Monopsony Power and the Transmission of Monetary Policy *

Bence Bardóczy[†] Gideon Bornstein[‡] Chiara Maggi[§] Sergio Salgado[¶]

July 10, 2024.

Preliminary & Incomplete

Abstract

This paper studies how labor market power affects the efficacy of monetary policy. First, we use administrative U.S. Census data to document that firms with high monopsony power—firms who account for more than 10% of the wage bill in their local labor market—are less responsive to monetary policy in terms of their overall wage bill. Second, we construct a heterogeneous oligopsonistic New-Keynesian model to study how the decline in labor market power over the past four decades affected the transmission of monetary policy. We show that wage stickiness is key to obtain the heterogeneous response across firms. One contribution of our paper is to develop a numerical approach to solve the model. Such task is non-trivial as each local labor market consists of a finite number of firms, and a law-of-large numbers cannot be invoked to eliminate the local uncertainty resulting from wage stickiness. We calibrate the model to match key features of the U.S. labor market. We find that the decline in labor market power since the 1980s has amplified the aggregate effect of monetary policy on output by about 18%.

*Tanvi Jindal and Paige Stevenson provided superb research assistance. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, the Federal Reserve Board of Governors, the Federal Reserve System, or the International Monetary Fund. All results have been reviewed to ensure that no confidential information is disclosed.

[†]Federal Reserve Board of Governors; bence.a.bardoczy@frb.gov

[‡]The Wharton School, University of Pennsylvania; gideonbo@wharton.upenn.edu

[§]International Monetary Fund; cmaggi@imf.org

[¶]The Wharton School, University of Pennsylvania; ssalgado@wharton.upenn.edu

1 Introduction

Over the past four decades, local labor markets in the U.S. have become less concentrated.[1] The Herfindahl-Hirschman index of local labor market wage bill has declined from 0.16 in 1980 to 0.11 in 2020 (Berger et al., 2022). Despite this decline, local labor markets are still highly concentrated. On average, 5% of the firms in a local labor market account for more than 30% of employment. This high degree of concentration can lead to a partial and heterogeneous pass-through of demand shocks to wages, a key element in the transmission of monetary policy to real outcomes.

In this paper, we study how labor market power shapes the transmission of monetary policy. To do that, we proceed in two steps. First, we use administrative U.S. Census data to document the heterogeneous response of firms to a monetary policy shock. We classify firms into low and high monopsony power groups, according to the share of the local labor market they employ. We then study how the wage bill of firms in each group dynamically responds to a high-frequency monetary policy shock. Our analysis shows that the response of low-monopsony power firms is about 50% larger. Second, we construct a heterogeneous oligopsonistic New-Keynesian model to study the aggregate implications of labor market power for the efficacy of monetary policy. The economy consists of a continuum of local labor markets, within each a finite number of heterogeneous firms employ workers. These oligopsonistic firms compete à la Bertrand and can adjust their nominal wages infrequently. We develop a numerical method to solve the model, and calibrate it to match key characteristics of the U.S. labor market. In the calibrated model, like in the data, high-monopsony power firms are less responsive to monetary policy. This is both because they face a lower labor supply elasticity and because the change in their markdown makes their wage less responsive to the rise in demand due to the policy. Quantitatively, we find that the decline in labor market concentration since 1980 has amplified the transmission of monetary policy to aggregate output by about 15%.

The dataset we use to study the heterogeneous response of firms to monetary policy is the Census' Longitudinal Employer-Household Dynamics (LEHD) merged to the Longitudinal Business Database (LBD). The dataset contains quarterly information on wage bill and employment of all establishments with employees in the U.S. between 1993 to 2019, including the county in which they are located, their age, industry, and the firm to which the establishment belongs. We follow Yeh, Macaluso, and Hershbein (2022) in defining a local labor market to be all workers working for firms in the same subsector (3 digit NAICS) in the same county. We classify a firm-county-subsector triplet as a high labor market power firm if it accounts for more than 10% of the total wage bill in the local labor market. Otherwise, it is classified as a low labor market power firm.

We use the local projection method (Jordà, 2005) to identify the dynamic effect of monetary policy on firms' wage bill. The monetary policy shock series is obtained from Jarociński and Karadi (2020), who identify high-frequency monetary policy shocks by studying the Fed funds rate futures market just before and after FOMC announcements. They further use a structural VAR approach to separate the monetary policy shock from the central bank information shock conveyed in the FOMC announcement. To study the response of wage bill across low- and high-labor market power firms, we first aggregate

[1] See Rossi-Hansberg, Sarte, and Trachter (2021) and Berger, Herkenhoff, and Mongey (2022).

the wage bill of all firms in a group to the state by subsector level. Such aggregation allows us to reduce the noise stemming from labor indivisibility. Our benchmark regression specification includes state-subsector-group fixed effects, lagged macroeconomic controls, as well as lagged values of the wage bill as suggested in Montiel Olea and Plagborg-Møller (2021). We allow the coefficients to vary by labor market power group, so that we can separately identify the effect of monetary policy on low and high labor market power firms.

Our main empirical finding is that high labor market power firms respond less to a change in monetary policy. The response of both low- and high-labor market power firms is hump shaped—an expansionary monetary policy increases the wage bill of both groups of firms with the peak occurring 8 quarters following the initial shock. The effect resides after about 16 quarters. Both the average response as well as the response at the peak is about 50% larger for low labor market power firms relative to high labor market power firms. This heterogeneous response holds even after controlling for the national size of firms as well as their age.

The heterogeneous response we find across firms suggests that presence and observed changes in labor market power over time can affect the transmission of monetary policy to real outcomes. To gauge at the aggregate effects of labor market power on monetary policy efficacy, we construct a structural general equilibrium model with heterogeneous oligopsonistic firms and nominal rigidities. The labor market structure builds on Berger et al. (2022). The economy consists of a continuum of local labor markets. Each local labor market contains a finite number of firms who differ along their productivity level. Workers in the economy have idiosyncratic tastes towards different local labor markets and towards different firms, resulting in a finite labor supply elasticity for each firm which endogenously varies with the wage bill share of the firm in the local market.

The oligopsonistic firms all produce a homogeneous good but set their wages taking into account the labor supply function they face. They compete à la Bertrand and can adjust their wage infrequently. When a firm adjusts its wage, it sets it so that the weighted average of its markdown, taking into account the probability its nominal wage may not be able to adjust in the future, is equal to the inverse of the labor supply elasticity. Since labor supply elasticity is lower for firms with a larger labor market share, these large firms set a larger markdown relative to competitive wages. The rest of our model is standard. Households' consumption basket consists of a variety of goods aggregate via CES preferences. Each good is produced by a monopolistic firm, who purchases its intermediate inputs from the oligopsonistic firms, and can adjust their price only infrequently. The monetary authority sets the nominal interest rate according to a Taylor rule.

Before quantifying the model, we use a special case of the model to show the importance of wage stickiness in driving the heterogeneous response across firms. We consider the limit case where the discount factor is equal to zero, so that optimal wage setting is static. We study the effect of a rise in demand, induced by an expansionary monetary policy, on a local labor market consisting of three firms: a low-, medium-, and high-productivity firms. We first show that absent wage stickiness, all firms adjust their wage by the same magnitude, resulting in a homogeneous response of labor and wage bill despite the heterogeneity in labor market power. When all firms can adjust their wage, their labor market share remains unchanged, they do not change their markdowns, so there is full passthrough from the demand shock to the wage. We then consider the case in which the medium-productivity firm cannot adjust its wage. When that is the case, there is a heterogeneous response between the low- and high-productivity firms. With one firm stuck at their previous nominal wage, the labor share of the other two firms goes up. Because the low productivity firm faces a higher labor supply elasticity, it responds more in terms of both employment and wage bill. Thus, wage stickiness is a key driver of the heterogeneous response across firms.

Solving the full version of the model brings two new challenges, relative to a standard heterogenous-agent New-Keynesian model such as Kaplan, Moll, and Violante (2018). First, the finite number of firms operating in each local labor market implies that a law of large numbers cannot be invoked to eliminate the idiosyncratic risk of wage stickiness. Due to the continuum of local labor markets, there is no aggregate uncertainty due to wage stickiness. There is, however, uncertainty within each local labor market. When a firm sets its wage, it needs to take into account the expected future wages of its competitors, which depends on the realization of the Calvo wage adjustment shocks. A firm that competes with 25 other firms in a local labor market, the median market size in our calibration, needs to consider more than 330 million combinations of which firms can or cannot adjust their wage over the next 10 periods. This is not computationally feasible. To overcome this challenge, we assume firms have perfect foresight regarding their competitors' wages. That is, firms know when their competitors will be able to adjust their wage. We show that with a large number of local labor markets, the random draw that defines whether or not a firm gets to adjust and when, does not numerically affect any aggregate outcome.[2]

The second computational challenge is that there is no closed-form solution to the optimal wage setting of the firm as a function of aggregate variables. Instead, we need to find the Nash equilibrium within each local labor market as a fixed point problem. As our calibration includes 3,000 local labor markets, finding the Nash equilibrium in each market is a computationally intensive task. Nonetheless, we show how one can use a nested sequence space Jacobian approach (Boppart, Krusell, and Mitman 2018; Auclert, Bardóczy, Rognlie, and Straub 2021) to solve the problem in a timely manner. Allowing us to estimate the model efficiently.

We calibrate the model to match key features of the U.S. labor market. The number of firms in each local labor market is calibrated to match the distribution of HHI across local labor markets in the U.S. The Pareto shape governing the dispersion of productivity levels across the oligopsonistic firms and the returns to scale of these firms are chosen to match the share of firms with a local labor market share above 10% as well as the share of total wage bill accounted for by these high labor market power firms. We calibrate the other structural parameters in the model to standard values in the literature.

The model is able to replicate well the heterogeneous response across low and high labor market power firms. In response to a monetary policy shock, the employment and wage bill of high labor market power firms go up relatively less for two reasons. First, the superelasticity of labor supply implied by the oligopsonistic competition in local labor markets results in a larger movement in markdowns for high labor market power firms. The change in the markdown results in a lower pass through of prices to wages, and the wage of high labor market power firms responds less than that of

[2] Note that wage stickiness breaks the block recursivity that Berger et al. (2022) exploit to solve the model dynamically.

low labor market power firms. Second, the labor supply elasticity of high labor market power firms is lower. This implies that for the same rise in the wage, the increase in employment for low labor market power firms is relatively larger. These two forces go in the same direction, resulting in a heterogeneous response across firms that is both qualitatively and quantitatively close to our empirical findings.

To study how labor market power shapes the transmission of monetary policy, we conduct two counterfactual experiments. The first experiment eliminates labor market power by imposing that firms target a markdown of 1 (wage equals the marginal product of labor) at all times. In such economy, the passthrough of demand shocks to wages and employment is not dampened by strategic wage setting. We compare the impulse response function to a monetary policy shock of our benchmark economy to that of the counterfactual economy. We find that eliminating oligopolistic competition results in a substantial increase in the efficacy of monetary policy. The cumulative effect of an expansionary monetary policy shock on aggregate output is about 32% larger in the economy with only low labor market power firms.

The second counterfactual experiment considers a more concentrated economy, corresponding to the average level of HHI that was present in U.S. local labor markets in the 1980s. In such economy, high labor market power firms account for a larger share of overall employment. Our analysis suggests that a monetary policy shock is about 15% more effective in 2019 at stimulating aggregate output than it was back in the 1980s. A related implication is that the slope of the Phillips curve is flatter in the benchmark economy relative to its slope in the counterfactual economy. Our structural analysis thus suggests that the decline in local labor market concentration have significantly increased the efficacy of monetary policy.

Literature Review. To be written.

2 Empirical Analysis

In this section, we evaluate whether monetary policy shocks have a differential impact on firms depending on their degree of labor market power. To this end, we combine two administrative data sources from the US Census Bureau. We use this data to construct measures of employment and wage bill within narrow labor markets and by firm characteristics such as age, firm size (sales and employment), and industry. We complement this data with macroeconomic time series, including measures of monetary policy shocks.

2.1 Data

Our primary source of data is the Census' Longitudinal Employer-Household Dynamics (LEHD) which is an employer-employee matched panel dataset constructed from unemployment insurance records. The LEHD contains quarterly data on employment and wages at the individual level and includes information on the worker (e.g., age and gender) and their employer (e.g., location of the job and narrow industry groups).

In our analysis, we use information from 1993q1 to 2019q4 for all states available at a particular point in time. This implies that the number of states increases over time—from 6 to 23—as new states are included in the sample.[3] The unit of observation in LEHD is an individual unemployment insurance account, which can be associated with multiple firms if an individual worker is employed in different firms in a given quarter.

