., $\chi''(1) = 0$. We refer to this reference as a steady state, an analogy to the dynamic versions of our model, and use $\{P_{ss}, Q_{ss}\}$ to denote the steady state.

2.2.1. Cyclicality of TFP Conditional on Demand Shocks

In response to an exogenous increase in the aggregate nominal spending, the comovement between the equilibrium TFP ψ^f and real spending e/P^* is characterized in the following Corollary 1.

Corollary 1. *Around the steady state, the equilibrium characterized in Proposition 1 features*

$$d\ln[\psi^f(Q^*)] = \beta^{macro}_{TFP} \cdot [d\ln(e) - d\ln(P^*)], \tag{13}$$

$$\text{where} \quad \beta^{macro}_{TFP} \equiv \frac{1}{\frac{1+\nu}{\varphi} - (\rho - 1)} \quad \text{and} \quad \varphi \equiv \left.\frac{d\ln[\psi^f(q)]}{d\ln(q)}\right|_{q=Q_{ss}}$$

Proof. See Appendix A.6. □

The elasticity β^{macro}_{TFP} captures our core mechanism. It increases with the love-for-variety parameter ρ and the elasticity of the matching function φ, but decreases with the curvature of search disutility ν. While a theoretical counterpart to β^{macro}_{TFP} can be found in Huo and Ríos-Rull (2020), our contribution is to analyze its implications for markups and inflation in the presence of nominal rigidities in Section 2.2, and to propose empirical measures for it in Section 2.3.

2.2.2. Cyclicality of Markups Conditional on Demand Shocks

In New Keynesian models, markups are typically countercyclical in response to demand shocks, as real wages are procyclical. Our proposed mechanism relaxes this constraint by introducing endogenous TFP through the macro elasticity β^{macro}_{TFP}, enabling both the markups and the real wages to be procyclical.

According to equation (13) in Corollary 1, the equilibrium gross markup, which is the ratio of price P^* over marginal cost of production $W/\psi^f(Q^*)$, satisfies

$$d\ln\left(\frac{P^*}{W/\psi^f(Q^*)}\right) = d\ln[\psi^f(Q^*)] - d\ln(W/P^*) = \left(\beta_{TFP}^{macro} - \frac{d\ln(W/P^*)}{d\ln(e/P^*)}\right)\cdot d\ln(e/P^*).$$

To highlight the possibility of procyclical markups, we consider a special case of fully sticky prices with $\chi''(1) \to +\infty$, and then we have the following Corollary 2:

Corollary 2. *Under the special case of fully sticky prices, markups conditional on shocks to the nominal spending e are procyclical, i.e., positively correlated with real spending $\frac{e}{P^*}$, iff*

$$\beta_{TFP}^{macro} > \frac{d\ln(W)}{d\ln(e)}. \tag{14}$$

Proof. See Appendix A.7. □

Condition (14) in Corollary 2 indicates that markups are procyclical as the rise of endogenous TFP outpaces that of the marginal cost of labor. When $\frac{d\ln(W)}{d\ln(e)}, \frac{d\ln(P^*)}{d\ln(e)} \in (0,1)$,

$$\frac{d\ln(W/P^*)}{d\ln(e/P^*)} = \frac{d\ln(W/e) - d\ln(P^*/e)}{d\ln(e/e) - d\ln(P^*/e)} < \frac{d\ln(W)}{d\ln(e)}.$$

Therefore, equation (14) is in fact a sufficient condition when prices are partially sticky.

Corollary 2 shows that markups can be procyclical as our proposed mechanism is sufficiently strong. Whether markups are indeed procyclical is a quantitative question.

2.2.3. Cyclicality of Inflation Conditional on Demand Shocks

Condition (12) in Proposition 1 is crucial for the understanding of inflation $\frac{P^*}{P^-}$. Log-linearizing condition (12) around the steady state yields the following Corollary 3.

Corollary 3. *Around the steady state, the equilibrium characterized in Proposition 1 features*

$$d\ln(P^*) = \frac{1}{(\beta_{markup}^{ss} + \varphi)\,\chi''(1)} \cdot \Big\{\underbrace{d\ln(W) - d\ln(P^*)}_{\textit{real wage}} + \underbrace{(\gamma - 1)\,d\ln[\psi^f(Q^*)]}_{\textit{occupancy rate}}\Big\}, \tag{15}$$

$$\textit{where} \quad \varphi \equiv \left.\frac{d\ln[\psi^f(q)]}{d\ln(q)}\right|_{q=Q_{ss}}, \quad \gamma \equiv \left.\frac{d\ln[1-\mathcal{E}(q)]}{d\ln[\psi^f(q)]}\right|_{q=Q_{ss}}, \quad \textit{and} \quad \beta_{markup}^{ss} \equiv \rho(1-\varphi) - 1.$$

Proof. See Appendix A.8. □

Equation (15) in Corollary 3 shows how inflation $d\ln(P^*)$ are linked to the changes in the real wage $d\ln(W) - d\ln(P^*)$ and the occupancy rate $d\ln[\psi^f(Q^*)]$. The case with no search frictions corresponds to $\varphi = 0$ and $d\ln[\psi^f(Q^*)] = 0$. Therefore, our search friction ($\varphi > 0$) makes the slope $\frac{1}{(\beta^{ss}_{markup}+\varphi)\,\chi''(1)}$ flatter, and also introduces an ambiguous term $(\gamma - 1)\,d\ln[\psi^f(Q^*)]$ in addition to the real wage. Now, we elaborate more on these two changes due to our search frictions.

Consider the slope term $\frac{1}{(\beta^{ss}_{markup}+\varphi)\,\chi''(1)}$. Here, $\chi''(1)$ captures the convexity of the price adjustment cost, which remains unaffected by search frictions. The term β^{ss}_{markup} represents the steady-state markup. Although it depends on search frictions through $1-\varphi$, we treat it as a calibration target across models. Thus, in model comparisons, β^{ss}_{markup} is also held invariant to search frictions. The remaining component, φ in the denominator of the slope, captures the effect of search frictions. To understand this component, we rewrite the slope term below using the steady-state markups $\beta^{ss}_{markup} = \rho\,(1-\varphi) - 1$ in Corollary 3:

$$\frac{1}{(\beta^{ss}_{markup}+\varphi)\,\chi''(1)} = \frac{1}{(\rho-1)\,(1-\varphi)\,\chi''(1)} = \frac{\frac{\rho}{\rho-1}}{\beta^{ss}_{markup}\,\chi''(1)}.$$

Here, $\frac{\rho}{\rho-1}$ is also the elasticity of substitution between goods varieties. As φ goes up, the same markup β^{ss}_{markup} implies a larger love-for-variety parameter ρ and a lower elasticity of substitution $\frac{\rho}{\rho-1}$, explaining why the slope is flatter with the presence of search frictions.

Consider the occupancy rate term $(\gamma-1)\,d\ln[\psi^f(Q^*)]$. We rewrite equation (15) in Corollary 3 as

$$d\ln(P^*) = \frac{1}{(\beta^{ss}_{markup}+\varphi)\,\chi''(1)} \cdot \Big\{ \underbrace{\gamma\, d\ln[\psi^f(Q^*)]}_{\text{desired gross markup}} - \underbrace{\{d\ln(P^*) + d\ln[\psi^f(Q^*)] - d\ln(W)\}}_{\text{actual gross markup}} \Big\},$$

which explicitly expresses the inflation pressure as the difference between the desired and actual markups. The occupancy rate term $(\gamma-1)\,d\ln[\psi^f(Q^*)]$ introduces two offsetting effects: a positive effect through the desired markup and a negative effect through the endogenous TFP in the actual markup.

2.3. Connecting the Mechanism to Observable Individual Behavior

The results in Corollaries 1-3 center on the macro elasticity β^{macro}_{TFP}, which is first introduced in Corollary 1. We establish a link between the macro elasticity β^{macro}_{TFP} derived from the representative-agent model and several observable micro elasticities derived from hypothetical idiosyncratic shocks to the individual agent around the steady state. The underlying assumption is that the idiosyncratic shocks do not have an aggregate effect around the steady state up to the first-order approximation.[3]

[3] Our method is analogous to linking the micro Frisch elasticity to the corresponding macro elasticity in the macro labor literature à la xxxxx.

2.3.1. Behaviors of Individual Households

Consider an individual household with an exogenous increase in total spending e. The following Lemma 5 precludes the equilibrium where households with different levels of spending e enter separate markets.

Lemma 5. *When households are heterogeneous in e, the equilibrium still features a single market.*

Proof. See Appendix A.9. □

Lemma 5 is a natural result of Lemma 1. Lemma 1 indicates that households have the same tradeoff between price p and tightness q, regardless of their total spending e. Therefore, it is impossible that a household strictly prefers market A over B, while another household has a reversed order of preference. As a result, we can analyze the individual household's response of search effort d^A to total spending e, taking as given the steady-state market $\{P_{ss}, Q_{ss}\}$, which yields the following Proposition 2.

Proposition 2. *Consider a nominal spending shock to an individual household around the steady state. The elasticity of search effort d^A to spending e in the equilibrium, denoted by β^{micro}_{search}, is given by*

$$\beta^{micro}_{search} \equiv \frac{\partial \ln[d^A(e, \{P_{ss}, Q_{ss}\})]}{\partial \ln(e)} = \frac{1}{(1+\nu) - (\rho - 1)}. \tag{16}$$

Proof. See Appendix A.10. □

β^{micro}_{search} in Proposition 2 describes the search behavior of the household. In particular, β^{micro}_{search} strictly increases in the love-for-variety parameter ρ, which captures the benefit of search, and strictly decreases in the convexity of search utility ν, which captures the cost of search. The difference between the benefit and the cost of search determines the elasticity β^{micro}_{search}.

2.3.2. Behaviors of Individual Firms

To model a firm with an exogenous increase in demand, we retain the representative-agent assumption for the household, but instead extend the aggregator (1) to

$$c^A \equiv \left(\int_\Phi d(\omega, p, q)\, \psi^h(q) \left[\omega\, c(\omega, p, q)\right]^{\frac{1}{\rho}} d\omega\, dp\, dq \right)^{\rho},$$

where ω denotes the taste of the representative household on the goods variety. Now, the market needs to be indexed by $\{\omega, p, q\}$. The tightness and demand functions become $\{\widetilde{q}(e, \overline{v}, \omega, p), \widetilde{c}(e, \overline{v}, \omega, p)\}$, where both the taste ω and the price p can affect the values of the two functions. Like Lemma 3, we can derive the counterpart Lemma 6 as follows.

Lemma 6. *The two functions $\widetilde{q}(e, \overline{v}, \omega, p)$ and $\widetilde{c}(e, \overline{v}, \omega, p)$ described above satisfy*

$$\psi^h\Big[\widetilde{q}(e, \overline{v}, \omega, p)\Big] \propto \overline{v}^{\frac{(1+\nu)-(\rho-1)}{(1+\nu)(\rho-1)}} \left(\frac{p}{\omega\, e}\right)^{\frac{1}{\rho-1}}, \tag{17}$$

$$\omega\, \widetilde{c}(e, \overline{v}, \omega, p) \propto \overline{v}^{-\frac{1}{\rho-1}} \left(\frac{p}{\omega\, e}\right)^{-\frac{\rho}{\rho-1}}. \tag{18}$$

Proof. See Appendix A.11. □

Conditions (17) and (18) in Lemma 6 are exactly conditions (7) and (8) in Lemma 3 with a change of variables from $(\widetilde{c}, p)$ to $(\omega\, \widetilde{c}, \frac{p}{\omega})$. We can use these two conditions to explore the responses of firms to variations in sales driven by ω. The main complexity comes from the endogenous movement of price p when the firm switches across markets due to a change of ω. Proposition 3 summarizes the results.

Proposition 3. *Consider a variation in ω for an individual firm, the elasticity of TFP from occupancy rate ψ^f to sales $p\, \widetilde{c}\, \psi^f$ around the steady state with $\omega = 1$, denoted by β^{micro}_{match}, is given by*

$$\beta^{micro}_{match} \equiv \frac{d\ln(\psi^f)}{d\ln(p) + d\ln(\widetilde{c}) + d\ln(\psi^f)} = \varphi \quad \text{where} \quad \varphi \equiv \left.\frac{d\ln[\psi^f(q)]}{d\ln(q)}\right|_{q=Q_{ss}}. \tag{19}$$

When prices are fully flexible, i.e., $\chi''(1) = 0$, the elasticity of price p to sales $p\, \widetilde{c}\, \psi^f$ around the steady state with $\omega = 1$, denoted by β^{micro}_{price}, is given by

$$\beta^{micro}_{price} \equiv \frac{d\ln(p)}{d\ln(p) + d\ln(\widetilde{c}) + d\ln(\psi^f)} = (\gamma - 1)\,\varphi \quad \text{where} \quad \gamma \equiv \left.\frac{d\ln[1 - \mathcal{E}(q)]}{d\ln[\psi^f(q)]}\right|_{q=Q_{ss}}. \tag{20}$$

Proof. See Appendix A.12. □

Equation (19) in Proposition 3 describes the matching outcomes of an individual firm. The intuition is straightforward. As φ rises, the congestion effect becomes weaker, and there will be more customers flowing into the market with a rise of ω. Note that equation (19) is not affected by the price stickiness because Lemma 6 guarantees that $d\ln(\psi^h)$ and $d\ln(p\, \widetilde{c})$ are both proportional to $d\ln(\frac{p}{\omega})$, so that there is no need separately consider how $d\ln(p)$ depends on $d\ln(\omega)$.

