MONETARY POLICY, MARKUP DISPERSION, AND AGGREGATE TFP *

Matthias Meier† and Timo Reinelt‡

July 31, 2019

Abstract

This paper studies a novel transmission mechanism of monetary policy, which explains why contractionary monetary policy shocks lower aggregate productivity. In models with firm-level heterogeneity in Calvo or Rotemberg price-setting frictions, firms with lower pass-through from marginal costs to prices optimally set higher markups. Contractionary monetary policy shocks further increase the relative markup of these firms. The consequential increase in markup dispersion lowers aggregate productivity. We provide empirical support for this mechanism by showing that (a) contractionary monetary policy shocks increase markup dispersion, (b) firms with higher pre-shock markups increase markups by more after the shock, and (c) markups are higher in industries with more rigid prices. A New Keynesian model, that is solved non-linearly to capture precautionary price setting, shows that our transmission channel is quantitatively important.

Keywords: Monetary Policy, aggregate productivity, precautionary price setting, markup dispersion, heterogeneous pass-through.

---

*A previous version of this paper was titled "The misallocation channel of monetary policy". We owe special thanks to Antonio Ciccone and Konrad Stahl for extremely helpful comments and suggestions. We also wish to thank numerous seminar participants for useful comments at various stages of this project. Funding by the German Research Foundation (DFG) through CRC TR 224 (Project C02) is gratefully acknowledged. Matthias Meier gratefully acknowledges financial support from 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 B6, 30-32, 68159 Mannheim, Germany; E-mail: `timo.reinelt@gess.uni-mannheim.de`.

1. INTRODUCTION

Understanding the nature of the monetary transmission channel is of central importance for both positive and normative questions in business cycle research. Commensurate with its importance, a vast empirical literature estimates the effects of monetary policy shocks.[1] Only few papers, however, study the response of aggregate productivity, which is surprising given the prominent role of productivity fluctuations in modern business cycle theory.[2] This literature shows that contractionary monetary policy shocks lower aggregate productivity, and explains it through responses in *average* firm-level productivity: e.g., utilization, fixed costs, and R&D. Instead, we show that markup dispersion *across* firms explains large aggregate productivity responses, even when average productivity is unchanged.

We propose a novel transmission mechanism of monetary policy, which critically builds on firm-level heterogeneity in the pass-through from marginal costs to prices. We show that firms with lower pass-through optimally set higher markups. This holds if the source of lower pass-through are stronger Calvo or Rotemberg price setting frictions. Because the profit function is asymmetric in a firm's price, firms have a precautionary motive to set a higher price when facing stronger price setting frictions. Low pass-through firms not only set higher markups. Likewise, their markup increases by more in response to a contractionary monetary policy shock. Hence, the shock increases the dispersion in markups across firms, which lowers aggregate productivity.[3]

Empirically, we identify US monetary policy shocks as high-frequency changes in federal funds futures around monetary policy announcements.[4] A one standard deviation monetary policy shock depresses aggregate total factor productivity (TFP) by 0.8% and even utilization-adjusted TFP by 0.4% two years after the shock. The responses are economically significant when compared to the associated 1% drop in aggregate output. Our central empirical finding is that contractionary monetary policy shock raises the dispersion in markups across firms within narrowly-defined industries and time. To compute markup dispersion, we use quarterly US firm-level

[1] E.g., Christiano et al. (1999), Romer and Romer (2004), and Gertler and Karadi (2015).

[2] An early contribution is Evans (1992), and more recently Christiano et al. (2005), Moran and Queralto (2018), and Garga and Singh (2019).

[3] This a well-known feature of the New Keynesian model, see, e.g., chapter 3 of Galí (2015). Relatedly, the adverse effects of higher markup dispersion on aggregate productivity are also discussed in a macro development literature on factor allocation, see, e.g., Hsieh and Klenow (2009).

[4] This identification strategy follows a recent literature, see, e.g., Gertler and Karadi (2015), Gorodnichenko and Weber (2016), and Nakamura and Steinsson (2018).

data from Compustat and follow the methodology in De Loecker and Warzynski (2012) and De Loecker et al. (2018). Through the lense of a model similar to Hsieh and Klenow (2009), we find that the estimated increase in markup dispersion can account for most of the estimated (utilization-adjusted) aggregate TFP response.

