Monetary Policy, Markup Dispersion, and Aggregate TFP*

Matthias Meier[†] Timo Reinelt[‡]

August 26, 2021

Abstract

Motivated by empirical evidence that monetary policy affects aggregate TFP, we study the role of markup dispersion for monetary transmission. Empirically, we show that the response of markup dispersion to monetary policy shocks can account for a significant fraction of the aggregate TFP response in the first two years after the shock. Analytically, we show that heterogeneous price rigidity can explain the response of markup dispersion if firms have a precautionary price setting motive, which is present in common New Keynesian environments. We provide empirical evidence on the relationship between markups and price rigidity in support of this explanation. Finally, we study the mechanism and its implications in a quantitative model.

Keywords: Monetary policy, markup dispersion, heterogeneous price rigidity, aggregate productivity.
JEL codes: E30, E50.

*A previous version was titled "The misallocation channel of monetary policy." First draft: April 2019. We owe special thanks to Klaus Adam, Antonio Ciccone, John Fernald, Konrad Stahl, Venky Venkateswaran, and our discussants Sergio de Ferra, Viven Lewis, Yuliyan Mitkov, Gašper Ploj, and Alexander Matusche for helpful comments and suggestions. We would also like to thank conference and seminar participants at the 2018 Midwest Macro, BiGSEM, the 2019 T2M, Minneapolis Fed, San Francisco Fed, Universities of Heidelberg, Leipzig, Tübingen, National Bank of Ukraine, DIW, CEPR Warsaw, CRC TR 224, Nazarbayev University, E1Macro-QMQM, FINPRO, RGS, CBRT-ECB, ENTER, ESEM, Vigo, the 2020 Oslo Macro, Cleveland Fed-ECB, University of Tübingen, Frankfurt–Mannheim Macro, VfS, and the 2021 SED for numerous helpful comments. We thank Michael Weber for generously sharing data on price adjustment frequencies with us. Matthias Meier and Timo Reinelt gratefully acknowledge financial support from the German Research Foundation (DFG) through CRC TR 224 (Project C02), the Stiftung Geld & Währung, and the UniCredit & Universities Foundation. The contribution of Matthias Meier has been prepared under the Lamfalussy Fellowship Program sponsored by the ECB. Any views expressed are only those of the authors and do not necessarily represent the views of the ECB or the Eurosystem.

[†] Universität Mannheim, Department of Economics, Block L7, 3–5, 68161 Mannheim, Germany; E-mail: m.meier@uni-mannheim.de

[‡] Universität Mannheim, Center for Doctoral Studies in Economics (CDSE), Block L7, 3–5, 68161 Mannheim, Germany; E-mail: timo.reinelt@gess.uni-mannheim.de

1 Introduction

We revisit one of the long-standing questions in macroeconomics: What are the channels through which monetary policy affects real economic outcomes? Our paper is motivated by empirical evidence that monetary policy shocks have sizable effects on measured aggregate productivity.[1] A potential explanation for fluctuations in measured aggregate TFP is changing resource misallocation across firms. The TFP-misallocation link has been widely studied in the macro-development literature (e.g., Hsieh and Klenow, 2009), and is well understood in the New Keynesian literature. While in New Keynesian models, misallocation is commonly captured by price dispersion, our preferred empirical measure of misallocation is dispersion in markups across firms. Markup dispersion is price dispersion when controlling for differences in marginal costs across firms.

We study the role of markup dispersion for monetary transmission by asking two questions: First, does markup dispersion respond to monetary policy shocks? Using US data, we document a significant response of markup dispersion, which can account for a significant fraction of the aggregate TFP response up to two years after the shock. Second, what explains the response of markup dispersion? We show analytically that heterogeneity in price setting frictions – in an otherwise standard New Keynesian framework – can explain the response of markup dispersion. The fundamental reason is that firms with stickier prices have a stronger precautionary price setting motive. This channel has testable implications, which, as we show, are supported empirically. Finally, we study the mechanism and its implications in a quantitative model.

We estimate the response of markup dispersion to monetary policy shocks based on quarterly balance-sheet data and high-frequency identified monetary policy shocks. A central contribution of this paper is to show that the dispersion of markups across firms (within industries) significantly increases after contractionary monetary policy shocks and decreases

[1] Using US data, we document that monetary policy shocks lower measured aggregate productivity, which reconfirms the evidence in Evans and Santos (2002), Christiano et al. (2005), Moran and Queralto (2018), Garga and Singh (2021), and Jordà et al. (2020).

after expansionary monetary policy shocks. The response is persistent and peaks about two years after the shock. We establish this empirical pattern for a host of markup measures, following, amongst others, De Loecker and Warzynski (2012) and Gutiérrez and Philippon (2017). To translate the estimated response of markup dispersion into an aggregate TFP response, we follow Hsieh and Klenow (2009) and Baqaee and Farhi (2020). The response of markup dispersion implies a response in aggregate TFP between -0.2% and -0.4% two years after a one standard deviation contractionary monetary policy shock. For comparison, the directly estimated empirical response of utilization-adjusted aggregate TFP is -0.4% at a two-year horizon. At more distant horizons, markup dispersion accounts for a decreasing fraction of the aggregate TFP response.

Our evidence sheds new light on the TFP effects of monetary policy. Strikingly, the estimated response of markup dispersion cannot be explained by a large class of New Keynesian models, at least when solved with standard perturbation methods. In many New Keynesian models, including medium-scale models (e.g., Christiano et al., 2005) and models with heterogeneous price rigidity (e.g., Carvalho, 2006), markup dispersion does not respond to monetary policy shocks up to a first-order approximation around the deterministic steady state. In the second-order approximation, markup dispersion responds, but counterfactually increases in response to both positive and negative shocks. In models with trend inflation (e.g., Ascari and Sbordone, 2014), markup dispersion decreases after contractionary and increases after expansionary monetary shocks, which contradicts our empirical evidence.

What can explain the response of markup dispersion to monetary policy shocks instead? We propose a novel mechanism that arises from heterogeneity in the severity of price setting frictions across firms. A sufficient condition for higher markup dispersion after a monetary tightening is that firms with higher markups have lower pass-through from marginal costs to prices, i.e., relatively strong price setting frictions. A contractionary monetary shock that lowers marginal costs increases the relative markup of low pass-through firms, which increases markup dispersion. Analogously, expansionary monetary shocks that raise marginal costs will lower markup dispersion. We show that a negative correlation between firm-level markup and pass-through can arise endogenously from heterogeneity in price-setting frictions. The types of price-setting frictions we consider are a Calvo (1983) friction, Taylor (1979) staggered price setting, Rotemberg (1982) convex adjustment costs, and Barro (1972) menu costs. The intuition for this negative correlation is a precautionary price setting motive. The firm profit function in the common New Keynesian environment is asymmetric, i.e., it penalizes markups below more than markups above the static optimal one. A higher reset markup provides insurance against low profits before the next price adjustment opportunity (Calvo/Taylor), or lowers the expected costs of future price re-adjustments (Rotemberg/Barro).[2] To summarize, heterogeneous price-setting frictions imply markup dispersion and hence TFP effects of monetary policy. Importantly, precautionary price setting is absent in the deterministic steady state. By extension, our transmission mechanism is absent in model with heterogeneous price-setting frictions when solved around the deterministic steady state.

We empirically test two implications of this transmission mechanism. First, precautionary price setting implies that firms with stickier prices charge higher markups. Second, the markups of firms with stickier prices should increase by relatively more. A caveat is that we do not observe firm-specific price adjustment frequencies. Instead, we capture variation in price adjustment frequencies across firms using price adjustment frequencies in five-digit industries together with the firm-specific sales composition across industries. We find that firms with stickier prices indeed have higher markups on average and increase their markups by more after monetary policy shocks. These two results hold when controlling for two-digit sector fixed effects, firm size, leverage, and liquidity.

Finally, we study the mechanism and its implications in a quantitative New Keynesian model with heterogeneous price rigidity. To capture precautionary price setting, we use non-linear solution methods to solve the model dynamics around the stochastic steady state, to which the economy converges in the presence of uncertainty but absent of shocks. We

[2]Relatedly, in a setup with homogeneous price setting frictions Fernandez-Villaverde et al. (2015) study precautionary price setting as a channel through which higher uncertainty leads to higher markups.

find that indeed firms with stickier prices set higher markups on average, and monetary policy shocks raise markup dispersion. Quantitatively, a one standard deviation contractionary monetary policy shock lowers aggregate TFP by -0.34%. We use the model to study two implications of our mechanism. Whereas a contractionary monetary shock increases aggregate markups in many New Keynesian models, the empirically estimated responses of aggregate markups in Nekarda and Ramey (2020) have the opposite sign. In our model, the aggregate markup falls if contractionary monetary shocks lower aggregate TFP sufficiently strongly. This argument extends to sector or firm-level markups if price rigidities are heterogeneous within sectors or firms such that sector or firm-level TFP responds to monetary policy. We further analyze the effectiveness of monetary policy when the endogenous TFP effects are ignored by the monetary authority. If the monetary authority attributes all TFP fluctuations to technology shocks, interest rates are adjusted less aggressively and monetary policy shocks lead to larger GDP fluctuations.

This paper is closely related to four branches of the literature. First, a growing literature studies the positive and normative implications of heterogeneous price rigidity, see, e.g., Aoki (2001), Carvalho (2006), Nakamura and Steinsson (2010), Eusepi et al. (2011), Carvalho and Schwartzman (2015), Pasten et al. (2020), and Rubbo (2020). We show that such heterogeneity gives rise to productivity effects of monetary policy. Similarly, Baqaee and Farhi (2017) show that negative money supply shocks lower aggregate TFP if sticky-price firms have exogenously higher ex-ante markups than flexible-price firms. We provide empirical evidence which supports this transmission channel and show that the rigidity–markup correlation can arise endogenously from differences in price rigidity.

Second, this paper relates to a literature that studies the productivity effects of monetary policy, e.g., Evans and Santos (2002), Christiano et al. (2005), Comin and Gertler (2006), Moran and Queralto (2018), Garga and Singh (2021), and Jordà et al. (2020). We confirm the empirical finding that monetary policy shocks lower aggregate productivity, but provide a novel explanation based on markup dispersion. In terms of alternative explanations, Christiano et al. (2005) show that variable utilization and fixed costs explain a relatively small fraction of the aggregate productivity response. Moran and Queralto (2018) and Garga and Singh (2021) show that R&D investment falls after monetary policy shocks, which may ultimately lower productivity. However, it is unclear whether the R&D response can explain a large response of aggregate productivity at short horizons. For example, Comin and Hobijn (2010) estimate that new technologies are adopted with an average lag of five years. Conversely, price rigidities are a more natural candidate for the effects at shorter horizons.

Third, our paper relates to a literature on the relation between inflation and price dispersion. Whereas we show that contractionary monetary policy shocks raise markup dispersion, Nakamura et al. (2018) document flat price dispersion across periods of high and low inflation since the 1970s. This suggests that long-lived changes in inflation have different effects than short-lived monetary policy shocks. For example, when trend inflation increases managers may schedule more frequent meetings to discuss price changes (Romer, 1990; Levin and Yun, 2007), while monetary policy shocks are less likely to trigger such responses.

Fourth, this paper relates to a growing literature that studies allocative efficiency over the business cycle. Eisfeldt and Rampini (2006) show that capital misallocation is countercyclical. Fluctuations in allocative efficiency may be driven by various business cycle shocks, e.g., aggregate demand shocks (Basu, 1995), aggregate productivity shocks (Khan and Thomas, 2008), uncertainty shocks (Bloom, 2009), financial shocks (Khan and Thomas, 2013), or supply chain disruptions (Meier, 2020). We relate to this literature by studying the transmission of monetary policy shocks through allocative efficiency. Interestingly, the effects of short- versus long-run changes in interest rates on allocative efficiency seem to differ in sign. Whereas we show that short-run expansionary monetary policy decreases misallocation, Gopinath et al. (2017) show that, in the case of Southern Europe, persistently lower interest rates have increased misallocation. Relatedly, Oikawa and Ueda (2018) study the long-run effects of nominal growth through reallocation across heterogeneous firms.

The remainder of this paper is organized as follows. Section 2 presents the main empirical evidence. Section 3 studies monetary transmission with heterogeneous price rigidity. Section 4 presents a quantitative model. Section 5 concludes and an Appendix follows.

2 Evidence on markup dispersion and TFP

In this section, we present novel empirical evidence that monetary policy shocks increase the markup dispersion across firms. We further show that aggregate TFP falls after monetary policy shocks and that a sizable share of this response can be accounted for by the response of markup dispersion.

2.1 Data

Firm-level markups. We use quarterly balance sheet data of publicly-listed US firms from Compustat. We estimate markups through a variety of methods. Our baseline method is the ratio estimator pioneered by Hall (1986, 1988) and more recently used in De Loecker and Warzynski (2012), De Loecker et al. (2020), Flynn et al. (2019) and Traina (2020). We further consider markups using the accounting profits and user cost approaches in Gutiérrez and Philippon (2017), Basu (2019) and Baqaee and Farhi (2020).

The ratio estimator of the markup can be obtained from the cost minimization problem. With a flexible input V_{it}, the markup μ_{it} of firm i in quarter t can be computed as

$$\mu_{it} = \frac{\text{output elasticity of } V_{it}}{\text{revenue share of } V_{it}}. \tag{2.1}$$

We assume that firms in the same two-digit-industry and quarter have a common output elasticity. All our subsequent empirical analysis focuses on differences of firm-level log markups from their industry-quarter average. Under our assumption, these markup differences do not depend on the output elasticities. Hence, our empirical results are not affected by challenges to identify output elasticities from revenue data, as recently emphasized by

Bond et al. (2020).[3] By controlling for industry-quarter fixed effects in log markups, we also difference out industry and time-specific characteristics such as differences in competitiveness and production technology.

Formally, we define differences of firm-level log markups from their industry-time average as $\hat{\mu}_{it} \equiv \log \mu_{it} - \frac{1}{N_{st}} \sum_{j \in \mathcal{J}_{st}} \log \mu_{jt}$, where $\mathcal{J}_{st}$ is the set of firms j in industry s, quarter t, and N_{st} is the cardinality of $\mathcal{J}_{st}$. Following De Loecker et al. (2020) we assume firms produce output using capital and a composite input of labor and materials, with the latter the flexible factor. We estimate the revenue share as the firm-quarter-specific ratio of costs of goods sold (`cogsq` in Compustat) to sales (`saleq`).

We further consider a host of alternative markup estimation methods in Section 2.4 below. First, we construct (non-ratio estimator) markups through an accounting profit approach and a user cost approach, following Gutiérrez and Philippon (2017) and Baqaee and Farhi (2020). Second, following Traina (2020), we add selling, general and administrative expenses (SGA) to the costs of goods sold in the baseline markup measure. Third, we estimate a four-digit industry-specific translog production technology, which implies variation in output elasticities within industry and time. Fourth, we estimate four-digit industry-quarter specific output elasticities through cost shares.

We consider all industries except public administration, finance, insurance, real estate, and utilities. We drop firm-quarter observations if sales, costs of goods sold, or fixed assets are reported only once in the associated year. We further drop observations if quarterly sales growth is above 100% or below -67% or if real sales are below 1 million USD. We finally drop the bottom and top 5% of the estimated markups. Appendix A.1 provides more details and summary statistics in Table 3. Our results are robust to alternative data treatments as we discuss toward the end of this section.

[3] Our baseline approach assumes the ratio estimator to be valid in principle. This excludes the case when the input is not perfectly flexible, or when its choice affects demand, see Bond et al. (2020). For robustness, we also consider non-ratio estimators of markups, see Section 2.4.

Monetary policy shocks. Using high-frequency data of federal fund future prices, we identify monetary policy shocks through changes of the future price in a narrow time window around FOMC announcements. The identifying restrictions are that the risk premium does not change and that no other macroeconomic shock materializes within the time window. We denote the price of a future by f, and by τ the time of a monetary announcement.[4] We use a thirty-minute window around FOMC announcements, as in Gorodnichenko and Weber (2016). Let $\Delta\tau^- = 10$ minutes and $\Delta\tau^+ = 20$ minutes, then monetary policy shocks are

$$\varepsilon_\tau^{\mathrm{MP}} = \mathrm{f}_{\tau+\Delta\tau^+} - \mathrm{f}_{\tau-\Delta\tau^-}. \tag{2.2}$$

To aggregate the shocks to quarterly frequency, we follow Ottonello and Winberry (2020). We assign daily shocks fully to the current quarter if they occur on the first day of the quarter. If they occur within the quarter, we partially assign the shock to the subsequent quarter. This procedure weights shocks across quarters corresponding to the amount of time agents have to respond. Formally, we compute quarterly shocks as

$$\varepsilon_t^{\mathrm{MP}} = \sum_{\tau \in \mathcal{D}(t)} \phi(\tau)\varepsilon_\tau^{\mathrm{MP}} + \sum_{\tau \in \mathcal{D}(t-1)} (1-\phi(\tau))\varepsilon_\tau^{\mathrm{MP}}, \tag{2.3}$$

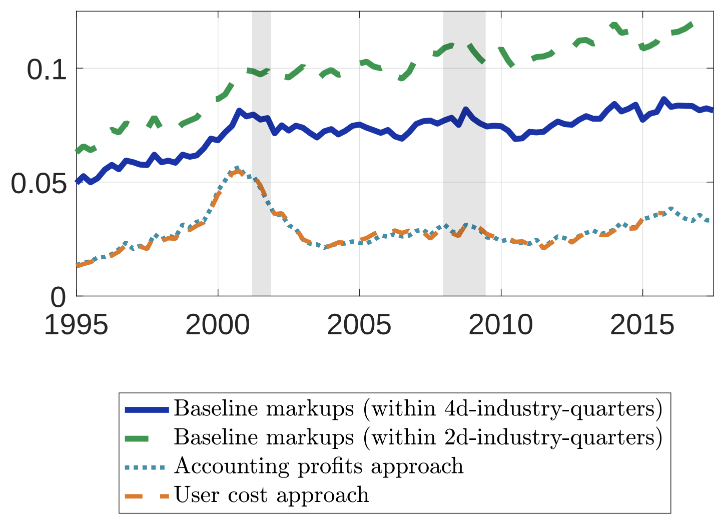
where $\mathcal{D}(t)$ is the set of days in quarter t and $\phi(\tau) =$ (remaining number of days in quarter t after announcement in τ) / (total number of days in quarter t).

As a baseline, we construct monetary policy shocks from the three-months ahead federal funds future, as in Gertler and Karadi (2015). Our baseline excludes unscheduled meetings and conference calls.[5] Following Nakamura and Steinsson (2018), our baseline further

[4] We obtain time and classification of FOMC meetings from Nakamura and Steinsson (2018) and the FRB. We obtain time stamps of the press release from Gorodnichenko and Weber (2016) and Lucca and Moench (2015).

[5] Unscheduled meetings and conference calls are often the immediate response to adverse economic developments. Price changes around such meetings may directly reflect these developments, which invalidates the identifying restriction. Non-scheduled meetings are also more likely to communicate private information about the state of the economy. Our results remain broadly robust when including these meetings.

Figure 1: Evolution of markup dispersion



Notes: The figure shows the evolution of markup dispersion for different markup measures from 1995Q1 to 2017Q3. Markup dispersion is the variance of log markups across firms, $\mathbb{V}_t(\hat{\mu}_{it})$, where $\hat{\mu}_{it}$ is the difference of a firm's log markup from the mean log markup across firms in the same industry-quarter. Baseline markups are constructed according to equation (2.1) assuming a common output elasticity for firms in the same 2d-industry-quarter. Further details on the accounting profits and user cost approaches are provided in Section 2.4.

excludes the apex of the financial crisis from 2008Q3 to 2009Q2.[6] The monetary policy shock series covers 1995Q2 through 2017Q3. We discuss alternative monetary policy shocks in Section 2.4. Table 4 in the Appendix reports summary statistics and Figure 8 (a) and (b) shows the shock series.

2.2 Markup dispersion

We estimate the response of markup dispersion to monetary policy shocks. Our baseline measure of markup dispersion is the cross-sectional variance $\mathbb{V}_t(\hat{\mu}_{it})$, where $\hat{\mu}_{it}$ denotes firm-level log markups in deviation from their respective industry-quarter mean. Recall that our baseline estimator of $\hat{\mu}_{it}$ does not depend on an estimator of the output elasticity under our assumption that firms within a two-digit industry-quarter have a common output elasticity. Figure 1 shows time series of markup dispersion for our baseline ratio estimator within four-

[6]We discard shocks during 2008Q3 to 2009Q2 and we do not regress post-2009Q2 outcomes on pre-2008Q3 shocks. Our results are robust to including this period.

digit-industry-quarters, the same estimator but within two-digit-industry-quarters, and for markups based on account profits and user costs.[7] Figure 8 (c) in the Appendix shows time series for further alternative markup dispersion, notably the ratio estimator when including SGA, the translog-based markups, and markups based on cost shares.

To estimate the effects of monetary policy shocks on markup dispersion, we use the local projection

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \varepsilon_t^{\mathrm{MP}} + \gamma_0^h \varepsilon_{t-1}^{\mathrm{MP}} + \gamma_1^h (y_{t-1} - y_{t-2}) + u_t^h, \tag{2.4}$$

for $h = 0, \ldots, 16$ quarters and where y_t is markup dispersion.[8]

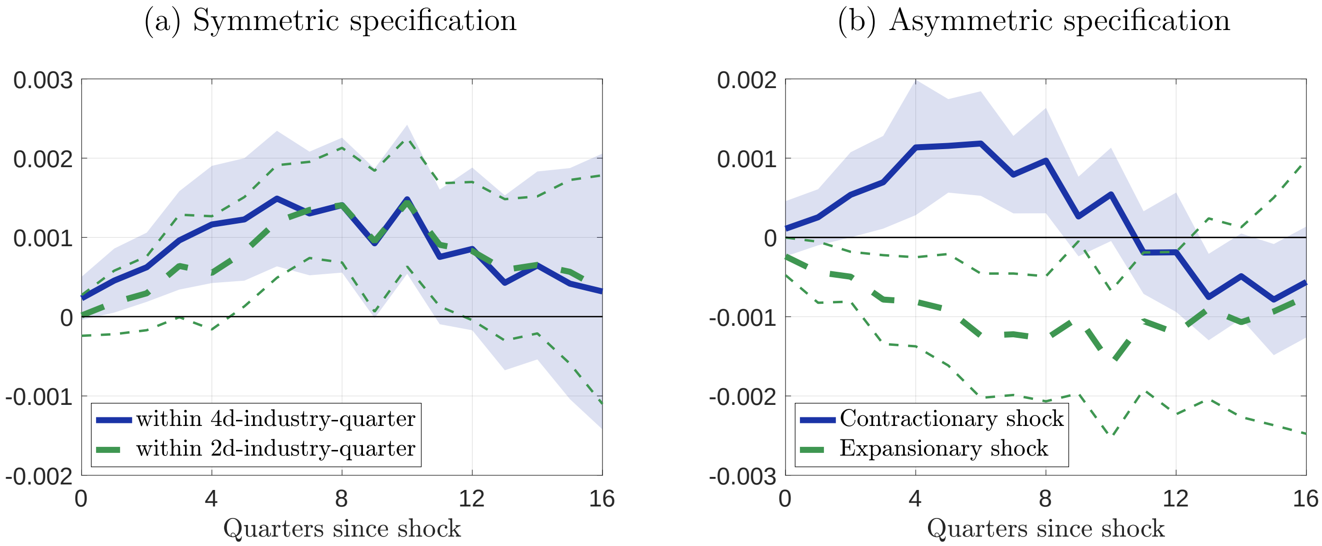
The central empirical finding of this paper is shown in panel (a) of Figure 2, which plots the response of markup dispersion, captured by the estimates of coefficients β^h. The key finding is that markup dispersion increases significantly and persistently. The response of markup dispersion peaks at about two years after the shock and reverts back to zero afterwards. Whether we compute markup dispersion within two-digit or four-digit industry-quarters changes this result by little.

The specification of (2.4) implicitly assumes that the effects of monetary policy shocks are symmetric in the sign of the shock. However, in a large class of New Keynesian models, solved via a second-order approximation, markup dispersion increases in response to both positive and negative shocks, cf. Figure 31 in the Appendix. So to investigate whether markup dispersion responds asymmetrically to shocks of different sign, we separately estimate the separate effects of contractionary and expansionary monetary policy shocks. To be precise, we replace $\varepsilon_t^{\mathrm{MP}}$ by the two sign-dependent shocks in specification (2.4). Panel (b) of Figure 2 shows the sign-dependent responses of (within 4-digit industry-quarter) markup dispersion. The evidence suggests that the responses are indeed symmetric in shock sign. While contractionary monetary policy shocks significantly increase markup dispersion, expansionary shocks

[7] Similar to De Loecker et al. (2020), our baseline markup dispersion has a positive time trend.

[8] Our results are practically unchanged when including one quarterly lead of the monetary policy shock.

Figure 2: Responses of markup dispersion to monetary policy shocks



Notes: The figure in panel (a) shows the responses of markup dispersion to a one standard deviation monetary policy shock, the coefficients β^h in (2.4). In panel (b) we allow for asymmetric effects by extending (2.4) to separately estimate the response to positive and negative shocks. Panel (b) shows the responses of (within 4-digit industry-quarter) markup dispersion to a one standard deviation contractionary and expansionary monetary policy shock, respectively. The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

significantly lower markup dispersion. In addition, the estimated magnitudes are comparable across shock sign. The results in panel (a) and (b) prove robust in a large number of dimensions, including alternative measures of markups, as we discuss in Section 2.4.