Equation (20) in Proposition 3 describes the desired price of a firm. The intuition is straightforward when we rewrite equation (20) as

$$\beta^{micro}_{price} \equiv \frac{d\ln(p)}{d\ln(p) + d\ln(\widetilde{c}) + d\ln(\psi^f)} = \frac{d\ln(p)}{d\ln(\psi^f)} \cdot \beta^{micro}_{match}.$$

$\frac{d\ln(p)}{d\ln(\psi^f)} = \gamma - 1$ according to equation (10) in Lemma 4, which captures the pricing effect of search.

2.3.3. Connecting to the Macro Elasticity

So far, we have collected a steady-state markup β^{ss}_{markup} in Corollary 3, and three micro-level elasticities $\{\beta^{micro}_{search}, \beta^{micro}_{match}, \beta^{micro}_{price}\}$ from Propositions 2 and 3. These four moments help identify the four parameters $\{\rho, \nu, \varphi, \gamma\}$, which allow us to map the four moments to the macro elasticity β^{macro}_{TFP} in Proposition 4.

Proposition 4. *The macro elasticity β^{macro}_{TFP} satisfies*

$$\beta^{macro}_{TFP} = \frac{\beta^{micro}_{search}\,\beta^{micro}_{match}}{1 + \beta^{micro}_{search}\,(\beta^{micro}_{match} + \beta^{ss}_{markup})}. \tag{21}$$

Proof. See Appendix A.13. □

Proposition 4 establishes via equation (21) that the aggregate effect of our search frictions captured by β^{macro}_{TFP} arises from two subcomponents. First, households increase their search efforts when spending more, as measured by β^{micro}_{search}. Second, a weaker congestion effect enhances the conversion of search to matches with firms, as measured by β^{micro}_{match}. β^{micro}_{price} does not appear in equation (21) because it primarily affects inflation through equation (15) in Corollary 3 but not the aggregate TFP.

3. Empirical Discipline for the Search–Occupancy Mechanism

This section empirically disciplines the mechanism that links household search and firm occupancy to measured productivity. We begin by quantifying four micro moments—$\beta^{\text{micro}}_{\text{search}}, \beta^{\text{micro}}_{\text{match}}, \beta^{\text{micro}}_{\text{price}}, \beta^{\text{ss}}_{\text{markup}}$—that map into the model's key structural parameters and jointly pin down the aggregate elasticity of endogenous TFP. We then provide sectoral evidence that offers external validation, emphasizing the contrast between industries with perishable and storable capacity and the observed comovement between occupancy and profit margins.

3.1. Empirical Measures of the Moments to Target

The four micro moments are drawn from a combination of existing empirical estimates and our own firm-level regressions. This subsection describes how each moment is quantified and explains how these moments discipline the aggregate elasticity of endogenous TFP through 21.

Elasticity of Search Activity to Spending: $\beta^{\text{micro}}_{\text{search}}$ A key ingredient of the mechanism is the elasticity with which higher consumer spending expands the set of varieties or merchants accessed. Two complementary studies provide quantitative discipline for this elasticity.

Dolfen et al. (2023) use a near-universe of U.S. Visa transactions from 2007 to 2017 and document that a 1% increase in household expenditures is associated with approximately a 0.6% increase in the number of merchants visited. Their measure captures a broad notion of variety—spanning goods and services—but partly reflects persistent heterogeneity in access and preferences rather than short-run adjustments in search intensity.

Michelacci et al. (2022) use longitudinal Nielsen Homescan barcode data to measure households' responses to income shocks during the 2008 tax rebate episode. They find that roughly 40% of the marginal propensity to consume reflects purchases of new products, with the remainder increasing quantities of existing items. Their estimate isolates true within-household adjustments but pertains primarily to nondurable goods, where inventory considerations weaken the link to firm-level TFP.

Despite differences in scope and identification, both studies point to a sizable extensive margin in consumer expenditures. We therefore adopt a conservative midpoint and set $\beta^{\text{micro}}_{\text{search}} = 0.50$, implying that roughly half of a marginal increase in spending leads consumers to visit new merchants or product locations—corresponding to an expansion in occupied production sites in our framework.

Elasticities of TFP and Desired Price with Respect to Firm-Level Demand: $\{\beta^{\text{micro}}_{\text{match}}, \beta^{\text{micro}}_{\text{price}}\}$ The firm-level TFP ψ^f and desired price p are not directly observed at quarterly frequency. We therefore use the sales-to-COGS ratio[4] in Compustat as a proxy for $\frac{p\psi^f}{W}$, which combines the firm's measured TFP and price relative to the wage.

To identify the two elasticities $\beta^{\text{micro}}_{\text{match}}$ and $\beta^{\text{micro}}_{\text{price}}$, we estimate the following panel regression:

$$\Delta \ln \frac{\text{Sales}}{\text{COGS}}_{f,j,t} = \alpha_f + \tau_{j,t} + \beta_1 \Delta \ln \text{Sales}_{f,j,t} + \left[\beta_2\, g(\text{flex}_j)\right] \Delta \ln \text{Sales}_{f,j,t} + \epsilon_{f,j,t}, \tag{22}$$

where f, j, and t index the firm, industry, and quarter. The variable flex_j denotes the industry-level quarterly price-adjustment frequency.[5]

Equation (22) embeds the restrictions implied by equations (19) and (20). We include firm fixed effects, α_f, to absorb all time-invariant components of profitability trends, and 6-digit industry–time

[4] COGS (cost of goods sold) primarily comprises costs that vary with output in the long run—materials, direct labor, and usage-based overhead. Following De Loecker et al. (2020), we refer to these as variable input costs. If full-occupancy production capacity is linear in variable inputs, then $d\ln(\text{Sales}/\text{COGS})$ around the steady state coincides exactly with $d\ln(p\psi^f/W)$ under demand shocks. This equivalence continues to hold when the model is augmented with fixed costs or convex adjustment costs for variable inputs. Appendix B.2 discusses robustness to these extensions and to the case of decreasing returns to scale.

[5] We use the 6-digit industry-level price-adjustment frequency constructed from the confidential microdata underlying the BLS Producer Price Index (PPI), as in Pasten et al. (2020).

fixed effects, $\tau_{j,t}$, to remove shocks common to firms within an industry–time cell. When input costs W vary primarily at the industry–time level, these fixed effects eliminate most of the variation in W, so the remaining variation in the dependent variable largely reflects movements in $\ln p\,\psi^f$.

Let $g(\text{flex}_j)$ be a monotone transformation of industry price flexibility, normalized such that $g(0) = 0$ and $g(1) = 1$. The restrictions in equations (19) and (20) imply that the elasticity $\frac{d\ln(p\,\psi^f)}{d\ln \text{Sales}}$ is affine in a properly specified transformation of flexibility, so that the regression recovers $\beta_1 + \beta_2\, g(\text{flex}_j)$ only if the chosen $g(\cdot)$ is consistent with this mapping. Because the model does not provide a closed-form expression for $g(\cdot)$, we consider several alternative monotone specifications and verify that the estimated moments β_1 and β_2 are stable across these choices, ensuring that identification does not hinge on a particular functional form.

We isolate demand-driven variation in firm sales growth using a shift–share seasonality instrument. For each firm, we estimate a firm-specific seasonal exposure from its first 20 quarterly observations and impose a 10-quarter quarantine gap so that the exposure is predetermined with respect to shocks realized in the estimation window.[6] The instrument interacts this firm-specific exposure with calendar-quarter indicators (seasonal dummies) in the estimation period. This preserves the shift–share structure: heterogeneity in firm exposures supplies the "share," while the quarter-to-quarter changes in the calendar indicators provide the "shift." Identification comes from differential responses of high- and low-exposure firms to the regular seasonal pattern after controlling for firm fixed effects and industry$\times$quarter fixed effects. Under the exclusion restriction that firm-specific supply shocks do not load on these calendar-quarter patterns beyond what is absorbed by the fixed effects, the resulting variation isolates demand-driven movements in sales.[7]

Table 1 reports the estimates used to identify the micro moments $\beta^{\text{micro}}_{\text{match}}$ and $\beta^{\text{micro}}_{\text{price}}$. Columns (1)–(3) present OLS specifications, and columns (4)–(6) report IV estimates that instrument for $\Delta \ln \text{Sales}$ and its interaction terms. Columns (2), (3), (5), and (6) incorporate linear or quadratic transformations of industry flexibility through $g(\text{flex}_j)$. The IV estimates closely mirror their OLS counterparts, and the coefficients on the interaction terms remain small and statistically indistinguishable from zero across specifications. A parsimonious summary of the results is $\beta^{\text{micro}}_{\text{match}} \approx 0.27$ and $\beta^{\text{micro}}_{\text{price}} \approx 0.00$, with the Kleibergen–Paap statistics indicating strong instrument relevance.

6 Our instrument is essentially the shift-share instrument with exogenous exposures in Goldsmith-Pinkham et al. (2020). Alfaro et al. (2024) and Gagliardone et al. (2025) also estimate the exposures as we do. app:seasonality-instruments provides more discussion about the details.

7 Appendices B.3 and B.4 describes the construction of the instrument, elaborates on the exclusion restriction, presents results for alternative instrument, and provides placebo tests.

Table 1: Baseline Regressions Used to Estimate the Micro Moments $\beta^{\text{micro}}_{\text{match}}$ and $\beta^{\text{micro}}_{\text{price}}$

	OLS regressions			IV regressions		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln$ Sales	0.309 (0.020)	0.225 (0.056)	0.272 (0.031)	0.266 (0.048)	0.286 (0.141)	0.266 (0.081)
flex $\times \Delta \ln$ Sales		0.181 (0.126)			-0.046 (0.286)	
flex$^2 \times \Delta \ln$ Sales			0.149 (0.124)			-0.001 (0.262)
KP Wald F				90.631	13.712	12.041
No. of Obs.	48,685	48,685	48,685	48,652	48,652	48,652

Notes: Standard errors are two-way clustered at firm and time levels. The sample covers U.S. Compustat firms from 1985Q1–2024Q4. Industries whose accounting structure or regulatory environment makes Sales/COGS an unreliable measure of variable profitability (such as agriculture, mining, utilities, finance, and public administration) are excluded. The results use sales-share weights to align Compustat with BEA industry composition. Kleibergen–Paap F-statistics indicate adequate instrument relevance across IV specifications. Additional details on sample construction and industry definitions are provided in Appendix B.1.

Average Markup of Firms:$\beta^{\text{ss}}_{\text{markup}}$ To measure the steady-state markup, we draw on the markup estimates in De Loecker et al. (2020). Because their baseline sample exhibits a sharp rise in markups in the post-2008 period, we truncate their series at 2007Q4 to match the endpoint of our VAR sample in Section 5.2. Using the pre-2008 period, the implied steady-state markup is $\beta^{\text{ss}}_{\text{markup}} = 0.34$.

Summary: Values of the Moments Based on Micro Observations The four empirical moments—$\beta^{\text{micro}}_{\text{search}}$ (household search elasticity), $\beta^{\text{micro}}_{\text{match}}$ (TFP-to-sales elasticity), $\beta^{\text{micro}}_{\text{price}}$ (price-to-sales elasticity), and $\beta^{\text{ss}}_{\text{markup}}$ (steady-state markup)—jointly identify the structural parameters $\{\rho, \nu, \varphi, \gamma\}$.Using the baseline moments reported in Table 2, the implied value is $\beta^{\text{macro}}_{\text{TFP}} = 0.10$.