We aggregate the employment and wage bill data in LEHD at the firm-county level within a quarter summing all employment and wage bill across all the establishments a firm has in a particular county and define a local labor market market to be all firms within a subsector (3 digit NAICS) in the county. We drop from our sample all firms with an NAICS3 code in national security (NAICS 600) or in personal services (NAICS 800) since these do not have information on sales in LBD. Our results do not change significantly if we consider these sectors in our sample. Then, for each firm in a local labor market, we calculate the share of the labor market wage bill it accounts for. We classify firms into low and high labor market power categories. A firm is considered to have high labor market power if it accounts for 10% or more of the total wage bill within the local labor market. Notice this implies that the same firm can be classified as having high labor market power in one labor market (e.g., a grocery chain is the only supermarket in a small town) but as having low labor market power in a different labor market (e.g., the same chain but in a big city). In our sample, about 6% of firm observations are classified as having high labor market power, accounting for an average of 32.5% of total employment and 34.9% of the total wage bill.

We combine the data from LEHD with information from the Longitudinal Business Database (LBD) which contains annual information on wage bill and employment for all firms in the private sector from 1978 to 2019 and revenue information starting in 1998. We use LBD to obtain the age of the firm and the levels of employment, wage bill, and revenue at the national level. We use this data to measure the share of firms as old (firms that are five years old or older) and the share of large (firms with national employment of 500 workers or more) within a local labor market.

As a final step, we aggregate the employment and wage bill of all firms classified as high (low) monopsony power within a labor market to the state-NAICS3 level within a quarter. This reduces the noise generated by large variations in employment and wage bill originating from small labor markets. This gives us an average sample of 150 thousand state-NAICS3-quarter observations, which we use for our empirical analysis. We combine this dataset with quarterly macroeconomic aggregates (e.g., GDP and unemployment) and a measure of monetary policy shocks obtained from Jarociński and Karadi (2020).

[3] More precisely, we have information on 6 states in 1993, including California. By 2000, this increased to 21, and 23 states in 2004. By the end of our sample, we have the following states: Arkansas, Arizona, California, Connecticut, Colorado, Kansas, Connecticut, Maine, South Carolina, North Dakota, New Jersey, New Mexico, Nevada, Ohio, Oklahoma, Pennsylvania, South Dakota, Tennessee, Utah Virginia, Washington, Wisconsin, Wyoming.

2.2 Econometric Specification and Results

To measure the impact of monetary shocks on the labor market and how that varies depending on the degree of labor market power, we run a series of local projections (Jordà (2005)) of the form

$$y^{s,i}_{t+k} = \alpha^{s,i} + \beta^k_1 MP_t + \Gamma^k_0(L)y^{s,i}_t + \Gamma^k_0(L)X_t + \epsilon^{s,i}_{t,k}, \tag{2.1}$$

where the dependent variable is the log wage bill in a particular state-NAICS3-quarter in $t+k$ quarters, MP_t is a monetary policy shock, and X_t contains a set of macroeconomic (GDP and unemployment). lagged values of $y^{s,i}_t$, and state and NAICS3 fixed effects. We run specification 2.1 for high and low monopsony firms, and combine all firms to obtain an average effect of monetary shock, and we plot β^k_1 for the corresponding specification.

Our baseline results are displayed in Figure 1 which shows the response of the log wage bill to an unexpected decrease in the fed funds rate of one standard deviation— the equivalent of 8 basis points.[4] The left panel shows a positive and statistically significant response of the wage bill to an expansionary monetary shock: on average, the wage bill increases by 0.5% after four quarters and reaches more than 1% increase eight quarters after the shock. The right panel of Figure 1 shows that the response varies significantly depending on whether firms have high or low monopsony power. In particular, firms with less market power tend to respond more than firms with high monopsony power: the wage bill among low monopsony power firms increases by 1.2% after a monetary stimulus of one standard deviation after eight quarters but increases by 0.8% for firms with high monopsony power.

One concern of our results is that low monopsony power firms are more responsive to monetary policy shocks because they are comprised of younger or smaller firms, which tend to be more responsive to aggregate shocks in general (Fort et al., 2013). To address this concern, we include in our regression the share of young firms within monopsony power groups within a labor market. As shown in the Appendix, controlling for the share of young or large firms does not significantly affect our results. In fact, controlling for firm characteristics tends to increase the difference in response between high and low monopsony firms as shown in Appendix Figure B.3.

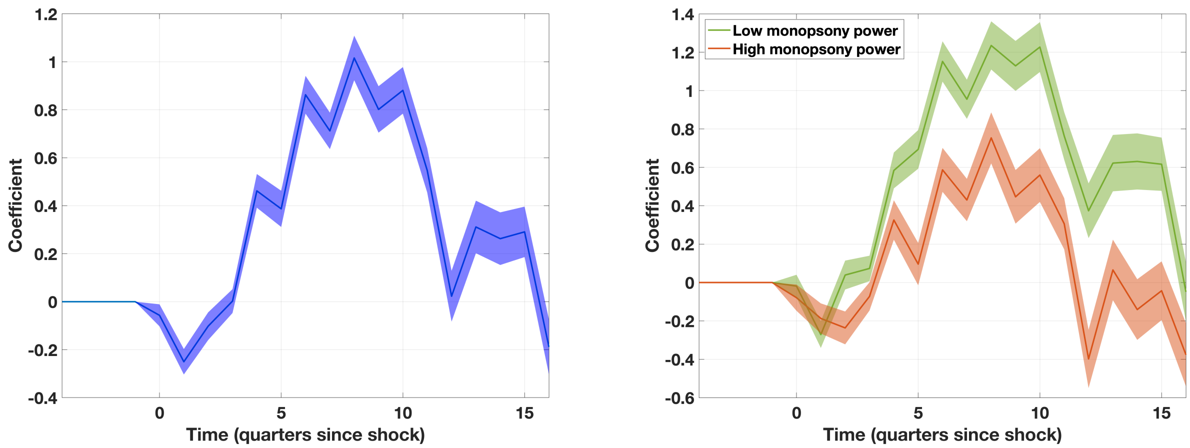
3 Model

Next, we present our oligopsonistic New Keynesian model with heterogeneous firms. To model imperfect competition, we build on the approach pioneered in the trade literature by Atkeson and Burstein (2008) and adapted to labor markets by Berger et al. (2022). Firms are atomistic on national markets but large on their local labor market, which gives rise to strategic interactions. The key feature of our model is wage stickiness. Wage stickiness is necessary to replicate our empirical results, but breaks block recursivity that made earlier models tractable.[5]

[4] To reduce the impact of outliers on our results, we trim our sample at the top and bottom 10% of the distribution of wage bill change between quarters $y^{s,i}_{t+k}$ and $y^{s,i}_t$.

[5] Berger et al. (2022) prove that their model is block recursive, meaning that local labor market equilibria are independent of aggregates. Wage stickiness breaks block recursivity, because it forces firms to play a dynamic game in which

Figure 1: Response of Log-Wage Bill to a Monetary Policy Shock



Note: This figure shows the response of log wage bill within a state-NAICS3 to a one-standard-deviation shock to the Fed Funds Rate. We plot the coefficient β_1^k for each horizon k. To reduce the impact of outliers, we trim the sample for each horizon k a the bottom and top 10% of the distribution of the change in $y_t^{s,i}$ between periods t and $t+k$. The regression includes an average of 150 thousand state-NAICS3-quarter observations accounting for 620 thousand firms. The shaded areas represent 95% confidence intervals. Standard errors clustered at the state-NAICS3 level.

Environment. Time is discrete and each period corresponds to a quarter. The economy consists of a representative household, a continuum of firms, and a central bank. There are two types of firms: producers and retailers. Producers inhabit a continuum of local labor markets $j \in [0,1]$, each with a finite number of firms indexed by $i \in \{1, 2, \ldots, M_j\}$. They hire labor from the household and produce a homogeneous intermediate good which they sell to retailers on a national market. Retailers differentiate the intermediate good into imperfectly substitutable varieties indexed by $k \in [0,1]$ which they sell to the household. Producers have monopsony power on the labor market. Retailers have monopoly power on the final goods market. We separate producers and retailers, and assume that retailers are not affected by the boundaries of local labor markets, for tractability.

3.1 Households

Setup. The representative household chooses sequences of consumption $\{c_{kt}\}$ from retailer k, labor supply $\{n_{ijt}\}$ to producer ij, and nominal bonds $\{B_t\}$ to maximize the present value of expected utility. Given initial wealth B_0, the household solves

$$\max_{c_{kt}, n_{ijt}, B_t} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(C_t, N_t), \tag{3.1}$$

where the aggregate consumption and labor supply indexes are given by

$$C_t = \left(\int_0^1 c_{kt}^{\frac{\epsilon-1}{\epsilon}} dk \right)^{\frac{\epsilon}{\epsilon-1}}, \qquad N_t = \left(\int_0^1 n_{jt}^{\frac{\zeta+1}{\zeta}} dj \right)^{\frac{\zeta}{\zeta+1}}, \qquad n_{jt} = \left(\sum_{i=1}^{M_j} n_{ijt}^{\frac{\eta+1}{\eta}} \right)^{\frac{\eta}{\eta+1}}, \tag{3.2}$$

aggregate fluctuations are payoff relevant.

with elasticities $\epsilon > 1$, and $\eta > \zeta > 0$. The ordering of the labor supply elasticities implies that it is easier to substitute labor within a local labor market than across markets. Utility maximization is subject to the following budget constraint every period

$$\int_0^1 p_{kt} c_{kt} dk + B_t = (1 + i_{t-1}) B_{t-1} + \int_0^1 \left(\sum_{i=1}^{M_j} W_{ijt} n_{ijt} \right) dj + \Pi_t, \tag{3.3}$$

where p_{kt} is the price of final good k, i_{t-1} is the nominal return on bonds between periods $t-1$ and t, W_{ijt} is the nominal wage of producer i on local labor market j, and Π_t is aggregate profits.

Optimality conditions. Goods demand is summarized by an aggregate Euler equation and retailer-specific demand curves

$$U_C(C_t, N_t) = \beta \mathbb{E}_t \left[\frac{1 + i_t}{\pi_{t+1}} U_C(C_{t+1}, N_{t+1}) \right], \qquad c_{kt} = \left(\frac{p_{kt}}{P_t} \right)^{-\epsilon} C_t, \tag{3.4}$$

where the aggregate price index, defined by $P_t C_t = \int p_{kt} c_{kt} dk$, and inflation can be expressed as

$$P_t = \left(\int_0^1 p_{kt}^{1-\epsilon} dk \right)^{\frac{1}{1-\epsilon}}, \qquad \pi_t = \frac{P_t}{P_{t-1}}. \tag{3.5}$$

Similarly, aggregate and firm-specific labor supply are given by

$$\frac{W_t}{P_t} = -\frac{U_N(C_t, N_t)}{U_C(C_t, N_t)}, \qquad n_{ijt} = \left(\frac{W_{ijt}}{W_{jt}} \right)^{\eta} \left(\frac{W_{jt}}{W_t} \right)^{\zeta} N_t, \tag{3.6}$$

where local and aggregate wage indexes are

$$W_t = \left(\int_0^1 W_{jt}^{1+\zeta} dj \right)^{\frac{1}{1+\zeta}}, \qquad W_{jt} = \left(\sum_{i=1}^{M_j} W_{ijt}^{1+\eta} \right)^{\frac{1}{1+\eta}}. \tag{3.7}$$

Note that both firm-specific labor supply and the two wage indexes are linearly homogeneous in wages. Therefore, these equations are valid for both nominal and real wages. Our convention is to use upper case letters to denote nominal variables and lower case letters to denote real variables. See appendix A.1 for derivations.

Labor market power. The CES labor supply system described by equations (3.6) and (3.7) implies that, as long as the elasticities η and ζ are finite, producers will have labor market power. Producer ij may offer a lower wage than its local competitors ($W_{ijt} < W_{jt}$) and national competitors ($W_{ijt} < W_t$) and it will still attract workers. Finite labor supply elasticities capture, in reduced form, the deeper reasons for firms' having labor market power: mobility costs, search costs, specialized human capital, and heterogeneous preferences over non-wage amenities (Berger et al., 2022).

3.2 Retailers.

Setup. There's a unit measure of retailers $k \in [0,1]$ who the buy homogeneous intermediate good for price M_t and turn it into differentiated final goods c_{kt} which they sell to households at price p_{kt}. Since varieties are imperfect substitutes, retailers have monopoly power and can set their own price. There is price stickiness à la Calvo: retailers can adjust their price with probability $1-\theta_p \in (0,1]$. The probability of adjustment is independently and identically distributed across periods and firms.