2.3 Aggregate productivity

Fluctuations in markup dispersion lead to changes in allocative efficiency of inputs across firms and thereby to fluctuations in aggregate TFP. To characterize this link, we build on Hsieh and Klenow (2009) and Baqaee and Farhi (2020). In a model with monopolistic competition and Dixit–Stiglitz aggregation, we can approximately express changes in aggregate TFP as

$$\Delta \log \mathrm{TFP}_t = -\frac{\eta}{2} \Delta \mathbb{V}_t(\log \mu_{it}) \; + \; \Big[\Delta \text{ exogenous productivity}\Big], \tag{2.5}$$

where η is the substitution elasticity between variety goods. The details of the derivation are provided in Appendix E.1.[9] An increase in the variance of log markups by 0.01 lowers aggregate TFP by $\frac{\eta}{2}$%. To provide some intuition for this link, first suppose firms are homogeneous. Aggregate output is maximal for given aggregate inputs if all firms produce the same quantity, which implies equal markups across firms. If instead firms have heterogeneous productivity and demand shifts, the efficient allocation of inputs is not homogeneous across firms, but still implies equal markups. Conversely, markup dispersion is associated with an allocation of inputs across firms that implies aggregate TFP losses.

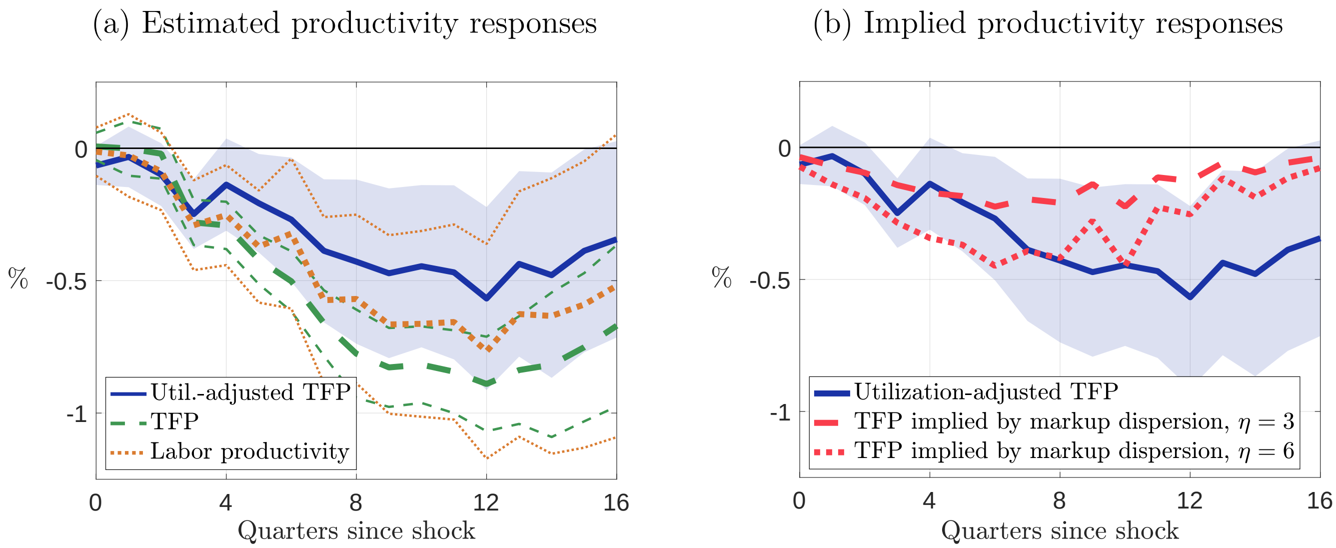
We empirically estimate the aggregate productivity response to monetary policy shocks and compare it with the implied productivity response according to equation (2.5) and the estimated response of markup dispersion in Figure 2(a). We consider aggregate TFP and utilization-adjusted aggregate TFP from Fernald (2014), as well as labor productivity, and estimate their responses to monetary policy shocks through equation (2.4).[10] Panel (a) of Figure 3 shows that the responses of all three aggregate productivity measures are significantly and persistently negative. At a two-year horizon, a one standard deviation monetary policy shock lowers aggregate TFP by 0.8%, labor productivity by 0.6% and utilization-adjusted aggregate TFP by 0.4%. For comparison, a monetary policy shock of the same magnitude raises the federal funds rate by up to 30 basis points and lowers aggregate output by about 1% at a two-year horizon, see Figure 10 in the Appendix. However, aggregate factor inputs respond little and thus aggregate TFP accounts for 50–80% of the output response at a two-year horizon.

We compute the implied TFP response by multiplying the estimated response of markup dispersion with $-\frac{\eta}{2}$%. Panel (b) of Figure 3 shows the implied response for $\eta = 6$, which

[9] In the calibrated New Keynesian model of Section 4, equation (2.5) is a close approximation to the joint behavior of aggregate TFP and markup dispersion, cf. panels (b) and (f) in Figure 6.

[10] Aggregate TFP is $\Delta \log \text{TFP} = \Delta y - w_k \Delta k - (1 - w_k)\Delta \ell$, with Δy real business output growth, w_k the capital income share, Δk real capital growth (based on separate perpetual inventory methods for 15 sub-categories of capital), $\Delta \ell$ the growth of hours worked plus growth in labor composition/quality. Utilization-adjustment follows Basu et al. (2006) and uses hours per worker to proxy factor utilization. Labor productivity is real output per hour in the nonfarm business sector. Figure 8 (d) in the Appendix shows the different aggregate productivity time series.

Figure 3: Aggregate productivity response to monetary policy shocks



Notes: Panel (a) shows the responses of aggregate productivity measures to a one standard deviation contractionary monetary policy shock. Panel (b) shows the imputed response of TFP, implied by the response of markup dispersion within four-digit industry-quarters, according to $\Delta \log \mathrm{TFP}_t = -\frac{\eta}{2}\Delta\mathbb{V}_t(\log \mu_{it})$, see equation (2.5), and using $\eta = 3$ and $\eta = 6$, respectively. Alongside, it shows the empirical response of utilization-adjusted TFP from panel (a). The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

corresponds to the estimate in Christiano et al. (2005), and $\eta = 3$, the assumption in Hsieh and Klenow (2009). The imputed TFP responses closely match the estimated TFP response within the first two years of the shock. This suggests that the response of markup dispersion is quantitatively important to understand the productivity effects of monetary policy.

An alternative explanation why aggregate productivity declines after monetary policy shocks is a reduction in R&D investment. In fact, Figure 9 in the Appendix shows that aggregate R&D expenditures fall after contractionary monetary policy shocks, which reconfirms the findings in Moran and Queralto (2018) and Garga and Singh (2021). Hence, there is scope for R&D to explain part of the aggregate TFP response. However, it is less clear how much of the short-run productivity response can be explained by R&D investment. The evidence on technology adoption suggests that R&D has rather medium-run than short-run productivity effects. For example, Comin and Hobijn (2010) estimates an average adoption lag of 5 years. A sluggish effect of R&D investment on aggregate productivity is consistent with the finding in Figure 3 (b) that markup dispersion accounts for a relatively small fraction of the TFP response 3–4 years after a monetary policy shock.

2.4 Robustness

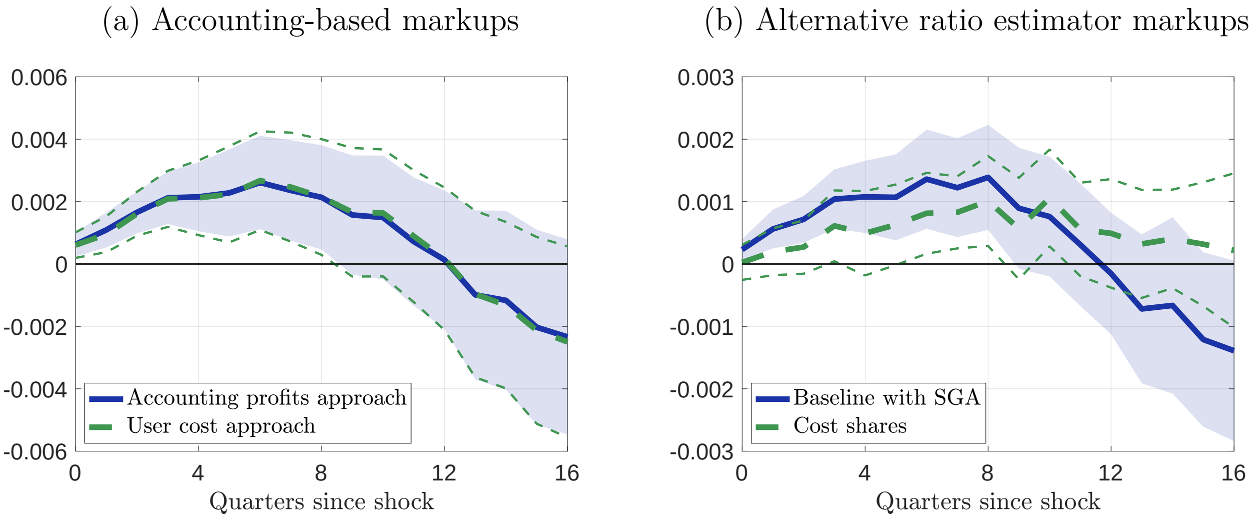
Markup estimation. We investigate the robustness of our empirical findings by considering a host of alternative markup measures. Our baseline results are robust to using these alternative markups. First, we construct (non-ratio estimator) markups through an accounting profit approach and a user cost approach, following Gutiérrez and Philippon (2017) and Baqaee and Farhi (2020). The accounting profit approach uses operating income after depreciation, which is sales (`saleq`) minus costs of goods sold (`cogsq`), selling, general and administrative expenses (`xsgaq`), and depreciation and amortization (`dpq`). We compute markups from these accounting profits via $(\text{accounting profit})_{it} = \left(1 - \mu_{it}^{-1}\right) \texttt{saleq}_{it}$. This is equivalent to constructing markups by dividing sales through the sum of costs considered in the accounting profits.[11]

For the user cost approach, we additionally subtract the firm's capital costs (excluding depreciation) from accounting profits as in Baqaee and Farhi (2020). We construct firm-level capital stocks k_{it} via a perpetual inventory method to property, plant and equipment, see Appendix A.1. The user cost of capital is $r_t = r_t^f + RP_{jt} - (1 - \delta_{jt})\Pi_{jt+1}^K$, where r^f is the risk-free real rate, RP_j the industry-specific risk premium, δ_j the industry-specific BEA depreciation rate, and Π_j^K is the industry-specific growth in the relative price of capital, based on data in Gutiérrez and Philippon (2017).[12] In general, the size of capital costs relative to total costs is modest with an average of 3.2%. This may explain the small differences

[11] We can therefore view this approach as resulting in 'average markups', instead of 'marginal markups'.

[12] The Gutiérrez and Philippon (2017) user cost is at annual frequency; we divide through by four to arrive at a quarterly rate. The data from Gutiérrez and Philippon (2017) ends in 2015, so that the time sample of user cost approach markups is shorter.

Figure 4: Response of dispersions in alternative markup measures



Notes: This figure shows the responses of markup dispersion to a one standard deviation monetary policy shock, coefficients β^h in (2.4). In panel (a), we consider the *accounting profits* and *user cost approach*, and we use markup dispersion within four-digit-industry-quarters. In panel (b), *baseline with SGA* considers markups that add SGA to the costs of goods sold, and we use markup dispersion within four-digit-industry-quarters. *Cost shares* considers a ratio estimator using four-digit-industry-quarter-specific cost shares as output elasticities, and we use markup dispersion within two-digit-industry-quarters. The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

between the accounting profit and user cost approaches.

Second, we construct a ratio estimator which adds selling, general and administrative expenses (SGA) to the costs of goods sold, following Traina (2020). Third, we estimate a four-digit-industry-specific translog production technology, which implies firm-quarter-specific output elasticities as in De Loecker et al. (2020). We then compute markups by combining output elasticities with revenue shares according to equation (2.1). Fourth, we compute four-digit-industry-quarter-specific cost shares to estimate output elasticities. Specifically, we follow De Loecker et al. (2020) and compute the industry-quarter median of costs of goods sold plus 3% of the capital stock (which approximates the user cost of capital by an annual rate of 12% that includes risk premium and depreciation) divided by sales. This is a valid estimator of the output elasticity if all factors are flexible.

Our results are robust to computing markups based on these alternative measures.[13] Figure 4 (a) shows the response of markup dispersion within four-digit-industry-quarters to monetary policy shocks when using the accounting profits and user cost approach. Figure 4 (b) shows the markup dispersion response when including SGA as well as for the cost share approach. In the Appendix we show additional results. Figure 11 shows the responses of all alternative markup dispersion measures within two-digit- and four-digit-industry quarters. Figure 12 shows the responses of all markup dispersion measures conditional on the sign of the monetary policy shock.

Firm-level data treatment. We show the robustness of our results under alternative data treatments. First, we keep firms with real sales growth above 100% or below -67%. Second, we keep small firms with real quarterly sales below 1 million 2012 USD. Third, instead of dropping the top/bottom 5% of the markup distribution per quarter, we drop the top/bottom 1%. Fourth, we condition on firms with at least 16 quarters of consecutive observations. Figure 13 shows that markup dispersion robustly increases after contractionary monetary policy shocks. Figure 14 shows the responses of markup dispersion remain symmetric in the sign of the monetary policy shock. A well-known recent trend is the delisting of public firms. We address the concern that this may affect our results in two ways. First, when only considering firms that are in the sample for at least 16 consecutive quarters, we find our results to be robust, as discussed above. Second, we estimate whether the number of firms in the sample responds to monetary policy shocks. Figure 15 shows that the response is insignificant and small.

Monetary policy shocks. We show that our results are robust to a variety of alternative monetary policy shock series. Similar to Nakamura and Steinsson (2018), we consider the first principal component of the current/three-month federal funds futures and the 2/3/4-quarters

[13] For the comparability of our results across markup measures, we include only firms in the robustness checks for which the baseline markup is non-missing after the data treatment steps. Additionally we trim the alternative markups at the 1% and 99% quantiles of the quarterly markup distributions.

ahead Eurodollar futures. We further address the concern that high-frequency future price changes may not only capture monetary policy shocks, but also release private central bank information about the state of the economy. To control for such information effects we employ two alternative strategies. First, following Miranda-Agrippino and Ricco (2018), we regress daily monetary policy shocks on internal Greenbook forecasts and revisions for output growth, inflation, and unemployment. Second, following Jarocinski and Karadi (2020), we discard daily monetary policy shocks if the associated high-frequency change in the S&P500 moves in the same direction. While our baseline series exclude unscheduled meetings and conference calls, which plausibly diminishes the role of information effects, we also reassess our results when including these events. A different concern may be that unconventional monetary policy drives our result. We address this by setting daily monetary policy shocks at Quantitative Easing (QE) announcements to zero. Figure 16 in the Appendix shows the response of markup dispersion for all monetary policy shock series. Figure 17 shows the sign-dependent responses of markup dispersion to monetary policy shock. Figure 18 in the Appendix shows the responses of aggregate productivity for all monetary policy shock series.

Great Recession. We exclude the apex of the Great Recession from 2008Q3 to 2009Q2 in our baseline estimations. However, our results do not depend on this choice. Moreover, the results are robust to using the Pre-Great Recession period until 2008Q2. Panels (d) and (e) of Figures 13 and 14 in the Appendix show that our results are robust across samples.

LP-IV. We revisit our main results with the LP-IV method (Stock and Watson, 2018). More precisely, we replace the monetary policy shocks $\varepsilon_t^{\mathrm{MP}}$ in equations (2.4) and (3.5) by the quarterly change in the one-year treasury rate and use $\varepsilon_t^{\mathrm{MP}}$ as an instrument. Figure 19 (a) and (b) in the Appendix shows that our results are robust to the LP-IV method.

Proxy SVAR. Additionally, we revisit our main results through a proxy SVAR model following Gertler and Karadi (2015).[14] Figure 20 in the Appendix shows the responses to monetary policy shocks in a VAR, including the one-year rate, (log) industrial production, (log) CPI, the excess bond premium of Gilchrist and Zakrajsek (2012), (log) TFP and the baseline measure of markup dispersion (within four-digit-industry-quarters). At a horizon between 1 and 5 quarters after the shock, the responses of TFP and markup dispersion are similar to our local projection results.

TFP measurement. Hall (1986) shows that the Solow residual is misspecified in the presence of market power. Hall shows that the correct Solow weights are not the income share for capital w_{kt} and labor $1 - w_{kt}$, but instead $\mu_t w_{kt}$ and $1 - \mu_t w_{kt}$, where μ_t is the aggregate markup. We examine the response of markup-corrected (utilization-adjusted) aggregate TFP to monetary policy shocks. We use the average markup series from De Loecker et al. (2020) to compute Hall's weights. Figure 21 (a) in the Appendix shows that the TFP response is barely different from Figure 3 (a). In response to expansionary monetary shocks, Figure 22 shows a significant increase of TFP, while the response to contractionary shocks is insignificant. We further investigate whether measurement error in quarterly TFP data is responsible for the effects of monetary policy. This problem was flagged for defense spending shocks by Zeev and Pappa (2015). We follow them in re-computing TFP using measurement error corrected quarterly GDP from Aruoba et al. (2016). Figure 21 (b) shows that measurement error corrected TFP also falls after monetary policy shocks. In addition, we show that Fernald's (2014) investment-specific and consumption-specific aggregate TFP series significantly falls after contractionary monetary policy shocks, see Figure 21 (c) and (d). Notably, the response of investment-specific TFP is significantly stronger than that of consumption-specific TFP.

[14] In contrast to the proxy SVAR model, both our baseline LP approach in (2.4) and the LP-IV approach are robust to non-invertibility, see Plagborg-Møller and Wolf (2021).

3 Heterogeneous price setting frictions

In this section, we characterize a novel mechanism through which firm heterogeneity in price setting frictions may explain why markup dispersion increases in response to contractionary monetary policy shocks, and decreases after expansionary ones. In addition, we provide empirical evidence in support of this mechanism, and discuss alternative mechanisms.

3.1 Sufficient condition

We first propose a sufficient condition for monetary policy shocks, which lower real marginal costs, to increase the dispersion of markups across firms. Let i index a firm and t time. A firm's markup is $\mu_{it} \equiv P_{it}/(P_t X_t)$, where P_{it} is the firm's price, P_t the aggregate price, and X_t real marginal cost. Let pass-through from marginal cost to price be defined as

$$\rho_{it} \equiv \frac{\partial \log P_{it}}{\partial \log X_t}. \tag{3.1}$$

This is the percentage price change in response to a percentage change in real marginal cost (without conditioning on price adjustment). The correlation between firm-level markup and firm-level pass-through is a key moment for the response of markup dispersion to shocks.

Proposition 1. *If* $\mathrm{Corr}_t(\rho_{it}, \log \mu_{it}) < 0$*, markup dispersion decreases in real marginal costs*

$$\frac{\partial \mathbb{V}_t(\log \mu_{it})}{\partial \log X_t} < 0,$$

and markup dispersion increases if $\mathrm{Corr}_t(\rho_{it}, \log \mu_{it}) > 0$*.*

Proof: See Appendix E.2.

Contractionary monetary policy shocks that lower real marginal costs increase the dispersion of markups if firms with higher markups have lower pass-through. While we focus on monetary policy shocks in this paper, in principle any shock that lowers real marginal costs will raise markup dispersion as long as markups and pass-through are negatively correlated across firms.

3.2 Precautionary price setting

We next show that firm-level heterogeneity in the severity of various price-setting frictions may explain a negative correlation between firm-level pass-through and markup. It follows from Proposition 1 that heterogeneous price-setting frictions can explain why contractionary monetary policy shocks raise markup dispersion.

Consider a risk-neutral investor that sets prices in a monopolistically competitive environment with an isoelastic demand curve and subject to adjustment costs:

$$\max_{\{P_{it+j}\}_{j=0}^{\infty}} \mathbb{E}_t \sum_{j=0}^{\infty} \beta^t \left[\left(\frac{P_{it+j}}{P_{t+j}} - X_{t+j} \right) \left(\frac{P_{it+j}}{P_{t+j}} \right)^{-\eta} Y_{t+j} - \text{adjustment cost}_{it+j} \right] \tag{3.2}$$

Adjustment costs differ across firms and may be deterministic or stochastic. This formulation nests the Calvo (1983) random adjustment, Taylor (1979) staggered price setting, Rotemberg (1982) convex adjustment costs, and Barro (1972) menu costs.

Importantly, the period profit (net of adjustment costs) is asymmetric in the price P_{it} and hence in the markup μ_{it}. Profits fall more rapidly for low markups than for high markups. This gives rise to a precautionary price setting motive: when price adjustment is frictional, firms have an incentive to set a markup above the frictionless optimal markup. Setting a higher markup today provides some insurance against low profits before the next price adjustment opportunity (Calvo/Taylor), or lowers the expected costs of future price re-adjustments (Rotemberg/Barro).

To characterize precautionary price setting, we study the problem in partial equilibrium. Analytically solving the non-linear price-setting problem with adjustment costs and aggregate uncertainty in general equilibrium is not feasible. We assume that aggregate price, real marginal costs, and aggregate demand, denoted by (P_t, X_t, Y_t), follow an i.i.d. joint log-normal process around the unconditional means $\bar{P}$, $\bar{X}$, and $\bar{Y}$. The (co-)variances of innovations are σ_k^2 and σ_{kl} for $k, l \in \{p, x, y\}$.

Calvo friction. Consider a Calvo (1983) friction, parametrized by a *firm-specific* price adjustment probability $1 - \theta_i \in (0, 1)$. The profit-maximizing reset price is

$$P_{it}^* = \frac{\eta}{\eta - 1} P_t X_t \frac{\mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j \theta_i^j \frac{X_{t+j}}{X_t} \left(\frac{P_{t+j}}{P_t} \right)^{\eta} \frac{Y_{t+j}}{Y_t} \right]}{\mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j \theta_i^j \left(\frac{P_{t+j}}{P_t} \right)^{\eta - 1} \frac{Y_{t+j}}{Y_t} \right]}, \tag{3.3}$$

and we denote the associated markup by μ_{it}^*. To isolate the role of uncertainty in price setting, we focus on the dynamics around the stochastic steady state, which is described by the unconditional means $(\bar{P}, \bar{X}, \bar{Y})$. The following proposition characterizes the precautionary upward price-setting bias – relative to the frictionless environment – as a function of θ_i, and establishes a condition under which firms with lower pass-through set higher markups.

Proposition 2. *If $P_t = \bar{P}$, $X_t = \bar{X}$, $Y_t = \bar{Y}$, and $(\eta - 1)\sigma_p^2 + \sigma_{py} + \eta\sigma_{px} + \sigma_{xy} > 0$, the firm sets a markup above the frictionless optimal one and the markup further increases the less likely price re-adjustment is,*

$$\mu_{it}^* > \frac{\eta}{\eta - 1} \quad \textit{and} \quad \frac{\partial \mu_{it}^*}{\partial \theta_i} > 0.$$

Pass-through ρ_{it} is zero with probability θ_i and positive otherwise. Expected pass-through, denoted by $\bar{\rho}_{it}$, of either a transitory or permanent change in X_t, falls monotonically in θ_i,

$$\frac{\partial \bar{\rho}_{it}}{\partial \theta_i} < 0.$$

If the above conditions are satisfied, then $\text{Corr}_t(\rho_{it}, \log \mu_{it}^*) < 0$.

Proof: See Appendix E.3.

A permanent decrease in real marginal costs leads to an permanent increase in the optimal reset price by the same factor. The pass-through is hence one for adjusting firms and zero for non-adjusting firms. A transitory decrease in real marginal costs increases the optimal reset price by less than the marginal cost change if the future reset probability is below one. The pass-through of adjusting firms is hence less than one and falling in price stickiness.

Staggered price setting. Consider Taylor (1979) staggered price setting and assume that firms adjust asynchronously and at different deterministic frequencies. Staggered price setting is a deterministic variant of the Calvo setup and yields very similar results.

Rotemberg friction. Consider the price-setting problem subject to Rotemberg (1982) quadratic price adjustment costs, parametrized by a *firm-specific* cost shifter $\phi_i \geq 0$, i.e., adjustment $\text{cost}_{it} = \frac{\phi_i}{2}\left(\frac{P_{it}}{P_{it-1}} - 1\right)^2$. The first-order condition for P_{it} is

$$\left[(1-\eta)\frac{P_{it}}{P_t} + \eta X_t\right]\left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t = \phi_i\left(\frac{P_{it}}{P_{it-1}} - 1\right)\frac{P_{it}}{P_{it-1}} - \phi_i\beta\mathbb{E}_t\left[\left(\frac{P_{it+1}}{P_{it}} - 1\right)\frac{P_{it+1}}{P_{it}}\right]. \tag{3.4}$$

The following proposition summarizes our analytical results.

Proposition 3. *If* $P_{t-1} = P_t = \bar{P}$, $X_t = \bar{X}$, $Y_t = \bar{Y}$, *and* $\frac{\sigma_{px}}{\sigma_p\sigma_x} > -1$, *then up to a first-order approximation of* (3.4) *around* $\phi_i = 0$, *it holds that*

$$\mu_{it} \geq \frac{\eta}{\eta - 1} \quad \text{and} \quad \frac{\partial\mu_{it}}{\partial\phi_i} \geq 0, \quad \text{with strict inequality if} \;\; \phi_i > 0.$$

If in addition $\eta \in (1, \tilde{\eta})$, *where* $\tilde{\eta} = 1 + (\exp\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\} - \exp\{\sigma_{px}\})^{-1}$, *the pass-through, of either a transitory or permanent change in* X_t, *falls monotonically in* ϕ_i,

$$\frac{\partial\rho_{it}}{\partial\phi_i} < 0.$$

If the above conditions are satisfied, then $\text{Corr}_t(\rho_{it}, \log\mu_{it}) < 0$.

Proof: See Appendix E.4.

Menu costs. Consider the price-setting problem subject to firm-specific menu costs. Due to the asymmetry of the profit function, price adjustment is more rapidly triggered for markups below the frictionless optimal markup than above. Thus, a higher reset markup may be optimal to economize on adjustment costs. Analytical results, however, are not available for the fully non-linear menu cost problem. Instead, we investigate this problem quantitatively. We find that markups increase in menu costs, consistent with precautionary price setting. Consequently, the correlation between pass-through and markup is negative. More details on calibration, solution, and results are provided in Appendix F.

3.3 Empirical evidence for the mechanism

We corroborate the mechanism by considering two testable implications. First, firms with higher markups adjust prices less frequently. Second, monetary policy shocks increase the relative markup of firms that adjust prices less frequently. We show that both implications are supported empirically.