Table 2: Values of the Target Moments and the Implied Macro Elasticity

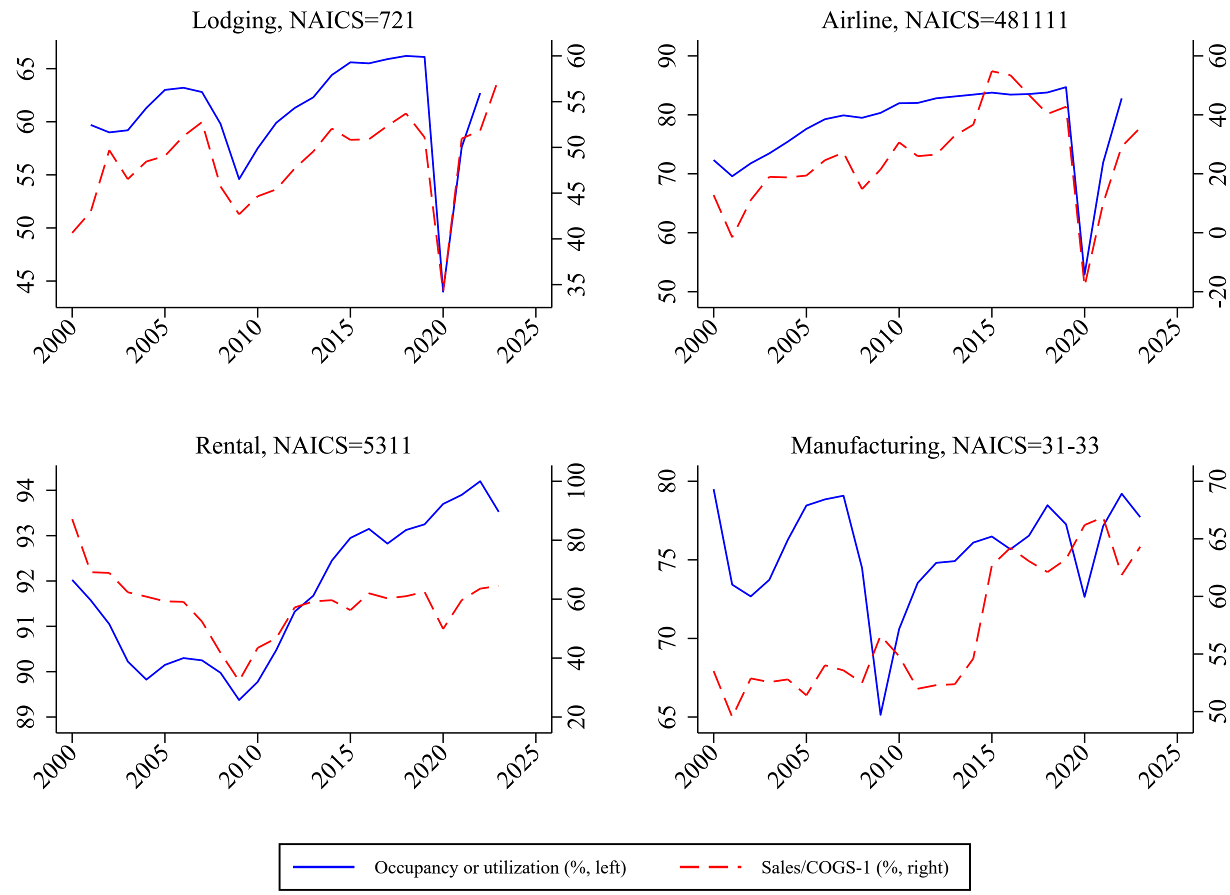
Moments	Value	Mapping	Sources
$\beta^{\text{micro}}_{\text{search}}$	0.50	$\frac{1}{(1+\nu)-(\rho-1)}$	Average of Michelacci et al. (2022) and Dolfen et al. (2023)
$\beta^{\text{micro}}_{\text{match}}$	0.27	φ	Our own estimates in Table 1
$\beta^{\text{micro}}_{\text{price}}$	0.00	$(\gamma-1)\,\varphi$	Our own estimates in Table 1
$\beta^{\text{ss}}_{\text{markup}}$	0.34	$\rho\,(1-\varphi)-1$	De Loecker et al. (2020) up to 2007Q4
$\beta^{\text{macro}}_{\text{TFP}}$	0.10	$\frac{1}{\frac{1+\nu}{\varphi}-(\rho-1)}$	$\frac{\beta^{\text{micro}}_{\text{search}}\,\beta^{\text{micro}}_{\text{match}}}{1+\beta^{\text{micro}}_{\text{search}}\,(\beta^{\text{micro}}_{\text{match}}+\beta^{\text{ss}}_{\text{markup}})}$

3.2. Additional Sector Evidence

We next bring sector-level evidence to bear on two implications of the mechanism. First, demand-induced variation in measured TFP should be visible in sectors where output is tied closely to real-time occupancy. Second, the strength of this relationship should vary systematically with perishability, generating larger values of $\beta^{\text{micro}}_{\text{match}}$ in sectors where idle capacity cannot be stored.

Comovement of TFP Proxies and Occupancy Figure 1 documents the time-series comovement between occupancy (or utilization) and the Sales/COGS ratio in sectors for which occupancy data are available: lodging (NAICS 721), airlines (NAICS 481111), and rental services (NAICS 5311). In all three sectors, the two series move closely together, especially around large aggregate shocks such as the 2008 recession and the COVID-19 collapse. Manufacturing, where idle capacity can be absorbed into inventories, provides a natural contrast and displays much weaker or even inverse comovement.

Figure 1: Comovement of TFP Proxies and Occupancy



Perishability and Cross-Sector Differences in $\beta^{\text{micro}}_{\text{match}}$. Table 3 splits the Compustat sample into goods-producing (NAICS ≤ 45) and service-producing (NAICS ≥ 48) industries. When output is storable, idle capacity adds to inventories rather than reducing contemporaneous measured TFP; in such sectors, $\beta^{\text{micro}}_{\text{match}}$ captures only the occupancy-induced component of TFP and is predicted to be near zero. Consistent with this implication, the IV estimates in columns (1)–(3) of Table 3 yield $\beta^{\text{micro}}_{\text{match}} \approx 0$ for goods sectors.

By contrast, sectors with perishable capacity exhibit substantially larger TFP elasticities with respect to sales, around 0.40 in columns (4)–(6). The price elasticity $\beta^{\text{micro}}_{\text{price}}$ remains indistinguishable from zero in both groups, in line with the aggregate results in Table 1.

Table 3: Regressions for $\beta^{\text{micro}}_{\text{match}}$ and $\beta^{\text{micro}}_{\text{price}}$ by Sectors

	Goods (NAICS≤45)			Service (NAICS≥48)		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln$ Sales	−0.007	0.014	0.007	0.430	0.392	0.383
	(0.034)	(0.065)	(0.047)	(0.072)	(0.206)	(0.125)
flex × $\Delta \ln$ Sales		−0.048			0.090	
		(0.103)			(0.420)	
flex^2 × $\Delta \ln$ Sales			−0.061			0.226
			(0.088)			(0.432)
KP Wald F	38.059	5.117	4.144	54.203	10.012	9.254
No. of Obs.	35,151	35,151	35,151	13,501	13,501	13,501

Note: Standard errors are two-way clustered at the firm and time levels. This table applies the same sample construction and specification as Table 1, splitting the data into goods-producing (NAICS≤45) and service-producing (NAICS≥48) industries. Kleibergen–Paap F-statistics are reported for instrument relevance across IV specifications. See Appendix B.5 for additional robustness results.

Takeaway. Across the time-series evidence in Figure 1 and the cross-sector regressions in Table 3, sectors with perishable capacity exhibit strong demand-driven variation in measured TFP and large values of $\beta^{\text{micro}}_{\text{match}}$, while goods sectors display neither pattern. This cross-sectional structure provides a clear empirical counterpart to the mechanism and reinforces the identification of $\beta^{\text{macro}}_{\text{TFP}}$ based on the four micro moments.

4. A Medium-Scale Model Suitable for Estimation

We now take our search model developed in Section 2 and empirically supported in Section 3 and transform it from a (compelling) story into a quantitative toolkit. This requires us to develop a dynamic environment with all of the features (i.e. bells and whistles) commonly used in medium-scale New Keynesian models for the purpose of structural estimation, as described, for instance, in the canonical model of Christiano et al. (2016). The features that we pose that are common to the literature are consumption habits, convex investment adjustment costs, Calvo-style nominal wage rigidity, fixed costs of production, and a Taylor Rule with interest rate smoothing.

We do have minor differences from Christiano et al. (2016). First, we use Rotemberg pricing instead of Calvo pricing in goods markets to achieve a more straightforward incorporation of our search frictions into the Phillips curve, avoiding the nuisance of having many prices in the market which complicates the search theory without any gain.[8] Second, we abstract from fiscal variables and working capital loans.

A feature that is standard in the literature is endogenous variable capital utilization: the intensity of capital usage is a costly voluntary decision that seemingly moves TFP. This feature is very different to ours and they should not be confused. Endogenous variable utilization moves TFP because capital and labor inputs are mismeasured. Shopping search frictions move TFP because household effort is not taken into account. We see this feature as an alternative mechanism to ours in so far as it is also capable of affecting measured productivity. So we want to compare both mechanisms. Our specifications have the property that a suitable choice of parameter values closes this mechanism while another parameter specification closes our mechanism which allows for a clean comparison.

To present all these standard features alongside our search frictions, we find it more convenient to use recursive language making the relevant objects functions of the aggregate state S which we will specify later. For expositional purposes, we now abstract from the stochastic trends in the layout of the model, assuming that all technology processes are stationary. We relax this assumption to address estimation issues as necessary. We highlight the changes on the firm side and the household sides relative to the standard setups in Sections 4.1 and 4.2, and briefly describe the rest of the model in Section 4.3. We provide a Section 4.4 to map our model variables to the NIPA account. Section 4.5 provides all required functional forms of the model for the estimation in Section 5.

4.1. Firms

In the medium-scale version of our model for estimation, several aspects of the firms need to be extended from the version in Section 2. Variable capital utilization and adjustment costs are incorporated into the household problem, where capital management is addressed. Firms, which are infinitely lived, hire labor and capital services every period for production via a constant returns-to-scale Cobb-Douglas function with fixed costs. The operations of the firms are subject to productivity shocks. The firm's sole dynamic concern is the costly price adjustment. Firms set prices given a demand of three components: household consumption, investment (both subject to search frictions), and firm's purchases of intermediate inputs from other firms (not subject to search frictions) to cover price adjustment costs.

We specify the production technology as a standard Cobb-Douglas function with a fixed cost ϑ every period:

$$F(k^s, n) = A\,(k^s)^{\alpha}\,(z^n\,n)^{1-\alpha} - \vartheta,$$

[8] In any case, the literature eliminates price dispersion when they log-linearize the model. In some literature where Calvo pricing is not convenient, such as the HANKs, Rotemberg pricing is also commonly used.

in which A is a normalization constant, k^s is rented capital services, n is hired labor services, and z^n is the level of neutral technology. The nominal marginal cost of production capacity $F(k^s, n)$ is $MC(S) = \frac{1}{A}\left[\frac{R^k(S)}{\alpha}\right]^\alpha \left[\frac{W(S)/z^n(S)}{1-\alpha}\right]^{1-\alpha}$ in which $W(S)$ is the nominal wage and $R^k(S)$ is the rental price of capital services.

The problem of the firm when specified recursively has as arguments the aggregate state and its own inherited price p^- from the previous period. We omit the description of the demand and tightness functions $\widetilde{c}(S, p)$ and $\widetilde{q}(S, p)$ as they respond to prices in the same way as in the static model of Section 2. We have

$$\Omega(S, p^-) = \max_p \left\{\psi^f\left[\widetilde{q}(S, p)\right]p - MC(S)\right\}\widetilde{y}(S, p) + [p - MC(S)]\,\widetilde{\chi}(S, p) \\ - MC(S)\,\vartheta - \chi(S, p, p^-) + \mathbb{E}\Big\{\mathcal{M}(S, S')\,\Omega(S', p)\Big\}. \quad (23)$$

Here, $\widetilde{q}(S, p)$ denotes the equilibrium tightness associated with price p as a firm choice. The household produces $\widetilde{y}(S, p)$ for consumption and investment, and $\widetilde{\chi}(S, p)$ for other firms' price adjustment costs, both of which depend on the firm's chosen price. Additionally, the firm incurs its own price adjustment costs, $\chi(S, p, p^-)$. The term $\mathcal{M}(S, S')$ refers to the stochastic discount factor.

4.2. Households

Households make decisions on total consumption and investment expenditures (with the search frictions having the same setup as in Section 2), as well as the level of capital services to extract from its capital stock —knowing that higher utilization rates accelerate depreciation— and on the quantity of bonds which are its means of saving.

Households purchase a variety of goods and aggregate them into consumption or investment goods using a CES aggregation technology with elasticity of substitution $\frac{\rho}{\rho-1}$ for both types of goods. The investment good is added to the capital stock, subject to capital adjustment costs that depend on the investment made in the previous period.[9] There is a union making the labor decisions, separability of the utility function into a search/consumption aggregate and a labor disutility term allows us to omit the specification of the latter. Households then care about a composite of an aggregate of consumption and aggregate search that we denote $m(c^A, d^A)$ and it is subject to (internal) habit formation. These considerations imply that the individual state of the household is its lagged consumption/search aggregate m^-, its stock of capital k, its previous period investment for capital accumulation x^{k-} (not including

[9] Appendix C.1 has additional details of how the of search frictions operate in medium-scale New Keynesian models.

the capital utilization costs), and its stock of bonds in real terms b. The household solves

$$v(S, m^{-}, x^{k-}, k, b) = \max_{\substack{c^A,\, d^A,\, m,\, b' \\ x^A,\, x^k,\, u,\, k'}} U\left[m\left(c^A, d^A\right) - \varsigma\, m^{-}\right] + \beta\, \mathbb{E}\Big\{v(S', m\left(c^A, d^A\right), x^k, k', b')\Big\} \tag{24}$$

$$\text{s.t.} \quad x^k = x^A - \widetilde{\delta}(u)\, k, \tag{25}$$

$$k' = (1-\delta)k + \left[1 - \mathcal{A}\left(\frac{x^k}{x^{k-}}\right)\right] x^k, \tag{26}$$

$$b' = w(S)L(S) + r^k(S)\, u\, k + \frac{R^{-}(S)}{\Pi(S)}\, b + T(S) - \Big\{d^A\, \psi^h\big[q(S)\big]\Big\}^{1-\rho}\Big(c^A + \frac{x^A}{z^x}\Big). \tag{27}$$

Here, u is the intensity of capital utilization, z^x is an investment-specific technology level that determines the rate of transformation between goods varieties for consumption purposes and those for investment purposes.