The mechanism we propose explains an increase in markup dispersion in response to contractionary monetary policy shocks, and a decrease in markup dispersion in response to expansionary shocks. In contrast, a large class of New Keynesian models cannot reconcile our empirical evidence on markup dispersion.[5] Our mechanism has further testable implications, which we corroborate empirically. First, precautionary price setting implies that firms with high average markups tend to be firms with more rigid prices. Hence, their markups should increase by more after a contractionary monetary policy shock. Second, precautionary price setting implies that markups are higher when prices are adjusted less frequently.

Finally, we investigate the quantitative relevance of our transmission mechanism in a simple New Keynesian model with a Calvo pricing friction that is heterogeneous across firms. Importantly, local solution techniques around the deterministic steady state will fail to capture our mechanism: In the deterministic steady state all firms charge the same markup. Instead, we use non-linear solution methods in order to capture markup differences in the stochastic steady state. Quantitatively, the model explains half of the empirically estimated response of utilization-adjusted TFP to monetary policy shocks. In a policy counterfactual, in which the monetary authority attributes the productivity response of monetary policy shocks to productivity shocks, aggregate output has a 10% higher standard deviation.

This paper is most closely related to three branches of literature. First, our paper relates to a growing literature that documents and studies the effects of heterogeneous price rigidity. Using CPI micro data, Bils and Klenow (2004) and Nakamura and Steinsson (2008), and using PPI micro data, Gorodnichenko and Weber (2016), document substantial dispersion in the frequency of price adjustment across sectors. Carvalho (2006) studies the role of such heterogeneity for the transmission of monetary policy, while Pasten et al. (2018) revisit this question in the presence of

[5]In standard New Keynesian models, price dispersion, and hence markup dispersion, is minimized in the deterministic steady state. Up to a first-order approximation around the deterministic steady state, shocks do not move markup dispersion. Up to a second-order approximation, markup dispersion increases (and aggregate TFP decreases) for both positive and negative shocks. If we add trend inflation to the model, markup dispersion is not minimized in the deterministic steady state. However, contractionary monetary policy shocks counterfactually lower markup dispersion and increase aggregate TFP, see Ascari and Sbordone (2014).

input-output networks. Gorodnichenko and Weber (2016) study asset price implications and Clayton et al. (2018) redistributive implications. Importantly, the previous literature has not studied the mechanism in our paper. Our mechanism is typically abstracted from by choice of solution method.

Second, the mechanism we propose is inspired by Fernandez-Villaverde et al. (2015). They propose precautionary price setting as transmission mechanism of uncertainty shocks. Because the profit function is asymmetric in a firm's price, firms optimally respond to higher uncertainty by setting higher prices and markups, when price adjustment is frictional. In contrast, we show that precautionary price setting generates markup differences across firms with differently severe price setting frictions. Relatedly, we contribute to Baqaee and Farhi (2018) who show that contractionary monetary shocks can lower aggregate TFP if firms with more rigid prices have higher pre-shock markups. We show that this correlation endogeneously arises from differences in price rigidity.

Third, by considering more recent identification schemes and time periods, our paper reconfirms previous findings on the aggregate productivity response to monetary policy shocks, e.g., Evans and Santos (2002), Christiano et al. (2005), Moran and Queralto (2018), Garga and Singh (2019). The causal explanations of lower aggregate productivity in these papers are based on lower average firm-level productivity. In contrast, our mechanism does not require any change in average productivity. Christiano et al. (2005) estimate a model with variable utilization and fixed costs, which, however, explains only a small fraction of their empirically estimated labor productivity response. Moran and Queralto (2018) and Garga and Singh (2019) provide empirical evidence for an adverse R&D response to contractionary monetary policy shocks. We see our paper as complementary to these prior contributions, with possibly very different policy implications. Interestingly, in models with endogenous entry and exit a la Hopenhayn (1992), contractionary shocks increase average and aggregate productivity.

The remainder of this paper is organized as follows. Section 2 presents the empirical evidence. Section 3 presents theoretical results, and Section 4 presents a quantitative model with results. Section 5 concludes and an Appendix follows.