For the subsequent empirical analysis, we use data on price adjustment frequencies together with the data described in Section 2. We observe average price adjustment frequencies over 2005–2011 for five-digit industries, computed in Pasten et al. (2020) from PPI micro data. We further use the Compustat segment files, which provide sales and industry codes of business segments within firms. The firm-specific sales composition across industries allows us to compute firm-specific price adjustment frequencies as sales-weighted average of industry-specific price adjustment frequencies. We expect this procedure to underestimate the true extent of heterogeneity across firms, which we expect will bias our subsequent regression coefficients toward zero because of attenuation bias.[15] For some firms,

[15] A sufficient condition for downward bias is that the error in the measured firm-specific price adjustment frequencies is independent of the true unobserved firm-specific price adjustment frequencies.

Compustat segment files are not available and for others, they report only one segment per firm. We can construct firm-specific price adjustment frequencies for 42% of firms. For the remaining firms, we use the price adjustment frequency of the five-digit industry they operate in.[16] More details are provided in Appendix A.4. To measure price rigidity, we consider both the price adjustment frequency and the implied price duration, defined as $-1/\log(1 - \text{price adjustment frequency})$.

Testable implication 1: Firms with stickier prices charge higher markups. We provide empirical evidence that firms with stickier prices tend to charge higher markups. To compare markups with average price adjustment frequencies and implied price durations for 2005–2011, we compute average firm-level markups over the same time period. Columns (1) and (3) of Table 1 show that firms, which have more rigid prices than other firms in the same two-digit industry, charge markups significantly above the industry average. The correlation is statistically significant for both implied price duration and price adjustment frequency as measures of price rigidity. While this correlation is consistent with precautionary price setting, it may reflect omitted factors. In columns (2) and (4) we control for firm-specific size, leverage, and liquidity, all averages over 2005–2011. The conditional correlations remain of the same sign and statistically significant at the 1% level. In Table 1 we have excluded firms for which price setting frictions are practically irrelevant, in particular, firms with a price adjustment frequency above 99% per quarter, which are about 3% of all firms. When including these, the relation between stickiness and markup remains positive, albeit somewhat less significant, see Table 5 in the Appendix. Note that we have not considered four-digit industry FE, because for many firms our measure of rigidity is based on the five-digit industry average, which limits the variation in rigidity measures within four-digit industries.

[16] Our results are robust when only using sectoral price adjustment frequencies.

Table 1: Markups and price stickiness

(a) Regressions of markups on implied price duration

	log(Markup)			
	Baseline		Accounting profits	User cost approach
Implied price duration	0.0537	0.0472	0.00711	0.00897
	(0.0180)	(0.0156)	(0.00300)	(0.00346)
Additional controls	No	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	3857	3857	3807	3799
Adjusted R^2	0.145	0.228	0.237	0.185

(b) Regressions of markups on price adjustment frequency

	log(Markup)			
	Baseline		Accounting profits	User cost approach
Price adjustment frequency	-0.391	-0.336	-0.0511	-0.0605
	(0.0999)	(0.0860)	(0.0199)	(0.0215)
Additional controls	No	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	3857	3857	3807	3799
Adjusted R^2	0.151	0.231	0.237	0.185

Notes: Regressions of firm-level markup on firm-level price adjustment frequency and implied price duration, respectively. The regressions with additional controls include firm-level size, liquidity, and leverage as regressors. Standard errors are clustered at the two-digit industry level and shown in parentheses.

Testable implication 2: Monetary policy shocks increase the relative markups of firms with stickier prices. We investigate whether contractionary monetary policy shocks increase the relative markup of firms with stickier prices. This is not necessarily the case if the average stickiness differs from the stickiness after monetary policy shocks, or if the marginal costs of firms with stickier prices respond differently from other firms.

We estimate panel local projections of firm-level log markups on the interaction between monetary policy shocks and firm-level price rigidity. We measure firm-level price rigidity by the price adjustment frequency or the implied price duration. Let Z_{it} denote a vector of firm-specific characteristics. We consider two specifications for Z_{it}: (i) including one of the two rigidity measures, and (ii) additionally including lags of firm size (log of total assets), leverage (total debt per total assets), and the ratio of liquid assets to total assets.[17] Our selection of controls is motivated by recent work in Ottonello and Winberry (2020) and Jeenas (2019), who study the transmission of monetary policy shocks through financial constraints. We use the panel local projection

$$y_{it+h} - y_{it-1} = \alpha_i^h + \alpha_{st}^h + B^h Z_{it}\varepsilon_t^{\mathrm{MP}} + \Gamma^h Z_{it} + \gamma^h (y_{it-1} - y_{it-2}) + u_{it}^h \tag{3.5}$$

for $h = 0, \ldots, 16$ quarters, in which we include two-digit-industry-time and firm fixed effects. To focus on the within-industry variation in the interaction between monetary policy shock and price rigidity, we subtract the corresponding two-digit industry mean from the measure of price rigidity. The main coefficients of interest are the coefficients in $\{B^h\}$ associated with price rigidity. These capture the relative markup increase for firms with stickier prices. Figure 5 shows the results. The markups of firms with stickier prices increase by significantly more after monetary policy shocks.[18] Firms with a price adjustment frequency one standard deviation above the associated two-digit-industry mean increase their markup by up to 0.2% more. Importantly, the estimates are almost identical when adding controls, see panel (b) of Figure 5. We additionally investigate the relative size response of firms with stickier prices. In particular, we consider firm-level sales market shares at the two-digit-industry-quarter level. As a relative increase in markup implies relatively lower demand, we expect that firms with stickier prices become relatively smaller after contractionary monetary policy shocks. Indeed, we find that firms with stickier prices lose market share after contractionary monetary policy shocks, as can be seen in panel (c) of Figure 5.[19]

[17] We subtract the firm-level mean from size, leverage and liquidity to focus on within-firm variation.

[18] Using Driscoll–Krayy standard errors yields almost the same confidence bands as in Figure 5.

[19] The response of dispersion in firm-level market shares increases after monetary policy shocks, similar to markup dispersion, see Appendix Figure 11 (f).

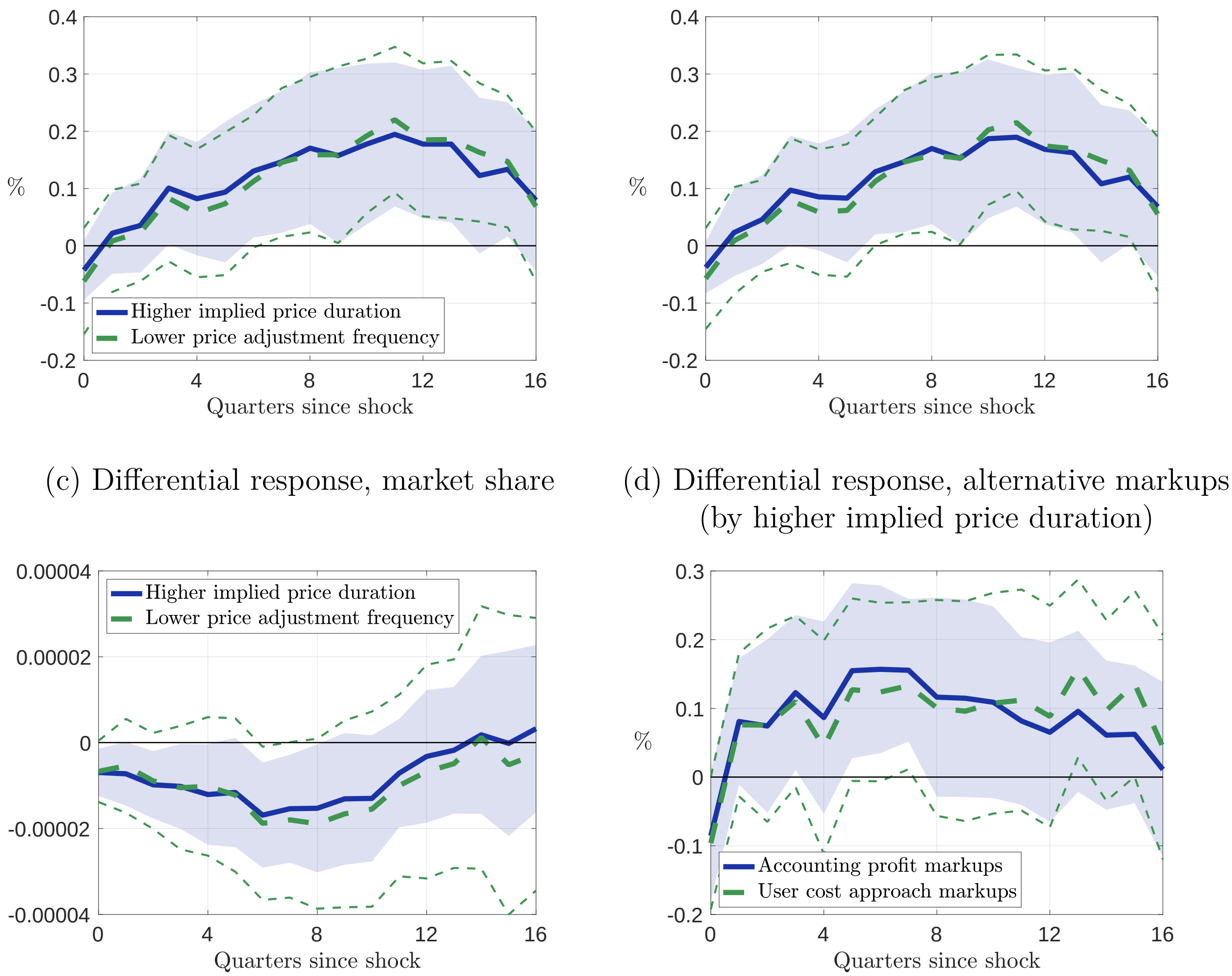
Figure 5: Relative markup and market share responses of firms with stickier prices

(a) Differential response, baseline markups

(b) ... controlling for size, leverage, liquidity

(c) Differential response, market share

(d) Differential response, alternative markups (by higher implied price duration)



Notes: The figures show the relative markup response of firms with a price adjustment frequency one standard deviation below (or with an implied price duration one standard deviation above) the two-digit-industry mean to a one standard deviation monetary policy shock. That is, we plot the appropriately scaled coefficients in B^h that are associated to price rigidity in the panel local projections (3.5). In panel (a), Z_{it} contains only price stickiness. In panel (b), Z_{it} also contains lagged log assets, leverage, and liquidity. The shaded and bordered areas indicate 90% error bands two-way clustered by firm and quarter.

Robustness. Our findings are robust along various dimensions, similar to the robustness checks in Section 2.4. We first consider alternative markup estimates based on the accounting profits and user costs approach. Table 6 shows the correlation between average markup and price rigidity. Figure 5 (d) shows the relative markup response of firms with stickier prices to monetary policy shocks. In the Appendix, we also estimate the relative markup responses when markups are based on cost shares, translog technology, or the baseline including SGA, see Figure 23. Second, we consider the role of alternative data treatments. Table 7 in the Appendix shows that the correlation between markups and price rigidity is robust across data treatments. Figure 24 shows that the relative markup response to monetary policy shocks is sensitive to removing outliers in the firm-level markups, but robust to other data treatments. Third, we consider alternative monetary policy shock series, see Figure 25 in the Appendix. Fourth, we consider an LP-IV setup as described in Section 2.4, see Figure 19 (c) in the Appendix. Finally, we include the apex of the Great Recession, see Figure 13 and Figure 24 (d) and (e).

3.4 Alternative mechanisms

A key condition to explain the response of markup dispersion to monetary policy shocks is a negative correlation between firm-level markups and pass-through (Proposition 1). We show that firm heterogeneity in price setting frictions can explain this correlation and we provide empirical evidence in support of this explanation. However, this does not preclude other mechanisms. In the following, we discuss three alternative mechanisms.

First, a non-isoelastic demand system as proposed by Kimball (1995) can explain a negative correlation between markup and pass-through and thus the response of markup dispersion.[20] Indeed, recent work by Baqaee et al. (2021) shows that under Kimball preferences (also applied, e.g., by Edmond et al., 2021), firms with a higher market share may have higher markups and lower pass-through. Even in the absence of heterogeneous price setting frictions, this environment can qualitatively explain our empirically estimated response of markup dispersion to monetary policy shocks. Second, a negative correlation between markup and pass-through can arise in an environment with oligopolistic competition and different elasticities of substitution across and within sectors, as proposed by Atkeson and Burstein (2008).

[20]The evidence for Kimball-type demand curves is mixed, however, see Klenow and Willis (2016).

Third, heterogeneity in pass-through across firms can arise from financial frictions. For example, markup dispersion may increase if contractionary monetary policy shocks increase by more the financing costs of firms with lower markups.[21]

4 Quantitative example

In this section, we investigate the transmission mechanism and its implications in a New Keynesian model with heterogeneous price rigidity.

4.1 Model setup

Our model setup builds on Carvalho (2006) and Gorodnichenko and Weber (2016). We discuss the model only briefly and relegate a formal description to Appendix G. An infinitely-lived representative household has additively separable preference in consumption and leisure, and discounts future utility by β. The intertemporal elasticity of substitution for consumption is γ and the Frisch elasticity of labor supply is φ. The consumption good is a Dixit–Stiglitz aggregate of differentiated goods with constant elasticity of substitution η.

The economy is populated by five types of monopolistically competitive intermediate goods firms. There is an equal mass of firms of each type. All firms produce differentiated output goods with the same linear technology in labor. The only ex-ante difference across firms is the exogenous price adjustment probability $1-\theta_k$, which is specific to type k. Firms set prices to maximize the value of the firm to the households. In contrast to Carvalho (2006) and the subsequent literature, which consider models with cross-sector differences in price rigidity, our model is a one-sector economy, in which price rigidity differs between firms. This speaks more directly to our empirical within-industry evidence. The monetary authority aims to stabilize inflation and the output gap. The output gap is defined as deviations of aggregate output from its natural level, defined as the flexible-price equilibrium

[21] On the interaction between financial frictions and price setting, see, for example, Gilchrist et al. (2017) and Kim (2020).

output. Monetary policy follows a Taylor rule with interest rate smoothing and is subject to monetary policy shocks, $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$.

In Section 4.4, we consider some variants of the baseline model in order to investigate the robustness of our quantitative findings. In particular, we consider real rigidities via firm-specific labor, Rotemberg price adjustment costs, and positive trend inflation.

4.2 Calibration and solution

A model period is a quarter. We set the elasticity of substitution between differentiated goods at $\eta = 6$, as estimated in Christiano et al. (2005). This is conservative when compared to $\eta = 21$ in Fernandez-Villaverde et al. (2015), who study precautionary price setting as transmission of uncertainty shocks. A higher η means more curvature in the profit function, hence more precautionary price setting, and larger TFP losses from markup dispersion. We use standard values for the discount factor β and the intertemporal elasticity of substitution γ. We set the former to match an annual real interest rate of 3%, and the latter to a value of 2. We use the estimates in Christiano et al. (2016) for the Taylor rule and set $\rho_r = 0.85$, $\phi_\pi = 1.5$, and $\phi_y = 0.05$.

The parameters which play a key role in this model are the price adjustment frequencies. For the five types of firms, we calibrate θ_k for $k = 1, \ldots, 5$ to match the empirical distribution of within-industry price adjustment frequencies based on Gorodnichenko and Weber (2016). They document mean and standard deviation of monthly price adjustment frequencies for five sectors. We first compute the value-added-weighted average of the means and variances. The monthly mean price adjustment frequency is 0.1315 and the standard deviation is 0.1131. Second, we fit a log-normal distribution to these moments. Third, we compute the mean frequencies within the five quintile groups of the fitted distribution. Finally, we transform the monthly frequencies into quarterly ones to obtain $\{\theta_k\}$.

We calibrate the Frisch elasticity of labor supply internally. The hours response to monetary policy shocks is small on impact, but larger at longer horizons, see Figure 10 in

Table 2: Calibration

Parameter		Value	Source/Target
Discount factor	β	$1.03^{-1/4}$	Risk-free rate of 3%
Elasticity of intertemporal substitution	γ	2	Standard
Elasticity of substitution between goods	η	6	Christiano et al. (2005)
Interest rate smoothing	ρ_r	0.85	Christiano et al. (2016)
Policy reaction to inflation	ϕ_π	1.5	Christiano et al. (2016)
Policy reaction to output	ϕ_y	0.05	Christiano et al. (2016)
Standard deviation of MP shock	σ_ν	0.00415	30bp effect on nominal rate
Frisch elasticity of labor supply	φ	0.1175	Relative hours response of 11.7%
Distribution of price adjustment frequencies			
Firm type k		Share	Price adjustment frequency $1-\theta_k$
1		0.2	0.0231
2		0.2	0.0678
3		0.2	0.1396
4		0.2	0.2829
5		0.2	0.8470

Notes: The distribution of price adjustment frequencies is chosen to match the within-sector distribution reported in Gorodnichenko and Weber (2016).

the Appendix. The utilization-adjusted TFP response is immediately negative but has a flatter profile at longer horizons. On average, the two responses have similar magnitude. The average difference of the hours response relative to the response of utilization-adjusted TFP, computed as the mean of $\frac{1\ -\ \text{response of hours in \%}}{1\ -\ \text{response of util-adj. TFP in \%}} - 1$ up to 16 quarters after the shock, is 11.7%. In the model, we compute the relative hours response in the same way and target 11.7% to calibrate the Frisch elasticity. Importantly, we do not directly target the absolute magnitude of the TFP response, but only a relative quantity. The calibrated Frisch elasticity is $\varphi = 0.1175$, which is low compared to the macroeconomics literature, but which is within the range of empirical estimates surveyed by Ashenfelter et al. (2010). The remaining parameter is the standard deviation of monetary policy shocks σ_ν, which we also calibrate internally. The target is the peak nominal interest rate response to a one standard deviation monetary policy shock of 30bp, see Figure 10. This yields $\sigma_\nu = 0.00415$.

For markup dispersion to arise from precautionary price setting, it is important to use an adequate model solution technique. We rely on local solution techniques, but, importantly, solve the model around its stochastic steady state. Whereas markup are the same across firms in the deterministic steady state, differences across firms may exist in the stochastic steady state. We apply the method developed by Meyer-Gohde (2014), which uses a third-order perturbation around the deterministic steady state to compute the stochastic steady state as well as a first-order approximation of the model dynamics around it.[22] In the stochastic steady state, precautionary price setting has large effects. Firms with the most rigid prices have 11.5% higher markups than firms with the most flexible prices.[23] As follows from Proposition 1, the negative correlation between markups and pass-through implies that contractionary monetary policy shocks increase markup dispersion and lower aggregate TFP.

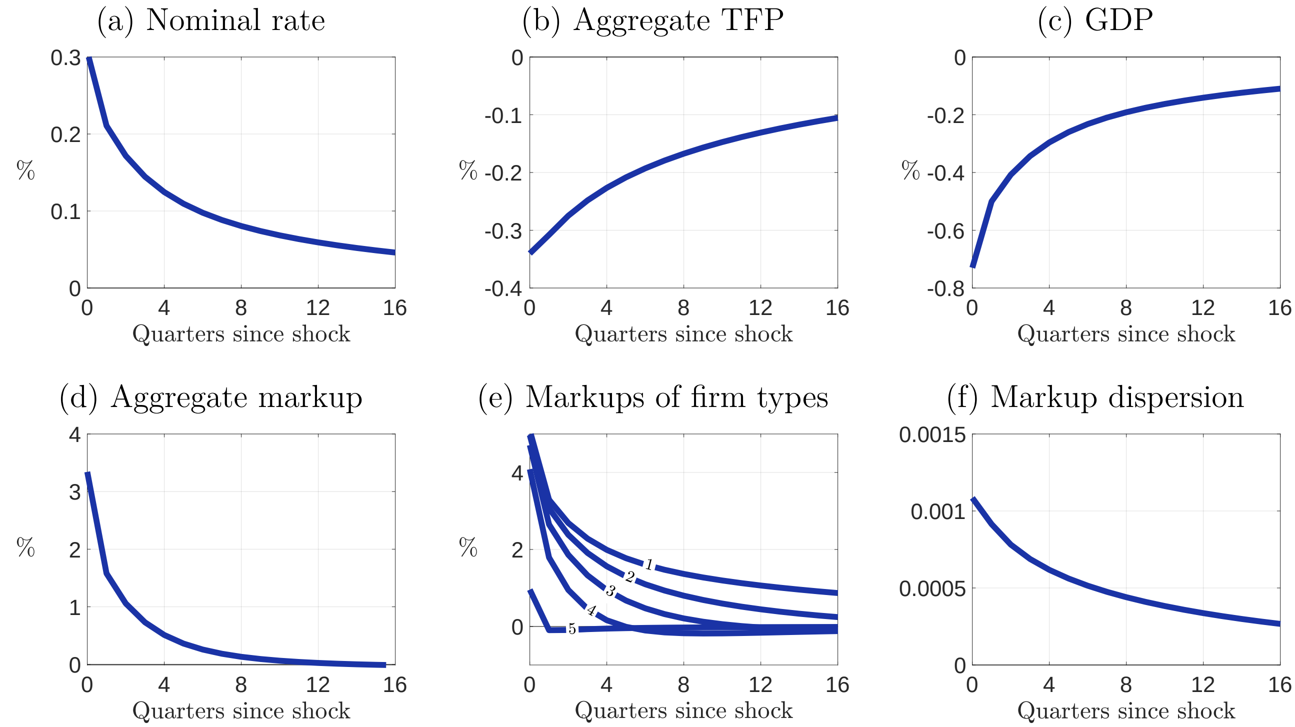
4.3 Results

Figure 6 shows the responses to a one-standard deviation monetary policy shock. The shock depresses aggregate demand and lowers real marginal costs. In response, firms want to lower their prices. For firms with stickier prices, however, pass-through is lower and on average their markups increase by more. Since firms with stickier prices have higher initial markups, markup dispersion increases. This worsens the allocation of factors across firms and thereby depresses aggregate TFP. The mechanism is quantitatively important. The increase in markup dispersion is about 75% of the peak empirical response, see Figure 2, and the model explains 60% of the peak empirical response in utilization-adjusted TFP,

[22] At an earlier stage of this paper, we have also solved the model globally using a time iteration algorithm for the case of two firm types with one of them having perfectly flexible prices. This yields very similar quantitative results compared to using the Meyer-Gohde (2014) algorithm. However, the computational costs of time iteration are exceedingly large for a more general setup with multiple firm types.

[23] Note that the only source of uncertainty in the stochastic steady state are monetary policy shocks. In principle, considering multiple shocks may increase or decrease the precautionary price setting motive. As Proposition 2 shows, precautionary price setting depends on the co-movement of prices, marginal costs, and aggregate demand. A sufficient condition for precautionary price setting is that all covariances between these variables are positive. This is commonly satisfied by monetary policy shocks, but, for example, not satisfied by technology shocks. Against this backdrop, if we add a technology shock to the model, which has the same effect on aggregate TFP as the monetary policy shock, we find very modest differences in precautionary price setting. The markup difference between top and bottom quintile increases from 11.52% to 11.57%. Similarly, the response to monetary policy shocks changes only marginally.

Figure 6: Model responses to monetary policy shocks



Notes: This figure shows responses to a one standard deviation contractionary monetary policy shock. In panel (e), the responses are the average markup responses of the firm types $k = 1, \dots, 5$, where $k = 1$ is the stickiest and $k = 5$ the most flexible type of firms.

see Figure 3. In addition, the responses show the frequency composition effect described by Carvalho (2006). The firms with flexible prices are quick to adjust. Hence, at longer horizons, the distribution of firms with non-adjusted prices is dominated by the stickier type of firms. This generates additional persistence in the responses.

In the model, contractionary monetary policy shocks raise markup dispersion and expansionary shocks lower markup dispersion, consistent with our empirical evidence. This response of markup dispersion critically depends on solving the model around the stochastic steady state, which allows us to capture precautionary price setting. In contrast, the deterministic steady state is characterized by zero markup dispersion. If we solve the model using a second-order approximation around the deterministic steady state, markup dispersion increases in response to both expansionary and contractionary monetary policy shocks, and irrespective of whether price rigidity is heterogeneous or homogeneous, see Figure 31 in the Appendix.

Even when capturing precautionary price setting, contractionary monetary policy shocks do not necessarily increase markup dispersion outside a local neighborhood around the stochastic steady state. After sufficiently large expansionary monetary policy shocks, markups of stickier firms may fall below the markups of more flexible firms. At this point, contractionary monetary policy may lower markup dispersion. We study the behavior of the model away from the stochastic steady state using a stochastic simulation of the model. The estimated response of markup dispersion on simulated data is similar and only somewhat smaller than the baseline response in Figure 6, see Appendix H.1 for details.

An important aspect of the monetary transmission channel in our model is the response of aggregate TFP. In contrast, traditional business cycle models assume that fluctuations in aggregate TFP are solely driven by exogenous technology shocks. This motivates us to examine the success of a Taylor rule in stabilizing output if the monetary authority in the model (mis-)perceives the aggregate TFP response to demand shocks as originating from technology shocks. Specifically, we construct a policy counterfactual, in which the only counterfactual element is natural output, and thus the output gap in the Taylor rule. Whereas model-consistent natural output responds to aggregate technology shocks but not to monetary policy shocks, counterfactual natural output responds to all changes in aggregate TFP.

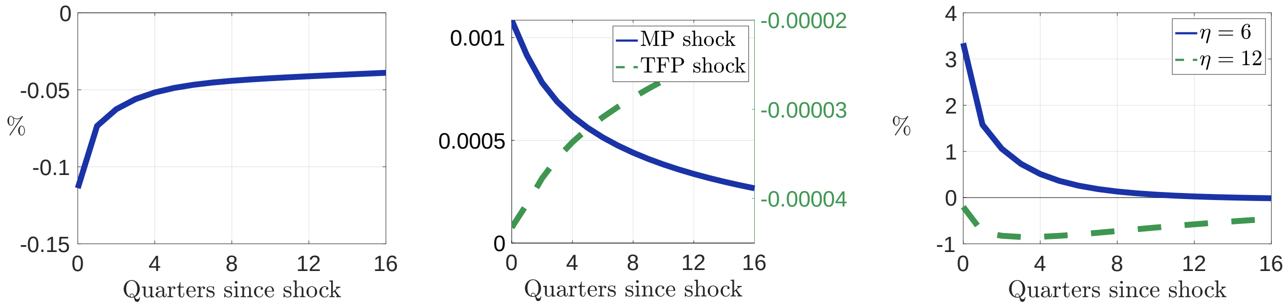
We then compare the effects of a monetary policy shock in the baseline and counterfactual model.[24] Panel (a) in Figure 7 shows the difference between the response of GDP in the counterfactual versus the baseline response.[25] Output drops by up to 0.11 percentage points more if the monetary authority attributes aggregate TFP fluctuations to technology shocks, and the response is markedly more persistent. In the counterfactual, the output gap response is dampened, which implies a less aggressive response of (systematic) monetary policy. This is similar to a lower Taylor coefficient on the output gap, and hence output falls by more. For further details and discussion, see Appendix H.2.