Households' objective (24) is standard, and constraints (25)-(27) describe the evolution of individual state variables. Equation (25) shows that increasing capital utilization intensity requires additional cost of maintenance $\widetilde{\delta}(u)$. We follow Christiano et al. (2016) so that this particular form of maintenance does not increase the investment adjustment costs. Equation (26) shows how investment increases capital next period, with $\mathcal{A}$ being the investment adjustment cost function convex in investment growth. Finally, equation (27) is the budget constraint in real terms, in which the sources of income include labor income earned at the real wage rate $w(S)$ from working the labor services determined by the union wage choice $L(S)$, capital income at real rental price $r^k(S)$ with the household being able to extract more capital services by pumping up its utilization; realized bond returns which depend on inflation $\Pi(S)$ transfers $T(S)$ including the government lump-sum transfer and the profits distributed by firms. The last term of the budget describes expenditures on consumption and investment, measured in units of consumption aggregates c^A. It is determined by the number of varieties $d^A\, \psi^h\big[q(S)\big]$ to the power of $1-\rho$ as the love-for-variety preference can transform fewer goods varieties into the same consumption aggregates when households search more.

4.3. Standard Parts of the Model

Calvo Wage Rigidity As in standard Calvo wage models, there is a unit measure of differentiated labor varieties supplied by the representative household, and a union in each labor variety that sets the wage monopolistically with a gross wage markup ρ_w. Every period, a random fraction θ_w of the wages cannot change. The union takes this rigidity into account when setting the wage dynamically on behalf of the household.

Monetary Policy Monetary policy follows a Taylor Rule that sets the nominal interest rate. As in Christiano et al. (2016), the shock in any period happens too late to affect current period choices. The innovation ϵ^R is paid to bondholders in the following period via the use of lump-sum taxes or subsidies. The interest rate that includes such innovation is $R(S, \epsilon^R)$, given by

$$R(S, \epsilon^R) = \rho_R \cdot \ln\left(\frac{R^-(S)}{R_{ss}}\right) + (1 - \rho_R) \cdot \left[\phi_\pi \cdot \ln\left(\frac{\Pi(S)}{\Pi_{ss}}\right) + \phi_y \cdot \ln\left(\frac{Y(S)}{Y_{ss}}\right)\right] + \epsilon^R, \tag{28}$$

where $R^-(S) = R(S^-, \epsilon^{R-})$. The i.i.d. innovation is Gaussian, satisfying $\epsilon^R \sim \mathcal{N}(0, \sigma_R)$, with σ_R capturing the size of the innovation.

Aggregate States They include technology shocks z^n and z^x, the inherited wage[10] and price W^- and P^-, and the realized interest rate from the prior period R, which incorporates the monetary innovation ϵ^{R-}. Due to its i.i.d. nature, the previous monetary policy shock ϵ^{R-} needs not be included separately in the state. Additionally, S contains the aggregate counterparts of the individual state variables: the aggregate habit stock M, last period's investment X^{k-} (which determines adjustment costs), the aggregate capital stock K. In summary,

$$S = (z^n, z^x, W^-, P^-, R, M, X^{k-}, K).$$

Equilibrium Equilibrium is standard. Agents maximize; the search friction is resolved as described in Section 2; capital services and bonds markets clear; the government balances its budget while following the Taylor rule.

4.4. National Income and Products Accounts

Consumption, Investment, and GDP In this economy, national accounting differs from the standard measures in several ways due to the love-for-variety preference. Aggregate consumption is not C^A (the effective consumption relevant to households), but rather $C = \psi^f(Q)\,\widetilde{c} = \psi^f(Q)^{1-\rho}\,C^A$. This discrepancy arises from the aggregation of product varieties and the search mechanisms. Similarly, measured investment is defined as $I = \psi^f(Q)^{1-\rho}\,X^k$. Maintenance investment for capital utilization cost is not part of GDP, as it is considered an intermediate input, as in Christiano et al. (2016).

[10] As is standard in the New Keynesian literature we substitute the wage dispersion for its lagged value because at log-linearization time it does not matter.

GDP Search frictions determine the relationship between actual and potential GDP. Actual GDP equals potential GDP multiplied by the occupancy rate ψ^f, which allows us to define the actual GDP as

$$Y = \psi^f \left[A\,(u\,K)^\alpha\,(z^n\,L)^{1-\alpha} - \vartheta - \text{price adjustment terms}\right] - \frac{(\psi^f)^{1-\rho}}{z^x}\,\widetilde{\delta}(u)\,K.$$

Note the two additional terms that appear relative to the steady state. The first captures the resources used for price adjustments. The last term reflects the extra depreciation resulting from capital utilization that differs from its steady-state level. Here, $\frac{1}{z^x}$ is relative price of investment, and $\widetilde{\delta}(u)$ is a depreciation function that satisfies $\widetilde{\delta}(1) = 0$.[11]

Three Types of Price Indices Defining price indices requires caution for two primary reasons. First, under love-of-variety preferences, transaction prices do not fully reflect utility-relevant prices. Second, investment-specific technology shocks drive a wedge between the GDP deflator and the Personal Consumption Expenditures (PCE) price index.

To address the first issue, we use the producer price p as our PCE index. This price is identical across firms in equilibrium and is the price paid by households, providing a consistent benchmark. We define the gross PCE inflation as $\Pi \equiv \frac{p}{p^-}$, which corresponds to the Federal Reserve's inflation target.

To address the second issue, we define the Investment Price Index as $\frac{p}{z^x}$, which measures the relative price of investment in terms of consumption varieties given investment-specific technology z^x. We then define the GDP deflator as a weighted average of the PCE price index and the Investment Price Index.

Accordingly, in both our VAR and our model, we deflate GDP, consumption, and wages by the GDP deflator, and investment further by the relative price of investment. This aligns the data and the model, capturing increases in real GDP caused by investment-specific technology improvement.

4.5. Functional Forms

Preferences We assume a log function over a GHH structure of consumption and search effort with an internal habit that enters additively, and an additively separable working disutility with constant Frisch elasticity $1/\xi$:

$$U(m, m^-) - V(\ell) = \ln(m - \varsigma\, m^-) - \eta\,\frac{\ell^{1+\xi}}{1+\xi} \quad \text{in which} \quad m \equiv c^A - \zeta\, Z\,\frac{(d^A)^{1+\nu}}{1+\nu},$$

[11] This specification of GDP leads to accounting rules which treat the additional depreciation from capital utilization as an intermediate input expense. In Section 7, we examine the case in which we classify this expense as an investment.

with c^A being the aggregate over varieties, and d^A being the aggregate of search effort in all markets. They are defined in equations (1) and (2). The scaling factor Z, which we call the technology diffusion factor following Christiano et al. (2016), satisfies

$$\ln(Z) = 0.99\,\ln(Z^-) + 0.01\,(1-\iota)\,\ln(z^y),$$

where z^y is the level of composite technology in the overall economy that has nearly zero impact on Z in the short run but a full impact in the long run. We will provide more details about z^y later on.

Variable Capital Utilization In addition to a constant depreciation rate δ, the intensity of capital utilization affects the depreciation of each unit of capital in amount

$$\widetilde{\delta}(u) \equiv \sigma_a\sigma_b/2\cdot(u-1)^2 + \sigma_b\cdot(u-1).$$

Note that in the steady state, $\widetilde{\delta}(1) = 0$, and the marginal depreciation is $\sigma_b \equiv \widetilde{\delta}_u(1)$, while $\sigma_a \equiv 1\cdot\widetilde{\delta}_{uu}(1)/\widetilde{\delta}_u(1)$ captures the elasticity of this marginal cost with respect to the capital utilization rate u. It is essential to distinguish between capital utilization and capacity utilization. In line with the definition provided by the Board of Governors of the Federal Reserve System, in which capacity refers to "the highest level of output a plant can sustain within the confines of its resources", we define capacity utilization rate in the model as: $util \equiv \frac{\psi^f\cdot(u^\alpha\cdot A\,K^\alpha L^{1-\alpha}-\vartheta)}{A\,K^\alpha L^{1-\alpha}-\vartheta}$. Under log-linearization, it becomes $d\ln(util) = d\ln(\psi^f) + (1+\vartheta\psi^f_{ss})\,\alpha\cdot d\ln(u)$.

Investment Goods Each of the varieties is used for both consumption and investment. The household exerts search effort, finds varieties, and purchases some quantity of each variety for consumption and some for investment. Investment goods are then an aggregate of varieties purchased by the household and allocated for investment purposes:

$$x^A = \Big(\int_\Phi d(p,q)\,\psi^h(q)\;x(p,q)^{\frac{1}{\rho}}\;dp\,dq\Big)^\rho.$$

Adjustment Costs to Investment Christiano et al. (2016) pose

$$\mathcal{A}\Big(\frac{x^k}{x^{k-}}\Big) \equiv \frac{1}{2}\cdot\Big\{\exp\Big[\sqrt{\widetilde{\mathcal{A}}}\,\Big(\frac{x^k}{x^{k-}}-\mu^k_{ss}\Big)\Big] + \exp\Big[-\sqrt{\widetilde{\mathcal{A}}}\,\Big(\frac{x^k}{x^{k-}}-\mu^k_{ss}\Big)\Big]\Big\} - 1,$$

In steady-state, $\frac{x^k}{x^{k-}} = (\mu^x_{ss})^{\frac{1}{1-\alpha}}\mu^y_{ss}$, we have $\mathcal{A}(\cdot) = \mathcal{A}'(\cdot) = 0$ and $\mathcal{A}''(\cdot) = \widetilde{\mathcal{A}}$.

Matching Function We pose a CRS-CES matching function, which can be specified directly in terms of varieties J (of which we assume there are 1) and search effort D, or simply in terms of market tightness q is

$$\psi(J, D) = B \cdot \left[(1-\varphi)\, J^{-\gamma} + \varphi\, D^{-\gamma}\right]^{-\frac{1}{\gamma}},$$

for some parameters $\{B, \varphi, \gamma\}$.

Adjustment Costs of Goods Prices The cost of changing the nominal price of a good is

$$\chi(S, p, p^{-}) = \frac{\kappa}{2} \cdot \left(\frac{p}{p^{-}} - \Pi_{ss}\right)^2 \cdot P(S) \cdot \psi^f\left[Q(S)\right] \cdot \widetilde{y}\left[S, P(S)\right].$$

where $\widetilde{y}\left[S, P(S)\right]$ is the quantity sold for consumption and investment purposes (not for price adjustment costs), and $P(S) \cdot \psi^f[Q(S)] \cdot \widetilde{y}[S, P(S)]$ normalizes the quadratic cost to a fraction of GDP.

Technological Progress While we have described the model as if it were stationary to avoid cumbersome notation, the actual technology processes follow stochastic trends. Following Christiano et al. (2016), we pose processes for a neutral technology level z^n, and for the investment-specific technology level z^x. Their gross growth rates $\mu^n = \frac{z^n}{z^{n-}}$ and $\mu^x = \frac{z^x}{z^{x-}}$ are assumed stationary and are described by

$$\begin{aligned}
\ln(\mu^n) - \ln(\mu^n_{ss}) &= \sigma_n \cdot \epsilon^n, \\
\ln(\mu^x) - \ln(\mu^x_{ss}) &= \rho_x \cdot \left[\ln(\mu^{x-}) - \ln(\mu^x_{ss})\right] + \sigma_x \cdot \epsilon^x.
\end{aligned}$$

Both innovations ϵ^n and ϵ^x are i.i.d. standard Gaussian. The overall technology level is $z^y = z^n \cdot (z^x)^{\frac{\alpha}{1-\alpha}}$.

5. Mapping the Model to Macroeconomic Data

This section takes the medium-scale versions of our search model to macroeconomic data. Our primary goal is to compare our search friction model with conventional models with and without the capital utilization intensity mechanism in terms of matching the empirical impulse responses à la Christiano et al. (2016). To do so reliably, we do not add degrees of freedom to the estimation of the search friction parameters. We calibrate all the parameters associated with our search frictions based on micro-evidence before estimation. These are parameters that determine the steady state properties. In contrast, the widely used capital utilization intensity typically introduces a free parameter, the elasticity of marginal utilization cost, σ_a, to improve the impulse response matching performance. We will show that even under this advantageous setup for capital utilization intensity, our search-based approach still delivers superior performance along multiple dimensions.