2. EMPIRICAL EVIDENCE

In this section, we show that contractionary monetary policy shocks lower aggregate productivity and raise the markup dispersion across firms (within industry and time). We further provide evidence which suggests that firm-level differences in price-setting frictions drive the response of markup dispersion.

2.1. *Identification of monetary policy shocks*

We identify monetary policy shocks using high-frequency futures prices based on the federal funds rate. A monetary policy shock is the price change in a narrow time window around a FOMC announcements. The identifying restrictions are that the risk premium is unchanged in that window, and that no shock other than the monetary announcement occurs in in or shortly before the time window. We denote the price of a future by f_τ, where τ is the time of a monetary announcement.[6] A monetary policy shock in τ is defined as

$$\varepsilon_\tau^{\mathrm{MP}} = \omega(\tau)(\mathrm{f}_{\tau+\Delta\tau^+} - \mathrm{f}_{\tau-\Delta\tau^-}), \tag{2.1}$$

where $\omega(\tau)$ is an adjustment function to account for the fact that the federal funds future settles on a month's *average* effective overnight federal funds rate.[7] We specify a thirty minute window around FOMC announcements, setting $\Delta\tau^- = 10$ minutes and $\Delta\tau^+ = 20$ minutes, as in Gorodnichenko and Weber (2016).

Because both the macro data and the firm-level data we use is at quarterly frequency, we aggregate the shocks following the approach in Ottonello and Winberry (2018). 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. In this way, we weight shocks across quarters corresponding to the amount of time firms have had to respond. Formally, let t denote quarters,

[6] We obtain time and classification of FOMC meetings from Nakamura and Steinsson (2018) until 2014 and from the Federal Reserve Board afterwards. We obtain time stamps of the FOMC press releases from Gorodnichenko and Weber (2016) and Lucca and Moench (2015).

[7] Only for the *current-month* federal funds future, future prices are differently predetermined by past federal funds rates at different days, so we apply an adjustment to make them more comparable. We use $\omega(\tau)$ = (total number of days in announcement month) / (remaining number of days in announcement month after meeting in τ), except if τ is within the last seven days of the month. We then use the next-month federal funds future price and $\omega(\tau) = 1$. For any federal funds future other than current-month, as well as for Eurodollar futures, the reference periods lies in the future, which obviates the adjustment, and we set $\omega(\tau) = 1$.

then the quarterly shocks is

$$\varepsilon_t^{\mathrm{MP}} = \sum_{\tau \in \mathcal{D}(t)} \phi(\tau)\varepsilon_\tau^{\mathrm{MP}} + \sum_{\tau \in \mathcal{D}(t-1)} (1-\phi(\tau))\varepsilon_\tau^{\mathrm{MP}}, \tag{2.2}$$

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).

To construct monetary policy shocks, we have purchased tick price data of federal funds futures and Eurodollar futures from the Chicago Mercantile Exchange. The data covers the period 1995Q2 to 2018Q3, which results in sample size $T = 96$. Our baseline monetary policy shock is constructed from the three-month ahead federal funds future, as in Gertler and Karadi (2015), and is restricted to scheduled FOMC meetings.[8] We discuss alternative monetary policy shock series at the end of this section. Table II in the Appendix reports summary statistics and Figure 5 shows the shock series.

2.2. *Aggregate productivity*

We show that monetary policy shocks have sizable effects on measures of aggregate productivity. To estimate the dynamic responses of aggregate productivity we use Jordà's (2005) local projections. Denoting (quarterly) aggregate productivity by y_t, our baseline specification is

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \varepsilon_t^{\mathrm{MP}} + \gamma_0^h \varepsilon_{t-1}^{\mathrm{MP}} + \gamma_1^h (y_{t-1} - y_{t-2}) + u_t^h, \tag{2.3}$$

which we estimate for $h = 0, \ldots, 16$. The coefficients $\{\beta^h\}_{h=0}^{16}$ are the cumulative response of productivity growth h periods after a monetary policy shock. We include one lag of the monetary policy shock to control for serial correlation, which arises from time aggregation in (2.2). Inference is based on Newey-West (1987) with $h+1$ lags for the horizon-h local projections, as in Ramey (2016). Following Nakamura and Steinsson (2018), we exclude the apex of the financial crisis from 2008Q3 to 2009Q2.[9]

[8]We exclude non-scheduled meetings and conference calls. These events are often the response to immediate adverse economic developments. Price changes around these events may reflect these developments, thus invalidating the identifying restriction. Similarly, non-scheduled meetings are more likely to release private central bank information about the state of the economy. Nonetheless, our results are broadly robust when including these events.