[24]We ensure the same interest rate response (30 bp) in baseline and counterfactual, by scaling up the size of the shock to 1.147 standard deviations in the counterfactual.

[25]Figure 28 in the Appendix provides further impulse responses for this counterfactual scenario.

Figure 7: Policy counterfactual and additional model results

(a) Differential GDP resp., policy counterfactual | (b) Markup dispersion resp. to technology shock | (c) Aggregate markup resp., low vs. high elasticity



Notes: Panel (a) shows the difference between the response to a monetary policy shock in the baseline model and the same model using a Taylor rule in which the output gap is computed by counterfactually assuming the TFP responses are driven by technology shocks. Panel (b) compares the response of markup dispersion to a monetary policy shock (left y-axis) with a technology shock (right y-axis). Panel (c) compares the response of the aggregate markup to a monetary policy shock for two values of the elasticity of substitution between differentiated goods.

Panel (b) in Figure 7 shows the response of markup dispersion to a negative technology shock with the size and persistence that matches the endogenous response of TFP to a monetary policy shock.[26] The behavior of markup dispersion helps to discriminate between productivity and monetary policy shocks. It increases after contractionary monetary policy shocks but decreases after contractionary productivity shocks. So, to avoid the cost of misattributing changes in aggregate TFP to technology shocks, the monetary authority could monitor changes in markup dispersion.

The fact that aggregate TFP responds to monetary policy shocks can change the sign of the (aggregate) markup response to monetary policy shocks. This relates to a recent debate. While monetary policy shocks raise markups in a large class of New Keynesian models, recent evidence in Nekarda and Ramey (2020) points in the opposite direction. Following Hall (1988), the aggregate markup in our model is

$$\mu_t = \frac{\text{TFP}_t}{W_t/P_t}, \tag{4.1}$$

[26]Figure 29 in the Appendix provides further impulse responses for the technology shock.

where W_t/P_t denotes the real wage. In standard New Keynesian models, tighter monetary policy reduces aggregate demand which lowers real marginal costs and, hence, markups increase. In contrast, equation (4.1) shows that the aggregate markup falls if aggregate TFP falls sufficiently strongly in response to tighter monetary policy. This argument extends to sectoral and even firm-level markups, if monetary policy shocks affect TFP at more disaggregated levels. In general equilibrium, an endogenous decline in aggregate TFP will feed back into real marginal costs, which also affects markups.

Panel (c) in Figure 7 shows the aggregate markup response to monetary policy shocks. In our baseline calibration with an elasticity of substitution $\eta = 6$ the aggregate markup raises. In some sense, that is because aggregate TFP does not fall strongly enough. We next compare our baseline results with the results when doubling the elasticity to $\eta = 12$. A larger η increases the misallocation costs of markup dispersion and thus the TFP loss after a monetary policy shock. For $\eta = 12$, the aggregate TFP response is almost twice as large, see Figure 30 in the Appendix. This is sufficient to explain lower aggregate markups after monetary policy shocks. Dynamically, the TFP loss leads to an increase in hours worked, which additionally increases marginal costs and lowers firm-level markups, reinforcing the effect on the aggregate markup.

4.4 Robustness to model variations

To investigate the robustness of our quantitative results, we analyze the effects of monetary policy shocks in a number of model variations. These include a model with real rigidities, a model with Rotemberg price adjustment, and a model with trend inflation.

Real rigidities We model real rigidities via firm-specific labor. In particular, households supply differentiated labor, which is firm-specific and immobile across firms. Under a condition similar to the one in Proposition 2, firms with more rigid prices optimally set higher prices. Proposition 1 then suggests that contractionary monetary policy shocks raise markup dispersion in this model. In a quantitative analysis of this model, we keep all model parameters unchanged except for the standard deviation of monetary policy shocks, which we re-calibrate to imply a 30bp increase of the nominal interest rate. Figure 32 in the Appendix shows that the response of markup dispersion and aggregate TFP to the shock is qualitatively the same as in the baseline model. Quantitatively, however, the responses are strongly amplified. For example, the peak decline in aggregate TFP is 0.87% compared to 0.34% in the baseline model. For details, see Appendix I.1.

Rotemberg price adjustment In the Rotemberg version of our model, we assume price adjustment costs of the form $\frac{\phi_i}{2}\left(\frac{P_{it}}{P_{it-1}}-1\right)^2$, as in Section 3.2. We assume ϕ_i differs across 5 quintile groups of firms similar to our Calvo model. We calibrate the ϕ_i for $i = 1, \ldots, 5$ to match the markups (in the stochastic steady state) of our baseline Calvo model. In addition, we re-calibrate the standard deviation of monetary policy shocks to match a 30 bp response of the nominal interest rate. We leave all other parameters unchanged. The Rotemberg model can exactly match the steady state markups of the baseline model. For example, firms in the most rigid quintile set 11.5% higher markups than firms in the most flexible quintile. This shows that quantitatively strong precautionary price setting motives are not per se limited to the Calvo model. At the same time, the calibrated ϕ_i are not unreasonably large in the sense that monetary policy shocks do not have larger real effects than our baseline Calvo model. In fact, monetary policy shocks generate a 2/3 smaller GDP and a 1/3 smaller TFP response, see Figure 33 in the Appendix.

Trend inflation We extend our baseline Calvo model by deterministic trend inflation as in Ascari and Sbordone (2014). We assume an annualized trend inflation of 2%. In this model, we only recalibrate the standard deviation of monetary policy shocks to match the 30 bp response of the nominal interest rate. We leave all other parameters unchanged. We find amplified markup differences across firms with differently rigid prices. In the stochastic steady state, firms with the most rigid prices have 19% higher markups than firms with the most flexible prices. Figure 34 in the Appendix further provides the impulse responses to a one-standard deviation monetary policy shock. On impact, monetary policy shocks generate an even larger increase in markup dispersion and thus drop in aggregate TFP.

5 Conclusion

This paper studies how markup dispersion matters for monetary transmission. Monetary policy shocks increase the dispersion of markups across firms if firms with stickier prices have higher pre-shock markups. Increased markup dispersion implies a change in the allocation of inputs across firms, which lowers measured aggregate TFP. Using aggregate and firm-level data, we document three new facts, which are consistent with this mechanism. First, firms that adjust prices less frequently have higher markups. Second, monetary policy shocks increase the relative markup of firms with stickier prices. Third, monetary policy shocks increase the markup dispersion across firms, and lower aggregate productivity. The empirically estimated magnitudes suggest that the response in markup dispersion is quantitatively important to understand the response of aggregate productivity. We show that an explanation for the negative correlation between markup and price stickiness are differences in price stickiness across firms. Firms with stickier prices optimally set higher markups for precautionary reasons. We show that our novel mechanism has implications for monetary policy and for the markup response to monetary policy shocks.

References

Aoki, K. (2001): "Optimal monetary policy responses to relative-price changes," *Journal of Monetary Economics*, 48, 55 – 80.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016): "Improving GDP measurement: A measurement-error perspective," *Journal of Econometrics*, 191, 384–397.

Ascari, G. and A. M. Sbordone (2014): "The Macroeconomics of Trend Inflation," *Journal of Economic Literature*, 52, 679–739.

Ashenfelter, O. C., H. Farber, and M. R. Ransom (2010): "Labor market monopsony," *Journal of Labor Economics*, 28, 203–210.

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

Baqaee, D., E. Farhi, and K. Sangani (2021): "The Supply-Side Effects of Monetary Policy," Working Paper 28345, National Bureau of Economic Research.

Baqaee, D. R. and E. Farhi (2017): "Productivity and Misallocation in General Equilibrium," *NBER Working Paper Series*.

——— (2020): "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

Barro, R. (1972): "A Theory of Monopolistic Price Adjustment," *Review of Economic Studies*, 39, 17–26.

Basu, S. (1995): "Intermediate Goods and Business Cycles: Implications for Productivity and Welfare," *The American Economic Review*, 85, 512–531.

——— (2019): "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence," *Journal of Economic Perspectives*, 33, 3–22.

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

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2020): "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," mimeo.

Calvo, G. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12, 383–398.

Carvalho, C. (2006): "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks," *The B.E. Journal of Macroeconomics*, 6, 1–58.

Carvalho, C. and F. Schwartzman (2015): "Selection and monetary non-neutrality in time-dependent pricing models," *Journal of Monetary Economics*, 76, 141–156.

Chow, G. and A.-l. Lin (1971): "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series," *The Review of Economics and Statistics*, 53, 372–75.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

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

Comin, D. and M. Gertler (2006): "Medium-Term Business Cycles," *American Economic Review*, 96, 523–551.

Comin, D. and B. Hobijn (2010): "An Exploration of Technology Diffusion," *American Economic Review*, 100, 2031–59.

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

De Loecker, J. and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–71.

Edmond, C., V. Midrigan, and D. Y. Xu (2021): "How costly are markups?" *mimeo*.

Eisfeldt, A. L. and A. A. Rampini (2006): "Capital reallocation and liquidity," *Journal of Monetary Economics*, 53, 369–399.

Eusepi, S., B. Hobijn, and A. Tambalotti (2011): "CONDI: A Cost-of-Nominal-Distortions Index," *American Economic Journal: Macroeconomics*, 3, 53–91.

Evans, C. L. and F. T. Santos (2002): "Monetary policy shocks and productivity measures in the G-7 countries," *Portugese Economic Journal*, 1, 47–70.

Fernald, J. (2014): "A quarterly, utilization-adjusted series on total factor productivity," *Federal Reserve Bank of San Francisco Working Paper Series*.

Fernandez-Villaverde, J., P. Guerron-Quintana, K. Kuester, and J. Rubio-Ramirez (2015):

"Fiscal Volatility Shocks and Economic Activity," *American Economic Review*, 105, 3352–84.

Flynn, Z., J. Traina, and A. Gandhi (2019): "Measuring Markups with Production Data," mimeo.

Garga, V. and S. R. Singh (2021): "Output hysteresis and optimal monetary policy," *Journal of Monetary Economics*, 117, 871–886.

Gertler, M. and P. Karadi (2015): "Monetary Policy Surprises, Credit Costs and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Gilchrist, S., R. Schoenle, J. Sim, and E. Zakrajšek (2017): "Inflation Dynamics during the Financial Crisis," *American Economic Review*, 107, 785–823.

Gilchrist, S. and E. Zakrajsek (2012): "Credit Spreads and Business Cycle Fluctuations," *The American Economic Review*, 102, 1692–1720.

Gopinath, G., S. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017): "Capital Allocation and Productivity in South Europe," *Quarterly Journal of Economics*, 132, 1915–1967.

Gorodnichenko, Y. and M. Weber (2016): "Are Sticky Prices Costly? Evidence from the Stock Market," *American Economic Review*, 106, 165–99.

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the US," Nber working paper series.

Hall, R. (1986): "Market Structure and Macroeconomic Fluctuations," *Brookings Papers on Economic Activity*, 17, 285–338.

——— (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, 921–47.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, CXXIV, 1403–1448.

Jarocinski, M. and P. Karadi (2020): "Deconstructing monetary policy surprises: the role of information shocks," *American Economic Journal: Macroeconomics*.

Jeenas, P. (2019): "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics," *mimeo*.

Jordà, Ò., S. Singh, and A. M. Taylor (2020): "The long-run effects of monetary policy," *mimeo*.

Khan, A. and J. K. Thomas (2008): "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics," *Econometrica*, 76, 395–436.

——— (2013): "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity," *Journal of Political Economy*, 121, 1055–1107.

Kim, R. (2020): "The Effect of the Credit Crunch on Output Price Dynamics: The Corporate Inventory and Liquidity Management Channel*," *The Quarterly Journal of Economics*, 136, 563–619.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27, 1241–1277.

Klenow, P. J. and J. L. Willis (2016): "Real Rigidities and Nominal Price Changes," *Economica*, 83, 443–472.

Levin, A. and T. Yun (2007): "Reconsidering the natural rate hypothesis in a New Keynesian framework," *Journal of Monetary Economics*, 54, 1344–1365.

Lucca, D. O. and E. Moench (2015): "The pre-FOMC announcement drift," *The Journal of Finance*, 70, 329–371.

Meier, M. (2020): "Supply Chain Disruptions, Time To Build, and the Business Cycle," *mimeo*.

Meyer-Gohde, A. (2014): "Risky linear approximations," mimeo.

Miranda-Agrippino, S. and G. Ricco (2018): "The Transmission of Monetary Policy Shocks," mimeo.

Moran, P. and A. Queralto (2018): "Innovation, productivity, and monetary policy," *Journal of Monetary Economics*, 93, 24–41, carnegie-Rochester-NYU Conference on Public Policy.

Nakamura, E. and J. Steinsson (2010): "Monetary Non-neutrality in a Multisector Menu Cost Model*," *The Quarterly Journal of Economics*, 125, 961–1013.

——— (2018): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133, 1283–1330.

Nakamura, E., J. Steinsson, P. Sun, and D. Villar (2018): "The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation*," *The Quarterly Journal of Economics*, 133, 1933–1980.

Nekarda, C. J. and V. A. Ramey (2020): “The Cyclical Behavior of the Price-Cost Markup,” *Journal of Money, Credit and Banking*, 52, 319–353.

Oikawa, K. and K. Ueda (2018): “Reallocation Effects of Monetary Policy,” *RIETI Discussion Paper Series 18-E-056*.

Ottonello, P. and T. Winberry (2020): “Financial Heterogeneity and the Investment Channel of Monetary Policy,” *mimeo*.

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.

Plagborg-Møller, M. and C. K. Wolf (2021): “Local Projections and VARs Estimate the Same Impulse Responses,” *Econometrica*, 89, 955–980.

Romer, D. (1990): “Staggered price setting with endogenous frequency of adjustment,” *Economics Letters*, 32, 205–210.

Rotemberg, J. J. (1982): “Monopolistic Price Adjustment and Aggregate Output,” *The Review of Economic Studies*, 49, 517–531.

Rubbo, E. (2020): “Networks, Phillips Curves and Monetary Policy,” mimeo.

Stock, J. H. and M. W. Watson (2018): “Identification and estimation of dynamic causal effects in macroeconomics using external instruments,” *The Economic Journal*, 128, 917–948.

Taylor, J. B. (1979): “Staggered Wage Setting in a Macro Model,” *American Economic Review*, 69, 108–113.

Traina, J. (2020): “Is Aggregate Market Power Increasing? Production Trends Using Financial Statements,” New working paper series, Stigler Center.

Woodford, M. (2003): *Interest and Prices*, Princeton University Press.

Zeev, N. B. and E. Pappa (2015): “Multipliers of unexpected increases in defense spending: An empirical investigation,” *Journal of Economic Dynamics and Control*, 57, 205–226.

Appendix

A Data construction and descriptive statistics

A.1 Firm-level balance sheet data

We use quarterly firm-level balance sheet data of listed US firms for the period 1995Q1 to 2017Q2 from Compustat. We delete duplicate firm-quarter observations. We use the NAICS industry classification and exclude firms in utilities (NAICS code 22), finance, insurance, and real estate (52 and 53), and public administration (99). We discard observations of sales (`saleq`), costs of goods sold (`cogsq`) and property, plant, and equipment (net PPE, `ppentq`, and gross PPE, `ppegtq`) and total assets (`atq`) that are weakly negative. We fill missing values of depreciation and amortization (`dpq`), selling, general and administrative expenses (`xsgaq`), debt in current liabilities (`dlcq`), long-term debt (`dlttq`) and cash and short-term investments (`cheq`) by zero. We discard observations of these same variables if they are strictly negative. We fill one-quarter gaps in the firm-specific series of these variables by linear interpolation. All variables are deflated using the GDP deflator, except PPE, which is deflated by the investment-specific GDP deflator. We construct a measure of the capital stock of firms using the perpetual inventory method: We initialize $K_{it_0} = \texttt{ppegtq}_{it_0}$ and recursively compute $K_{it} = K_{it-1} + (\texttt{ppentq}_{it} - \texttt{ppentq}_{it-1})$. We drop firm-quarter observations if sales, costs of goods sold, or fixed assets are only reported once in the associated year. We further drop observations if quarterly sales growth is above 100% or below -67% or if real sales are below 1 million USD. Table 3 shows descriptive statistics for our baseline sample.

Table 3: Summary statistics for Compustat data

	mean	sd	min	max	count
Sales	632.22	3067.46	1.00	132182.15	329173
Fixed assets	987.38	5490.96	0.00	273545.97	326223
Variable costs	439.58	2317.01	0.13	104456.86	329173
Total Assets	2716.05	13374.72	0.00	559922.78	326632

Notes: Summary statistics for Compustat data. All variables are in millions of 2012Q1 US$.

A.2 Monetary policy shocks

We construct high-frequency identified monetary policy shocks as described in Subsection 2.1. Table 4 reports summary statistics for shock series and Figure 8 shows the time series.

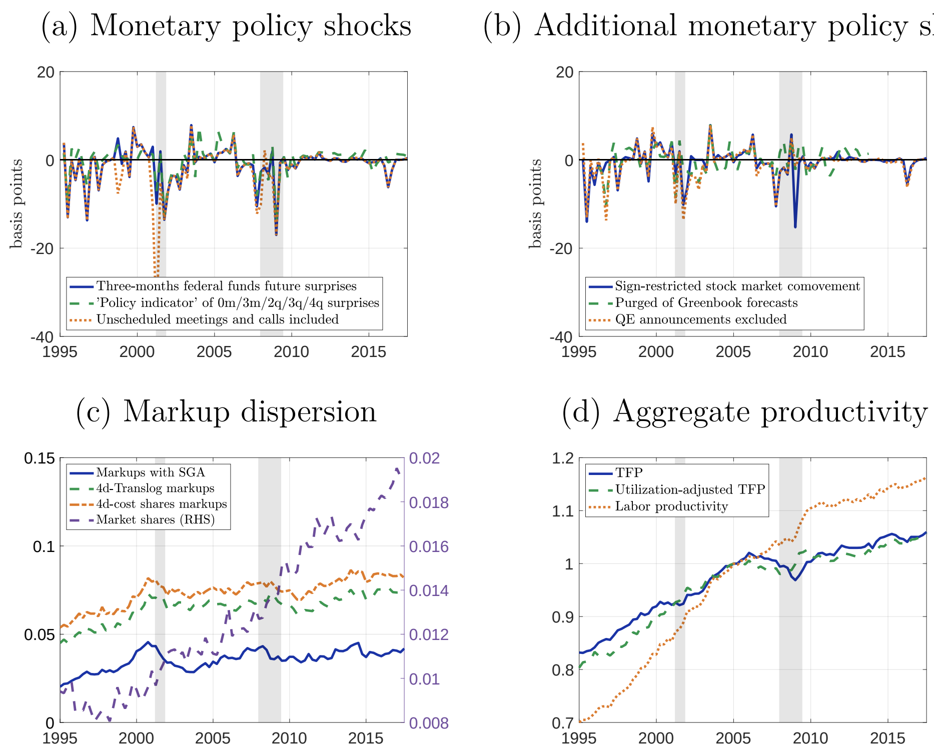
Table 4: Summary statistics of monetary policy shocks

	mean	sd	min	max	count
Three-month Fed funds future surprises	-1.00	4.06	-17.01	7.87	94
... unscheduled meetings and conference calls included	-1.84	5.70	-38.33	7.86	94
... purged of Greenbook forecasts	-0.00	3.10	-10.47	7.98	71
... sign-restricted stock market comovement	-0.52	3.47	-15.27	7.87	94
... QE announcements excluded	-0.83	3.72	-13.71	7.87	94
'Policy indicator' surprise	-0.05	3.43	-14.13	7.45	94

Notes: Summary statistics for monetary policy shocks in basis points.

A.3 Time series plots

Figure 8: Monetary policy shocks, aggregate productivity, and markup dispersion



Notes: Panel (a) and (b) show monetary policy shock series. Panel (c) plots markup dispersion measures within four-digit-industry-quarters in addition those in Figure 1 . Productivity measures in panel (d) are in logs and normalized to 1 in 2005Q1. Shaded gray areas indicate NBER recession dates.

A.4 Data on price rigidity

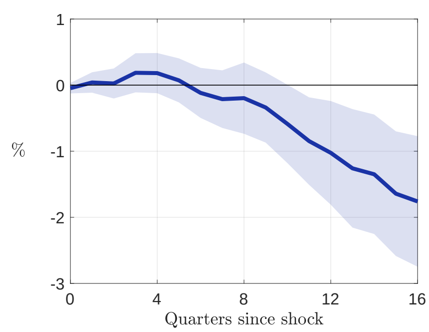
To maximize firm-level variation in price rigidity, we weight average industry-level price adjustment frequency with firms' industry sales from the Compustat segment files. Industry-level price adjustment frequency is based on Pasten et al. (2020). We define the implied price duration as $-1/\log(1 - \text{price adjustment frequency})$.

We obtain firms' yearly industry sales composition using the operation segments and, if these are not available, the business segments from the Compustat segments file. We drop various types of duplicate observations: In case of exact duplicates, we keep one. In case there are different source dates or more than one accounting month per year, we keep the observation with the newest source dates or the later accounting month, respectively. We drop segment observations for firm-years if the industry code is not reported. If only some segment industry codes are missing, we assign the firm-specific industry code to the segments with missing industry code.

We then compute every firm's average price rigidity over segments weighted by sales. In case we do not observe the five-digit-industry-level price stickiness for all segments or we observe only one segment, we use the five-digit price rigidity measure associated to the firm's general five-digit industry code. Note that even in this case, there is variation across firms within four-digit industries. Our sample comprises 8,091 unique firms. For 1,891 firms (23%), we can compute a segment-based price stickiness level in some year. For firm-years with segment-based price stickiness, the mean (median) number of segments is 2.36 (2) with a standard deviation of 0.67.

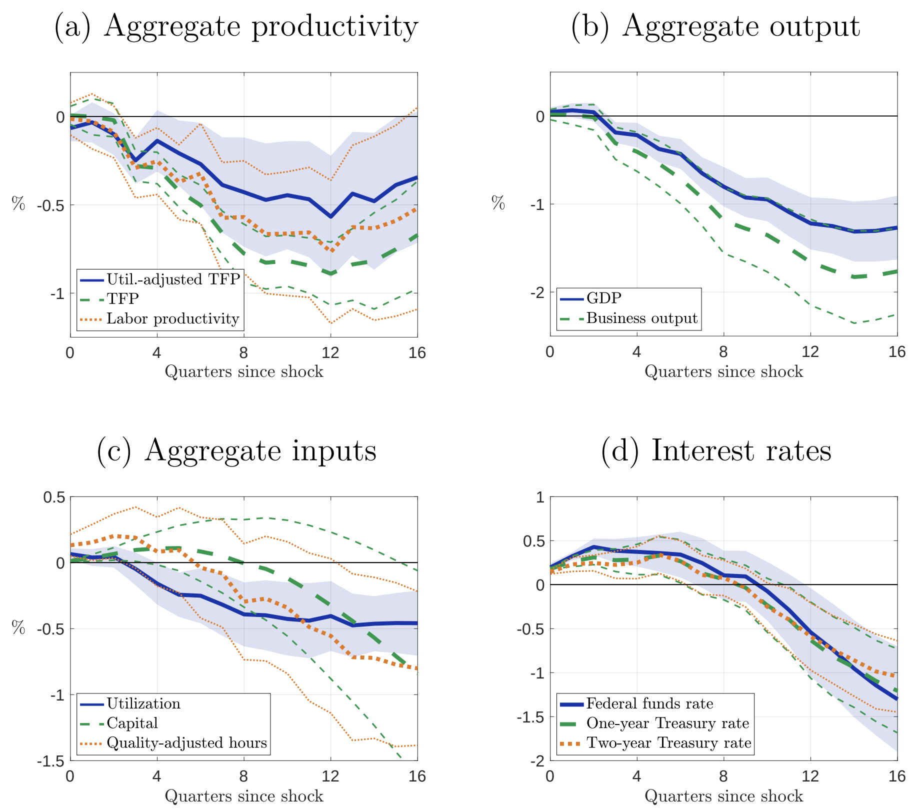
B Additional empirical results

Figure 9: Aggregate R&D response to monetary policy shock



Notes: This figure shows the response of aggregate R&D investment to monetary policy shocks obtained from local projections as in equation (2.4). The shaded area indicate one standard error bands based on the Newey–West estimator.