We estimate four versions of the model for comparison, starting with the one that incorporates no search or capital utilization ("None"). We then introduce search friction alone ("Search Alone"), that stands as our baseline in the second economy. Capital utilization alone is the third version ("u Alone"). The fourth version combines both search friction and capital utilization ("Search and u"). All versions of the model can be specified with the same functional forms choosing parameter values that turn on or off the various ingredients. They are calibrated to the same steady-state aggregate targets and are then estimated to replicate the properties of the same empirical impulse response functions. We start by summarizing the model calibration in Section 5.1. Section 5.2 briefly describes our structural VAR for empirical impulse response before we proceed with the estimation in Section 5.3.

5.1. Calibration and Targets

There are a total of 29 parameters to specify, of which 17 are calibrated to target steady-state moments, while the remaining 12 will be estimated via impulse response matching. Table 4 presents the values of the 17 calibrated parameters alongside the corresponding targets (this association is not always one-to-one, so for those that are not we associate the targets to the parameters that we think speak the most about each other). We now turn to describe them in groups.

Table 4: Calibration to Target Steady-State Moments in the Four Economies

Parameter (quarterly model)	Symbol	None	Search Alone	u Alone	Search and u	Steady-state (annual data)	Target
				Search Frictions Parameters and Targets			
Convexity of search cost	ν	∞	1.836	∞	1.836	Elast. of search activity to spending	0.50 or 0
Share of search in matching	φ	0	0.270	0	0.270	Elast. of sales-to-COGS to sales	0.27 or 0
Complementary in matching	γ	1	1	1	1	Elast. of desired price to sales	0
Love-for-variety preference	ρ	1.341	1.836	1.341	1.836	Average firm markup	34.%
Efficiency of matching fun.	B	1	0.814	1	0.814	Occupancy rate	0.814 or 1.
				Other Frictions			
Gross wage markup	ρ_w	1.200	1.200	1.200	1.200	Wage markup	20.%
Prob. of no wage changes	θ_w	0.750	0.750	0.750	0.750	Wage duration	1.
				Growth Rates			
BGP growth of neutral tech.	μ^n_{ss}	.3%	.3%	.3%	.3%	Output growth per capita	1.7%
BGP growth of invest. tech.	μ^x_{ss}	.3%	.3%	.3%	.3%	Capital growth per capita	2.9%
				Aggregate Production Properties			
Discount factor	β	0.997	0.997	0.997	0.997	Real interest rate	3.0%
Capital depreciation rate	δ	0.025	0.025	0.025	0.025	Investment share	26.1%
Capital share in production	α	0.301	0.301	0.301	0.301	Capital-output ratio	2.
Fixed cost share in GDP	$B\vartheta$	0.278	0.278	0.278	0.278	Labor share	66.7%
Marginal cost of utilization	σ_b	0.036	0.036	0.036	0.036	Utilization rate	1.
				Normalization Parameters and Targets			
Efficiency of production fun.	A	0.858	0.892	0.858	0.892	Output level	1.
Weight of labor in utility	η	0.735	1.320	0.716	1.303	Working hours	1.
Weight of search in utility	ζ	0.485	1.211	0.485	1.211	Market tightness	1.
				Some Non Targeted Properties			
Profit share in GDP		4.8%	4.8%	4.8%	4.8%		-

Search Frictions Parameters and Targets (5) As discussed in Section 3.1 we use micro elasticities estimates of search activity to spending (.5), of sales-to-COGS (cost of goods sold) ratio(.32) and of desired prices to sales (-.15) together with the average firm markup (34%) to discipline the model.[12] One additional target is needed to solve the model which is relevant for the steady state but not for the cyclical behavior and this is the average capacity utilization. For this we use the average of the series "Capacity Utilization: Total Index" from the Board of Governors of the Federal Reserve System (US)

[12] Recall that the elasticity of desired prices to sales was effectively indistinguishable from zero and in the robustness section we explore this value.

from 1967 to 2007. In the economies without search the elasticities are zero and the occupancy rate is 1. The parameters involved are $\{\nu, \varphi, \gamma, \rho, B\}$.[13] The cases with no search are nested when all three micro elasticities approach zero. Recall that we choose to target the same markup across all versions of the model, which requires the search models to have a higher taste for varieties parameter. This ensures that all economies have the same profit share and fixed costs.[14] Once we specify these targets the only other object left in the model that is related to search is the weight in the utility function, but this is just a units parameter with no relevance for fluctuations. Given these targets, the estimation process of the two search economies has no new parameters to improve the fits.

Other Frictions (2) The labor market is also subject to wage setting frictions a la Calvo with wages changed once a year on average and an average markup of 20% (as in Christiano et al. (2016)) sufficient to ensure that there are no problems with workers' compliance (Huo and Ríos-Rull (2020)).

Secular Growth Rates (2) The annual long-term growth rates of real GDP and capital per capita are 1.7% and 2.9%, respectively, which are taken from Christiano et al. (2016). To implement the balanced growth path, we set the two trends of the technology μ^n_{ss} and μ^x_{ss} to satisfy $2.9\%/4 = \ln(\mu^x_{ss}) + \ln(\mu^n_{ss})$, and $1.7\%/4 = \frac{\alpha}{1-\alpha}\ln(\mu^x_{ss}) + \ln(\mu^n_{ss})$.

Aggregate Production Proerties (5) We set the long-term annual real interest rate to 3%. We target a capital-output ratio of 2, a labor share of 2/3, and an investment share of 26.1%. In particular, the investment share is defined as the sum of private domestic investment share in GDP (BEA: A006RE) and the durable goods share in GDP (BEA: DDURRE) for the period 1951-2007. We set the steady-state capital use intensity at 1. This pins down parameters $\{\beta, \delta, \alpha, \vartheta, \sigma_b\}$.

Normalization (3) There are various parameters that have no relevance for the findings since they are just unit parameters or they play no role given the log-linearization solution approach that we follow. We normalize steady state output, labor and market tightness all to 1, all of which determine efficiency in the production function, and weights of search and labor in the utility function. There are two additional parameters that play no role in the estimation, the efficiency of the matching function and the marginal cost of capital utilization, but we have already described them in the search and production targets.

Other Untargeted Steady-State Moments There are some properties of the steady states that we did not target but that are relevant when thinking of an aggregate economy. Here we report the implied

[13] In Appendix C.1, we show that the mapping from these empirical moments to the four parameters of search does not change when we extend the static model in Section 2 to the full-fledged model in Section 4.

[14] In medium-scale New Keynesian models à la Christiano et al. (2016), the markups are estimated and hence different across models. The strength of the fixed cost channel in driving the productivity movement also varies across models.

steady-state profit share (4.8%, which is slightly higher than but still close to various calculations in De Loecker et al. (2020))

5.2. Structural VAR

Our structural VAR regression primarily follows the procedure of Christiano et al. (2010, 2016), with 14 observed aggregate time series, 2 lags as controls, and 3 identified structural shocks. In the regression stage, invertible linear combinations of the aggregate time series are constructed to make the regressors stationary. In the impulse response stage, we invert the linear transformation to back out the dynamics of the original aggregate time series. Still, there are two differences from Christiano et al. (2016).

First, we include the impulse response of both labor productivity and labor share to highlight two crucial empirical patterns that are often ignored in medium-scale New Keynesian models: that labor productivity is procyclical and that the labor share is countercyclical (except for the first few periods after a monetary policy shock). The models that we explore and especially the mechanisms that we are interested in (the shopping search friction and factor intensity variability) have implications for these variables and we think that they should be taken into account.

Second, for consistency between the model and the data, we define real GDP, real consumption, and real wages all as their nominal counterparts denominated by the GDP deflator. The real investment is denominated by both the GDP deflator and the relative price of investment.[15]

5.3. Impulse Response Matching

There are 11 additional parameters to be determined (12 in the case of capacity utilization), the values of which are obtained through estimation by matching the impulse responses of various time series to the three shocks we introduce. A key feature of these shocks is that, given the model specifications, they are observable, which allows us to construct their impulse responses within the model and compare them with the counterparts in the data.[16] These parameters are standard in medium-scale New Keynesian models, and not directly related to our search frictions, thus we choose the same prior distributions as in Christiano et al. (2016) for most of them. There are three exceptions. First, we calibrated the fixed costs ϑ and the love for variety parameter ρ to ensure that the role of fixed costs in shaping productivity movements is the same in all versions of the model. In contrast, Christiano et al. (2016) directly estimate the markup parameter ρ and solve out the value of ϑ to ensure zero steady-state profit

[15] More details about the VAR regression are provided in Section 5.2.

[16] For further details, see Christiano et al. (2016). Our estimation essentially follows the "Calvo wage" version of models presented in that study, with minor differences in steady-state calibration.

which is a slightly different calibration strategy. Second, we directly estimate the slope of the Phillips curve (NKPC) on real marginal cost $\widetilde{\kappa}$, as we adopt Rotemberg pricing to replace Calvo pricing. Third, we shrink the prior of the inverse of Frisch elasticity ξ aiming to obtain an estimate that is centered in the prior interval to heavily penalize implausible Frisch elasticity values.

The estimation results are summarized in Table 5, presenting the priors and posteriors of parameters. The posteriors of parameters are generally consistent across specifications, with two notable exceptions: a smaller size of neutral technology shocks σ_n and a larger Taylor rule coefficient on GDP, ϕ_y, when the search friction is included.[17] The first property points to the search friction as providing an amplification mechanism, as smaller shocks are needed to account for the variation of aggregate variables. We like this feature as now the variance of the technology shocks needed to account for their impulse response is less than half of what is needed in the standard model and less than two thirds of what the capacity utilization needs. Our search mechanism also reveals a more dovish Central Bank: it responds more aggressively to output.

We also report the cyclicality of markups (regressing log changes of markups on that of real GDP) when all three shocks hit the economies. Search frictions and capital utilization both make markups more procyclical, but none of the estimates are significantly different from zero, which is consistent with the lack of agreement in the literature about their actual behavior.

What we think is the highlight of Table 5 are the goodness of fit measures comparing the performance of all four versions of the model. We use two types of measures. The first one is the log marginal likelihood, a Bayesian metric that evaluates the model performance weighted over the prior distribution. The second is the log likelihood computed at the posterior mode, which excludes the influence of priors, thereby providing a direct measure of the distance between the model and the data. This distance is further decomposed by the three structural shocks, allowing for a more detailed diagnostic.

The ranking of models is consistent across all measures, with "Search and *u*" performing the best, followed by "Search Alone", then "*u* Alone", and finally "None". The fair comparisons are those between "None" and "Search Alone" and between "*u* alone" and "Search and *u*" since these pairs have the same number of parameters. In both cases,the search economy does a lot better than its counterpart without search. This is the case even when we compare the Search Alone with the u Alone economies with the Search Alone doing much better and more than satisfying the Kass and Raftery (1995) criteria for very strong evidence of one model over another.

Economically, the improved likelihood reflects the ability of the search mechanism to amplify demand

[17] Other parameters all have overlapping 95% credible intervals in theposterior distribution when comparing "Search Alone" with "None", or "*u* Alone" with "Search and *u*".

shocks through occupancy-driven total factor productivity (TFP). This allows the model to match key empirical responses of output, labor productivity, and labor share to monetary and technology shocks, as shown in Figures 2 and 4.