A retailer that cannot adjust its price has no decision to make. It produces just enough to satisfy household demand at the current prices. A retailer that can reset its price will maximize the present value of expected profits until the next price adjustment

$$\max_{p^*_{kt},\{c_{kt+\tau},y_{kt+\tau}\}} \mathbb{E}_t \sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} \left(p^*_{kt} c_{kt+\tau} - M_{t+\tau} c_{kt+\tau}\right) \tag{3.8}$$

subject to the demand curve

$$c_{kt+\tau} = \left(\frac{p^*_{kt}}{P_{t+\tau}}\right)^{-\epsilon} C_{t+\tau}, \tag{3.9}$$

where $R_{t,t+\tau}$ is the nominal discount factor between periods t and $t+\tau$

$$R_{t,t+\tau} = \begin{cases} 1 & \text{for } \tau = 0, \\ \prod_{s=0}^{\tau-1} \frac{1}{1+i_{t+s}} & \text{for } \tau > 0. \end{cases} \tag{3.10}$$

Optimality condition. Although the retailer's problem is standard, it is a useful to consider it due to its similarity to the novel problem of the oligopsonistic producers. Price adjusters are symmetric and set the same price

$$p^*_t = \frac{\epsilon}{\epsilon-1} \frac{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} P^\epsilon_{t+\tau} C_{t+\tau} M_{t+\tau}}{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} P^\epsilon_{t+\tau} C_{t+\tau}}. \tag{3.11}$$

Equation (3.11) shows that all retailers have the same market power, captured by a *desired markup* that is constant and equal to $\epsilon/(\epsilon-1)$. The firm sets p^*_t to achieve this markup over the expected average marginal cost until the next price adjustment. The marginal cost τ periods ahead is discounted with the nominal discount factor $R_{t,t+\tau}$, with the probability θ_p^τ that the price is still in effect, and weighted by expected sales.

Loglinearizing the first order condition (3.11) in conjunction with the price index (3.5) yields the textbook New Keynesian Phillips curve

$$\hat{\pi}_t = \frac{(1-\theta_p)(1-\beta\theta_p)}{\theta_p} \hat{m}_t + \beta \hat{\pi}_{t+1}, \tag{3.12}$$

where $\hat{\pi}_t$ and $\hat{m}_t$ denote of inflation and real marginal cost in log-deviations from the zero-inflation steady state. See appendix A.2 for derivations.

Role of price stickiness. Wage stickiness is necessary to match the *heterogeneity* in the wagebill responses of low-, and high monopsony firms. And, conditional on wage stickiness, price stickiness is necessary to match the *average* wagebill response. If prices were flexible, an expansionary demand shock would cause real wages to fall in equilibrium. Since market power lies with firms, who take labor supply curves as given, equilibrium employment would have to fall, too. However, we find that expansionary monetary policy shocks raise both wages and employment. Allowing prices to be slightly stickier than wages is crucial to matching the full set of estimated impulse responses.

3.3 Producers

Setup. There is a continuum of producers who are ex-ante heterogeneous in two dimensions. First, each firm operates on a specific local labor market $j \in [0,1]$, along with finitely many local competitors $i \in \{1, 2, \ldots, M_j\}$. Thus, each firm faces different local competitors. Second, firms differ in their productivity $z_{ij} \in (0, \infty)$ which is drawn from a location invariant distribution $F(z)$ once and for all.

This market structure implies that producers are infinitesimal with respect to the macroeconomy, and will take as given the (expected) sequences of the aggregate wage $\{W_t\}$ and labor supply $\{N_t\}$. On their own labor market however, they engage in Bertrand competition. Producer ij internalizes that the nominal wage it sets, W_{ijt}, affects not only its own employment n_{ijt}, but the local wage index W_{jt} as well. We assume that producers take the current and future wages of their competitors as given. This structure arises if producers commit to a wage strategy that depends on aggregate variables and all individual Calvo realizations, but does not depend explicitly on competitors' wages.[6]

Finally, producers face nominal wage adjustment costs à la Calvo. They can reset their wage with probability $1-\theta_w \in (0,1]$ that is independently and identically distributed over time and across firms. A producer that cannot adjust its wage has no decision to make. It has to accommodate labor supply at the current wages. A producer that can reset its wage will maximize the present value of expected profits until the the next wage adjustment

$$\max_{W^*_{ijt}, \{W_{jt+\tau}, n_{ijt+\tau}\}} \mathbb{E}_t \sum_{\tau=0}^{\infty} \theta_w^\tau R_{t,t+\tau} \left(M_{t+\tau} z_{ij} n^\alpha_{ijt+\tau} - W^*_{ijt} n_{ijt+\tau} \right), \tag{3.13}$$

where $\alpha \in (0,1)$, subject to the firm-specific labor supply curve and local wage index

$$n_{ijt+\tau} = \left(\frac{W^*_{ijt}}{W_{jt+\tau}} \right)^\eta \left(\frac{W_{jt+\tau}}{W_{t+\tau}} \right)^\zeta N_{t+\tau}, \qquad W_{jt+\tau} = \left((W^*_{ijt})^{1+\eta} + \sum_{i' \neq i} W^{1+\eta}_{i'jt+\tau} \right)^{\frac{1}{1+\eta}}. \tag{3.14}$$

[6] In dynamic game theory, this equilibrium concept is referred to as "open loop Nash equilibrium" (see, for example, Fershtman and Kamien, 1987) and is often used for its tractability over the Markov perfect equilibrium (MPE). In the context of monetary policy transmission under oligopolistic competition, Wang and Werning (2022) analyze the MPE with homogenous oligopolistic firms and find that the difference in impulse response functions between the open-loop equilibrium and MPE is quantitatively negligible. We're currently working on studying the difference between the two equilibrium concepts in our environment, focusing on special cases where MPE is computationally tractable, i.e., few firms per market inhabiting ex-ante symmetric markets.

Optimality condition. In appendix A.3, we derive the optimal reset wage

$$W_{ijt}^{*}=\frac{\sum_{\tau=0}^{\infty}\theta_w^{\tau}R_{t,t+\tau}n_{ijt+\tau}\left[\eta-(\eta-\zeta)s_{ijt+\tau}\right]M_{t+\tau}z_{ij}\alpha n_{ijt+\tau}^{\alpha-1}}{\sum_{\tau=0}^{\infty}\theta_w^{\tau}R_{t,t+\tau}n_{ijt+\tau}\left[1+\eta-(\eta-\zeta)s_{ijt+\tau}\right]}, \tag{3.15}$$

where $s_{ijt}\equiv W_{ijt}n_{ijt}/(W_{jt}n_{jt})$ is the wage bill share of firm ij on market j in period t. Compare this expression to the well-known price setting equation of retailers (3.11). Firm ij sets W_{ijt}^{*} to implement a *desired markdown* with respect to the marginal revenue product of labor $M_{t+\tau}z_{ij}\alpha n_{ijt+\tau}^{\alpha-1}$. It considers the expected present value of MRPL until the next wage adjustment, discounted and weighted by employment. The crucial difference from monopolistic competition is that the desired markdown

$$\mu_{ijt}\equiv\frac{\epsilon_{ijt}}{1+\epsilon_{ijt}}\equiv\frac{\eta-(\eta-\zeta)s_{ijt}}{1+\eta-(\eta-\zeta)s_{ijt}} \tag{3.16}$$

may vary both across firms and over time. Given the ordering $\eta>\zeta>0$, larger (more productive) firms who are dominant on their local labor market set lower wages. Equation (3.16) highlights that we can interpret the markdown in terms of an *equilibrium labor supply elasticity*, ϵ_{ijt}, which captures market power of firm ij both within its local labor market and on the national labor market. Equation (3.16) also clarifies that the relevant measure of labor market power for the purposes of wage setting (and by extension labor demand and production) is the wage bill share. This result guides our classification of firms by labor market power in our empirical analysis.

The mapping between the markdown $\mu_{ijt}<1$ and the primitive labor supply elasticities η,ζ depends on the form of imperfect competition. We assume Bertrand competition, implying that firms condition their best response on the wages of their competitors. In the presence of wage stickiness, Bertrand is more natural than Cournot which assumes the firms consider their competitors' employment instead.[7]

3.4 Central bank

The central bank sets the nominal rate according to a standard Taylor rule

$$i_t=i_{ss}+\phi_\pi(\pi_t-1)+\varepsilon_t. \tag{3.17}$$

The ex-post real return on nominal bonds satisfies the Fisher equation

$$1+r_t=\frac{1+i_{t-1}}{1+\pi_t}. \tag{3.18}$$

[7] Berger et al. (2022) work with Cournot competition and show that the equilibrium elasticity is a harmonic mean

$$\epsilon_{ijt}^{Cournot}=\left[\frac{1}{\eta}-\left(\frac{1}{\eta}-\frac{1}{\zeta}\right)s_{ijt}\right]^{-1}$$

3.5 Equilibrium

Given a sequence of monetary policy shocks $\{\varepsilon_t\}$, sequences of Calvo shocks for each producer $\{\delta_{ijt}\}$, and initial wealth B_{-1}, the equilibrium is a sequence of allocations $\{B_t, C_t, N_t, n_{jt}, n_{ijt}\}$ and prices $\{i_t, r_t, \pi_t, w^*_{ijt}, w_{ijt}, w_{jt}, w_t, m_t\}$ such that

- Households and producers optimize;
- Local labor markets are in Nash equilibrium with real wages and wage indexes[8] satisfying (3.7);
- Inflation satisfies the loglinearized Phillips curve[9] (3.12);
- The nominal rate satisfies the Taylor rule and the real rate satisfies the Fisher equation;
- Goods and bond markets clear

$$Y_t \equiv \int_0^1 \sum_{i=1}^{M_j} z_{ij} n_{ijt}^{\alpha} di = C_t, \qquad B_t = 0. \tag{3.19}$$

4 Special Case

In this section, we study a special case of our model to show the importance of wage stickiness in driving the heterogeneous response across different firms. Let the firm-specific labor supply curve be

$$n_{ij} = \left(\frac{w_{ij}}{w_j}\right)^{\eta} \left(\frac{w_j}{w}\right)^{\zeta} N, \tag{4.1}$$

where n_{ij} is the workers willing to work at firm i in market j, w_{ij} is the wage offered by the firm, w_j is the market wage given by $\left(\sum_i w_{ij}^{1+\eta}\right)^{\frac{1}{1+\eta}}$, w is the aggregate wage and N is aggregate labor supply.

We assume that in each market there are finite number of firms. Firms therefore internalize how their individual wage affects the market wage. We further assume that there are continuum of markets so that firm-level wages do not affect the aggregate wage and labor supply.

Firms use a decreasing-returns-to-scale production technology with labor being the sole input. They are heterogeneous along their productivity level z_{ij}, and are all subject to an aggregate demand D. The firm problem is

$$\max_{w_{ij}, n_{ij}, w_j} \quad D z_{ij} n_{ij}^{\alpha} - w_{ij} n_{ij}, \tag{4.2}$$

$$\text{s.t.} \quad n_{ij} = \left(\frac{w_{ij}}{w_j}\right)^{\eta} \left(\frac{w_j}{w}\right)^{\zeta} N, \tag{4.3}$$

$$w_j = \left(\sum_{i=1}^{M} w_{ij}^{1+\eta}\right)^{\frac{1}{1+\eta}}. \tag{4.4}$$

[8] The wage indexes are stated in terms on nominal wages, but they apply to real wages $w_t \equiv W_t/P_t$ as well.

[9] Working directly with the loglinearized Phillips curve lets us omit the second-order effects of price dispersion.

Solving the firm's problem, we obtain a wage setting equation

$$w_{ij} = \frac{\epsilon_{ij}}{1+\epsilon_{ij}} \alpha D z_{ij} n_{ij}^{\alpha-1}, \tag{4.5}$$

where

$$\epsilon_{ij} = \eta - (\eta - \zeta) s_{ij} = (1 - s_{ij})\eta + s_{ij}\zeta, \tag{4.6}$$

with $s_{ij} \equiv (w_{ij}/w_j)^{(1+\eta)} \in (0,1)$ being the wagebill share of firm i in sector j. The larger is the wagebill share of the firm, the lower is its markdown. Note that (4.5) only makes sense for $s_{ij} \in [0,1]$. This is always satisfied here, because it is the wagebill share.

We can use the labor supply constraint to rewrite this optimality condition as

$$w_{ij} = \frac{\epsilon_{ij}}{1+\epsilon_{ij}} z_{ij} \left(w_{ij}^{\eta} w_j^{\zeta-\eta} \right)^{\alpha-1} \tilde{D}, \tag{4.7}$$

where $\tilde{D} = \alpha D w^{-\zeta(\alpha-1)} N^{\alpha-1}$.