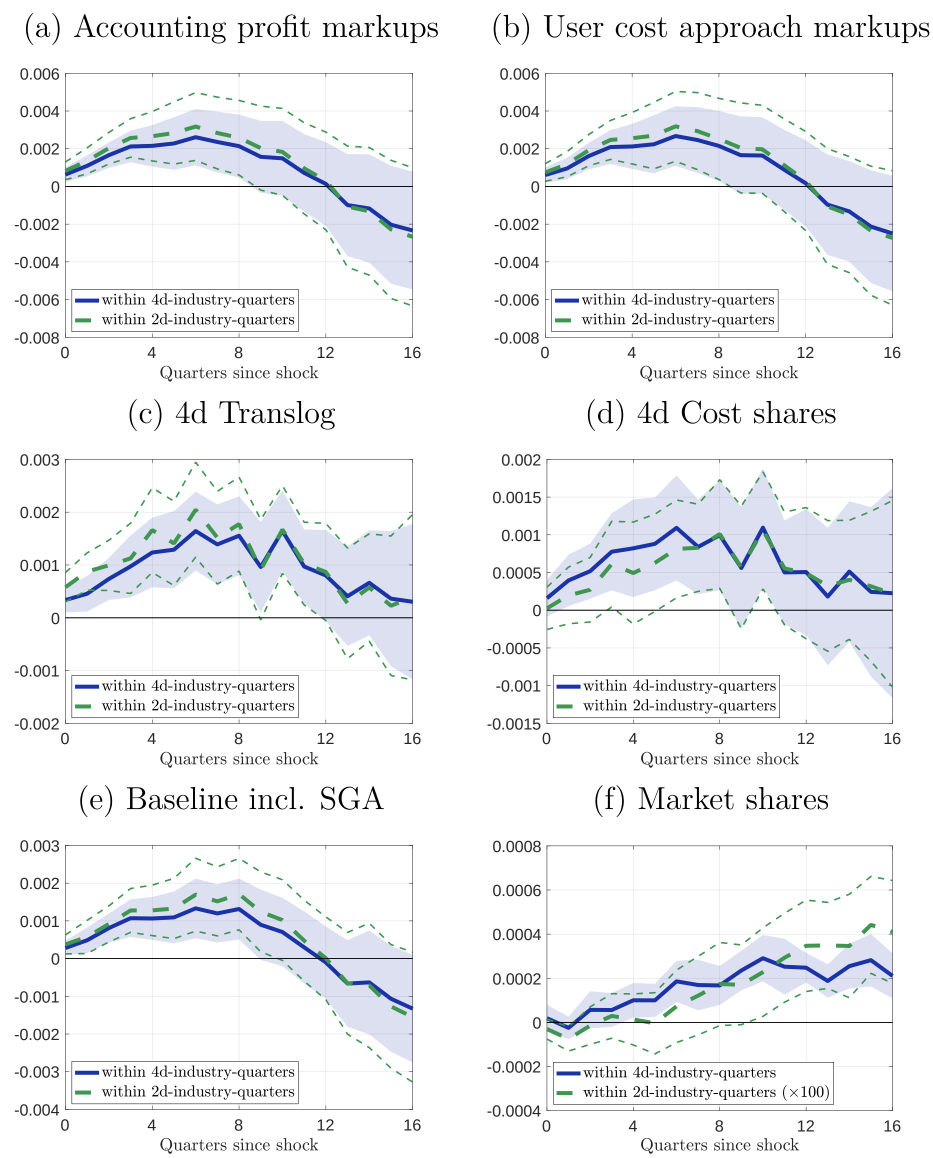
Figure 10: Macroeconomic responses to monetary policy shocks



Notes: This figure shows macroeconomic responses to monetary policy shocks obtained from local projections as in equation (2.4). The local projections in Panel (d) are estimated in levels rather than log differences. The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

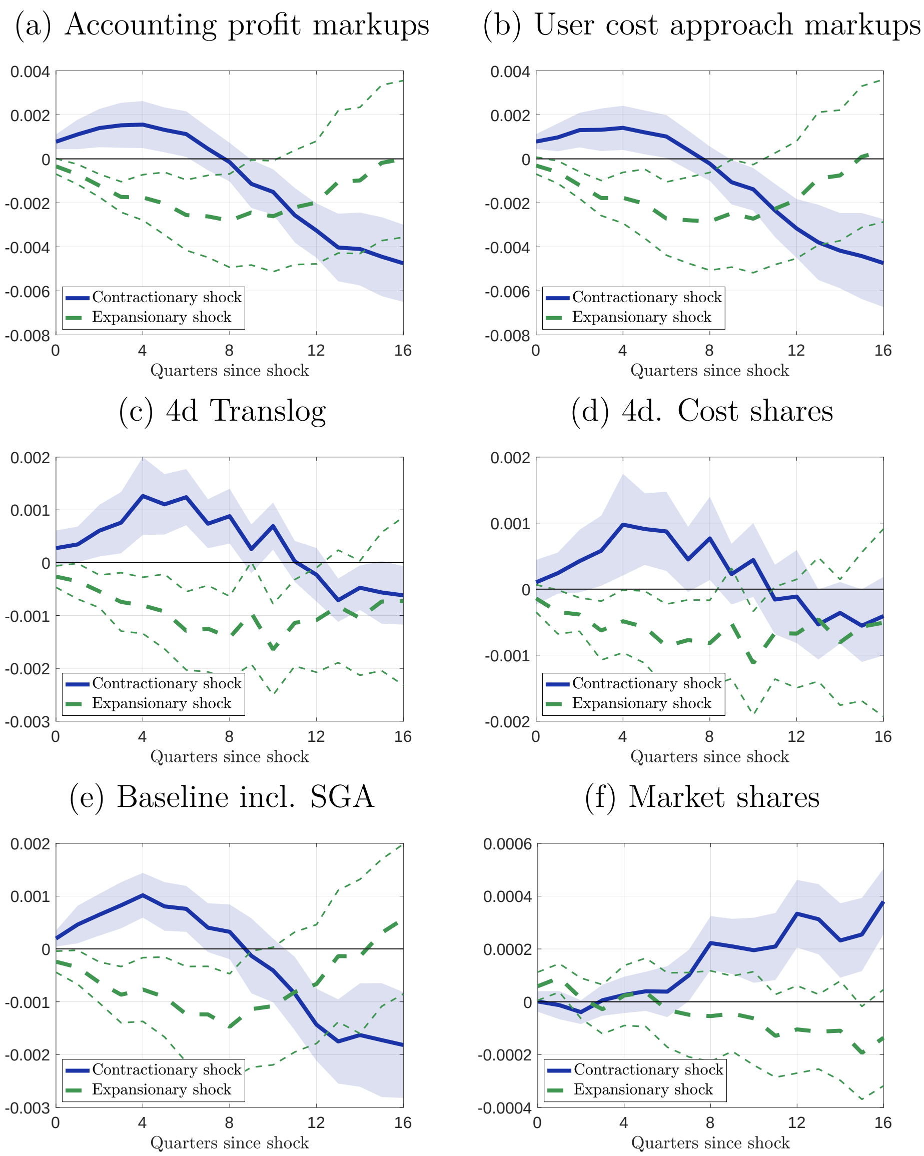
C Robustness of evidence in Section 2

Figure 11: Responses of markup dispersion



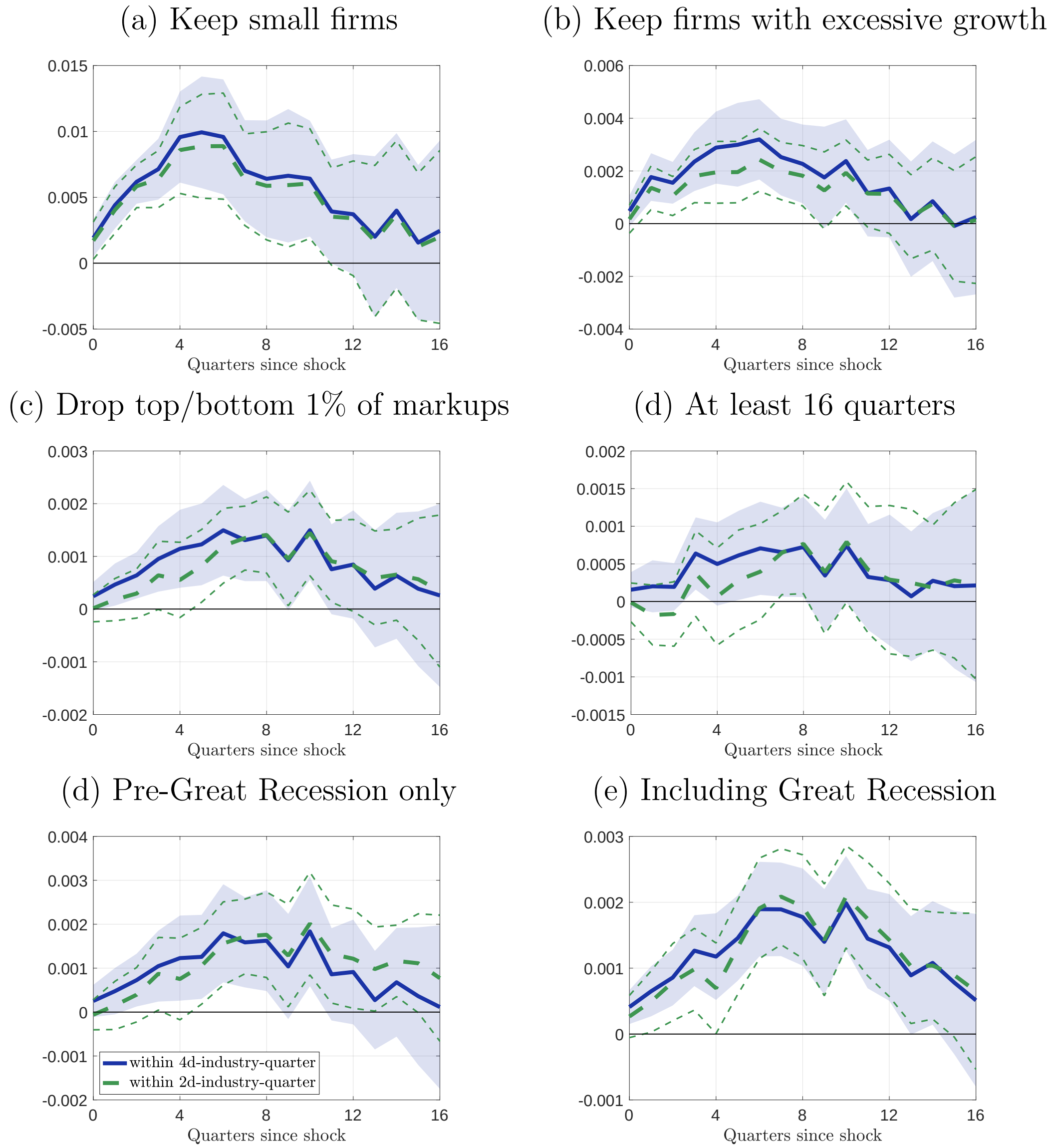
Notes: This figure shows the responses of markup dispersion to monetary policy shocks obtained from local projections as in equation (2.4). Markup dispersion is measured within two-digit and four-digit industry-quarters based on different markup measures, see Section 2.4 for details. Market share dispersion is computed as the variance of firm-level sales over total sales within two-digit and four-digit industry-quarters. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 12: Asymmetric responses of markup dispersion



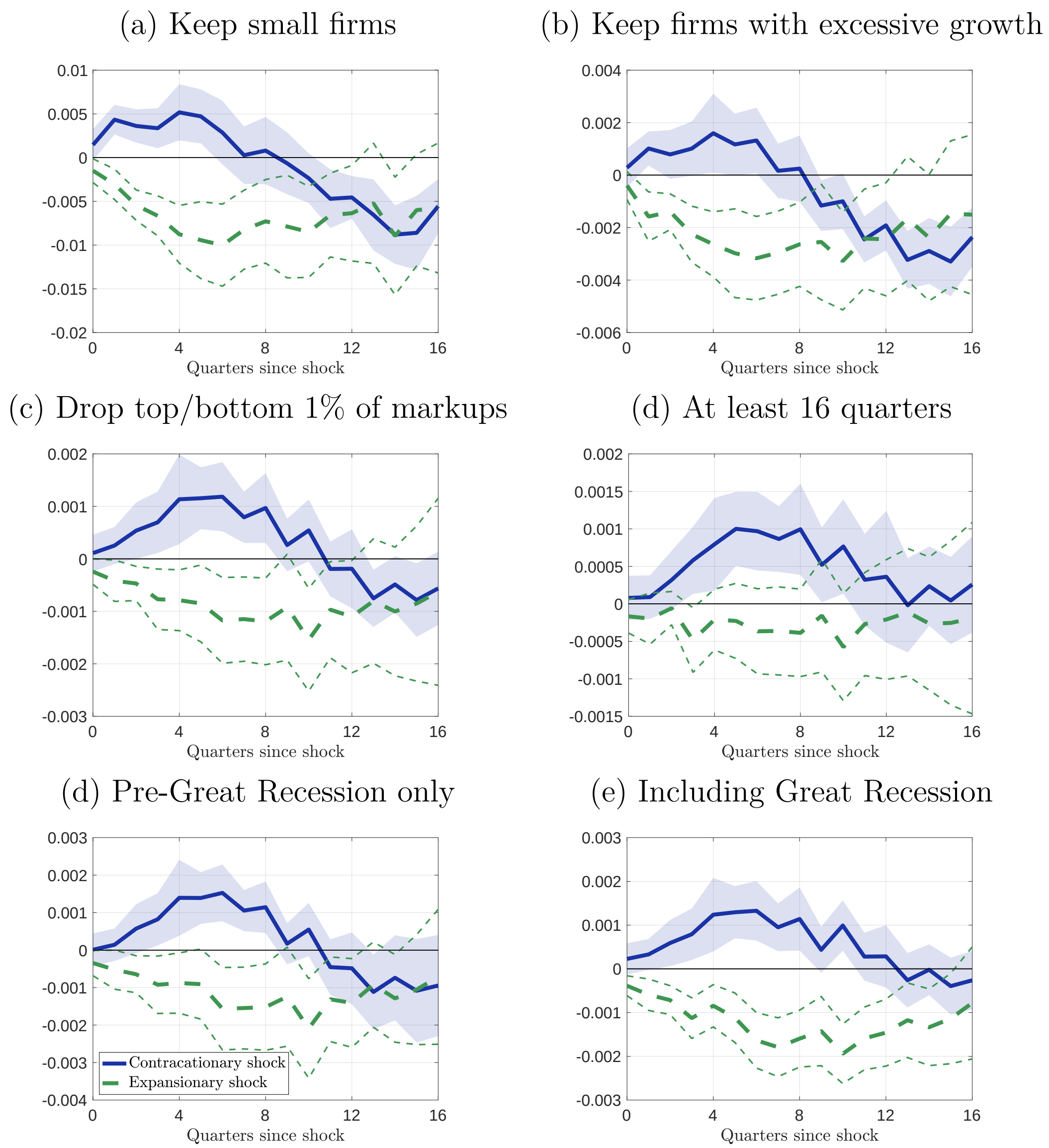
Notes: This figure shows the asymmetric responses to monetary policy shocks obtained from local projections extending specification (2.4) to separately estimate the response to positive and negative shocks. Markup dispersion is measured within two-digit-industry-quarters. Panels (a)–(e) use various markup measures and panel (f) uses market shares within two-digit-industry-quarters, see Section 2.4 for details. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 13: Responses of markup dispersion under alternative data treatments



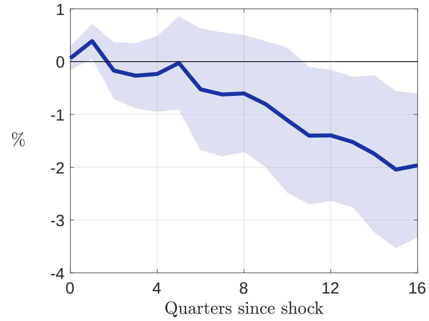
Notes: This figure shows the responses to monetary policy shocks obtained from local projections as in equation (2.4). Markup dispersion is measured within two-digit and four-digit industry-quarters using the baseline markup measure. See Section 2.4 for details on the different data treatments. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 14: Asymmetric markup dispersion responses for alternative data treatments



Notes: This figure shows the responses to monetary policy shocks obtained from local projections extending specification (2.4) to separately estimate the response to positive and negative shocks. Markup dispersion is measured within four-digit industry-quarters using the baseline markup measure. See Section 2.4 for details on the different data treatments. The shaded and bordered areas indicate one standard error bands based on Newey–West.

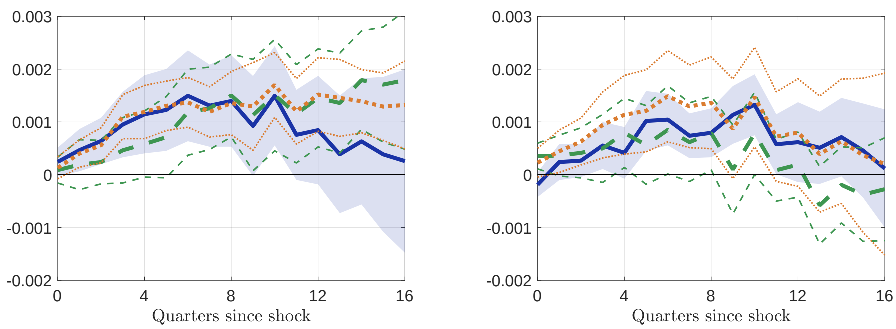
Figure 15: Response of firm-level observations after monetary policy shocks



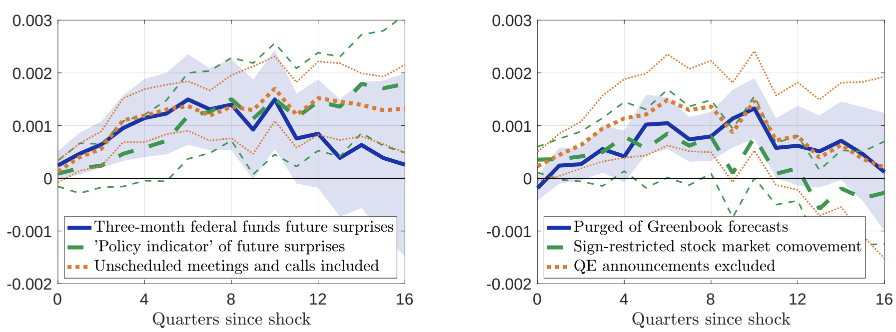
Notes: This figure shows the response of the number of firm-level observations in our sample to monetary policy shocks obtained from local projections as in equation (2.4). The shaded area is a one standard error band based on Newey–West.

Figure 16: Responses of markup dispersion for alternative monetary policy shocks

(a) within 2d-industry-quarter



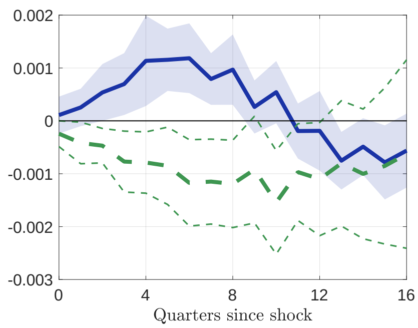
(b) within 4d-industry-quarter



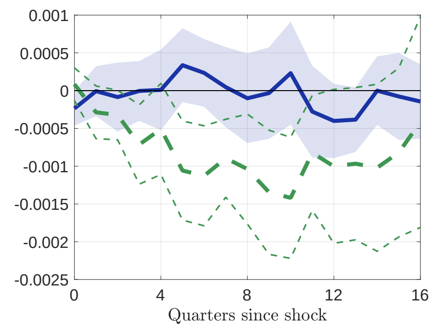
Notes: This figure shows the responses to monetary policy shocks obtained from local projections as in equation (2.4). Markup dispersion is measured within four-digit industry-quarters using the baseline markup measure. See Section 2.4 for details on the different monetary policy shocks. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 17: Asymmetric markup dispersion responses for alternative monetary policy shocks

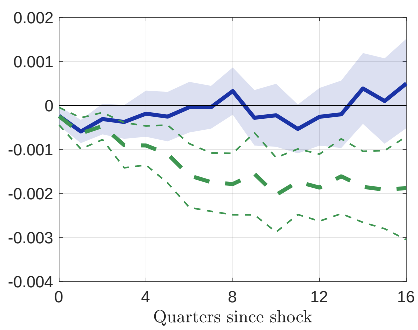
(a) Baseline



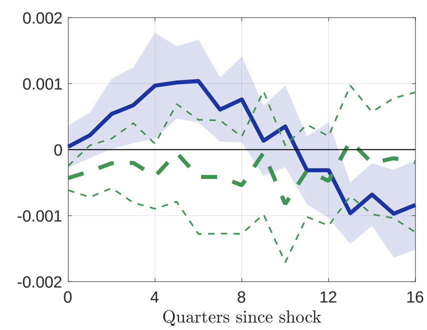
(b) Purged of Greenbook forecasts



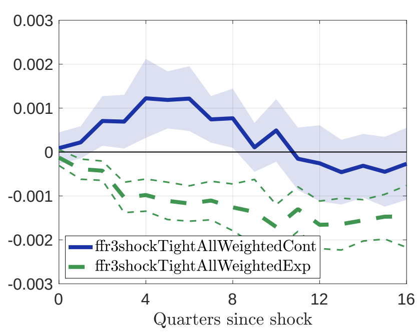
(c) ‘Policy indicator’ of future surprises



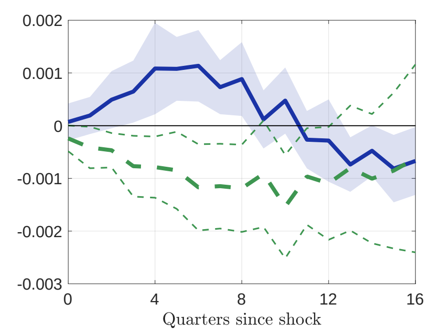
(d) Sign-restricted stock market comovement



(e) Unscheduled meetings and calls included



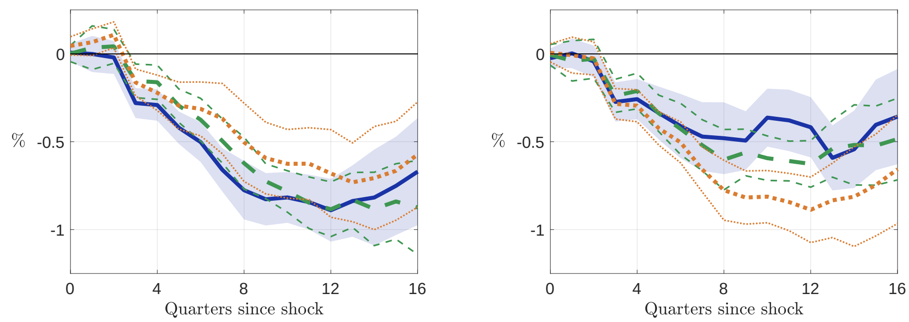
(f) QE announcements excluded



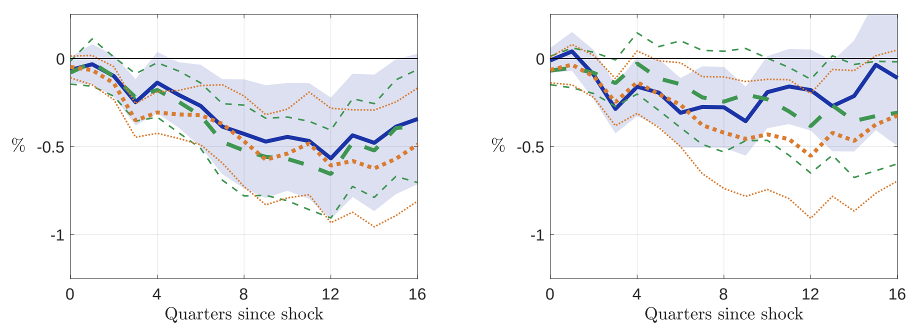
Notes: This figure shows the asymmetric responses to monetary policy shocks obtained from local projections extending specification (2.4) to separately estimate the response to positive and negative shocks. Markup dispersion is measured within four-digit industry-quarters using the baseline markup measure. See Section 2.4 for details on the different monetary policy shocks. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 18: Aggregate productivity responses for alternative monetary policy shocks