[9]Dropping the height of the financial crisis is not critical for our results.

As measures of aggregate productivity, we consider aggregate TFP and utilization-adjusted aggregate TFP from Fernald (2014), as well as labor productivity.[10] In addition, we estimate the response of aggregate output measures, aggregate factor inputs, and various interest rates. We use specification (2.3) in all cases.

Figure 1(a) shows the estimated responses of aggregate productivity to a one-standard deviation contractionary monetary policy shock. A monetary policy shock of this magnitude raises the federal funds rate by up to 30 basis points, see Figure 1(d). Panel (a) has three main takeaways. First, tighter monetary policy lowers aggregate productivity. The response is statistically and economically significant with a decline of 0.4-0.8% two years after the shock. Compared to the response of aggregate output, see panel (b), TFP accounts for about 50-80% of the output response at a two-year horizon. Second, the response of aggregate productivity builds up gradually, and peaks only three years after the shock. Third, the differences across productivity measures are relatively small. Utilization adjustment explains at most half of the aggregate TFP response.

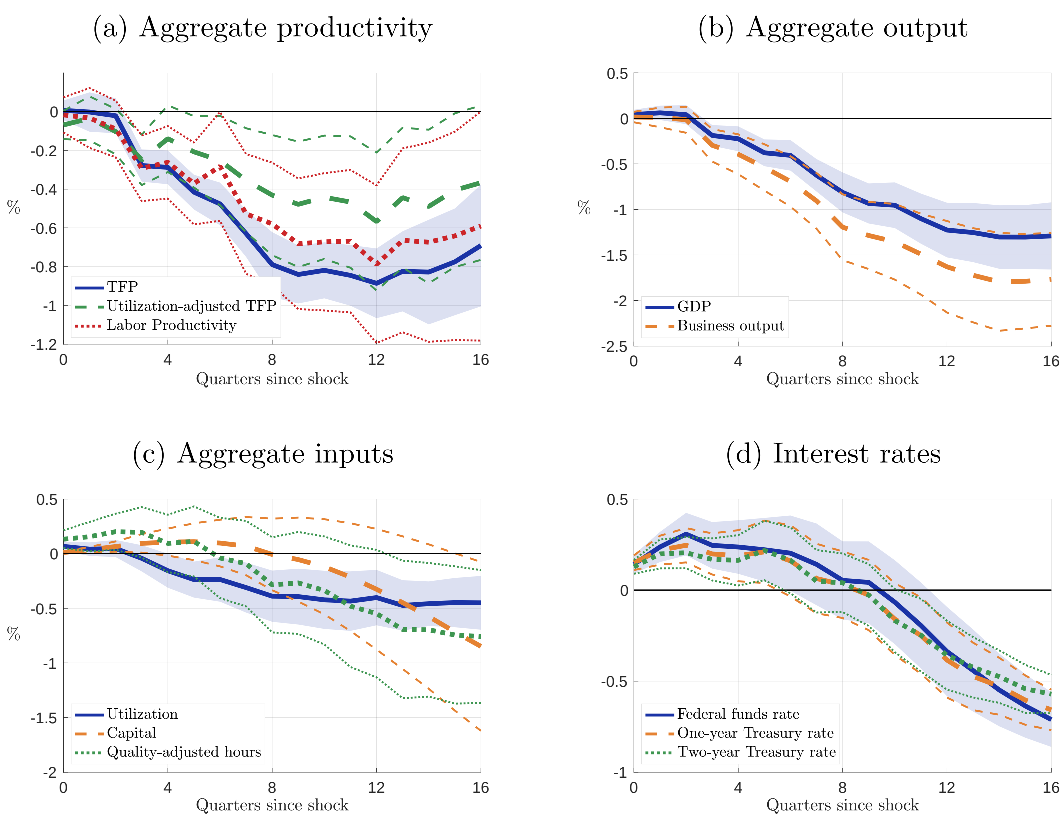
In addition, the responses of aggregate business output and GDP are significant, persistent, and economically meaningful with declines of 1.2% ans 0.8% two years after the shock, respectively. In contrast, the responses of capital and quality-adjusted hours are insignificant for the most part. This highlights even more the critical role of the aggregate TFP response for understanding the real effects of monetary policy. Panel (d) shows that the response of the federal funds rate is fairly comparable to the one-year and two-year treasury rate responses.