Consider an economy that exists for two periods (0 and 1), in which every market contains three firms, $i \in \{s, m, l\}$. We assume that $z_s < z_m < z_l$. Equation (4.7) together with the definition of w_j implies a system of four equations which pin down the equilibrium wage levels. In period 0, we assume that $\tilde{D} = 1$ and that firms can flexibly choose their wages. Let w_s^0, w_m^0, w_l^0 be the optimal wages of the three firms in this period.[10]

Suppose that in period 1, aggregate demand goes up so that $\tilde{D}_1 > 1$. The following two propositions show that only with wage stickiness we shall observe a heterogeneous response across firms.

Proposition **1.** *If all firms can adjust their wages, then $s_{ij}^1 = s_{ij}^0$ for all i and j. That is, all firms respond in the same way to the rise in demand.*

Intuitively, all firms maintain the same level of markdown and pass through the rise in demand to their wages. As the relative wages of these firms remain stable, maintaining the same level of markdown is indeed optimal.

Proposition **2.** *Suppose that firm m cannot adjust its wage so that $w_m^1 = w_m^0$. Then $\frac{s_{lj}^1}{s_{lj}^0} < \frac{s_{sj}^1}{s_{sj}^0}$ for all j. That is, the small firm responds more than the large firm.*

5 Quantitative Analysis

5.1 Numerical Algorithm

Most preference-based models of market power (Atkeson and Burstein, 2008; Berger et al., 2022) are tractable because they imply that strategic interactions are static. Nominal rigidities however, necessarily introduce dynamics. When a firm sets its wage, it has to consider future developments in

[10] We implicitly assume that firms believe that $\tilde{D}$ will not change in the future.

both aggregate and local conditions, since it may not be able respond to them fully when they are realized. In this subsection, we present a novel numerical strategy to solve general equilibrium models with dynamic oligopoly or oligopsony.

Challenges. The challenges are two. First, firms and workers interact on a very large number of markets.[11] Second, idiosyncratic uncertainty affects local equilibria.

The challenge of working with a large number markets is straightforward. Equilibrium prices are characterized implicitly, as solutions to a fixed point problem. Having many markets means having to deal with a large-scale fixed point problem. Note that this challenge is not present in typical HANK models, in which the distribution affects aggregate demand or supply but agents still interact on a small number of national markets. We have to both keep track of a high-dimensional distribution and solve for many equilibrium prices.

To understand the challenge posed by idiosyncratic uncertainty, consider the problem of finding the local equilibrium on a single market. Equation (3.15) shows that the the optimal wage of a single firm ij in a single period t depends on expectations of aggregate variables $\{R_{t,t+\tau}, M_{t+\tau}\}_{\tau=0}^{\infty}$ as well as the expectations of the local wage index $\{W_{jt+\tau}\}_{\tau=0}^{\infty}$. Recall that the local wage index is given by

$$W_{jt+\tau} = \left(\sum_{i=1}^{M_j} W_{ijt+\tau}^{1+\eta}\right)^{\frac{1}{1+\eta}}, \qquad W_{ijt+\tau} = \begin{cases} W_{ijt+\tau}^{*} & \text{if firm } ij \text{ can adjust in } t+\tau, \\ W_{ijt+\tau-1} & \text{otherwise.} \end{cases} \tag{5.1}$$

Since M_j is finite, the realizations of Calvo shocks to individual firms affect the local wage index. For concreteness, let's assume that there are 100 firms on the market, which is the mode in our calibrated model. That implies $2^{99} \approx 6 \cdot 10^{29}$ possible combinations of which competitors will adjust their wage next period. Evaluating this expectation exactly is impossible even for a single firm, not to mention finding the mutual best responses of 100 firms. And that is just local equilibrium on a single market.

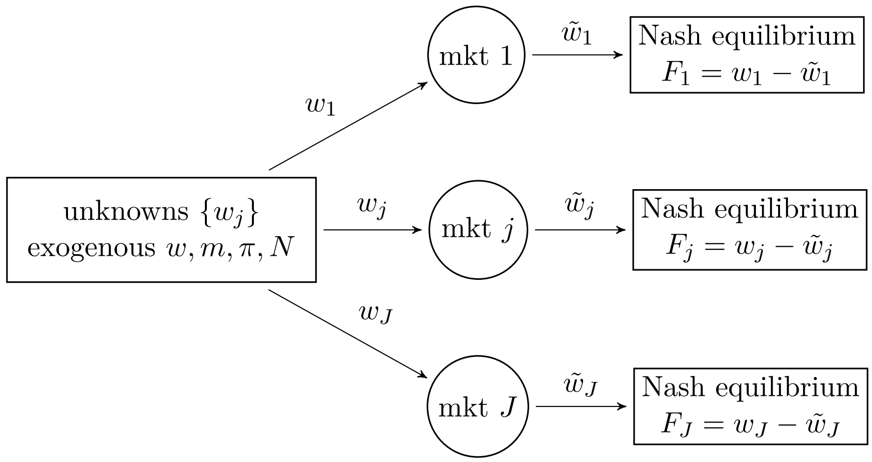
Solution. Since preserving idiosyncratic uncertainty in full is impossible, we will assume some of it away. Specifically, we assume that firms have perfect foresight of the local wage index $\{W_{jt+\tau}\}_{\tau=0}^{\infty}$, but are uncertain about their own Calvo shocks. This approach gives us enough tractability to proceed, while it preserves precautionary behavior with respect to agents' own idiosyncratic shocks. This is the weakest assumption we need to apply the sequence-space Jacobian method (Auclert et al., 2021) to our oligopsonistic economy.

The modularity of the SSJ method enables us to overcome the challenge of many markets. The key property of the model we exploit is that firms on market j don't compete directly with firms on market j'. This can be seen clearly from the labor supply curve

$$n_{ijt} = \left(\frac{W_{ijt}}{W_{jt}}\right)^{\eta} \left(\frac{W_{jt}}{W_t}\right)^{\zeta} N_t. \tag{5.2}$$

[11] Our calibrated model features about 100,000 firms, distributed over 3,000 markets. Solving transition dynamics over 100 quarters means characterizing $100 \times 100,000$ firm-level wages, 100×3000 local wages plus 100×2 aggregate prices.

Figure 2: DAG Representation of Local Equilibria Conditional on Aggregates



Firm ij considers how its wage W_{ijt} compares to the local wage W_{jt} and to the aggregate wage W_t. Local wages on other markets matter only through their effect on the aggregate wage. Given this insight, we solve the model in two steps. In the inner layer, we solve for Nash equilibria on local labor markets independently of each other and conditional on aggregates. In the outer layer, we solve for general equilibrium on national markets.

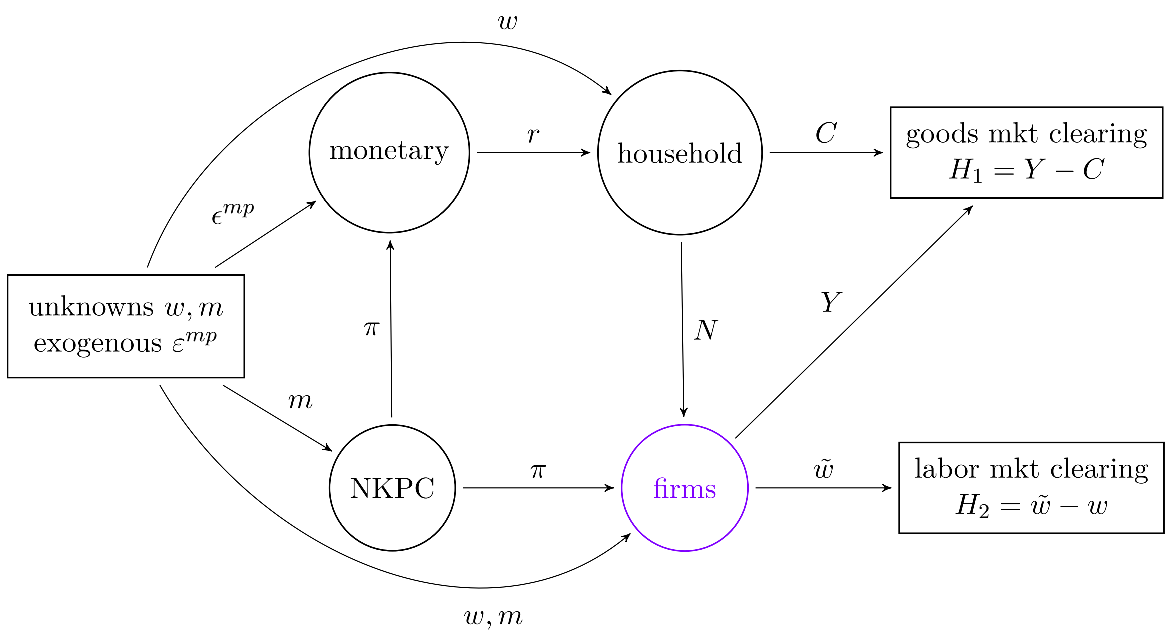
Applying the sequence-space Jacobian method. Let's start with the description of the inner layer. The theoretical model has a continuum of local markets and runs forever. In practice, we use a large but finite number of markets, J, and assume that the economy returns to steady state after some finite number of periods T. For every market $j = 1, 2, \ldots, J$, we implement a function $F_j : \mathbb{R}^{T+4T} \to \mathbb{R}^T$ which maps the sequence of the local wage index $\{w_{jt}\}_{t=1}^T$ and sequences of aggregates $\{m_t, \pi_t, w_t, N_t\}_{t=1}^T$ into a sequence of residuals $\{w_{jt} - \tilde{w}_{jt}\}_{t=1}^T$, where $\tilde{w}_{jt}$ is the local wage that emerges from the optimal decisions and Calvo shocks of firms on market j taking inputs as given. Nash equilibrium is the fixed point $w_{jt} = \tilde{w}_{jt}$ for all $t = 1, \ldots, T$. By the implicit function theorem, the Jacobians of Nash equilibrium wages with respect to aggregates are given by

$$\frac{dw_j}{dx} = -\left(\frac{\partial F_j}{\partial w_j}\right)^{-1} \frac{\partial F_j}{\partial x} \qquad \text{for } x \in \{m, \pi, w, N\}. \tag{5.3}$$

In sum, we can give a general characterization of local equilibria conditional on aggregate fuctuations around the steady state by manipulating $T \times T$ matrices instead of $JT \times JT$ matrices which would be infeasible for large J. Figure 2 visualizes the directed acyclic graph (DAG) representing this inner layer.

The market-specific Jacobians of the inner layer may be consolidated into Jacobians of aggregate

Figure 3: DAG Representation of General Equilibrium



wage and output as

$$\frac{d\tilde{w}}{dx} = w_{ss}^{-\zeta}\left(\frac{1}{J}\sum_{j=1}^{J} w_{j,ss}^{\zeta}\frac{dw_j}{dx}\right), \qquad \frac{dY}{dx} = \frac{1}{J}\sum_{j=1}^{J}\frac{dY_j}{dx} \qquad \text{for } x \in \{m, \pi, w, N\}. \tag{5.4}$$

Equipped with these Jacobians of the firm block, we can write down an implicit function that characterizes aggregate prices $\{m_t, w_t\}_{t=1}^{T}$ conditional on monetary policy shocks $\{\varepsilon_t^{mp}\}_{t=1}^{T}$. Figure 3 shows the DAG representation of this outer layer. The solution, given by the implicit function theorem, is analogous to that of the inner layer

$$\begin{bmatrix} dw \\ dm \end{bmatrix} = -\begin{bmatrix} \frac{\partial H_1}{\partial w} & \frac{\partial H_1}{\partial m} \\ \frac{\partial H_2}{\partial w} & \frac{\partial H_2}{\partial m} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial H_1}{\partial \varepsilon^{mp}} \\ \frac{dH_2}{d\varepsilon^{mp}} \end{bmatrix} d\varepsilon^{mp}. \tag{5.5}$$

The impulse responses of other endogenous variables can be obtained by forward accumulation of block Jacobians along the DAG, as explained by Auclert et al. (2021).

The detailed description of Jacobian construction is relegated to appendix B.1. In practice, we choose $J = 3,000$ markets and verify that the solution is not sensitive to increasing the number of markets further. Similarly, we verify that the truncation horizon $T = 100$ is sufficiently high that increasing it has negligible impact on our experiments. Figure B.4 illustrates the Jacobians of local equilibrium wages obtained by (5.3) and of the aggregate wage obtained by aggregating markets according to (5.4). Figure B.5 shows that the impact of Calvo draws washes out by aggregation across sufficiently high number of markets.