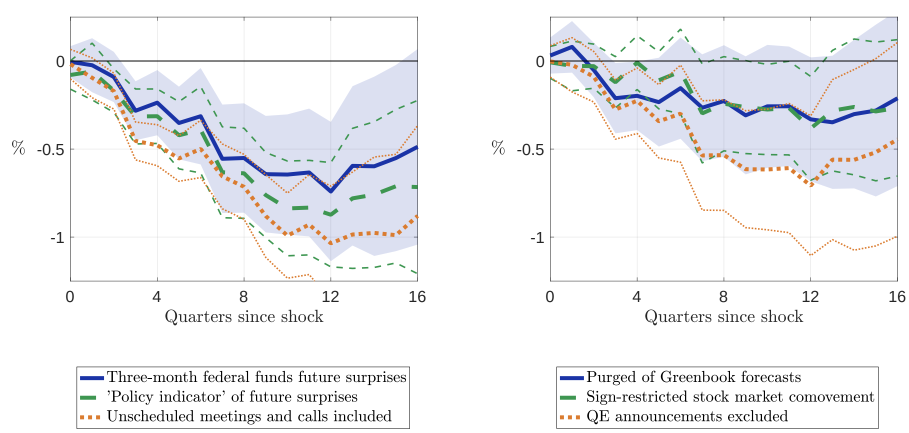
(a) TFP



(b) Utilization-adjusted TFP

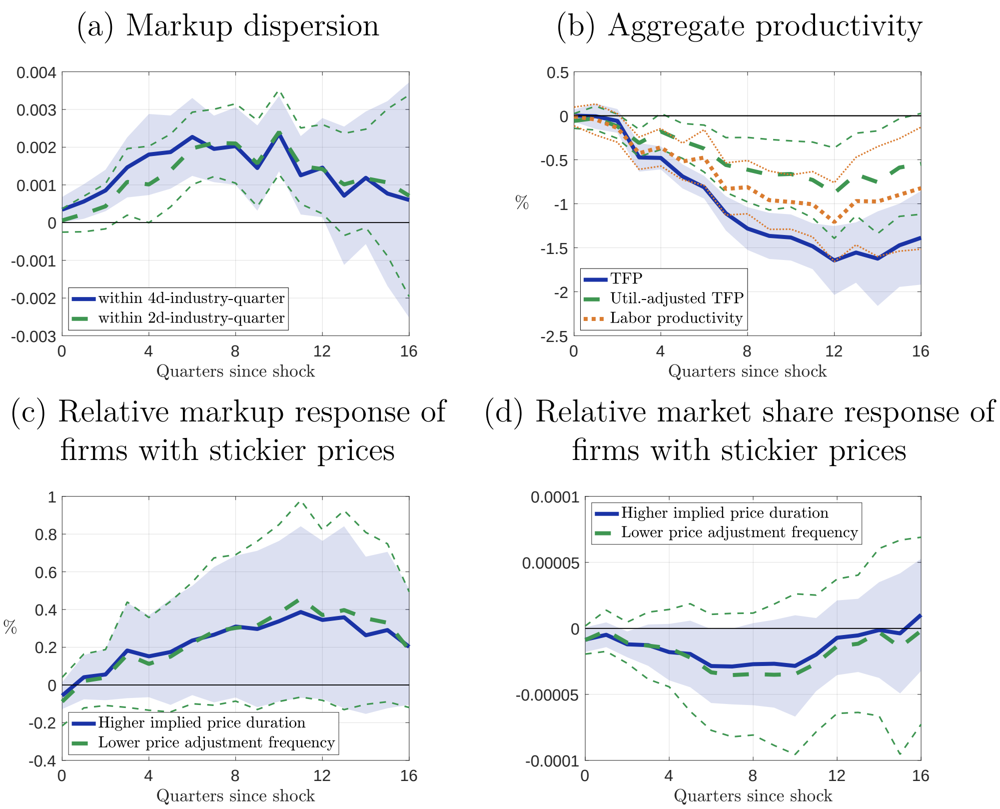


(c) Labor productivity



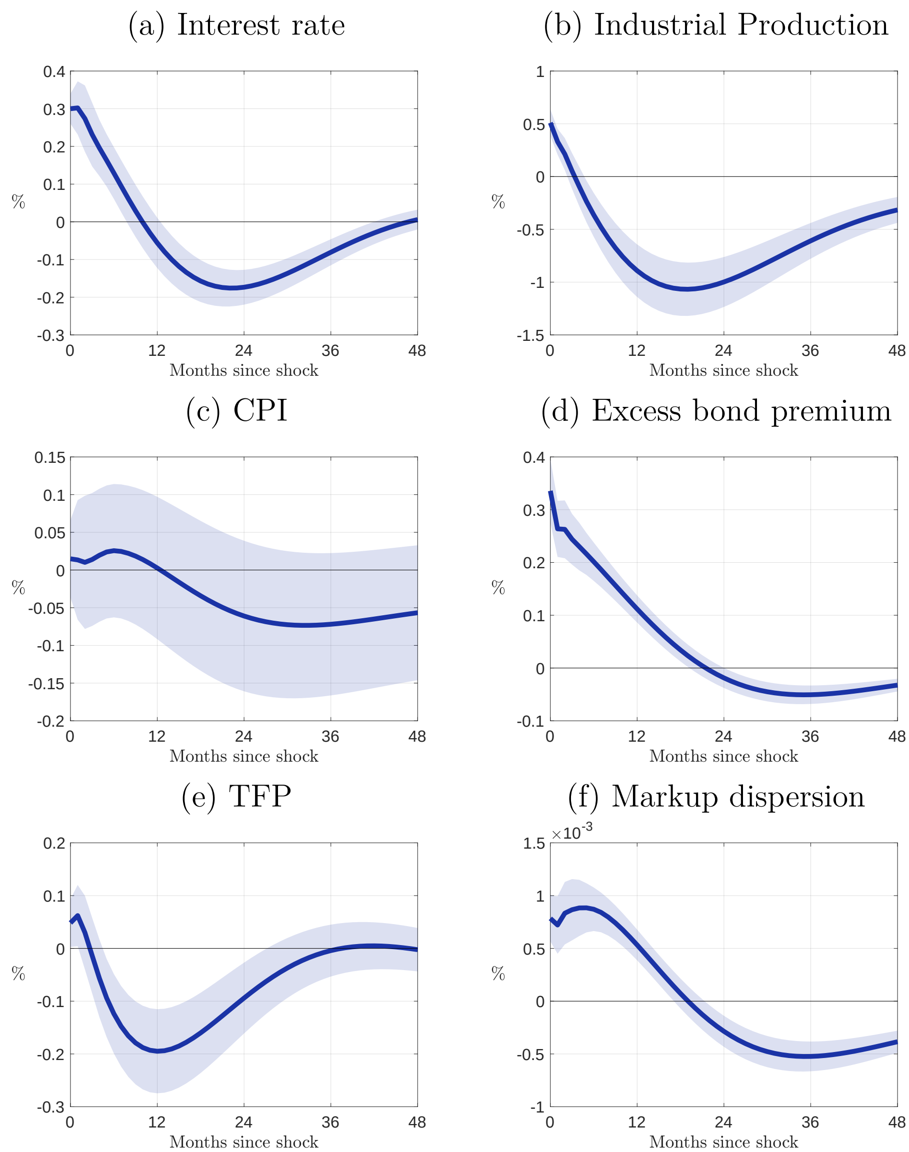
Notes: This figure shows the responses of aggregate productivity to monetary policy shocks obtained from local projections as in equation (2.4). The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 19: Main results using LP-IV



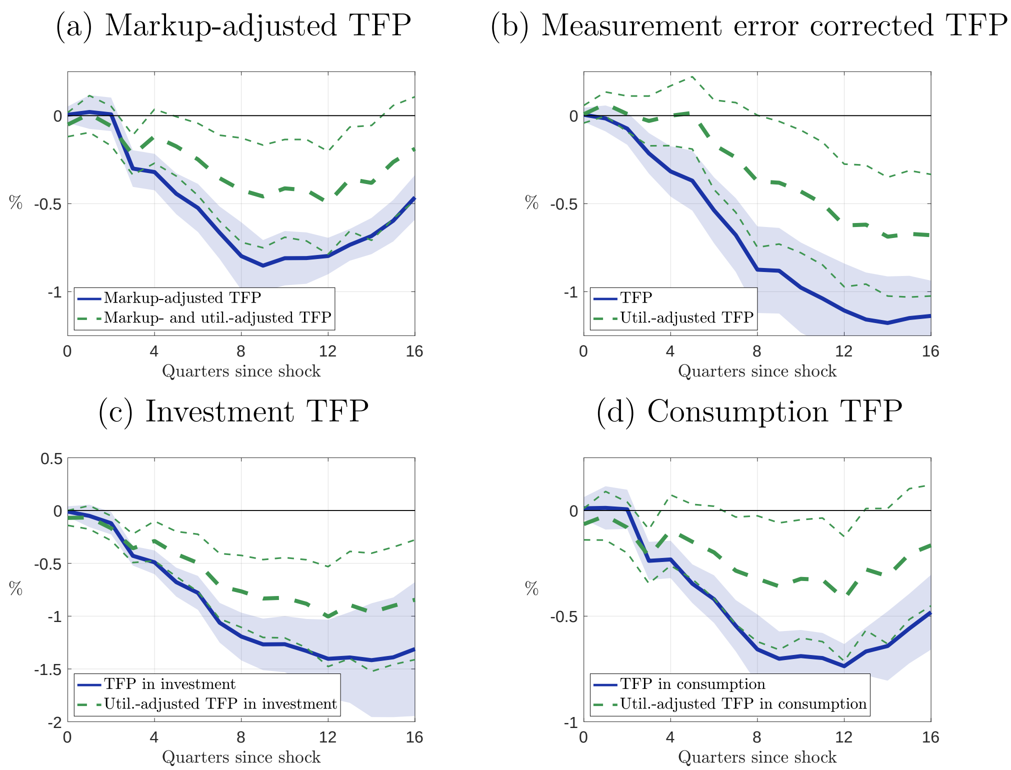
Notes: This figure shows the responses to monetary policy shocks obtained from local projections with instrumental variables (LP-IV), $y_{t+h} - y_{t-1} = \alpha^h + \beta^h \Delta R_t + \gamma_1^h (y_{t-1} - y_{t-2}) + u_t^h$, (in panels (a) and (b)) and analogues of the panel local projections (in panels (c) and (d), with firm-level), where the changes in the one-year Treasury rate, ΔR_t, (and the interactions thereof with price stickiness, respectively) are instrumented with the monetary policy shocks $\varepsilon_t^{\mathrm{MP}}$ (and the interactions of monetary policy shocks with price stickiness, respectively). The shaded and bordered areas in panels (a) and (b) indicate a one standard error band based on Newey–West, and in panels (c) and (d) they indicate a 90% error band two-way clustered by firms and quarters.

Figure 20: Proxy SVAR results



Notes: This figures shows the responses to a monetary policy shock, which raises the interest rate by 30bp, based on proxy SVAR similar to Gertler and Karadi (2015). The VAR is estimated at monthly frequency with three lags, including the one-year rate, (log) industrial production, (log) CPI, the excess bond premium of Gilchrist and Zakrajsek (2012), (log) TFP and the baseline measure of markup dispersion (within four-digit-industry-quarters). TFP and markup dispersion are interpolated to monthly frequency using the procedure of Chow and Lin (1971). Shaded areas are one-standard error bands from a wild bootstrap-after-bootstrap.

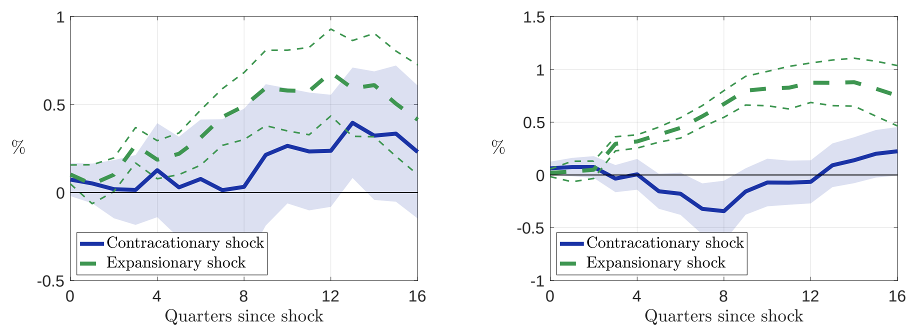
Figure 21: Further productivity responses



Notes: Responses to monetary policy shocks obtained from local projections as in equation (2.4). Investment-TFP and Consumption-TFP are from Fernald (2014). Markup-corrected TFP is constructed following Hall (1988) using the average markup estimated by De Loecker et al. (2020). Measurement error corrected TFP is constructed using measurement error corrected GDP from Aruoba et al. (2016), total hours from the BLS, and capital stock and output elasticities from Fernald (2014). The utilization-adjusted measure subtracts utilization from Fernald (2014). The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 22: Asymmetric responses of (util.-adjusted) TFP to monetary policy shocks

(a) Util.-adjusted TFP (b) TFP



Notes: This figure shows the responses of productivity to monetary policy shocks obtained from local projections extending specification (2.4) to separately estimate the response to positive and negative shocks. TFP and utilization-adjusted TFP are from Fernald (2014). The shaded and bordered areas indicate one standard error bands based on Newey–West.

D Robustness of evidence in Section 3

Table 5: Regressions of markup on price stickiness incl. all price adjustment frequencies

(a) Regressions of markups on implied price duration

	log(Markup)			
	Baseline		Accounting profits	User cost approach
Implied price duration	0.0433 (0.0197)	0.0362 (0.0176)	0.00686 (0.00290)	0.00937 (0.00336)
Additional controls	No	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	4014	4014	3961	3952
Adjusted R^2	0.107	0.184	0.239	0.166

(b) Regressions of markups on price adjustment frequency

	log(Markup)			
	Baseline		Accounting profits	User cost approach
Price adjustment frequency	-0.244 (0.144)	-0.185 (0.134)	-0.0450 (0.0180)	-0.0629 (0.0194)
Additional controls	No	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	4014	4014	3961	3952
Adjusted R^2	0.103	0.180	0.239	0.167

Notes: Regression of firm-level markup (averaged over 2005–2011) on firm-level price adjustment frequency and implied price duration, respectively, when including firms with price adjustment frequencies above 99%. Standard errors are clustered at the two-digit industry level and shown in parentheses.

Table 6: Regressions of markup on price stickiness for alternative markup series

(a) Regressions of markups on implied price duration

	log(Markup)		
	4d Translog	4d cost shares	Baseline incl. SGA
Implied price duration	0.0248	0.0512	0.00714
	(0.0117)	(0.0144)	(0.00455)
Additional controls	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes
Observations	3785	3826	3813
Adjusted R^2	0.126	0.287	0.238

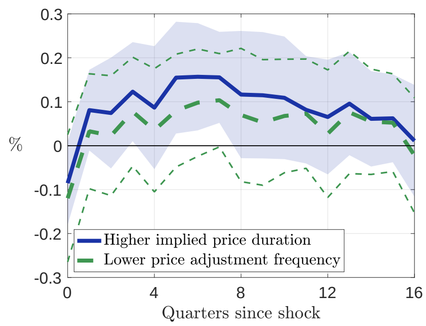
(b) Regressions of markups on price adjustment frequency

	log(Markup)		
	4d Translog	4d cost shares	Baseline incl. SGA
Price adjustment frequency	-0.182	-0.370	-0.0428
	(0.0924)	(0.0760)	(0.0363)
Additional controls	Yes	Yes	Yes
2-digit industry FE	Yes	Yes	Yes
Observations	3785	3826	3813
Adjusted R^2	0.128	0.292	0.237

Notes: Regression of firm-level markup (averaged over 2005–2011) on firm-level price adjustment frequency and implied price duration, respectively. For details on the different markup measures, see Section 2.4. Standard errors are clustered at the two-digit industry level and shown in parentheses.

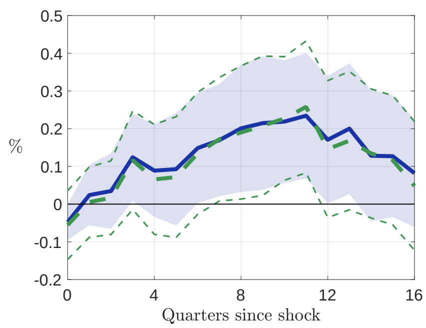
Figure 23: Relative markup and market share response of firms with stickier prices for alternative markup measures

(a) Accounting profits markups

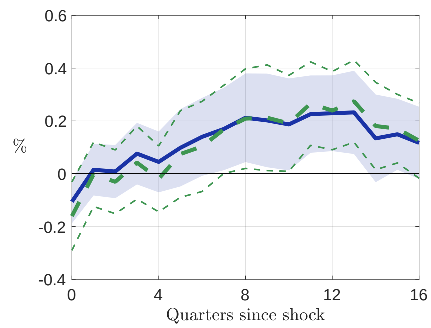


(b) User cost markups

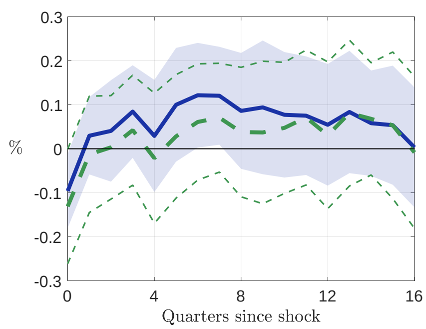
(c) 4d Translog



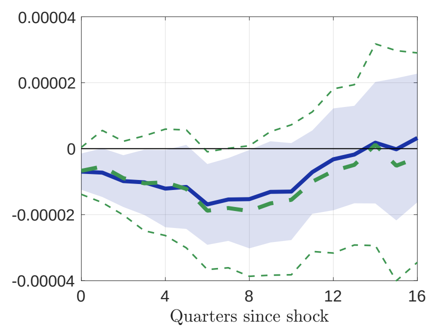
(d) 4d. Cost shares



(e) Baseline w. SGA



(f) Market shares



Notes: The figures show the response to a one standard deviation contractionary monetary policy shock of the (log) firm-level markup (or market share) of firms with a price adjustment frequency one standard deviation below mean (or with an implied price duration one standard deviation above mean) from panel local projections as in equation (3.5). Panels (a)–(e) use different markup measures and panel (f) uses market shares within two-digit-industry-quarters; see Section 2.4 for details. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

Table 7: Regressions of markup on price stickiness under alternative data treatments

(a) Keep small firms

	log(Markup)			
Implied price duration	0.0417 (0.0143)	0.0485 (0.0162)		
Price adjustment frequency			-0.301 (0.0681)	-0.352 (0.0810)
Additional controls	No	Yes	No	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	4395	4389	4395	4389
Adjusted R^2	0.167	0.195	0.169	0.197

(b) Keep firms with excessive growth

	log(Markup)			
Implied price duration	0.0522 (0.0161)	0.0473 (0.0142)		
Price adjustment frequency			-0.385 (0.0854)	-0.337 (0.0768)
Additional controls	No	Yes	No	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	4208	4160	4208	4160
Adjusted R^2	0.128	0.193	0.134	0.196

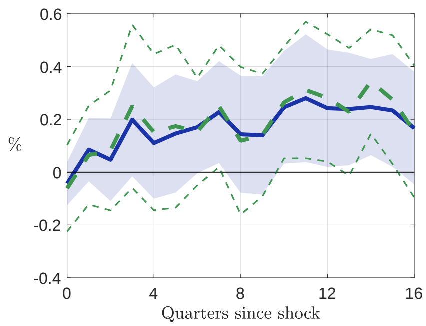
(c) Drop top/bottom 1% of markups

	log(Markup)			
Implied price duration	0.0537 (0.0180)	0.0472 (0.0155)		
Price adjustment frequency			-0.391 (0.0999)	-0.336 (0.0860)
Additional controls	No	Yes	No	Yes
2-digit industry FE	Yes	Yes	Yes	Yes
Observations	3857	3857	3857	3857
Adjusted R^2	0.145	0.228	0.151	0.231

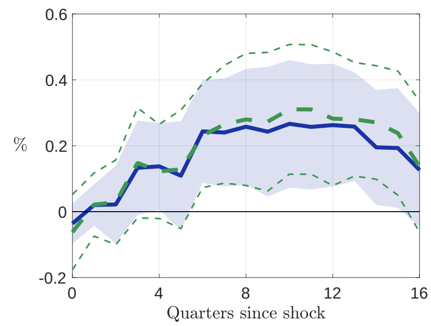
Notes: Regression of baseline firm-level markup (averaged over 2005–2011) on firm-level price adjustment frequency and implied price duration, respectively. See Section 2.4 for details on the different data treatments. Standard errors are clustered at the two-digit industry level and shown in parentheses.

Figure 24: Relative markup response of firms with stickier prices under alternative data treatments

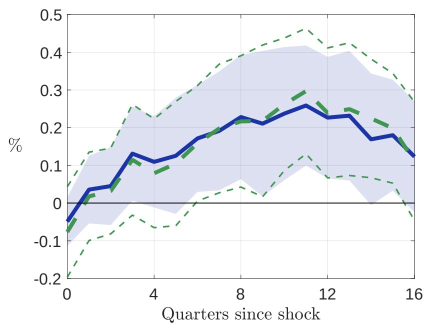
(a) Keep small firms



(b) Keep firms with excessive growth

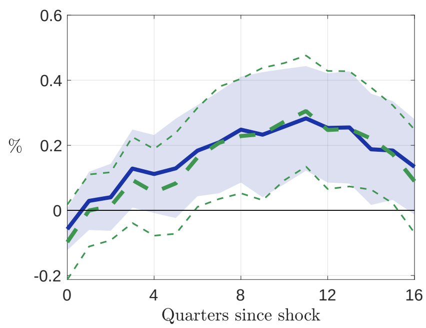


(c) Drop top/bottom 1% of markups

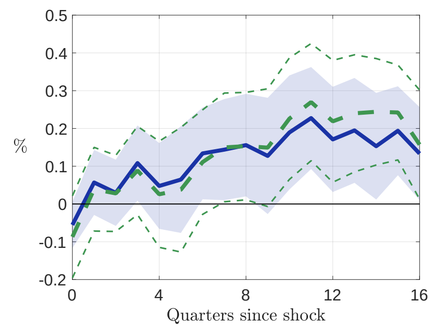


(d) At least 16 quarters

(d) Pre-Great Recession only



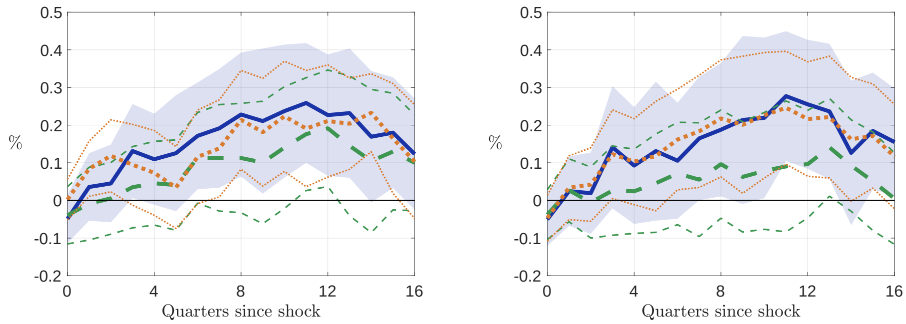
(e) Including Great Recession



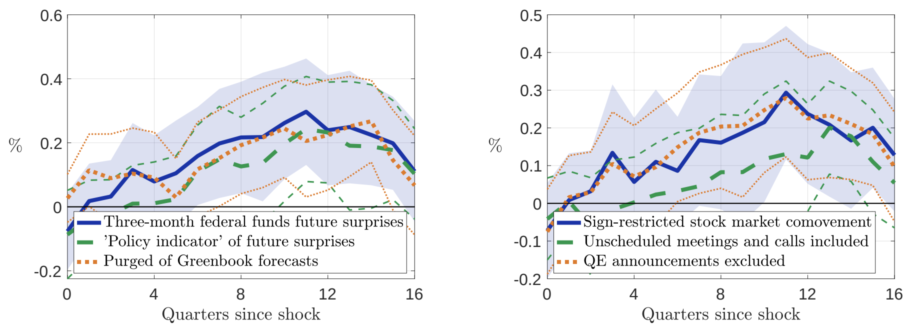
Notes: The figures show the response to a one standard deviation contractionary monetary policy shock of the firm-level markup of firms with a price adjustment frequency one standard deviation below mean (or with an implied price duration one standard deviation above mean) from panel local projections as in equation (3.5). The regressions include interactions with lagged log assets, leverage, and liquidity and their interactions with the monetary policy shock. See Section 2.4 for details on the different data treatments. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

Figure 25: Relative markup response of firms with stickier prices for alternative monetary policy shocks

(a) Higher implied price duration



(b) Lower price adjustment frequency



Notes: The figures show the response to a one standard deviation contractionary monetary policy shock of the firm-level markup of firms with a price adjustment frequency one standard deviation below mean (or with an implied price duration one standard deviation above mean) from panel local projections as in equation (3.5). The regressions include interactions with lagged log assets, leverage, and liquidity and their interactions with the monetary policy shock. See Section 2.4 for details on the different monetary policy shocks. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

E Proofs

E.1 Markup dispersion and aggregate TFP

Consider a continuum of monopolistically competitive firms that produce variety goods Y_{it}. Firms employ a common constant-returns-to-scale production function $F(\cdot)$ that transforms a vector of inputs L_{it} into output subject to firm-specific productivity shocks $Y_{it} = A_{it}F(L_{it})$. The cost minimization problem yields that firm-specific $X_{it} = X_t/A_{it}$, where X_t denotes a common marginal costs term. Aggregate GDP is the output of a final good producer, which aggregates variety goods using a Dixit–Stiglitz aggregator $Y_t = (\int Y_{it}^{(\eta-1)/\eta} di)^{\eta/(\eta-1)}$. The cost minimization problem of the final good producer yields a demand curve for variety goods $Y_{it} = (P_{it}/P_t)^{-\eta} Y_t$, where P_t is an aggregate price index. Variety good producers choose prices to maximize period profits

$$\max_{P_{it}} \left(\tau_{it} P_{it} - X_{it}\right) Y_{it} \qquad \text{s.t.} \quad Y_{it} = (P_{it}/P_t)^{-\eta} Y_t,$$

where τ_{it} is a *markup wedge* in the spirit of Hsieh and Klenow (2009) and Baqaee and Farhi (2020). This wedge may be viewed as a shortcut for price rigidities. Profit maximization yields a markup $\mu_{it} = P_{it}/X_{it} = \frac{1}{\tau_{it}} \frac{\eta}{\eta-1}$. We compute aggregate TFP as a Solow residual by

$$\log \mathrm{TFP}_t = \log \left(\int Y_{it}^{(\eta-1)/\eta} di \right)^{\eta/(\eta-1)} - \log \int \frac{Y_{it}}{A_{it}} di.$$

This Solow residual has a model consistent Solow weight of one for the aggregate cost term. If we (a) apply a second-order approximation to $\log \mathrm{TFP}_t$ in $\log A_{it}$ and $\log \tau_{it}$, or if we (b) assume that A_{it} and τ_{it} are jointly log-normally distributed, we obtain

$$\log \mathrm{TFP}_t = -\frac{\eta}{2} \mathbb{V}_t(\log \mu_{it}) + \mathbb{E}_t(\log A_{it}) + \frac{\eta - 1}{2} \mathbb{V}_t(\log A_{it}).$$

Wedges τ_{it} drive markup dispersion and distort the economy away from allocative efficiency. Firms with high τ_{it} charge lower markups and use more inputs than socially optimal, and vice versa for low τ_{it}. This misallocation across firms results in lower aggregate TFP.