Table 5: Results of the Bayesian Impulse Response Matching Estimation

		None	Search Alone	u Alone	Search and u
	Prior Distribution $\mathcal{D}$,**Mode**,[2.5-97.5%]	Posterior Distribution **Mode**,[2.5-97.5%]			
Convexity of util. cost, σ_a	$\mathcal{G}$,**0.32**,[0.09-1.24]	$+\infty$	$+\infty$	**0.25**,[0.13-0.41]	**0.39**,[0.18-0.59]
Slope of NKPC on mc, $\widetilde{\kappa}$	$\mathcal{G}$,**0.10**,[0.06-0.14]	**0.07**,[0.05-0.09]	**0.04**,[0.02-0.06]	**0.07**,[0.05-0.10]	**0.09**,[0.06-0.12]
Inverse of Frisch, ξ	$\mathcal{G}$,**0.99**,[0.81-1.21]	**0.52**,[0.43-0.62]	**0.47**,[0.41-0.55]	**0.77**,[0.63-0.93]	**0.63**,[0.53-0.84]
Internal habit, ς	$\mathcal{B}$,**0.50**,[0.21-0.79]	**0.77**,[0.73-0.81]	**0.75**,[0.69-0.80]	**0.72**,[0.68-0.76]	**0.75**,[0.70-0.81]
Investment adj. cost, $\widetilde{\mathcal{A}}$	$\mathcal{G}$,**7.50**,[4.57-12.4]	**6.65**,[4.57-9.20]	**6.51**,[4.09-8.29]	**5.35**,[3.49-6.87]	**4.63**,[3.34-6.24]
Taylor rule: inflation, ϕ_π	$\mathcal{G}$,**1.69**,[1.42-2.01]	**2.59**,[2.28-2.87]	**2.13**,[1.88-2.41]	**2.17**,[2.01-2.60]	**2.13**,[1.98-2.53]
Taylor rule: GDP, ϕ_y	$\mathcal{G}$,**0.08**,[0.03-0.22]	**0.01**,[0.00-0.03]	**0.17**,[0.13-0.20]	**0.01**,[0.00-0.02]	**0.06**,[0.03-0.09]
Taylor rule: smoothing, ρ_R	$\mathcal{B}$,**0.76**,[0.37-0.94]	**0.82**,[0.80-0.85]	**0.83**,[0.80-0.85]	**0.79**,[0.75-0.81]	**0.76**,[0.73-0.80]
Std. monetary policy, $400\sigma_R$	$\mathcal{G}$,**0.65**,[0.56-0.75]	**0.69**,[0.59-0.74]	**0.68**,[0.59-0.73]	**0.67**,[0.61-0.75]	**0.66**,[0.61-0.75]
Std. neutral tech., $100\sigma_n$	$\mathcal{G}$,**0.08**,[0.03-0.22]	**0.43**,[0.40-0.46]	**0.30**,[0.26-0.32]	**0.37**,[0.34-0.39]	**0.27**,[0.24-0.29]
Std. investment tech., $100\sigma_x$	$\mathcal{G}$,**0.08**,[0.03-0.22]	**0.31**,[0.26-0.37]	**0.26**,[0.22-0.32]	**0.26**,[0.22-0.32]	**0.25**,[0.20-0.30]
Investment tech persistence, ρ_x	$\mathcal{B}$,**0.78**,[0.53-0.92]	**0.51**,[0.39-0.60]	**0.50**,[0.37-0.59]	**0.49**,[0.35-0.58]	**0.45**,[0.33-0.57]
Log marginal likelihood	–	**204.5**	**295.1**	**274.9**	**346.8**
Log likelihood at mode	–	**267.6**	**355.5**	**317.7**	**383.5**
– monetary policy shocks	–	98.7	121.0	117.4	143.4
– neutral technology shocks	–	106.7	140.8	135.2	171.6
– inv.-specific tech. shocks	–	62.2	93.7	65.1	68.4
Markup cyclicality (3 shocks)	–	**−0.136**	**−0.053**	**−0.014**	**+0.081**
	–	[−0.311, +0.004]	[−0.197, +0.076]	[−0.145, +0.087]	[−0.007, +0.153]

Note: The posterior distribution is simulated using a totally of 100,000 draws with the first 50% of each chain burned. The unconditional cyclicality of markups is measured by regressing the log difference of gross markups on the log difference of real output, using annual data simulated at the posterior mode. We simulate it 200 times, using a sample of 25 years, which is the sample length of Burstein et al. (2025), to construct the 95% 2.5th and 97.5th percentiles.

We can see directly the models performances, in terms of their impulses responses vis a vis the data in Figures 2 to 4 that show impulse responses of all model versions against the data counterpart, as well as the untargeted responses of capital utilization, occupancy rate, TFP, and gross markup.

In the case of monetary policy shocks, the primary improvement is observed in the impulse responses of labor productivity and labor share. Specifically, the "Search and u" economy successfully generates procyclical labor productivity and countercyclical labor share, without inducing countercyclical inflation.

In contrast, economies without search frictions typically generate weaker responses in labor productivity and procyclical responses in the labor share.

For neutral technology shocks, although all economies feature countercyclical labor shares, the two that incorporate search frictions align more closely with the data.

Regarding investment-specific technology shocks, the economies with search frictions produce more substantial fluctuations in hours and consumption, thereby mitigating the "comovement puzzle" discussed in Justiniano et al. (2010). The responses of labor share also align more with the data.

Overall, these results confirm that the search mechanism significantly enhances the performance of standard medium-scale New Keynesian models in terms of matching aggregate time series data. While the widely used capital utilization mechanism remains useful, it is less effective in comparison.

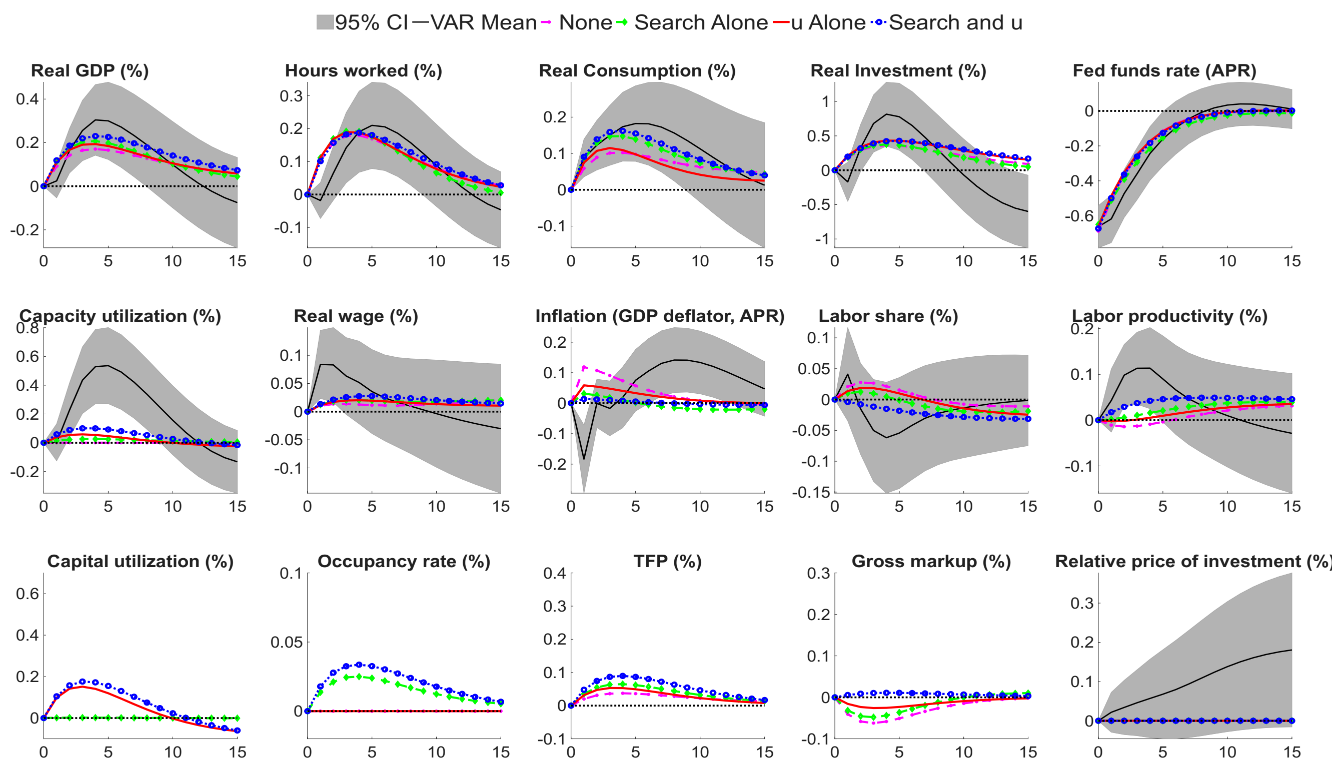


Figure 2: Impulse Response Matching for Monetary Policy Shocks

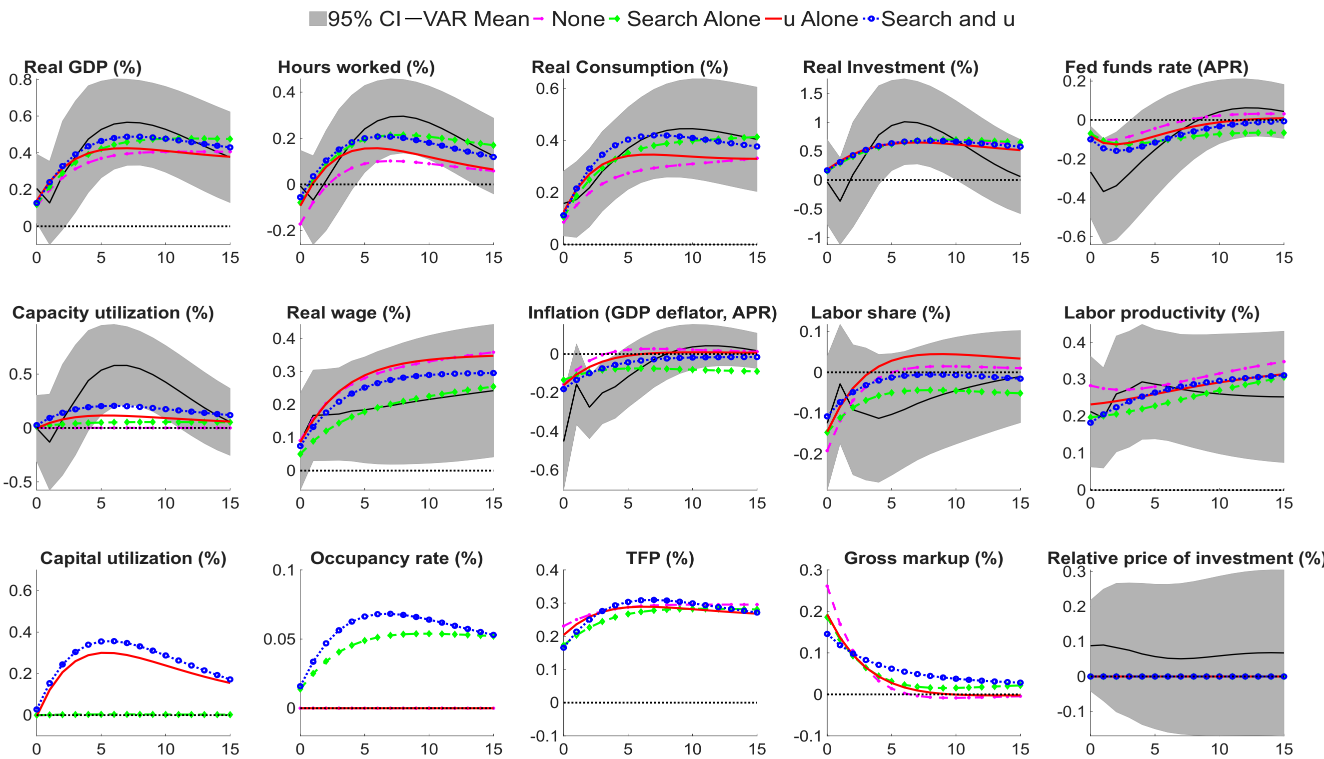


Figure 3: Impulse Response Matching for Neutral Technology Shocks

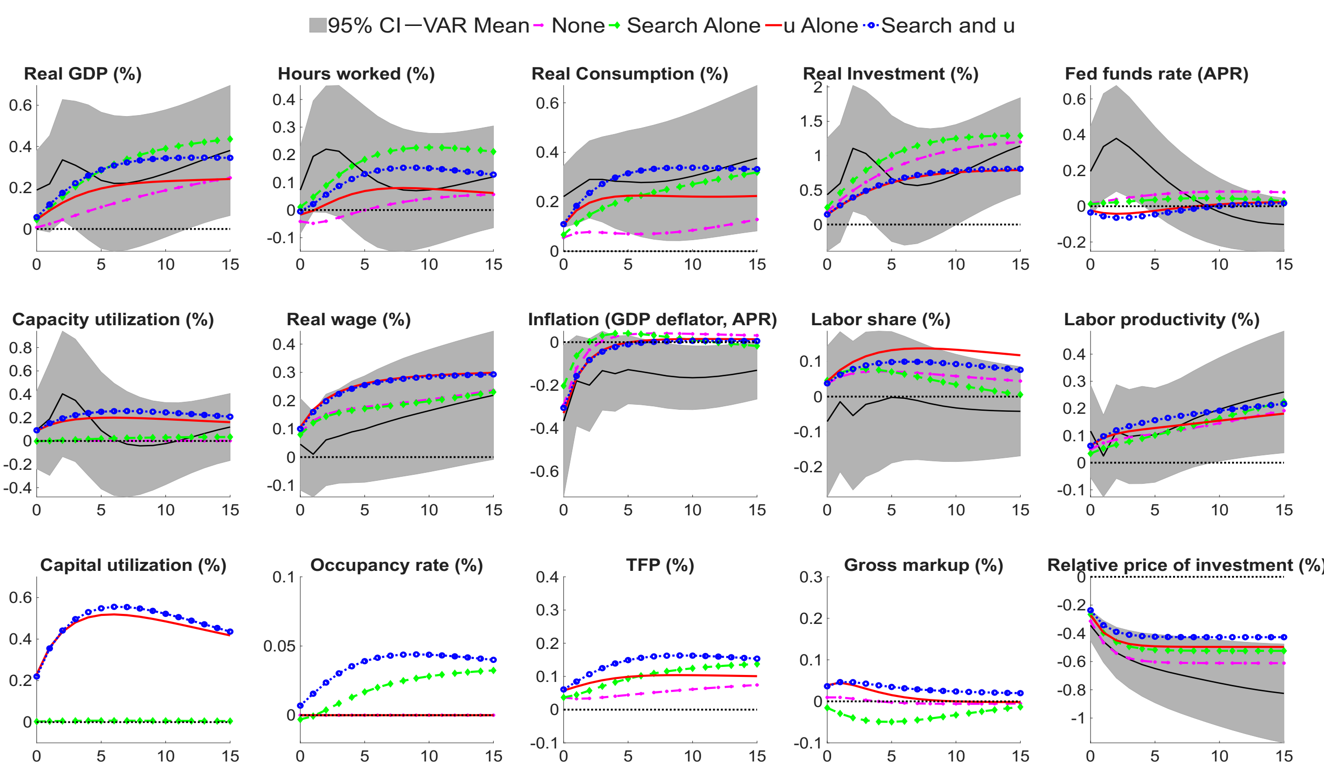


Figure 4: Impulse Response Matching for Investment-Specific Technology Shocks

6. Implications of the Estimated Model

This section quantitatively assesses the aggregate implications of TFP, markup, and inflation cyclicality in our estimated medium-scale model, demonstrating the economic significance of Corollaries 1-3 when all macroeconomic forces operate simultaneously.