5.2 Calibration

We distinguish two sets of parameters. First, parameters that are common in New Keynesian models. These we set to conventional values. Second, parameters of the oligopsonistic firm block. These we calibrate internally to match moments of the distribution of firms and local labor markets. Table 1 summarizes the calibration. Table 2 shows the fit of targeted moments.

Fixed parameters. We assume that the representative household has separable preferences over consumption and labor aggregates

$$U(C_t, N_t) = \frac{C_t^{1-\gamma}}{1-\gamma} - \varphi \frac{N_t^{1+\nu}}{1+\nu}, \tag{5.6}$$

and set the elasticity of intertemporal substitution $1/\gamma = 0.5$, and the Frisch elasticity $1/\nu = 0.5$. We set the quarterly discount factor to $0.98^{1/4}$, implying an annual real interest rate of 2%. We assume zero inflation in steady state. We set the elasticity of substitution between retail goods to $\epsilon = 7$, whose only role is to pin down the relative price of intermediate goods in steady state at $m = (\epsilon - 1)/\epsilon$.

For now, we set the primitive labor supply elasticities to $\eta = 3.74$ and $\zeta = 0.76$ following Berger et al. (2019). Nominal wages are reset once a year on average, $\theta_w = 0.75$. Prices are slightly more sticky, $\theta_p = 0.85$. The Taylor rule coefficient on inflation is $\phi_\pi = 1.5$. Finally, we assume that the monetary policy shock follows and AR(1) process with quarterly autocorrelation $\rho_\varepsilon = 0.7$. The standard deviation of innovations is irrelevant, since we linearize the model around $\sigma_\varepsilon = 0$.

Table 1: Summary of parameters

Parameter		Value	Parameter		Value
A. Fixed parameters			*B. Internally calibrated parameters*		
β	Discount factor	$0.98^{1/4}$	α	Returns to scale	0.79
$1/\gamma$	EIS	0.50	a_z	Productivity dist shape	3.01
$1/\nu$	Frisch elasticity	0.50	ξ_m	Market size distribution shape	$[0.79, 0.99]$
η	Within market elasticity	3.74	σ_m	Market size distribution scale	$[23.06, 15.55]$
ζ	Across market elasticity	0.76			
ϵ	Retail goods elasticity	7.00			
θ_p	Price stickiness	0.85			
θ_w	Wage stickiness	0.75			
ϕ_π	Taylor rule coefficient	1.5			
φ	Labor disutility	0.11			
ρ_ϵ	Persistence of MP shock	0.7			

Internally calibrated parameters. We estimate five parameters by targeting eight cross-sectional moments. The estimated parameters include the returns to scale α. We assume that productivity

Table 2: Targeted Moments

Moment	Source	Data	Model
A. High-monopsony firms' share			
Population	LEHD	0.06	0.06
Wage bill	LEHD	0.34	0.41
Employment	LEHD	0.32	0.24
B. Local HHI			
Mean	LEHD	0.23	0.20
Standard deviation	LEHD	0.29	0.18
10th percentile	LEHD	0.01	0.04
50th percentile	LEHD	0.08	0.15
90th percentile	LEHD	0.77	0.50

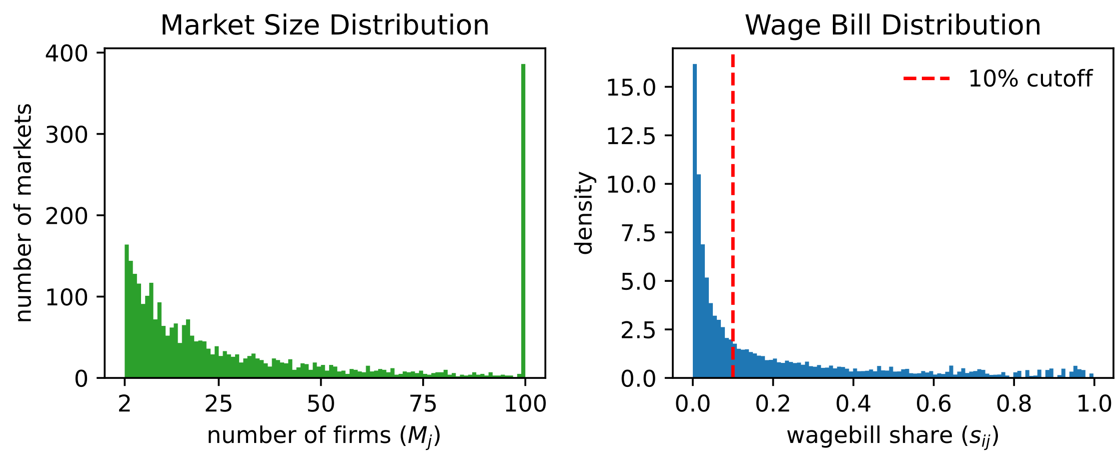
is Pareto distributed with shape a_z, and market size M_j is a mixture of two generalized Pareto distributions with shape ξ_m and scale σ_m. The theoretical model has a unit measure of local labor markets. When we solve the model numerically, we work with $3,000$ markets. Given our calibration of the market size distribution, we end up with about $100,000$ firms in total. We normalize aggregate labor supply to $N = 1$ and back out the labor disutility parameter that justifies this as an equilibrium outcome.

Comparisons between low-, and high-monopsony firms are central to our analysis. Panel A. of table 2 shows that only 6% of firms are high-monopsony—defined as having a wage bill share above 10%— but they account for about a third of the aggregate employment and wage bill. The model matches the population share, but overstates the wage bill share and understates the employment share. This is because there are only two ways in which a firm can acquire large market share in our model. First, by having few local competitors. Second, by having high productivity, in which case it will also pay a high wage, albeit at a wider markdown below the marginal revenue product of labor. All in all, the model does a reasonably good job if matching the footprint of high-monopsony firms.

Wage bill-weighted local Herfindahl indexes are the key measure of labor market concentration in the model. Panel B. of table 2 shows five moments of the HHI distribution across markets. The average market is highly concentrated, with a mean HHI above 0.2. This is the value we would observe with just five equally sized firms. Variation between markets is large, as shown by the high standard deviation and fat tails of the distribution. Our model matches the average HHI and implies substantial dispersion, though the latter still falls fort of the data.

Figure 4 shows the resulting distribution of market size and wage bill share. The market size distribution is bimodal. Duopsonies and effectively competitive markets with a 100 firms are the two most common. The wage bill distribution is concentrated at the bottom (most firms have low market power, but there's significant mass throughout the unit interval. Note that wage bill share (and HHI) above 0.5 is possible only due to dispersion ion firm-level productivity.

Figure 4: Steady-State Distributions



5.3 Labor Market Power and the Efficacy of Monetary Policy

First, we demonstrate that our calibrated model replicates the differential responses of low- and high labor market firms to a monetary policy shock. Second, we use the model to perform counterfactual experiments that uncover the role of oligopsony in the transmission of monetary policy. Table 3 summarizes our results. We find that labor market power substantially dampens the real effects of monetary policy, while having limited impact on its effect on inflation.

Heterogeneous passthrough. Figure 5 shows the model-implied impulse response of the real wage bill to an expansionary monetary policy shock, disaggregated by labor market power at the firm level. The response of low monopsony firms—which account for less than 10% of the local wage bill—is about 1.5 times higher than the response of high monopsony firms. The intuition for the result is the same as in the stylized model of 4. Quantitatively, the difference between the relative responses is close to our empirical estimates from the Census data shown by figure 1. Matching the hump-shaped pattern is a matter of quantitative refinements that we haven't pursued yet.[12]

No market power. In the first counterfactual exercise, we eliminate labor market power entirely by imposing that the desired markdown is 1 for all firms. We keep the distribution of firms, firm-specific labor supply functions, and wage stickiness the same as in the benchmark economy.

Table 3 shows that eliminating labor market power amplifies the impact of monetary policy on aggregate output by 26–32%. Cumulative output response is defined as the present value of output deviations from steady state. In contrast, the effect of on inflation dynamics is modest. Inflation response on impact rises slightly, while long-run change in the price level is 9% lower. Appendix figure B.6 shows the corresponding impulse responses.

[12]Partial indexation of wages and information frictions are features that may generate hump shaped wage and employment responses.

Figure 5: Response of Wage Bill to a Monetary Policy Shock

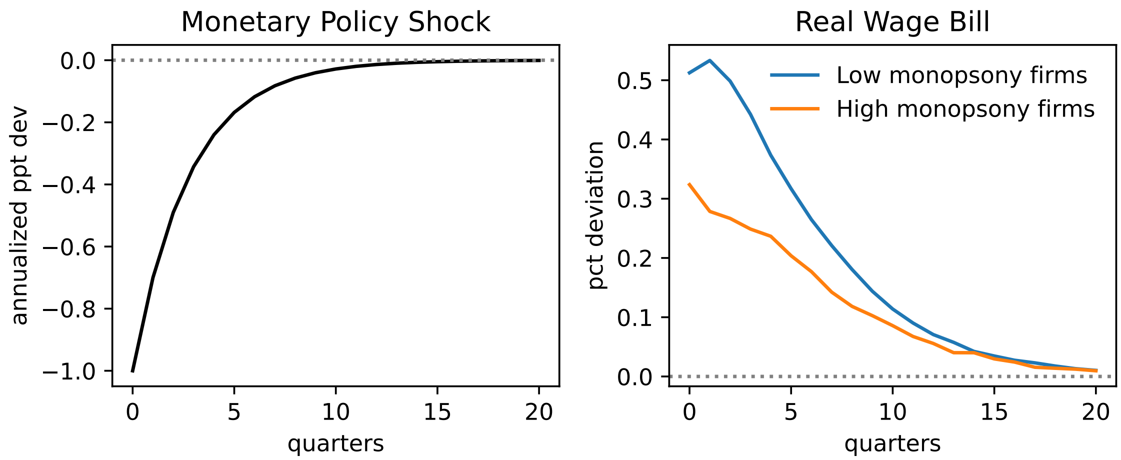


Table 3: Summary of Counterfactual Exercises

	Benchmark	No market power	Higher concentration
Mean local HHI	0.20	0.07	0.28
Aggregate markdown	1.33	1.00	1.35
Output response: impact	1.00	1.26	0.90
Output response: cumulative	1.00	1.32	0.85
Inflation response: impact	1.00	1.02	0.99
Inflation response: cumulative	1.00	0.91	1.04

Note: Output and inflation responses are to the same AR(1) monetary policy shock with $\rho_\epsilon = 0.7$. We normalize impact and cumulative responses separately by their value in the benchmark economy.

Historical trends in labor market power. Our second counterfactual exercise is aimed at gauging the potential of historical trends in local concentration to affect monetary policy transmission. To this end, we recalibrate the shape parameter of the productivity distribution to increase the average wage bill-weighted HHI by 8 basis points, from 0.2 to 0.28. This change is close in magnitude to that reported in the literature, and suggested by our own preliminary analysis of the data.

The last column of table 3 shows that this relatively modest increase in concentration dampens output response to monetary policy by 10–15%. Similarly to the previous exercise, the impact on inflation response is much smaller, only -1–4%. Appendix figure B.7 shows the corresponding impulse responses.

In sum, our analysis suggests that the decline in local labor market concentration has significantly increased the efficacy of monetary policy.

6 Conclusion

To be written.

References

Atkeson, A. and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031. 3, 5.1

Auclert, A., B. Bardóczy, M. Rognlie, and L. Straub (2021): "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 89, 2375–2408. 1, 5.1, 5.1

Berger, D., K. Herkenhoff, and S. Mongey (2019): "Labor Market Power," . 5.2

——— (2022): "Labor Market Power," *American Economic Review*, 112, 1147–93. 1, 1, 2, 3, 5, 3.1, 7, 5.1

Boppart, T., P. Krusell, and K. Mitman (2018): "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 89, 68–92. 1

Fershtman, C. and M. I. Kamien (1987): "Dynamic duopolistic competition with sticky prices," *Econometrica: Journal of the Econometric Society*, 1151–1164. 6

Fort, T. C., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2013): "How firms respond to business cycles: The role of firm age and firm size," *IMF Economic Review*, 61, 520–559. 2.2

Jarociński, M. and P. Karadi (2020): "Deconstructing monetary policy surprises—the role of information shocks," *American Economic Journal: Macroeconomics*, 12, 1–43. 1, 2.1

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American economic review*, 95, 161–182. 1, 2.2

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary policy according to HANK," *American Economic Review*, 108, 697–743. 1

Montiel Olea, J. L. and M. Plagborg-Møller (2021): "Local projection inference is simpler and more robust than you think," *Econometrica*, 89, 1789–1823. 1

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2021): "Diverging Trends in National and Local Concentration," *NBER Macroeconomics Annual*, 35, 115–150. 1

Wang, O. and I. Werning (2022): "Dynamic Oligopoly and Price Stickiness," *American Economic Review*, 112, 2815–49. 6

Yeh, C., C. Macaluso, and B. Hershbein (2022): "Monopsony in the US Labor Market," *American Economic Review*, 112, 2099–2138. 1

Online Appendix

A Mathematical Derivations

A.1 Household's problem

The representative household solves

$$\max_{c_{kt}, n_{ijt}, B_t} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(c_t, n_t) \tag{A.1}$$

$$\text{s.t.} \quad \int_0^1 p_{kt} c_{kt} dk + B_t = (1 + i_{t-1}) B_{t-1} + \int_0^1 \left(\sum_{i=1}^{M_j} W_{ijt} n_{ijt} \right) dj + \Pi_t \tag{A.2}$$

$$c_t = \left(\int_0^1 c_{kt}^{\frac{\epsilon-1}{\epsilon}} dk \right)^{\frac{\epsilon}{\epsilon-1}} \tag{A.3}$$

$$n_t = \left(\int_0^1 n_{jt}^{\frac{\zeta+1}{\zeta}} dj \right)^{\frac{\zeta}{\zeta+1}} \tag{A.4}$$

$$n_{jt} = \left(\sum_{i=1}^{M} n_{ijt}^{\frac{\eta+1}{\eta}} \right)^{\frac{\eta}{\eta+1}} \tag{A.5}$$

We solve this problem in four steps.