E.2 Proof of Proposition 1

Denote by $\mathbb{V}_t(\cdot)$, $\mathrm{Cov}_t(\cdot)$, $\mathrm{Corr}_t(\cdot)$ respectively the cross-sectional variance, covariance, correlation operator. The cross-sectional variance of the log markup is

$$\mathbb{V}_t(\log \mu_{it}) = \int (\log P_{it} - \log P_t - \log X_t)^2 di - \left[\int (\log P_{it} - \log P_t - \log X_t) di \right]^2.$$

The derivative w.r.t. $\log X_t$ is

$$\frac{\partial \mathbb{V}_t(\log \mu_{it})}{\partial \log X_t} = 2\int \log(\mu_{it})\rho_{it}di - 2\int \log(\mu_{it})di \int \rho_{it}di = 2\mathrm{Cov}_t(\rho_{it}, \log \mu_{it}).$$

Hence, the markup variance falls in $\log X_t$ if $\mathrm{Corr}_t(\rho_{it}, \log \mu_{it}) < 0$. $\square$

E.3 Proof of Proposition 2

We assume that

$$\log \begin{pmatrix} P_t/\bar{P} \\ X_t/\bar{X} \\ Y_t/\bar{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} -\frac{\sigma_p^2}{2} \\ -\frac{\sigma_x^2}{2} \\ -\frac{\sigma_y^2}{2} \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & & \\ \sigma_{px} & \sigma_x^2 & \\ \sigma_{py} & \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right).$$

Define $\tilde{\theta}_i \equiv \frac{\beta\theta_i}{1-\beta\theta_i}$, as well as

$$C_{it} \equiv \mathbb{E}_t \left[\frac{X_{t+1}}{X_t} \left(\frac{P_{t+1}}{P_t} \right)^{\eta} \frac{Y_{t+1}}{Y_t} \right],$$
$$D_{it} \equiv \mathbb{E}_t \left[\left(\frac{P_{t+1}}{P_t} \right)^{\eta-1} \frac{Y_{t+1}}{Y_t} \right],$$
$$\Psi_{it} \equiv \frac{1+\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}},$$

which allows us to rewrite the first-order condition in (3.3) as

$$P_{it}^* = \frac{\eta}{\eta-1} P_t X_t \Psi_{it}.$$

The terms C_{it} and D_{it} can be simplified

$$C_{it} = \frac{\bar{X}\bar{P}^{\eta}\bar{Y}}{X_t P_t^{\eta} Y_t} \exp\left\{ \eta(\eta-1)\frac{\sigma_p^2}{2} + \eta\sigma_{px} + \eta\sigma_{py} + \sigma_{xy} \right\},$$
$$D_{it} = \frac{\bar{P}^{\eta-1}\bar{Y}}{P_t^{\eta-1} Y_t} \exp\left\{ (\eta-1)(\eta-2)\frac{\sigma_p^2}{2} + (\eta-1)\sigma_{py} \right\}.$$

Since $\tilde{\theta}_i \in (0,1)$, we obtain $\Psi_{it} > 1$ when $P_t = \bar{P}$ and $X_t = \bar{X}$, if

$$(\eta-1)\sigma_p^2 + \sigma_{py} + \eta\sigma_{px} + \sigma_{xy} > 0.$$

Under this condition, we obtain $\mu_{it}^* > \frac{\eta}{\eta-1}$. Under the same condition, we further obtain

$$\frac{\partial \Psi_{it}}{\partial \tilde{\theta}_i} = \frac{C_{it} - D_{it}}{(1+\tilde{\theta}_i D_{it})^2} > 0, \quad \text{and hence} \quad \frac{\partial \Psi_{it}}{\partial \theta_i} > 0.$$

We next study the pass-through of a transitory or permanent change in X_t. Consider first a *transitory* change in X_t away from $\bar{X}$. The expected pass-through is

$$\bar{\rho}_{it} = (1-\theta_i)\frac{\partial \log P_{it}}{\partial \log X_t} = (1-\theta_i)\left(1+\Phi_{it}\right), \quad \text{where} \;\; \Phi_{it} = \frac{\partial \log \Psi_{it}}{\partial \log X_t}$$

and

$$\Phi_{it} = \frac{\tilde{\theta}_i \frac{\partial C_{it}}{\partial \log X_t}(1+\tilde{\theta}_i D_{it}) - (1+\tilde{\theta}_i C_{it})\tilde{\theta}_i \frac{\partial D_{it}}{\partial \log X_t}}{(1+\tilde{\theta}_i D_{it})^2}\Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}}\Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i C_{it}} < 0.$$

Hence pass-through becomes

$$\bar{\rho}_{it} = \frac{1-\theta_i}{1+\tilde{\theta}_i C_{it}} \in (0,1).$$

In addition, the pass-through falls in θ_i,

$$\frac{\partial \bar{\rho}_{it}}{\partial \theta_i} = -(1+\Phi_{it}) + (1-\theta_i)\frac{\partial \Phi_{it}}{\partial \theta_i} < 0.$$

We next examine a *permanent* change in X_t, which is a change in $\bar{X}$ (starting in period t). At $P_t = \bar{P}$ and $X_t = \bar{X}$,

$$\frac{\partial \log P_{it}^*}{\partial \log \bar{X}} = 1.$$

Expected pass-through is then $\bar{\rho}_{it} = 1-\theta_i$ and hence $\frac{\partial \bar{\rho}_{it}}{\partial \theta_i} < 0$. □

E.4 Proof of Proposition 3

Let us first define

$$C_{it} = \left(\frac{P_{it}}{P_{i,t-1}} - 1\right)\frac{P_{it}}{P_{i,t-1}},$$
$$D_{it} = \mathbb{E}_t\left[\left(\frac{P_{i,t+1}}{P_{it}} - 1\right)\frac{P_{i,t+1}}{P_{it}}\right],$$

such that we can re-write the first-order condition in equation (3.4) more compactly as

$$(1-\eta)\left(\frac{P_{it}}{P_t}\right)^{1-\eta} Y_t + \eta X_t \left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t = \phi_i(C_{it} - D_{it}).$$

Further define $\bar{\phi}_i = 0$ and denote by an upper bar any object that is evaluated at $\bar{\phi}_i$, such as the price P_{it}, which is $\bar{P}_{it} = \frac{\eta}{\eta-1} P_t X_t$. In addition,

$$\bar{C}_{it} = \left(\frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} - 1 \right) \frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} = (\Pi_{pt}\Pi_{xt})^2 - \Pi_{pt}\Pi_{xt},$$
$$\bar{D}_{it} = E_t \left[\left(\frac{\bar{P}_{i,t+1}}{\bar{P}_{it}} - 1 \right) \frac{\bar{P}_{i,t+1}}{\bar{P}_{it}} \right] = \frac{\exp\left\{ \frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px} \right\}}{(\Pi_{pt}\Pi_{xt})^2} - \frac{\exp\{\sigma_{pw}\}}{\Pi_{pt}\Pi_{xt}}.$$

We next use a first-order approximation of the first-order condition at $\bar{\phi}_i$ and with respect to ϕ_i and $\log P_{it}$. Denoting $\mathrm{dlog}P_{it} = \log P_{it} - \log \bar{P}_{it}$ and $\mathrm{d}\phi_i = \phi_i$, we obtain

$$(1-\eta)^2 \left(\frac{P_{it}}{\bar{P}_t} \right)^{1-\eta} Y_t \mathrm{dlog}P_{it} - \eta^2 X_t \left(\frac{\bar{P}_{it}}{P_t} \right)^{-\eta} Y_t \mathrm{dlog}P_{it} = (\bar{C}_{it} - \bar{D}_{it}) \mathrm{d}\phi_i.$$

This yields

$$\Psi_{it} \equiv \frac{\mathrm{dlog}P_{it}}{\mathrm{d}\phi_i} = \frac{\bar{D}_{it} - \bar{C}_{it}}{(\eta-1)^\eta \eta^{1-\eta} X_t^{1-\eta} Y_t},$$

and hence $\log P_{it} \approx \log \bar{P}_{it} + \Psi_{it} d\phi_i$. For $\phi_i > 0$, the markup is above the frictionless one if $P_{it} > \bar{P}_{it}$, which holds if $\Psi_{it} > 0$. For $P_t = \bar{P}$ and $X_t = \bar{X}$, $\Psi_{it} > 0$ if

$$\sigma_p^2 + \sigma_x^2 + 2\sigma_{px} > 0,$$

for which a sufficient condition is that the correlation

$$\rho_{px} \equiv \frac{\sigma_{px}}{\sigma_p \sigma_x} > -1.$$

Under the same condition, $\frac{\partial P_{it}}{\partial \phi_i} > 0$.

We next study the pass-through of a transitory or permanent change in X_t. The pass-through is

$$\rho_{it} = 1 + \frac{\partial \Psi_i}{\partial \log X_t} d\phi_i.$$

We next examine the conditions under which pass-through falls in ϕ_i, i.e., conditions under which

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0,$$

which is equivalent to examining the conditions for

$$\frac{\partial \bar{D}_{it}}{\partial \log X_t} - \frac{\partial \bar{C}_{it}}{\partial \log X_t} + (\eta - 1)(\bar{D}_{it} - \bar{C}_{it}) < 0.$$

Consider first a *transitory* change in X_t away from $\bar{X}$,

$$\frac{\partial \bar{C}_{it}}{\partial \log X_t} = 2(\Pi_{pt}\Pi_{xt})^2 - \Pi_{pt}\Pi_{xt},$$
$$\frac{\partial \bar{D}_{it}}{\partial \log X_t} = -2(\Pi_{pt}\Pi_{xt})^{-2}\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} + (\Pi_{pt}\Pi_{xt})^{-1}\exp\left\{\sigma_{px}\right\}.$$

For $P_t = \bar{P}$ and $X_t = \bar{X}$, we obtain

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0 \qquad \text{if} \quad \eta < \tilde{\eta}^{\text{transitory}} = 2 + \frac{1 + \exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\}}{\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} - \exp\left\{\sigma_{px}\right\}}$$

We next consider a *permanent* change, for which we have

$$\frac{\partial \bar{C}_{it}}{\partial \log X_t} = 2(\Pi_{pt}\Pi_{wt})^2 - \Pi_{pt}\Pi_{wt}, \quad \frac{\partial \bar{D}_{it}}{\partial \log X_t} = 0.$$

For $P_t = \bar{P}$ and $X_t = \bar{X}$, we obtain

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0 \qquad \text{if} \quad \eta < \tilde{\eta}^{\text{permanent}} = 1 + \frac{1}{\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} - \exp\left\{\sigma_{px}\right\}}$$

It always holds that $\eta^{\text{permanent}} < \eta^{\text{transitory}}$ and we define $\tilde{\eta} \equiv \eta^{\text{permanent}}$. □

F Menu cost model

To study the presence of precautionary price setting in menu cost models, we proceed numerically. Consider the partial equilibrium menu cost model

$$V(p, Z) = \mathbb{E}_\xi[\max\{V^A(Z) - \xi, V^N(Z)\}]$$
$$V^A(Z) = \max_{p^*}\left\{\left(\frac{p^*}{P} - X\right)\left(\frac{p^*}{P}\right)^{-\eta} + \beta\mathbb{E}_Z\left[V(p^*, Z')\right]\right\}$$
$$V^N(p, Z) = \left(\frac{p}{P} - X\right)\left(\frac{p}{P}\right)^{-\eta} + \beta\mathbb{E}_Z\left[V(p, Z')\right]$$

where p is the price a firm sets and Z denote a vector of the aggregate state variables price level (P), aggregate demand (Y), and marginal costs (X). The firm chooses to adjust prices in the presence of the menu cost ξ.

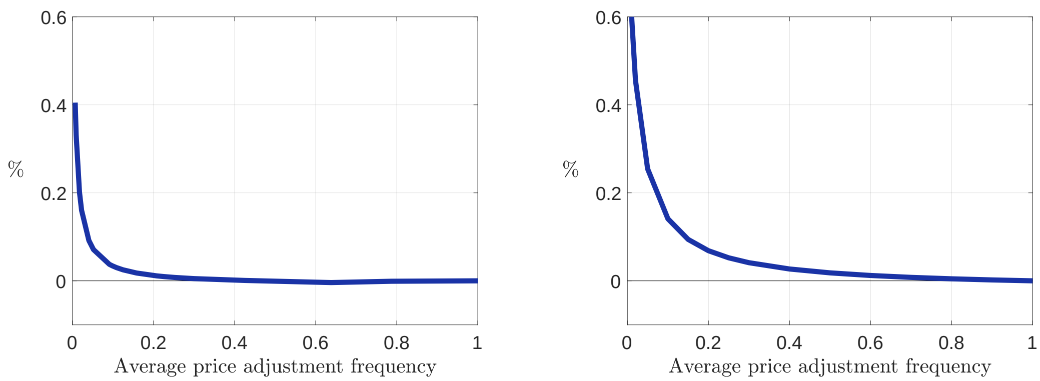
We set $\eta = 6$ and $\beta = 1.03^{-1/4}$. We solve the model using value function iteration with off-grid interpolation with respect to p using cubic splines as basis function. To solve accurately for differences in p^* that arise from small differences in ξ requires a fine grid for both p and Z. To alleviate the numerical challenge, we assume ξ is stochastic and drawn from an iid exponential distribution, parametrized by $\bar{\xi}$. Results change only little when using a uniform distribution.

We assume 200 grid points on a log-spaced grid for p. To capture aggregate uncertainty in Z, we first estimate a first-order Markov process for Z in the data and then discretize it using a Tauchen procedure. In the univariate case, when only allowing for inflation uncertainty, the precautionary price setting was accurately captured starting from about 49 grid points for Z. Discretizing a three-variate VAR with 49 grid points for each variable is costly. Even more importantly, the state space, on which to solve the model, becomes very large. We therefore proceed with the univariate case. We estimate an AR(1) on quarterly post-1984 data of the log CPI and apply the Tauchen method with 49 grid points.

Figure 26: Precautionary price setting under menu costs and Calvo

(a) Menu cost

(b) Calvo



Notes: The figures show percentage difference between the dynamic optimal price relative to the frictionless optimal one.

We solve the stationary equilibrium of the menu cost and Calvo model for a vector of different $\bar{\xi}$, which imply different equilibrium price adjustment frequencies. Figure 26 plots the price setting policy p^* at the unconditional mean of Z for different average price adjustment frequencies. We compare menu costs in panel (a) with Calvo in panel (b). The figures shows that precautionary price setting exists and is amplified by the degree of price-setting friction in a menu cost environment. Compared to Calvo, menu costs generate somewhat muted precautionary price setting.

G Details on the Quantitative New Keynesian Model

G.1 Model

We assume a representative infinitely-lived household who maximizes

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} - \frac{N_t^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} \right),$$

subject to the budget constraints $P_t C_t + R_t^{-1} B_t \leq B_{t-1} + W_t N_t + D_t$ for all t, where C_t is aggregate consumption, P_t an aggregate price index, B_t denotes one-period discount bounds purchased at price R_t^{-1}, N_t employment, W_t the nominal wage, and D_t aggregate dividends. We impose the solvency constraint $\lim\limits_{T\to\infty} \mathbb{E}_t[\Lambda_{t,T} \frac{B_T}{P_T}] \geq 0$ for all t, where $\Lambda_{t,T} = \beta^{T-t}(C_T/C_t)^{-\frac{1}{\gamma}}$ is the stochastic discount factor. The final output good Y_t is produced with a Dixit–Stiglitz aggregator

$$Y_t = \left(\int_0^1 Y_{it}^{\frac{\eta-1}{\eta}} \mathrm{d}i \right)^{\frac{\eta}{\eta-1}},$$

where η is the elasticity of substitution between differentiated goods $\{Y_{it}\}$. Each intermediate good i is produced by a monopolistically competitive intermediate good firm i. The unit measure of differentiated goods is split equally across K different types of intermediate goods firms producing the differentiated goods. The firm types are indexed by $k = 1, \ldots, K$ and firms are ordered according to their type such that firms indexed $i \in [0, 1/K)$ belong to type $k = 1$ and firms indexed $i \in ((k-1)/K, k/K]$ belong to type $k = 2, \ldots, K$. Firms across the K types are ex-ante identical except for differences in their exogenous price reset probability $1 - \theta_k$. Intermediate goods are produced with technology $Y_{it} = A_t N_{it}$, where A_t is a common technology shifter, which follows $\log A_t = \rho_a \log A_{t-1} + \varepsilon_{a,t}$ and $\varepsilon_{a,t} \sim \mathcal{N}(0, \sigma_a^2)$ are technology shocks. Real marginal costs are hence $mc_t = w_t/A_t$. Final good aggregation implies an isoelastic demand schedule for intermediate goods given by $Y_{it} = (P_{it}/P_t)^{-\eta} Y_t$, where P_{it} is the firm-level price and P_t the aggregate price index

$$P_t = \left[\int_0^1 P_{it}^{1-\eta} di \right]^{\frac{1}{1-\eta}} = \left[\frac{1}{K} \sum_{k=1}^{K} P_{kt}^{1-\eta} \right]^{\frac{1}{1-\eta}}.$$

P_{kt} denotes the firm type-k specific price index

$$P_{kt} = \left[\int_{(k-1)/K}^{k/K} P_{it}^{1-\eta} di \right]^{\frac{1}{1-\eta}} = \left[(1-\theta_k) \tilde{P}_{kt}^{1-\eta} + \theta_k P_{kt-1}^{1-\eta} \right]^{\frac{1}{1-\eta}},$$

where $\tilde{P}_{kt}$ is the optimal reset price, which solves maximizes the value of the firm to its shareholder

$$\max_{P_{it}} \sum_{j=0}^{\infty} \theta_i^j \mathbb{E}_t \left[\mu_{t,t+j} \left(\frac{P_{it}}{P_{t+j}} - mc_{t+j} \right) \left(\frac{P_{it}}{P_{t+j}} \right)^{-\eta} Y_{t+j} \right],$$

where μ_t is the marginal utility of consumption $\mu_t = C_t^{-1/\gamma}$. The monetary authority follows a Taylor rule to stabilize inflation, $\Pi_t = P_t/P_{t-1}$, and fluctuations in output, Y_t, around its natural level, denoted $\tilde{Y}_t$, subject to monetary policy shocks ν_t,

$$R_t = R_{t-1}^{\rho_r} \left[\frac{1}{\beta} (\Pi_t)^{\phi_\pi} \left(\frac{Y_t}{\tilde{Y}_t} \right)^{\phi_y} \right]^{1-\rho_r} \nu_t, \quad \log \nu_t \sim \mathcal{N}(0, \sigma_\nu^2).$$

G.2 Equilibrium conditions

$$\frac{\tilde{P}_{kt}}{P_t} = \frac{\eta}{\eta - 1} \frac{\mathbb{E}_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta} mc_{t+j}}{\mathbb{E}_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta-1}} \quad \forall k = 1, \dots, K \qquad \text{(reset price)}$$

$$1 = \frac{1}{K} \sum_{k=1}^{K} \left(\frac{P_{kt}}{P_t} \right)^{1-\eta} \qquad \text{(aggregate price index)}$$

$$\frac{P_{kt}}{P_t} = \left[(1 - \theta_k) \left(\frac{\tilde{P}_{kt}}{P_t} \right)^{1-\eta} + \theta_k \Pi_t^{\eta-1} \left(\frac{P_{kt-1}}{P_{t-1}} \right)^{1-\eta} \right]^{\frac{1}{1-\eta}} \quad \forall k = 1, \dots, K \quad \text{(type } k \text{ price index)}$$

$$S_t = \frac{1}{K} \sum_{k=1}^{K} S_{kt} \qquad \text{(aggregate price dispersion)}$$

$$S_{kt} = (1 - \theta_k) \left(\frac{\tilde{P}_{kt}}{P_t} \right)^{-\eta} + \theta_k \Pi_t^{\eta} S_{kt-1} \quad \forall k = 1, \dots, K \qquad \text{(type } k \text{ price dispersion)}$$

$$Y_t = \frac{A_t}{S_t} N_t \qquad \text{(aggregate output)}$$

$$TFP_t = \frac{Y_t}{N_t} \qquad \text{(TFP)}$$

$$mc_t = \frac{w_t}{A_t} \qquad \text{(marginal cost)}$$

$$N_t^{\frac{1}{\varphi}} C_t^{\frac{1}{\gamma}} = w_t \qquad \text{(intratemporal optimality)}$$

$$C_t^{-\frac{1}{\gamma}} = \beta \mathbb{E}_t \left[\frac{R_t}{\Pi_{t+1}} C_{t+1}^{-\frac{1}{\gamma}} \right] \qquad \text{(intertemporal optimality)}$$

$$C_t = Y_t \qquad \text{(resource constraint)}$$

$$R_t = R_{t-1}^{\rho_r} \left[\frac{1}{\beta} (\Pi_t)^{\phi_\pi} \left(\frac{Y_t}{\tilde{Y}_t} \right)^{\phi_y} \right]^{1-\rho_r} \nu_t \qquad \text{(Taylor rule)}$$

H Additional model responses

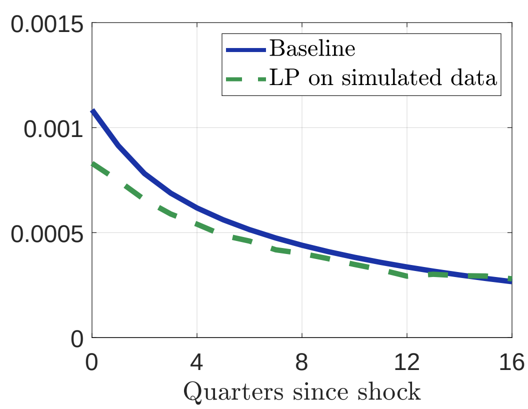
H.1 Stochastic simulation of the model

In our baseline model, contractionary monetary policy shocks raise markup dispersion and expansionary shocks lower markup dispersion. This response of markup dispersion critically depends on solving the model around the stochastic steady state, which allows us to capture precautionary price setting. However, even when capturing precautionary price setting, contractionary monetary policy shocks do not increase markup dispersion in all states of the world outside a local neighborhood around the stochastic steady state. In particular, after sufficiently large expansionary monetary policy shocks, the average markup of stickier firms may fall below the average markup of more flexible firms. At this point, a contractionary monetary policy shock may lower markup dispersion.

We investigate this possibility through a large stochastic simulation of our model. We simulate 50,000 firms for 10,000 periods and find that the average markup of the stickiest quintile of firms is below the average markup of the most flexible quintile of firms in 11.1% of the periods.

We further investigate by how much this occasional inverted order of markups affects the average response of markup dispersion to monetary policy shocks. We compute a time series of markup dispersion and project it on the simulated monetary policy shock series using our empirical framework in equation (2.4). The estimated average response of markup dispersion is similar but a bit smaller than the baseline response, see Figure 27.

Figure 27: Response of markup dispersion



Notes: This figure shows impulse responses to a one standard deviation monetary policy shock. *Baseline* replicates the response of markup dispersion implied by the model solution and as shown in Figure 6. The dashed line is the response of markup dispersion when applying our empirical setup in equation (2.4) to a large stochastic simulation of the model. We simulate the markups of 50,000 firms, equally distributed across the five groups of price stickiness, over 10,000 periods.

H.2 Policy counterfactual

In the following, we provide some details on a counterfactual experiment, in which the monetary authority in the model (mis-)perceives the aggregate TFP response to monetary policy shocks as originating from technology shocks.

The natural level of output in the absence of price setting frictions is

$$\tilde{Y}_t = \left[\frac{\eta}{\eta-1} A_t^{1+\frac{1}{\varphi}}\right]^{\frac{1}{\frac{1}{\varphi}+\frac{1}{\gamma}}} .$$

In the counterfactual, we assume the monetary authority mis-attributes observed fluctuations in TFP to exogenous productivity shocks A_t, which we implement by using the following counterfactual (cf) natural output definition

$$\tilde{Y}_t^{\mathrm{cf}} = \left[\frac{\eta}{\eta-1} TFP_t^{1+\frac{1}{\varphi}}\right]^{\frac{1}{\frac{1}{\varphi}+\frac{1}{\gamma}}} .$$

Since monetary policy shocks lower TFP_t, counterfactual natural output $\tilde{Y}_t^{\mathrm{cf}}$ falls, while natural output $\tilde{Y}_t$ remains constant.

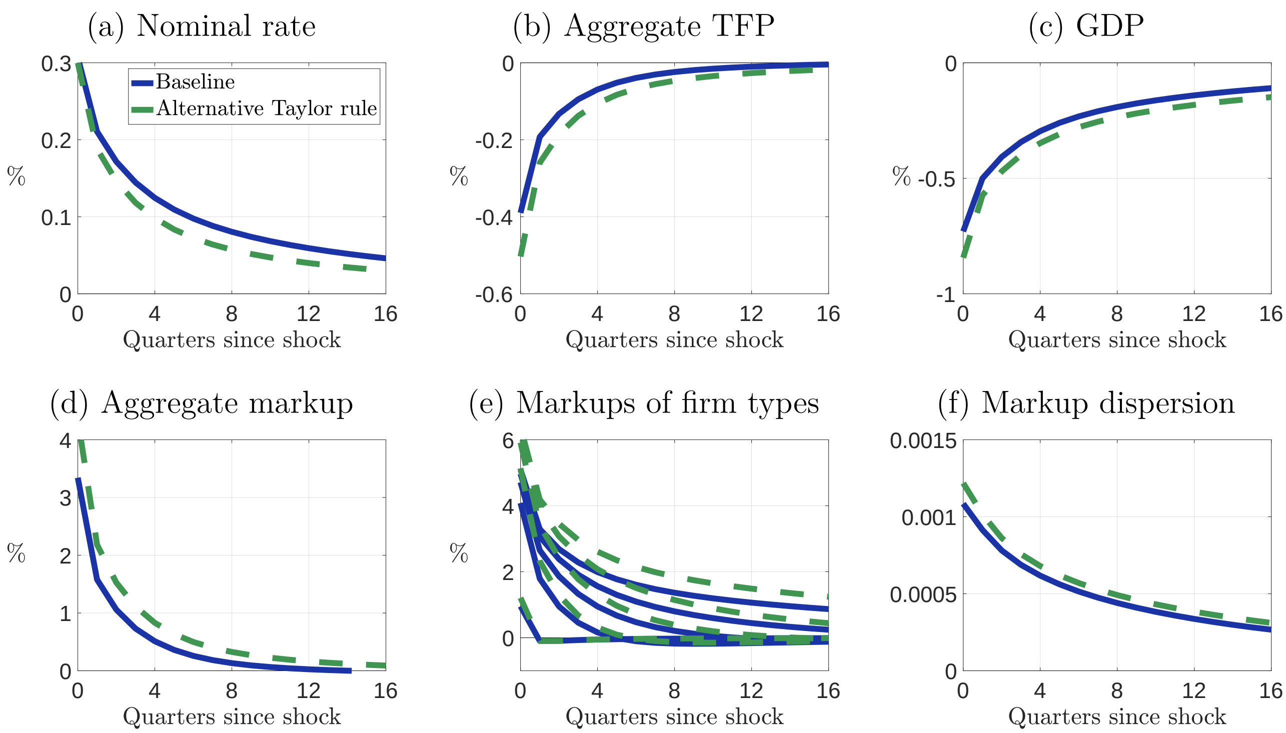
Let us define the systematic component of monetary policy in the baseline and counterfactual Taylor rule as

$$\bar{R}_t = \frac{1}{\beta}\left(\Pi_t\right)^{\phi_\pi}\left(\frac{Y_t}{\tilde{Y}_t}\right)^{\phi_y}, \qquad \text{and} \quad \bar{R}_t^{\mathrm{cf}} = \frac{1}{\beta}\left(\Pi_t\right)^{\phi_\pi}\left(\frac{Y_t}{\tilde{Y}_t^{\mathrm{cf}}}\right)^{\phi_y} .$$

The systematic component of monetary policy sets a lower nominal interest rate in response to lower inflation and output gaps. In the counterfactual, the responsiveness of the systematic component to lower output gaps is dampened because $\tilde{Y}_t^{\mathrm{cf}}$ falls as well. The counterfactual Taylor rule is hence similar to a Taylor rule with a smaller coefficient ϕ_y. This may lead to large output and inflation responses to monetary policy shocks. In addition, because the response of Y_t to a shock converges more quickly to the response of $\tilde{Y}_t^{\mathrm{cf}}$ than to $\tilde{Y}_t$, we can think of the implicit ϕ_y in the counterfactual as falling in the forecast horizon. This explains the more persistent effects in the counterfactual.