6.1. Occupancy Rate and TFP in the Transmission of Monetary Policy Shocks

We decompose the response of Federal Funds Rate shocks to output into multiple supply-side channels, including the movements of TFP driven by occupancy rate, costly capital utilization intensity, fixed cost of production, capital input, and, labor input. Figure 5 demonstrates the decomposition results for two versions of our estimated models: the "Search Alone" model and the "Search and *u*" model.

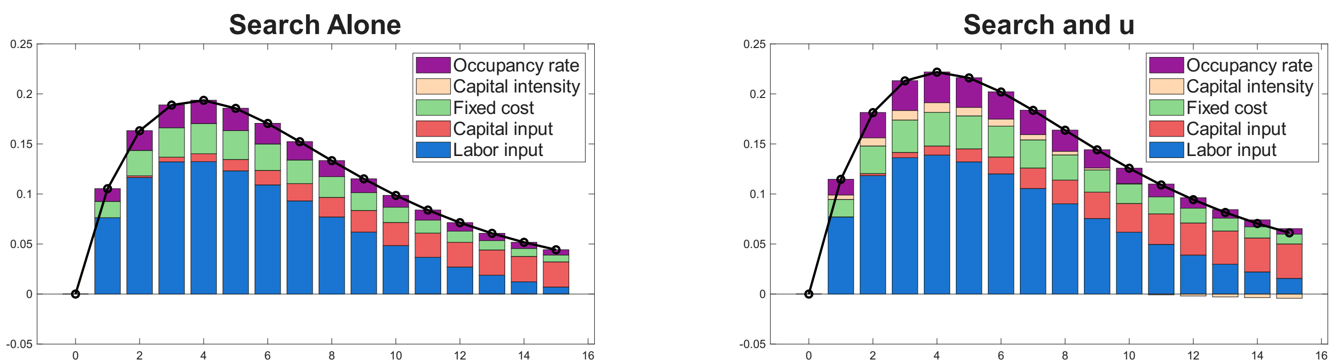


Figure 5: Transmission of Monetary Policy Shocks

In both models, labor input accounts for more than half of the output response, while capital input starts to play a mild role after one year. The occupancy rate is nearly as important as the fixed cost of production, and capital utilization intensity only plays a minor role.[18]

In the "Search Alone" model, we find the occupancy rate accounts for 10% of the output response across all horizons, which is consistent with the prediction of Corollary 1 with $\beta^{macro}_{TFP} = 0.10$, the implied value from micro evidence in Section 3.1.

In the "Search and *u*" model, the contribution of occupancy rate to output is roughly 12% around the one-year horizon, and drops to about 8% after three years. The slightly higher number 12% reflects a mild issue when the cost of capital utilization is treated as an intermediate input.[19]

[18] The reason why capital utilization intensity only plays a minor role is due to the omission of its cost in GDP accounting, due to the assumption that utilization cost is a part of intermediate inputs rather than investment. In Appendix F.5, we also estimate the model in which utilization cost is treated as investment. In this model, capital utilization intensity drops immediately after an expansionary monetary policy shock, so as to reduce investment adjustment cost.

[19] Appendix E.1 provides more details about this issue.

6.2. Comparison of Markup Cyclicality with the Literature

This section examines the cyclicality of markups in our estimated model relative to the existing literature. We focus our comparison on two notable prior studies: Nekarda and Ramey (2020) and Burstein et al. (2025), which explore the cyclicality of markups conditional on different structural shocks and at various aggregation levels, respectively.

6.2.1. Conditional Cyclicality of Markups

We compare the conditional markup cyclicality in our estimated medium-scale model with the empirical findings of Nekarda and Ramey (2020). Both of us measure the gross markup with the inverse of labor share, and consider the monetary policy, neutral technology, and investment-specific technology shocks. The comparison is summarized in the following Table 6, which indicates that our search frictions improve the standard "u Alone" model by turning the markup cyclicality conditional on monetary policy shocks closer to the empirical findings of Nekarda and Ramey (2020).

Table 6: Conditional Cyclicality of Markups (Measured as Inverse Labor Share)

	Monetary policy shocks	Neutral tech. shocks	Invest. tech. shocks
Nekarda and Ramey (2020)	Procyclical	Procyclical	Countercyclical
u Alone Model	Countercyclical	Procyclical	Countercyclical
Search and u Model	Weakly procyclical	Procyclical	Countercyclical

6.2.2. Bottom-up Cyclicality of Markups

Burstein et al. (2025) measure markup cyclicality by regressing firm-level markups on sector-level output, sector-level markups on sector-level output, and sector-level markups on aggregate-level output. Before comparing our model-implied markup cyclicality with their empirical results, several considerations must be addressed to ensure an apples-to-apples comparison.

The first consideration is the measure of markup cyclicality. Burstein et al. (2025) measure markups based on intermediate inputs following the approach of Bils et al. (2018), while we use the cost of goods sold (COGS) following De Loecker et al. (2020), which include both intermediate input and labor costs. Moreover, Burstein et al. (2025) estimate the elasticity of the production function taking advantage of separate price and quantity data, while we assume that the elasticity of production function is constant. Our measuring approach and underlying assumption reflect the limitation of Compustat data. However, Compustat covers firms in service sectors at a quarterly frequency, which is necessary for our studies.

To ensure that our measure of markup cyclicality is comparable to Burstein et al. (2025) we replicate their cyclicality regressions using our measure of markups in Compustat data, restricting our sample to manufacturing sectors and aggregating the quarterly data to annual ones. The comparison of regression results is in the following Table 7.

Table 7: Replicating Burstein et al. (2025) with Compustat Manufacturing Sectors

	Firm on sector		Sector on sector		Sector on aggregate	
	Burstein	Our results	Burstein	Our results	Burstein	Our results
$\Delta \ln(Y)$	−0.024	0.004	0.160	0.080	−0.239	0.050
	(0.009)	(0.011)	(0.040)	(0.027)	(0.116)	(0.026)
$\Delta \ln(Y) \times$ market share	0.280	0.182				
	(0.040)	(0.090)				
Fixed effects	None	None	2 Way	2 Way	Sector	Sector
No. of sectors	275	86	275	86	275	86
No. of obs.	8,051,767	94,599	6,875	2,921	6,875	2,921

Note: Our sample range is 1984-2019, longer than 1995-2019 in Burstein et al. (2025). To ensure sufficient number of observations within sectors in Compustat, we choose a 4-digit division rather than 5-digit in Burstein et al. (2025).

Table 7 shows that the markup cyclicality patterns are generally consistent between the two datasets. The only exception is the sector-on-aggregate regression, which exhibits a marginally negative coefficient in Burstein et al. (2025) but a marginally positive one in our results.[20]

We now address the second consideration regarding the relevant levels of aggregation for comparison with our representative-agent model. We regress the aggregate, sector, and firm level gross markups on the aggregate real sales in the full sample of Compustat to measure the unconditional markup cyclicality at three aggregation levels in Table 8. The coefficients are all weakly positive.

Table 8: Cyclicality of Markups in Compustat at Various Aggregation Levels

	Aggregate on agg.	Sector on agg.	Firm on agg.
$\Delta \ln(Y)$	0.036	0.037	0.002
	(0.050)	(0.032)	(0.019)
Fixed effects	No	Sector	Firm
No. of obs.	35	9,313	232,296

[20] Neither the model of Burstein et al. (2025) nor those of ours generates strong countercyclical markups in aggregate.

The third consideration is the measure of unconditional markup cyclicality in our estimated models, which only considers three structural shocks, but not all possible shocks that drive the business cycles. As shown in Table 5, the aggregate elasticity of markup to GDP in models in the "u and Search" is positive but statistically insignificant, conditional on the three structural shocks combined.

6.3. Flattening of the Phillips Curve

As an extension of equation (15) in Corollary 3, the log-linearized Phillips curve in our "Search and u" model and "Search Alone" model is

$$d\log(\Pi) = \frac{1}{(\beta^{ss}_{markup} + \varphi)\,\chi''(1)} \cdot \Big[d\log(mc) + (\gamma - 1)d\log(\psi^f)\Big] + \beta\mathbb{E}\Big[d\log(\Pi')\Big].$$

In contrast, the Phillips curve in our "u Alone" model and "No Search or u" model is

$$d\log(\Pi) = \frac{1}{\beta^{ss}_{markup}\,\chi''(1)} \cdot d\log(mc) + \beta\mathbb{E}\Big[d\log(\Pi')\Big].$$

In our baseline calibration, the sensitivity of inflation to the real marginal cost of the composite inputs would reduce from $\frac{1}{0.34\chi''(1)}$ to $\frac{1}{(0.34+0.27)\chi''(1)}$, almost by a half, with the presence of our search frictions. In other words, our search frictions can reduce the price adjustment cost by nearly half, when generating the same sensitivity of inflation to cost pressures.

Holding the convexity of the price adjustment cost function $\chi''(1)$ constant, we assess the evolution of the Phillips curve over time, conditional on the change of the economic structure regarding the overall average markups and the composition of non-service and service sectors. We run the same regressions as Table 1 but split the sample across time.[21]The regression coefficients is as low as 0.14 before 2005 and as higher as 0.35 since 2015. The corresponding sensitivity of inflation to cost pressures is $\frac{1}{(0.34+0.14)\chi''(1)}$ and $\frac{1}{(0.34+0.35)\chi''(1)}$, which is reduced by 30%. This finding is aligned with the muted reaction of inflation to cost pressures in Del Negro et al. (2020).[22]

7. Robustness to Alternative Model Specifications

In this section, we demonstrate that the seemingly extreme assumptions of our search model, introduced for tractability, are innocuous when the model is calibrated to match target moments in micro evidence. These assumptions include the GHH preference that precludes the wealth effects in the search decision,

[21] Appendix B.5 provides the details.

[22] We try no to discuss the role of γ as the corresponding regression coefficients in Table 1 are insignificant. Once taking into account the change of γ over time, the main message does not change.

the preinstalled inputs that force the idle inputs to be wasted, and the need to search all varieties every period, which exaggerates the exposure of the product market to search frictions.

7.1. Wealth Effects in Search Decisions

By assuming GHH preference, we ensure the following nice property in the household's optimal search even in our medium-scale models, as a generalization of equation (16) in Proposition 2:

$$\beta^{micro}_{search} = \frac{1}{(1+\nu)-(\rho-1)},$$

which does not capture the mechanism that drives the richer household to search less as in Kaplan and Menzio (2016). A potential concern is that the size of β^{macro}_{TFP} may become smaller if the restriction of the GHH preference is relaxed to allow for some wealth effects.

To address this concern while at the same time maintaining the clarity of our model mechanisms, we generalize the instantaneous utility function to

$$\ln(h-\varsigma h^{-})-\eta\,\frac{\ell^{1+\xi}}{1+\xi}\quad\text{in which}\quad h\equiv\frac{(c^{A})^{1-\iota}}{1-\iota}-\zeta\,Z\,\frac{(d^{A})^{1+\nu}}{1+\nu},$$

where $\iota=0$ corresponds to the original GHH preference and $\iota\in(0,1)$ captures the new wealth effect. This preference implies the following new relation for the elasticity of search activity to spending:

$$\beta^{micro}_{search} = \frac{1}{\frac{1+\nu}{1-\iota}-(\rho-1)}.$$

Here, the wealth effect ι dampens the incentive for search if other parameter values are unchanged.

However, in our calibration, β^{micro}_{search} is kept unchanged and the cost of search parameter ν will adjust downwards as the wealth effect parameter ι becomes larger (still remains smaller than 1). As a result, the following relation as in equation (21) of Proposition 4 is kept unchanged:

$$\beta^{macro}_{TFP} = \frac{\beta^{micro}_{search}\,\beta^{micro}_{match}}{1+\beta^{micro}_{search}\,(\beta^{micro}_{match}+\beta^{ss}_{markup})}.$$

7.2. Partial Waste of Idle Inputs

The assumption that all inputs are preinstalled implies that once a production location is not occupied, all associated inputs are wasted. This assumption is not realistic because firms that primarily sell goods may use inventory to alleviate the waste of inputs.