Step 1. Derive retailer-specific demand curves and aggregate price index from a static expenditure minimization problem.

$$P_t C_t = \min_{\{c_{kt}\}} \int_0^1 p_{kt} c_{kt} dk \qquad \text{s.t.} \left(\int_0^1 c_{kt}^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} = \bar{U} \tag{A.6}$$

Let λ_t denote the Lagrange multiplier on the constraint. The FOC is

$$p_{kt} = \lambda_t C_t^{\frac{1}{\epsilon}} c_{kt}^{-\frac{1}{\epsilon}} \tag{A.7}$$

Integrating across varieties yields

$$\int_0^1 p_{kt} c_{kt} dk = \lambda_t C_t^{\frac{1}{\epsilon}} \int_0^1 c_{kt}^{\frac{\epsilon-1}{\epsilon}} dk = \lambda_t C_t \tag{A.8}$$

Combining this last last expression with the definition of the price index as $P_t C_t = \int_0^1 p_{kt} c_{kt} dk$ implies that $P_t = \lambda_t$. Substitute this into the FOC to get the retailer-specific demand curve

$$c_{kt} = \left(\frac{p_{kt}}{P_t} \right)^{-\epsilon} C_t \tag{A.9}$$

Substitute the demand curve back into the FOC to express the price index

$$p_{kt}c_{kt} = p_{kt}^{1-\epsilon}P_t^{\epsilon}C_t \tag{A.10}$$

$$\int_0^1 p_{kt}c_{kt}dk = P_t^{\epsilon}C_t\int_0^1 p_{kt}^{1-\epsilon}dk \tag{A.11}$$

$$P_tC_t = P_t^{\epsilon}C_t\int_0^1 p_{kt}^{1-\epsilon}dk \tag{A.12}$$

$$P_t = \left(\int_0^1 p_{kt}^{1-\epsilon}dk\right)^{\frac{1}{1-\epsilon}} \tag{A.13}$$

Step 2. Derive local labor-market specific labor supply and aggregate wage index from static income maximization problem.

$$W_tN_t = \max_{\{n_{jt}\}}\int_0^1 W_{jt}n_{jt}dj \qquad \text{s.t.} \left(\int_0^1 n_{jt}^{\frac{\zeta+1}{\zeta}}dj\right)^{\frac{\zeta}{\zeta+1}} = N_t \tag{A.14}$$

Let λ_t denote the Lagrange multiplier on the constraint. The FOC is

$$W_{jt} = \lambda_t N_t^{-\frac{1}{\zeta}} n_{jt}^{\frac{1}{\zeta}} \tag{A.15}$$

Integrating across local labor markets yields

$$\int_0^1 W_{jt}n_{jt}dj = \lambda_t N_t^{-\frac{1}{\zeta}}\int_0^1 n_{jt}^{\frac{\zeta+1}{\zeta}}dj = \lambda_t N_t \tag{A.16}$$

Combining this last last expression with the definition of the aggregate wage index as $W_tN_t = \int_0^1 W_{jt}n_{jt}dj$ implies that $W_t = \lambda_t$. Substitute this into the FOC to get the local labor market-specific labor supply curve

$$n_{jt} = \left(\frac{W_{jt}}{W_t}\right)^{\zeta} N_t \tag{A.17}$$

Substitute the labor supply curve back into the FOC to express the aggregate wage index

$$W_{jt}n_{jt} = W_{jt}^{1+\zeta}W_t^{-\zeta}N_t \tag{A.18}$$

$$\int_0^1 W_{jt}n_{jt}dj = W_t^{-\zeta}N_t\int_0^1 W_{jt}^{1+\zeta}dj \tag{A.19}$$

$$W_tN_t = W_t^{-\zeta}N_t\int_0^1 W_{jt}^{1+\zeta}dj \tag{A.20}$$

$$W_t = \left(\int_0^1 W_{jt}^{1+\zeta}dj\right)^{\frac{1}{1+\zeta}} \tag{A.21}$$

Step 3. Derive firm-specific labor supply and local wage indexes from static income maximization problem.

$$W_{jt} n_{jt} = \max_{\{n_{ijt}\}} \sum_{i=1}^{M_j} W_{ijt} n_{ijt} \qquad \text{s.t.} \left(\sum_{i=1}^{M} n_{ijt}^{\frac{\eta+1}{\eta}} \right)^{\frac{\eta}{\eta+1}} = n_{jt} \tag{A.22}$$

Analogously to the second step, we obtain

$$n_{ijt} = \left(\frac{W_{ijt}}{W_{jt}} \right)^{\eta} n_{jt} \tag{A.23}$$

$$W_{jt} = \left(\sum_{i=1}^{M_j} W_{ijt}^{1+\eta} \right)^{\frac{1}{1+\eta}} \tag{A.24}$$

Combining (A.17) and (A.23) imply the firm-specific labor supply curve in the main text

$$n_{ijt} = \left(\frac{W_{ijt}}{W_{jt}} \right)^{\eta} \left(\frac{W_{jt}}{W_t} \right)^{\zeta} N_t \tag{A.25}$$

Step 4. Derive aggregate consumption, labor supply, and savings choices from dynamic utility maximization problem.

$$\max_{C_t, N_t, B_t} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(C_t, N_t) \tag{A.26}$$

$$\text{s.t. } P_t C_t + B_t = (1 + i_{t-1}) B_{t-1} + W_t N_t + \Pi_t \tag{A.27}$$

Let λ_t denote the Lagrange multipliers on the budget constraint for all $t \geq 0$. The FOCs are

$$[C_t] : \qquad 0 = U_C(C_t, N_t) - \lambda_t P_t \tag{A.28}$$

$$[N_t] : \qquad 0 = U_N(C_t, N_t) + \lambda_t W_t \tag{A.29}$$

$$[B_t] : \qquad 0 = -\lambda_t + \beta \mathbb{E}_t[(1 + i_t)\lambda_{t+1}] \tag{A.30}$$

Combining the FOCs yields the aggregate Euler equation and labor supply curves

$$U_C(C_t, N_t) = \beta \mathbb{E}_t \left[\frac{1 + i_t}{\pi_{t+1}} U_C(C_{t+1}, N_{t+1}) \right], \qquad \frac{W_t}{P_t} = -\frac{U_N(C_t, N_t)}{U_C(C_t, N_t)} \tag{A.31}$$

A.2 Retailers

Firm k solves

$$\max_{p_{kt}^*,\{c_{kt+\tau},y_{kt+\tau}\}} \sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} \left(p_{kt}^* c_{kt+\tau} - M_{t+\tau} y_{kt+\tau}\right) \tag{A.32}$$

$$\text{s.t. } c_{kt+\tau} = y_{kt+\tau} \tag{A.33}$$

$$c_{kt+\tau} = \left(\frac{p_{kt}^*}{P_{t+\tau}}\right)^{-\epsilon} C_{t+\tau} \tag{A.34}$$

Substitute the constraints into the objective function

$$\sum_{\tau=0}^{\infty} R_{t,t+\tau} C_{t+\tau} P_{t+\tau}^{\epsilon} \left[(p_{kt}^*)^{1-\epsilon} - M_{t+\tau}(p_{kt}^*)^{-\epsilon}\right]$$

The FOC wrt p_t^* is

$$0 = \sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} C_{t+\tau} P_{t+\tau}^{\epsilon} \left[(1-\epsilon)(p_{kt}^*)^{-\epsilon} + \epsilon M_{t+\tau}(p_{kt}^*)^{-\epsilon-1}\right] \tag{A.35}$$

$$0 = \sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} C_{t+\tau} P_{t+\tau}^{\epsilon} \left[(1-\epsilon) p_{kt}^* + \epsilon M_{t+\tau}\right] \tag{A.36}$$

$$p_{kt}^* = \frac{\epsilon}{\epsilon-1} \frac{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} P_{t+\tau}^{\epsilon} C_{t+\tau} M_{t+\tau}}{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} P_{t+\tau}^{\epsilon} C_{t+\tau}} \tag{A.37}$$

Given that Calvo shocks are iid and the price index is given by

$$P_t = \left(\int_0^1 p_{kt}^{1-\epsilon} dk\right)^{\frac{1}{1-\epsilon}} \tag{A.38}$$

we can write the law of motion for the price level as

$$P_t = \left[(1-\theta_p)(p_{kt}^*)^{1-\epsilon} + \theta_p P_{t-1}^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}} \tag{A.39}$$

Phillips curve. Divide (A.37) and (A.39) with P_t to eliminate nominal variables

$$\frac{p_{kt}^*}{P_t} = \frac{\epsilon}{\epsilon-1} \frac{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} \pi_{t,t+\tau}^{\epsilon+1} C_{t+\tau} m_{t+\tau}}{\sum_{\tau=0}^{\infty} \theta_p^\tau R_{t,t+\tau} \pi_{t,t+\tau}^{\epsilon} C_{t+\tau}} \tag{A.40}$$

$$\frac{p_t^*}{P_t} = \left(\frac{1-\theta_p \pi_{t-1,t}^{\epsilon-1}}{1-\theta_p}\right)^{\frac{1}{1-\epsilon}} \tag{A.41}$$

where $m_t \equiv M_t/P_t$ denotes real marginal cost. We can express inflation recursively as

$$\frac{K_t}{F_t} = \left(\frac{1-\theta_p \pi_{t-1,t}^{\epsilon-1}}{1-\theta_p}\right)^{\frac{1}{1-\epsilon}} \tag{A.42}$$

$$K_t = \epsilon C_t m_t + \theta_p \frac{\pi_{t,t+1}^{\epsilon+1}}{1+i_t} K_{t+1} \tag{A.43}$$

$$F_t = (1-\epsilon) C_t + \theta_p \frac{\pi_{t,t+1}^{\epsilon}}{1+i_t} F_{t+1} \tag{A.44}$$

Loglinearize these equations, assuming $\pi_{ss} = 1$

$$\hat{K}_t - \hat{F}_t = \frac{\theta_p}{1-\theta_p} \hat{\pi}_{t-1,t} \tag{A.45}$$

$$\hat{K}_t = (1-\beta\theta_p)\left(\hat{C}_t + \hat{m}_t\right) + \beta\theta_p \left[(\epsilon+1)\hat{\pi}_{t,t+1} + \hat{K}_{t+1} - \beta d i_t\right] \tag{A.46}$$

$$\hat{F}_t = (1-\beta\theta_p)\hat{C}_t + \beta\theta_p \left[\epsilon\hat{\pi}_{t,t+1} + \hat{F}_{t+1} - \beta d i_t\right] \tag{A.47}$$

Combine these to get the textbook NKPC

$$\frac{\theta_p}{1-\theta_p}\hat{\pi}_{t-1,t} = (1-\beta\theta_p)\hat{m}_t + \beta\theta_p \left[\hat{\pi}_{t,t+1} + \frac{\theta_p}{1-\theta_p}\hat{\pi}_{t,t+1}\right] \tag{A.48}$$

$$\hat{\pi}_{t-1,t} = \frac{(1-\theta_p)(1-\beta\theta_p)}{\theta_p}\hat{m}_t + \beta\hat{\pi}_{t,t+1} \tag{A.49}$$