To quantify the implications of counterfactual natural output, we compare the macroeconomic effects of monetary policy shocks that raise the nominal interest rate by 30 bp in the baseline and counterfactual model. We keep all model parameters, including the variance of monetary policy shock unchanged in the counterfactual, but scale up the size of the shock in the counterfactual such that the nominal interest rate increases by 30bp. The required size of the monetary policy shock in the counterfactual corresponds to $1.147 \cdot \sigma_\nu$. Figure 28 in Appendix H compares the responses to a monetary policy shock in the baseline model and the counterfactual exercise. Both GDP and markup dispersion respond by more on impact and more persistently in the counterfactual.

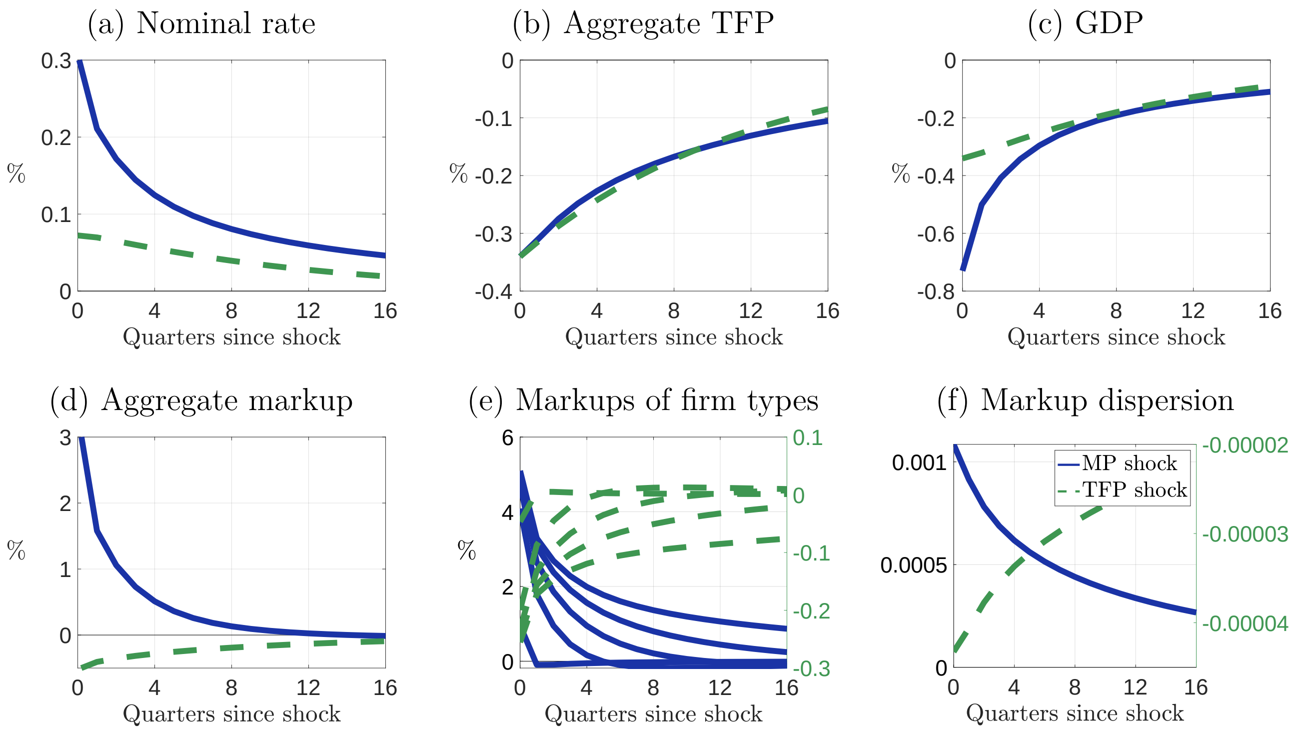
Figure 28: Model responses to monetary policy shocks under alternative Taylor rule



Notes: This figure shows impulse responses to a one standard deviation monetary policy shock. *Baseline* corresponds to the model in the main text. In particular, the monetary authority follows a Taylor rule, which reacts to fluctuation in the output gap. The gap is defined relative to natural output (the level prevailing under flexible prices), which is unchanged after monetary policy shocks. *Alternative Taylor rule* describes a policy counterfactual in which the monetary authority computes natural output as if the responses of aggregate TFP to monetary policy shocks were driven by technology shocks.

H.3 Technology shocks

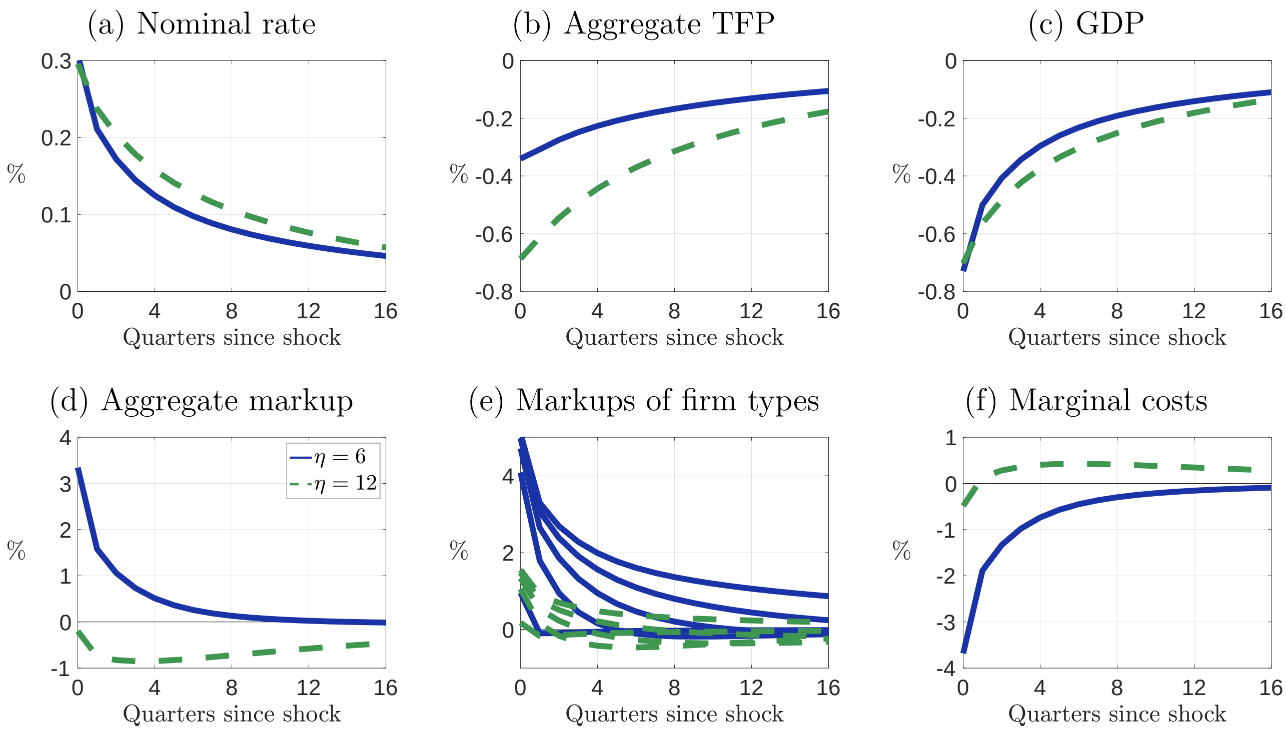
Figure 29: Model responses to technology shocks



Notes: This figure compares the impulse responses to a one standard deviation monetary policy shock to those to a technology shock. The persistence of TFP and the technology shock size are calibrated to match the shape of the TFP response to monetary policy shocks.

H.4 Elasticity of substitution

Figure 30: Model responses to monetary policy shocks when varying the elasticity of substitution

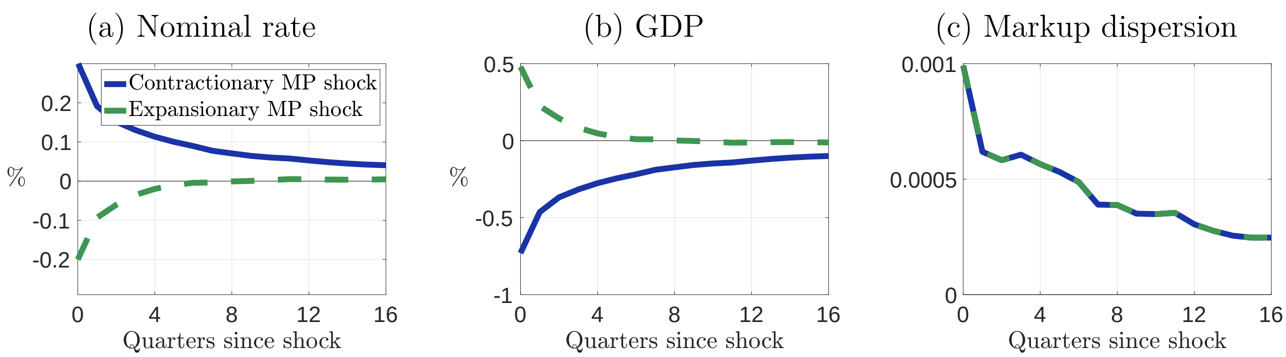


Notes: This figure shows impulse responses to a one standard deviation monetary policy shock for two values of the elasticity of substitution between variety goods η. The value 6 corresponds to our baseline calibration and the value 12 corresponds to an intermediate value of elasticities considered in the literature (e.g., Fernandez-Villaverde et al., 2015). The standard deviation of monetary policy shocks is re-calibrated to match the response of the nominal rate of 30bp.

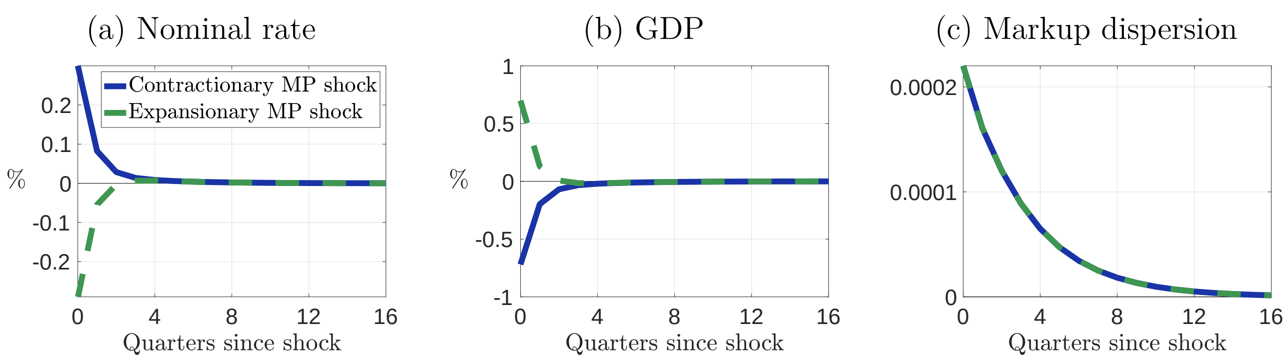
H.5 Solving the model around the deterministic steady state

Figure 31: Model responses to monetary policy shocks around the deterministic steady state

(i) Heterogeneous price rigidity



(ii) Homogeneous price rigidity



Notes: This figure shows responses to a one standard deviation monetary policy shock, when solving the model through a second-order approximation around the deterministic steady state.

I Robustness to model variations

I.1 Real rigidities (firm-specific labor)

In this section, we consider a variant of the baseline model, in which labor is firm-specific. We find that the presence of such real rigidity amplifies the TFP effects of monetary policy.

In particular, we integrate the assumptions on differentiated labor supply from Woodford (2003) into our model of heterogeneous price rigidity. Households supply differentiated labor inputs N_{it} specific to each differentiated goods producer with preferences

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} - \int_0^K \frac{N_{it}^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}} di \right).$$

Wages may differ across the differentiated labor inputs and we assume that the producers of differentiated goods take wages as given.

Compared to the model with homogeneous labor supply, described by its equilibrium conditions in Appendix G.2, the main difference is the marginal cost, which is now firm-specific. For a firm of type k, which adjusts its price in period t, period $t+j$ marginal costs (conditional on no price re-adjustment after period t) are given by

$$mc_{k,t+j|t} = \frac{w_{k,t+j|t}}{A_{t+j}}.$$

Using the intratemporal optimality condition, production technology, and demand curve, we obtain

$$mc_{k,t+j|t} = \left(\left(\frac{\tilde{P}_{kt}}{P_t} \frac{P_t}{P_{t+j}} \right)^{-\eta} \frac{Y_{t+j}}{A_{t+j}} \right)^{\frac{1}{\varphi}} \frac{C_{t+j}^{\frac{1}{\gamma}}}{A_{t+j}} = \left(\frac{\tilde{P}_{kt}}{P_t} \right)^{-\frac{\eta}{\varphi}} \Pi_{t,t+j}^{\frac{\eta}{\varphi}} \tilde{mc}_{t+j}.$$

where $\tilde{mc}_{t+j}$ collects aggregate variables in period $t+j$. The optimal reset price of a type k firm is

$$\frac{\tilde{P}_{kt}}{P_t} = \frac{\eta}{\eta-1} \frac{E_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta} mc_{kt+j|t}}{E_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta-1}},$$

which we can rewrite, using the above expression for $mc_{kt+j|t}$, as

$$\left(\frac{\tilde{P}_{kt}}{P_t} \right)^{1+\frac{\eta}{\varphi}} = \frac{\eta}{\eta-1} \frac{E_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta+\frac{\eta}{\varphi}} \tilde{mc}_{t+j}}{E_t \sum_{j=0}^{\infty} \theta_k^j \beta^j \mu_{t+j} Y_{t+j} \Pi_{t,t+j}^{\eta-1}}.$$

This reset price may feature precautionary price setting. In fact, when applying the partial equilibrium framework used in Proposition 2, one can show that the reset price increases in θ_k if

$$\left[\tilde{\eta}(\tilde{\eta}-1) - (\eta-1)(\eta-2)\right] \sigma_p^2 + \left[\tilde{\eta} - (\eta-1)\right] \sigma_{py} + \tilde{\eta}\sigma_{px} + \sigma_{xy} > 0,$$

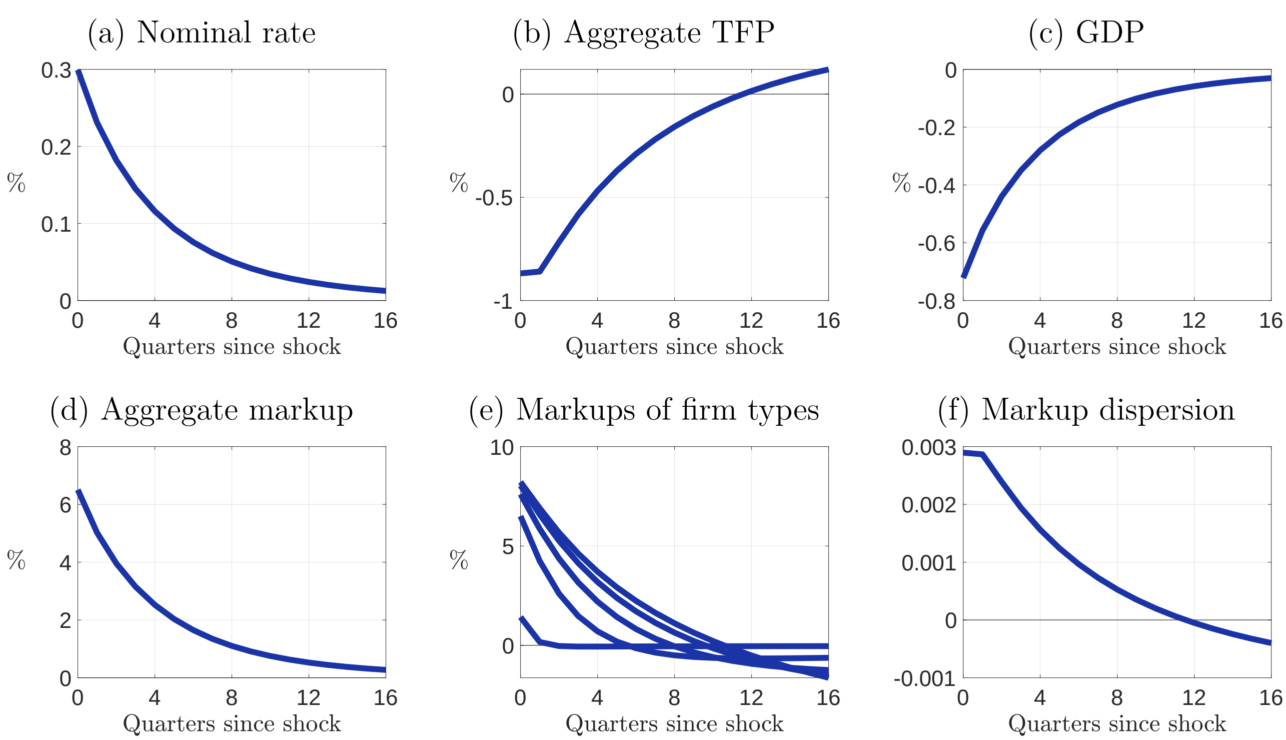
where $\tilde{\eta} = \eta + \frac{\eta}{\varphi}$. Since marginal costs fall in the reset price, the markup also increases in θ_k if the above condition is met. This condition is similar to the setup with homogeneous labor. A sufficient condition for precautionary price setting is that aggregate prices have positive variance, and that the covariances between prices, aggregate demand, and the aggregate component of real marginal costs, are non-negative.

Importantly, if the above condition is satisfied, markups are negatively correlated with pass-through. Hence, Proposition 1 implies that markup dispersion increases in response to monetary policy shocks that lower real marginal costs. Aggregate TFP, computed as the Solow residual of an aggregate production function, then falls. The derivation of the Solow residual is identical to the model with homogeneous labor supply (and independent from assumptions on labor supply). If we aggregate labor input across firms, substitute in the production technology for differentiated goods ($Y_i = AN_i$), and the CES demand curve for differentiated goods, we obtain

$$Y_t = \frac{A_t}{S_t} N_t = TFP_t \cdot N_t.$$

Figure 32 below quantifies the response of markup dispersion and aggregate TFP to monetary policy shocks in this model. For comparability with the baseline (homogeneous labor) model, we keep all model parameters unchanged except for the standard deviation of monetary policy shocks, which we re-calibrate to imply a 30bp increase of the nominal interest rate.

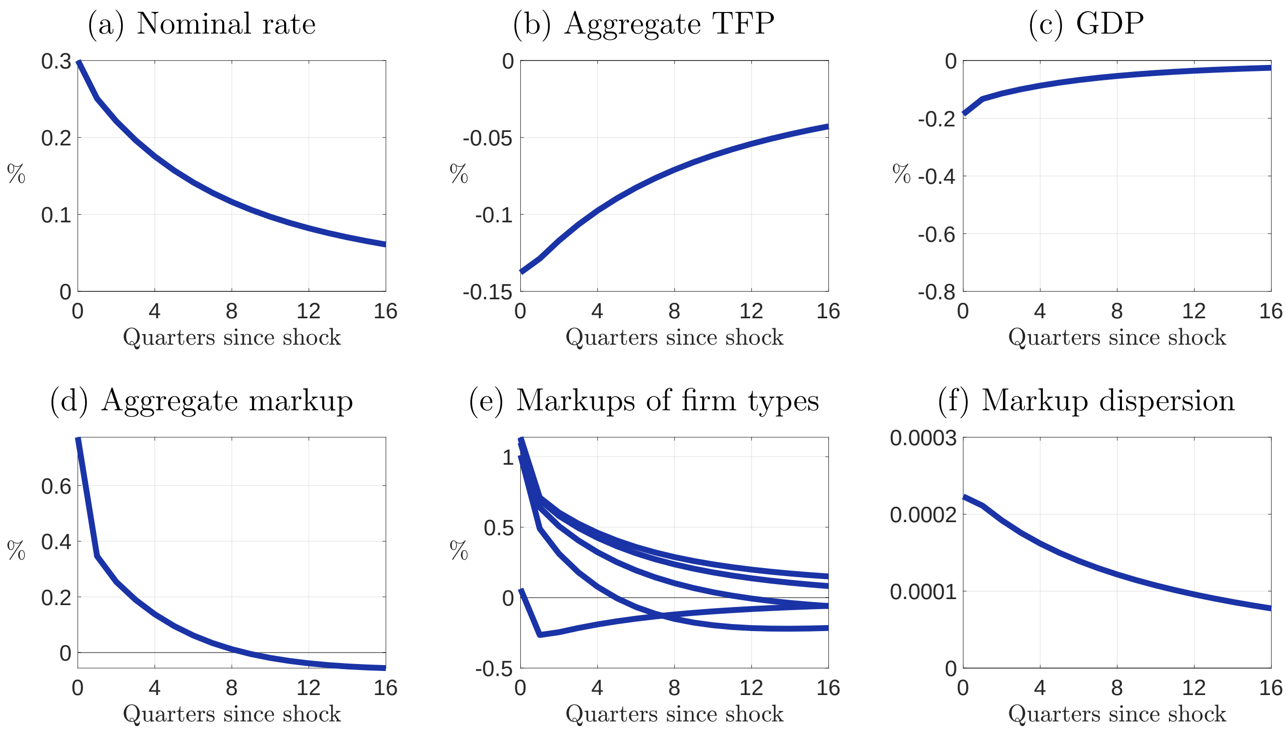
Figure 32: Model responses to monetary policy shocks when assuming specific labor



Notes: This figure shows responses to a one standard deviation contractionary monetary policy shock.

I.2 Rotemberg adjustment costs

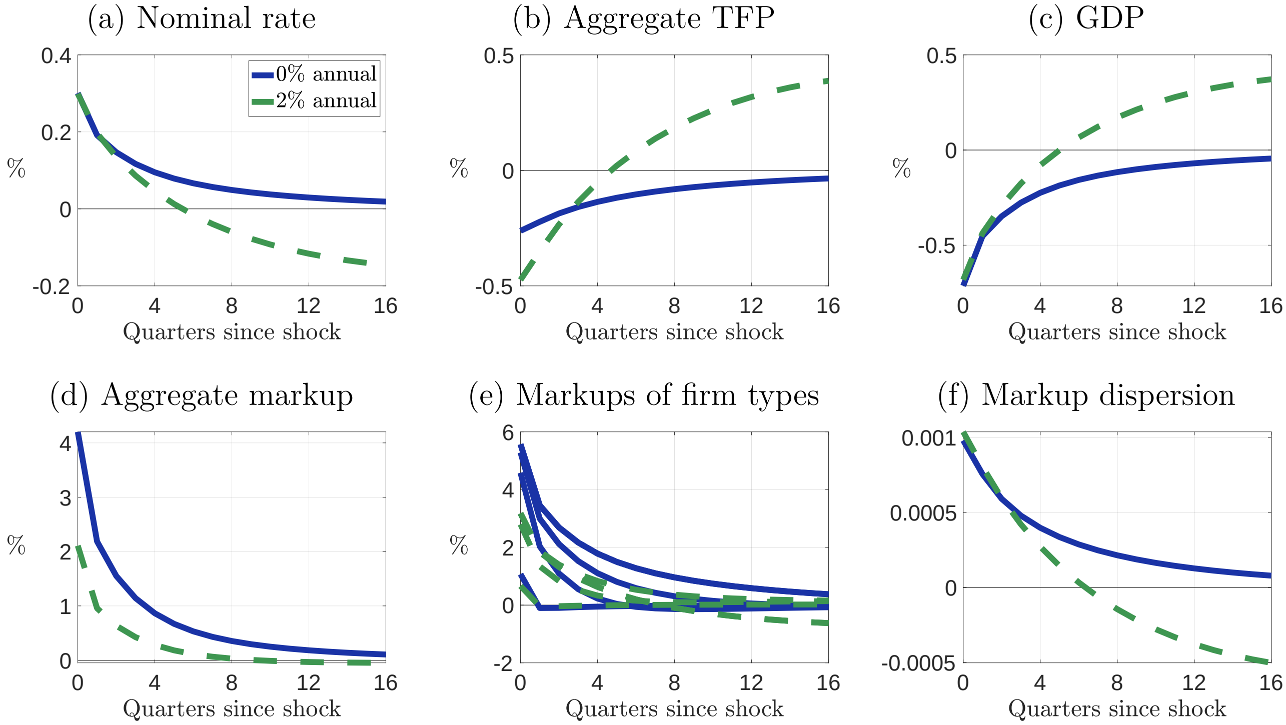
Figure 33: Model responses to monetary policy shocks when assuming Rotemberg adjustment costs



Notes: This figure shows responses to a one standard deviation contractionary monetary policy shock.

I.3 Trend inflation

Figure 34: Model responses to monetary policy shocks when assuming trend inflation



Notes: This figure shows responses to a one standard deviation contractionary monetary policy shock. We set θ_1 to the value of θ_2, because the price setting problem is not well defined for the baseline value of θ_1 when annual trend inflation is 2%.