To model an intermediate case between the full waste and no waste of idle inputs, we introduce a parameter $\varpi \in (0,1)$ to capture this fraction, which leads to the following cost of inputs in production:

$$\Big[1-(1-\varpi)(1-\psi^f)\Big]\cdot MC\cdot y.$$

As a result, we have $\frac{\psi^f}{1+(\varpi^{-1}-1)\psi^f_{ss}}$ as the TFP from the occupancy rate, and then the following relation as a generalization of equation (19) in Proposition 3:

$$\beta^{micro}_{match}=\frac{1}{1+(\varpi^{-1}-1)\psi^f_{ss}}\cdot\frac{d\ln(\psi^f)}{d\ln(p)+d\ln(\tilde{c})+d\ln(\psi^f)}=\frac{\varphi}{1+(\varpi^{-1}-1)\psi^f_{ss}},$$

which dampens how firm-side matches boost the firm-level TFP.

In calibration, if we keep β^{micro}_{match} unchanged, then a drop in the fraction of wasted inputs ϖ would be compensated by smaller congestion effect of search $1-\varphi$. As a result, the relation in equation (21) of Proposition 4 for β^{macro}_{TFP} is still unchanged.

7.3. Limited Exposure to Search Frictions

Now consider the more complex issue, where not all transactions are subject to search frictions. Households have a fraction of the varieties automatically found in each period and firms have a fraction of the locations automatically filled with the households. Let $\mathcal{I}$ be the number of varieties found by the household and o be the fraction of locations occupied in a firm, then the searching technology becomes

$$\mathcal{I}=(1-\varrho)\mathcal{I}_{ss}+d^A\psi^h,$$

and the matching technology becomes

$$o=(1-\varrho)o_{ss}+\Big[1-(1-\varrho)o_{ss}\Big]\psi^f.$$

In calibration, if we keep both β^{micro}_{search} and β^{micro}_{match} unchanged, a drop in the parameter of search friction exposure ϱ will lead to adjustment in other parameters, which implies a macro elasticity β^{macro}_{TFP} satisfying

$$\beta^{macro}_{TFP}>\frac{\beta^{micro}_{search}\,\beta^{micro}_{match}}{1+\beta^{micro}_{search}\,(\beta^{micro}_{match}+\beta^{ss}_{markup})},$$

which is strictly higher than than the value indicated by equation (21) in Proposition 4.

7.4. Taking Stock

Table 9 below summarize all of the details about the establishment of the mapping from the target moments $\{\beta^{micro}_{search}, \beta^{micro}_{match}, \beta^{micro}_{price}, \beta^{ss}_{markup}\}$ to the macro elasticity β^{macro}_{TFP}

Table 9: Theoretical Results for Alternative Specifications of Our Search Frictions

	Baseline	$\iota \in (0,1)$	$\varpi \in (0,1)$	$\varrho \in (0,1)$
β^{micro}_{search}	$\frac{1}{(1+\nu)-(\rho-1)}$	$\frac{1}{\frac{1+\nu}{1-\iota}-(\rho-1)}$	$\frac{1}{(1+\nu)-(\rho-1)}$	$\frac{1}{1+\frac{\nu}{\varrho}-(\rho-1)}$
β^{micro}_{match}	φ	φ	$\frac{\varphi}{1+(\varpi^{-1}-1)\,\psi^f_{ss}}$	$\frac{\varrho\,\varphi}{\varrho+(1-\varrho)\,(1-\varphi)}$
β^{micro}_{price}	$(\gamma-1)\,\varphi$	$(\gamma-1)\,\varphi$	a complex expression	$\frac{\varrho\,\varphi}{\varrho+(1-\varrho)\,(1-\varphi)} \cdot \frac{\gamma-1}{\varrho+(1-\varrho)\,(1-\varphi)}$
β^{ss}_{markup}	$\rho\,(1-\varphi)$	$\rho\,(1-\varphi)$	$\rho\,(1-\varphi)+\frac{(\varpi^{-1}-1)\,\psi^f_{ss}}{1+(\varpi^{-1}-1)\,\psi^f_{ss}}\,\varphi$	$\frac{\rho\,(1-\varphi)}{\varrho+(1-\varrho)\,(1-\varphi)}$
β^{macro}_{TFP}	$\frac{1}{\frac{1+\nu}{\varphi}-(\rho-1)}$	$\frac{1}{\frac{1}{\varphi}\frac{1+\nu}{1-\iota}-(\rho-1)}$	$\frac{1}{1+(\varpi^{-1}-1)\,\psi^f_{ss}} \cdot \frac{1}{\frac{1+\nu}{\varphi}-(\rho-1)}$	$\frac{1}{1+\frac{1+\nu-\varphi}{\varrho\,\varphi}-(\rho-1)}$
			$\beta^{macro}_{TFP} = \frac{\beta^{micro}_{search}\beta^{micro}_{match}}{1+\beta^{micro}_{search}(\beta^{micro}_{match}+\beta^{micro}_{markup})}$	even larger

We re-estimate the "Search Alone" model on top of the three alternative specifications of search frictions introduced above, and summarize the results in Table 10. The results indicate that none of these alternative specifications changes the message we deliver in our model comparisons: the "Search Alone" model consistently outperforms the "*u* Alone" model under both the prior and the mode of the parameters.[23]

Table 10: Estimation Results for Alternative Specifications

	u Alone	Search Alone	$\iota = 0.5$	$\varpi = 0.5$	$\varrho = 0.5$
Log marginal likelihood	**289.2**	**313.8**	**324.9**	**310.0**	**316.2**
Log likelihood at mode	**329.8**	**369.6**	**375.3**	**367.5**	**372.3**
– monetary policy shocks	115.8	123.8	128.3	123.9	126.1
– neutral technology shocks	136.2	138.9	140.8	145.1	139.5
– inv.-specific tech. shocks	77.8	106.9	106.2	98.4	106.8

[23] In Appendix F, we provide more robustness checks of Bayesian estimation and they all exhibit consistent results.

8. Conclusion

We propose that search frictions in the goods market generate a powerful nexus between demand and measured productivity. When households increase their search effort in response to rising expenditures, they raise the occupancy of firms' pre-installed production capacity, thereby endogenously boosting total factor productivity (TFP). This mechanism provides a micro-founded explanation for the procyclicality of productivity and inverse labor share conditional on demand shocks, and helps alleviate the issue of countercyclical markups in New Keynesian models.

Our framework recasts the concept of capaicity utilization. It moves beyond the traditional focus on the intensive margin of capital utilization to encompass the extensive margin of capacity occupancy, which is particularly relevant in service sectors where output is perishable and matching is crucial. This perspective implies that aggregate supply is not merely a constraint but is co-determined by the demand side through their effect on matching efficiency.

The implications extend beyond business cycle analysis. Our model offers a new channel for the transmission of macroeconomic policy, in which demand stimuli can increase productive efficiency. It also provides a coherent explanation for the documented flattening of the Phillips curve, as the search friction dampens the pass-through of cost pressures to inflation.

This research opens several productive avenues for future work. The framework could be extended to incorporate heterogeneous agents, exploring how inequality in wealth or access to information affects aggregate search efficiency and macroeconomic outcomes. The rise of e-commerce platforms, which drastically reduce search costs and alter matching technologies, presents a natural application for our model to study structural change and market power. Finally, the interaction between product and labor market frictions remains a promising area for synthesizing two fundamental sources of macroeconomic inefficiency.

In conclusion, by formally integrating search and matching into a general equilibrium setting, this paper provides a unified explanation for key business cycle phenomena and offers a refined toolkit for understanding the modern economy, where the efficiency of production is increasingly contingent on the effectiveness of match-making.

References

Alfaro, I., N. Bloom, and X. Lin (2024): "The Finance Uncertainty Multiplier," *Journal of Political Economy*, 132, 577–615.

Bai, Y., J.-V. Ríos-Rull, and K. Storesletten (2025): "Demand Shocks as Technology Shocks," *Review of Economic Studies, forthcoming*.

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

Basu, S. and C. L. House (2016): "Chapter 6 - Allocative and Remitted Wages: New Facts and Chanllenges for Keynesian Models," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 297–354.

Bils, M., P. J. Klenow, and B. A. Malin (2018): "Resurrecting the Role of the Product Market Wedge in Recessions," *American Economic Review*, 108, 1118–1146.

Boehm, C. E. and N. Pandalai-Nayar (2022): "Convex Supply Curves," *American Economic Review*, 112, 3941–3969.

Burdett, K. and K. L. Judd (1983): "Equilibrium Price Dispersion," *Econometrica*, 51, 955–969.

Burstein, A., V. M. Carvalhoz, and B. Grassi (2025): "Bottom-up Markup Fluctuations," *Quarterly Journal of Economics, forthcoming*.

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

Christiano, L. J., M. S. Eichenbaum, and M. Trabandt (2016): "Unemployment and Business Cycles," *Econometrica*, 84, 1523–1569.

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "Chapter 7 - DSGE Models for Monetary Policy Analysis," *Handbook of Monetary Economics*, 3, 285–367.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

Del Negro, M., G. E. Primiceri, M. Lenza, and A. Tambalotti (2020): "What's Up with the Phillips Curve?" *Brookings Papers on Economic Activity*, 2, 285–338.

Dolfen, P., L. Einav, P. J. Klenow, B. Klopack, J. D. Levin, L. Levin, and W. Best (2023): "Assessing the Gains from E-Commerce," *American Economic Journal: Macroeconomics*, 15, 342–370.

Gagliardone, L., M. Gertler, S. Lenzu, and J. Tielens (2025): "Anatomy of the Phillips Curve: Micro Evidence and Macro Implications," *American Economic Review*, forthcoming.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110, 2586–2624.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988a): "Investment, Capacity Utilization, and the Real Business Cycle," *American Economic Review*, 78, 402–417.

——— (1988b): "Investment, Capacity Utilization and the Real Business Cycle," *aer*, 78, 402–417.

Huo, Z. and J.-V. Ríos-Rull (2020): "Demand Induced Fluctuations," *Review of Economic Dynamics*, 37, S99–S117.

Huo, Z. and J.-V. Ríos-Rull (2020): "Sticky Wage Models and Labor Supply Constraints," *American Economic Journal: Macroeconomics*, 12, 284–318.

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

Kaplan, G. and G. Menzio (2016): "Shopping Externalities and Self-Fulfilling Unemployment Fluctuations," *Journal of Political Economy*, 124, 771–825.

Kass, R. E. and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Li, N. (2021): "An Engel Curve for Variety," *Review of Economic Statistics*, 103, 72–87.

Michaillat, P. and E. Saez (2015): "Aggregate Demand, Idle Time, and Unemployment," *Quarterly Journal of Economics*, 130, 507–569.

Michelacci, C., L. Paciello, and A. Pozzi (2022): "The Extensive Margin of Aggregate Consumption Demand," *Review of Economic Studies*, 89, 909–947.

Nekarda, C. J. and V. A. Ramey (2020): "The Cyclicality of the Price-Cost Markup," *Journal of Money, Credit, and Banking: 50th Anniversary Issue*, 52, 319–353.

Nord, L. (2023): "Shopping, Demand Composition, and Equilibrium Prices," *Working Paper*.

Pasten, E., R. Schoenle, and M. Weber (2020): "The Propagation of Monetary Policy Shocks in a Heterogeneous Production Economy," *Journal of Monetary Economics*, 116, 1–22.

Smets, F. and R. Wouters (2007): "Shocks and Frictions in US Business Cycle; A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Appendix (in preparation ...)

Contents

A. Proofs and Discussion for Section 2

A.1. Proof of Lemma 1

A.2. Proof of Lemma 2

A.3. Proof of Lemma 3

A.4. Proof of Lemma 4

A.5. Proof of Proposition 1

A.6. Proof of Corollary 1

A.7. Proof of Corollary 2

A.8. Proof of Corollary 3

A.9. Proof of Lemma 5

A.10. Proof of Proposition 2

A.11. Proof of Lemma 6

A.12. Proof of Proposition 3

A.13. Proof of Proposition 4

A.14. Discussing the Assumption in Section 2 about the Capacity of a Location

B. Explaining the Empirical Work for Section 3

B.1. Data Clearing and Description

B.2. Using $\frac{Sales}{COGS}$ to Proxy $\frac{p\psi^f}{W}$

B.3. Seasonality Bartik Instruments

B.4. Leave-one-out IV and Placebo Test

B.5. Regressions for Robustness

C. Technical Details for the Full Model in Section 4

C.1. Search Frictions in Medium-Scale Models for Section 4.2

C.2. Equilibrium Conditions

D. Technical Details for the Estimation Procedure in Section 5

D.1. Solving the Steady State

D.2. Accounting Issue in VAR

D.3. Estimation Procedure

E. Explaining the Implications for Section 6

E.1. Decomposing the Transmission of Monetary Policy Shocks

E.2. Understanding Markups in the Model

E.3. Deriving the Phillips Curve

F. More Details and Specifications of the Robustness Analysis in Section 7

F.1. Wealth Effects in Search Decisions

F.2. Partial Waste of Idle Inputs

F.3. Limited Exposure to Search Frictions

F.4. Alternative Matching Functions

F.5. Utilization Cost as Investment

F.6. Introducing Working Capital Loans

F.7. Removing Two Targets in Estimation