A.3 Producers

A firm that is allowed to reset its wage solves

$$\max_{W^*_{ijt}, \{W_{jt+\tau}, n_{ijt+\tau}\}} \quad \sum_{\tau=0}^{\infty} \theta_w^\tau R_{t,t+\tau} \left(M_{t+\tau} z_{ij} n_{ijt+\tau}^{\alpha} - W^*_{ijt} n_{ijt+\tau}\right) \tag{A.50}$$

$$\text{s.t.} \qquad n_{ijt+\tau} = \left(\frac{W^*_{ijt}}{W_{jt+\tau}}\right)^{\eta} \left(\frac{W_{jt+\tau}}{W_{t+\tau}}\right)^{\zeta} N_{t+\tau} \tag{A.51}$$

$$W_{jt+\tau} = \left((W^*_{ijt})^{1+\eta} + \sum_{i' \neq i} W_{i'jt+\tau}^{1+\eta}\right)^{\frac{1}{1+\eta}} \tag{A.52}$$

FOC. Let $\mu_{ijt+\tau}$ denote the Lagrange multiplier on labor supply, and $\nu_{ijt+\tau}$ denote the Lagrange multiplier on the wage index. The FOCs are

$$[W^*_{ijt}]: \qquad \sum_{\tau=0}^{\infty} \theta_w^\tau R_{t,t+\tau} W^*_{ijt} n_{ijt+\tau} = \sum_{\tau=0}^{\infty} \left[\eta n_{ijt+\tau}\mu_{ijt+\tau} - W_{jt+\tau}^{-\eta}\left(W^*_{ijt}\right)^{1+\eta} \nu_{ijt+\tau}\right] \tag{A.53}$$

$$[W_{jt+\tau}]: \qquad (\eta - \zeta) n_{ijt+\tau}\mu_{ijt+\tau} = \nu_{ijt+\tau} W_{jt+\tau} \tag{A.54}$$

$$[n_{ijt+\tau}]: \qquad \theta_w^\tau R_{t,t+\tau}\left(M_{t+\tau} z_{ij} \alpha n_{ijt+\tau}^{\alpha-1} - W^*_{ijt}\right) = \mu_{ijt+\tau} \tag{A.55}$$

Combine all three

$$
\begin{aligned}
\sum_{\tau=0}^{\infty} \theta_w^{\tau} R_{t,t+\tau} W_{ijt}^{*} n_{ijt+\tau} &= \sum_{\tau=0}^{\infty} \left[\eta n_{ijt+\tau} \mu_{ijt+\tau} - \left(\frac{W_{ijt}^{*}}{W_{jt+\tau}} \right)^{1+\eta} \nu_{ijt+\tau} W_{jt+\tau} \right] \\
&= \sum_{\tau=0}^{\infty} n_{ijt+\tau} \mu_{ijt+\tau} \left[\eta - (\eta - \zeta) \left(\frac{W_{ijt}^{*}}{W_{jt+\tau}} \right)^{1+\eta} \right] \\
&= \sum_{\tau=0}^{\infty} \theta_w^{\tau} R_{t,t+\tau} \left(M_{t+\tau} z_{ij} \alpha n_{ijt+\tau}^{\alpha} - W_{ijt}^{*} n_{ijt+\tau} \right) \left[\eta - (\eta - \zeta) \left(\frac{W_{ijt}^{*}}{W_{jt+\tau}} \right)^{1+\eta} \right]
\end{aligned}
$$

which we can rearrange to get a markdown formula

$$
W_{ijt}^{*} = \frac{\sum_{\tau=0}^{\infty} \theta_w^{\tau} R_{t,t+\tau} n_{ijt+\tau} \left[\eta - (\eta - \zeta) \left(\frac{W_{ijt}^{*}}{W_{jt+\tau}} \right)^{1+\eta} \right] M_{t+\tau} z_{ij} \alpha n_{ijt+\tau}^{\alpha-1}}{\sum_{\tau=0}^{\infty} \theta_w^{\tau} R_{t,t+\tau} n_{ijt+\tau} \left[1 + \eta - (\eta - \zeta) \left(\frac{W_{ijt}^{*}}{W_{jt+\tau}} \right)^{1+\eta} \right]} \tag{A.56}
$$

Wage bill share. Next, we prove that $\left(\frac{W_{ijt}}{W_{jt}} \right)^{1+\eta}$ is the wage bill share of firm ij in period t.

$$
s_{ijt} \equiv \frac{W_{ijt} n_{ijt}}{\sum_{l=1}^{M_j} W_{ljt} n_{ljt}} = \frac{W_{ijt} n_{ijt}}{\sum_{l=1}^{M_j} W_{ljt} n_{ijt} \left(\frac{W_{ijt}}{W_{ljt}} \right)^{-\eta}} = \frac{W_{ij}^{1+\eta}}{\sum_l W_{ljt}^{1+\eta}} = \frac{W_{ijt}^{1+\eta}}{W_{jt}^{1+\eta}} \tag{A.57}
$$

The first equality is the definition of the local wage bill share. The second equality uses the ratio of labor supply curves to firms i and l. The third equality just collects and cancels terms. The fourth equality uses the expression of the local wage index.

A.4 Special case

Proof of Proposition 1. Proof. Conjecture that $w_{ij}^1 = A w_{ij}^0$ for all i and j. Then $\frac{w_{ij}^0}{w_j^0} = \frac{w_{ij}^1}{w_j^1}$ and $\epsilon_{ij}^1 = \epsilon_{ij}^0$. Equation (4.7), can be simplified to

$$
A w_{ij}^0 = \frac{\epsilon_{ij}^0}{1 + \epsilon_{ij}^0} z_{ij} \left(\frac{w_{ij}^0}{w_j^0} \right)^{\eta(\alpha-1)} \left(w_j^1 \right)^{\zeta(\alpha-1)} \tilde{D}_1, \tag{A.58}
$$

Using our conjecture, we have $w_j^1 = A w_j^0$, so that the equation above can be written as

$$
A w_{ij}^0 = \frac{\epsilon_{ij}^0}{1 + \epsilon_{ij}^0} z_{ij} \left(\frac{w_{ij}^0}{w_j^0} \right)^{\eta(\alpha-1)} A^{\zeta(\alpha-1)} \left(w_j^0 \right)^{\zeta(\alpha-1)} \tilde{D}_1. \tag{A.59}
$$

Using that equation (4.7) holds with equality in period 0, we obtain

$$
A = \tilde{D}_1^{\frac{1}{1+\zeta(1-\alpha)}}. \tag{A.60}
$$

Indeed, this validates our conjecture that the value A does not vary across the three firms. And we obtain $s^1_{ij} = s^0_{ij}$ for all i and j. ■

Proof of Proposition 2. Proof. Suppose by contradiction that $\Delta \ln w_{jl} \geq \Delta \ln w_{js}$. We can rewrite equation (4.7) as follows:

$$s_{ij}^{\frac{1+(1-\alpha)\eta}{1+\eta}} = \frac{\eta - (\eta-\zeta)s_{ij}}{1+\eta-(\eta-\zeta)s_{ij}} \alpha D z_{ij} \left(\frac{w_j}{w}\right)^{-(1-\alpha)\zeta} N^{-(1-\alpha)} w_j^{-1} \tag{A.61}$$

Divide by period 0 equation

$$\Delta s_{ij}^{\frac{1+(1-\alpha)\eta}{1+\eta}} = \Delta \frac{\eta - (\eta-\zeta)s_{ij}}{1+\eta-(\eta-\zeta)s_{ij}} \Delta D \left(\Delta \frac{w_j}{w}\right)^{-(1-\alpha)\zeta} \Delta w_j^{-1} \tag{A.62}$$

We want to show that if $\Delta s_{lj} \geq \Delta s_{hj}$, then $\Delta \frac{\eta-(\eta-\zeta)s_{lj}}{1+\eta-(\eta-\zeta)s_{lj}} < \Delta \frac{\eta-(\eta-\zeta)s_{hj}}{1+\eta-(\eta-\zeta)s_{hj}}$. This would lead to a contradiction. Let x_0, x_1, y_0, y_1 be the wage shares in period 0 and 1 of the large firm (x) and small firm (y). Let $\gamma = \eta - \zeta > 0$. Let's write the longer version:

$$\frac{\eta - \gamma x_1}{\eta - \gamma x_0} \frac{1+\eta-\gamma x_0}{1+\eta-\gamma x_1} < \frac{\eta - \gamma y_1}{\eta - \gamma y_0} \frac{1+\eta-\gamma y_0}{1+\eta-\gamma y_1}. \tag{A.63}$$

We can rearrange this to

$$1 - \frac{\gamma(x_1 - x_0)}{(\eta-\gamma x_0)(1+\eta-\gamma x_1)} < 1 - \frac{\gamma(y_1 - y_0)}{(\eta-\gamma y_0)(1+\eta-\gamma y_1)}. \tag{A.64}$$

So the condition holds if and only if

$$\frac{x_1 - x_0}{y_1 - y_0} \frac{\eta - \gamma y_0}{\eta - \gamma x_0} \frac{1+\eta-\gamma y_1}{1+\eta-\gamma x_1} > 1. \tag{A.65}$$

Recall that we've assumed $\frac{x_1}{x_0} \geq \frac{y_1}{y_0}$, which implies $\frac{x_1-x_0}{y_1-y_0} \geq \frac{x_0}{y_0}$. So it suffices to show that

$$\frac{x_0}{y_0} \frac{\eta - \gamma y_0}{\eta - \gamma x_0} \frac{1+\eta-\gamma y_1}{1+\eta-\gamma x_1} > 1. \tag{A.66}$$

Rearranging, we obtain that the condition holds if

$$\frac{\eta/y_0 - \gamma}{\eta/x_0 - \gamma} \frac{1+\eta-\gamma y_1}{1+\eta-\gamma x_1} > 1. \tag{A.67}$$

The first fraction is strictly greater than one if and only if $x_0 > y_0$, and the second fraction is strictly greater than one if and only if $x_1 > y_1$. Lemma 1 shows that indeed these two conditions hold.
In addition, we've implicitly assumed that $(y_1 - y_0) > 0$. That is, that $\Delta s_{lj} > 1$. To see that this is the case, recall that we've assumed $\Delta s_{lj} \geq \Delta s_{hj}$. Suppose by contradiction that $\Delta s_{lj} \leq 1$. Our assertion that $\Delta s_{lj} \geq \Delta s_{hj}$, then implies $\Delta s_{hj} \leq 1$. Since $w^1_m = w^0_m$, Δs_{lj} and Δs_{hj} are lower than one then imply that $\Delta w_{lj} \leq 1$ and $\Delta w_{hj} \leq 1$. This implies that $\Delta w_j \leq 1$. Now consider equation (A.62). The RHS consists of the multiplication of four terms. $\Delta w_j \leq 1$ implies that the third and fourth terms are

both greater than one. ΔD is strictly greater than one. And if $s^1_{lj} < s^0_{lj}$, then also the first term is greater than one. We then get a contradiction, because the RHS then implies that $\Delta s_{lj} > 1$. Thus, we have that Δs_{lj} is greater than 1, which concludes our proof as it leads to a contradiction. Therefore, we have that $\Delta \ln w_{lj} < \Delta \ln w_{hj}$. ■

LEMMA **1.** *Consider two firms in sector j with $z_1 > z_2$. Then, we have that $n_1 > n_2$.*

Proof. From the labor supply curve, we have

$$\frac{n_1}{n_2} = \left(\frac{w_1}{w_2}\right)^{\eta}.$$

So we need to show that $w_1 > w_2$, which is equivalent to $s_1 > s_2$. Rearranging the optimal wage setting equation (4.7), we obtain

$$s_{ij}^{\frac{1+(1-\alpha)\eta}{1+\eta}} = \frac{\eta - (\eta-\zeta)s_{ij}}{1+\eta-(\eta-\zeta)s_{ij}} \alpha D z_{ij} \left(\frac{w_j}{w}\right)^{-(1-\alpha)\zeta} N^{-(1-\alpha)} w_j^{-1}$$

So that

$$\left(\frac{s_1}{s_2}\right)^{\frac{1+(1-\alpha)\eta}{1+\eta}} = \frac{1-\frac{1}{1+\eta-(\eta-\zeta)s_1}}{1-\frac{1}{1+\eta-(\eta-\zeta)s_2}} \frac{z_1}{z_2}. \tag{A.68}$$

Suppose by contradiction then $s_2 \geq s_1$. Then, we have that

$$\frac{1}{1+\eta-(\eta-\zeta)s_1} \leq \frac{1}{1+\eta-(\eta-\zeta)s_2},$$

so that

$$\frac{1-\frac{1}{1+\eta-(\eta-\zeta)s_1}}{1-\frac{1}{1+\eta-(\eta-\zeta)s_2}} \geq 1.$$

However, because $\frac{z_1}{z_2} > 1$, equation (A.68) implies that $s_1 > s_2$ because the RHS of the equation is strictly greater than one. A contradiction. Thus, $s_1 > s_2$. ■

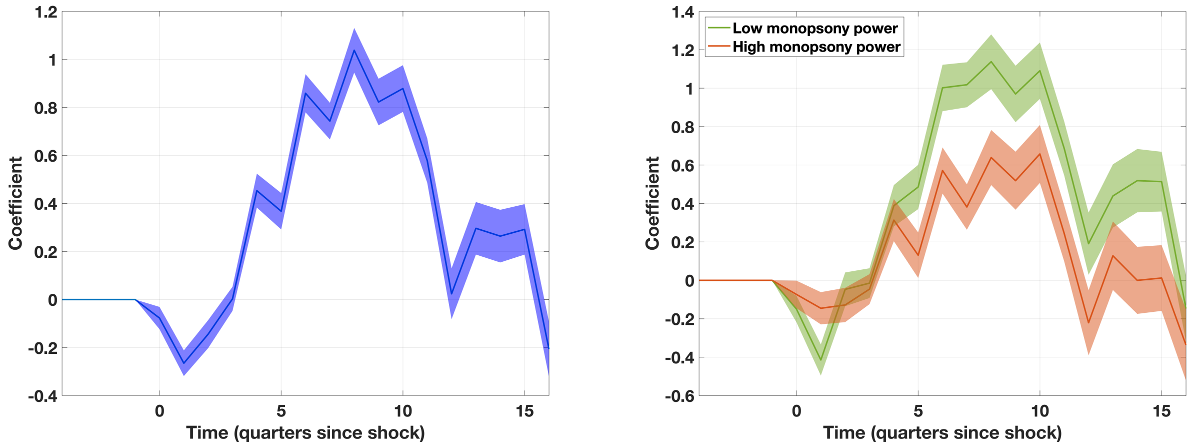
B Numerical Algorithm

B.1 Computing the Jacobians of the firm block

To be written.

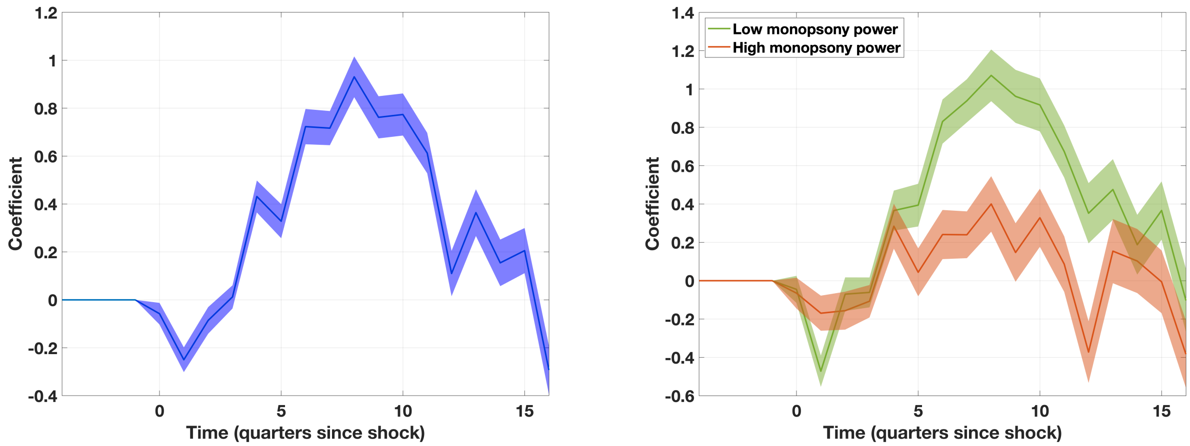
C Additional Figures and Tables

Figure B.1: Response of Log-Wage Bill to a Monetary Policy Shock with Age Controls



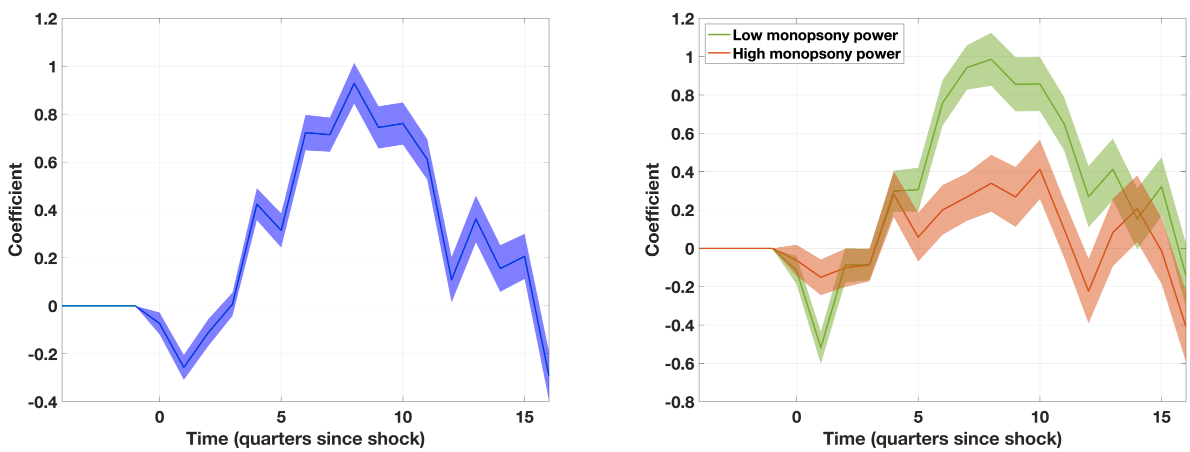
Note: This figure shows the response of log wage bill within a state-NAICS3 to a one-standard-deviation shock to the Fed Funds Rate. The regression includes about 160 thousand states-NAICS3-quarter observations. We plot the coefficient β_1^k for each horizon k. To reduce the impact of outliers, we trim the sample for each horizon k a the bottom and top 10% of the distribution of the change in $y_t^{s,i}$ between periods t and $t+k$. The shaded areas represent 95% confidence intervals. Standard errors clustered at the state-NAICS3 level.

Figure B.2: Response of Log-Wage Bill to a Monetary Policy Shock with Size Controls



Note: This figure shows the response of log wage bill within a state-NAICS3 to a one-standard-deviation shock to the Fed Funds Rate. The regression includes 160 thousand states-NAICS3-quarter observations. We plot the coefficient β_1^k for each horizon k. To reduce the impact of outliers, we trim the sample for each horizon k a the bottom and top 10% of the distribution of the change in $y_t^{s,i}$ between periods t and $t+k$. The shaded areas represent 95% confidence intervals. Standard errors clustered at the state-NAICS3 level.

Figure B.3: Response of Log-Wage Bill to a Monetary Policy Shock with Age and Size Controls



Note: This figure shows the response of log wage bill within a state-NAICS3 to a one-standard-deviation shock to the Fed Funds Rate. The regression includes xxx states-NAICS3-quarter observations. We plot the coefficient β_1^k for each horizon k. To reduce the impact of outliers, we trim the sample for each horizon k a the bottom and top 10% of the distribution of the change in $y_t^{s,i}$ between periods t and $t+k$. The shaded areas represent 95% confidence intervals. Standard errors clustered at the state-NAICS3 level.

Figure B.4: Column 10 of Jacobians of Local and Aggregate Wages

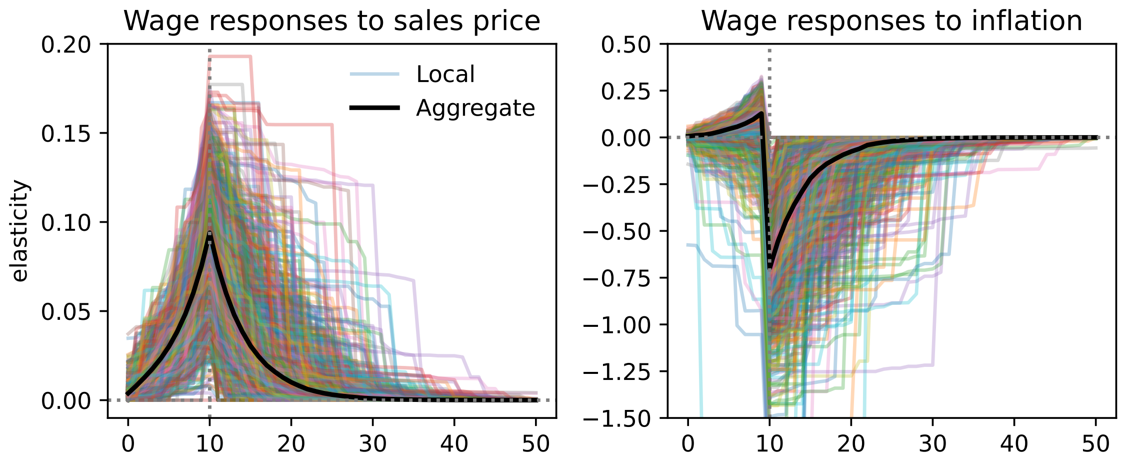


Figure B.5: Jacobians of Aggregate Wage for Two Sets of Calvo Shocks

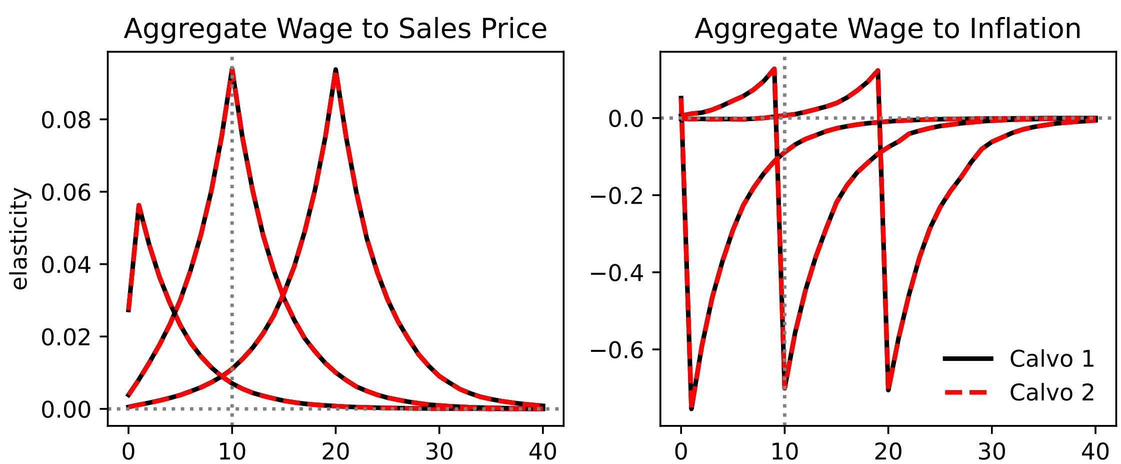


Figure B.6: Local Concentration and the Aggregate Impact of a Monetary Policy Shock I.

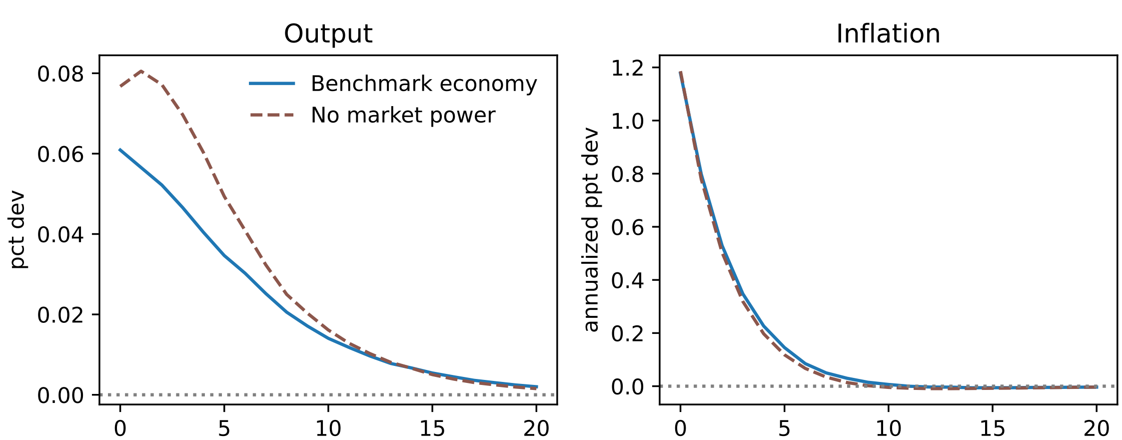


Figure B.7: Local Concentration and the Aggregate Impact of a Monetary Policy Shock II.