2.3. ***Markup dispersion***

We next examine the role of dispersion in markups across firms to explain the aggregate productivity responses. Using firm-level data, we show that markup dispersion (within industry and time) increases in response to contractionary monetary policy shocks. The increase can account for a large fraction of the aggregate TFP response.

[10] Aggregate TFP is $\Delta\texttt{TFP} = \Delta y - w_k \Delta k - (1 - w_k)\Delta \ell$, with Δy real business output growth, w_k the capital income share, Δk real capital growth (based on PIMs for 15 types 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 5(c) in the Appendix plots the different aggregate productivity series.

Figure 1: Responses of aggregate productivity to a monetary policy shock



Notes: The plots show the responses of aggregate productivity to a one-standard deviation contractionary monetary policy shock, i.e., coefficients β^h in equation (2.3). The shaded area is a one standard error band based on the Newey-West estimator.

Motivation

We focus on markup dispersion for two reasons. First, nominal rigidities are at the core of the New Keynesian model. Rigidities, and heterogeneity therein, naturally give rise to markup dispersion. Second, factor allocation is potentially important for productivity differences. As shown by Hsieh and Klenow (2009) and Bayer et al. (2018), factor productivity dispersion is tightly linked to markup dispersion, and the latter can account for large cross-country differences in aggregate TFP. Importantly, *aggregate* TFP can vary in markup dispersion even if *average* firm-level TFP remains unchanged.

Consider a mass of monopolistically competitive firms that produce variety goods Y_{it} with CRS technology. Aggregate output is a Dixit-Stiglitz aggregator with substitution elasticity η. Firms set prices P_{it} to maximize $(\tau_{it}P_{it} - MC_t)\,Y_{it}$ subject to demand for Y_{it}, where MC_t is marginal costs and τ_{it} is a markup wedge. This wedge may be viewed as a shortcut for price rigidities. The markup is then $\mu_{it} = \tau_{it}^{-1}\frac{\eta}{\eta-1}$ and aggregate TFP is given by[11]

$$\texttt{TFP}_t = -\frac{\eta}{2}\mathbb{V}_t[\log\mu_{it}] \; + \; \Big[\text{aggregate exogenous productivity}\Big], \tag{2.4}$$

where $\mathbb{V}_t[\log\mu_{it}]$ is the variance of log markups across firms. 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.

Evidence

Empirically, we use balance sheet data from Compustat to estimate the response of markup dispersion to monetary policy shocks.[12] To estimate firm-level markups, we adopt the approach of De Loecker and Warzynski (2012). If firms have a flexible factor of adjustment, say X, then cost minimization implies that the markup, μ_{it},

[11]We compute aggregate TFP as the Solow residual. The aggregate TFP equation follows from a second-order approximation, which is exact if τ_{it} are log-normally distributed across firms.

[12]Using Compustat has two central advantages over most other firm-level data sets: First, it is available at quarterly frequency, whereas most other firm-level data has lower time frequency. With annual firm-level data for example we would need to aggregate monetary policy shocks to annual frequency, which would severely dilute the informativeness of the shocks. Second, while Compustat only contains listed firms, it does cover all sectors. In contrast, there is excellent establishment level data, which, however, only covers the manufacturing sector.

of firm i in quarter t can be computed as

$$\mu_{it} = \frac{\text{output elasticity of } X_{it}}{\text{revenue share of } X_{it}}. \tag{2.5}$$

There are various reasons why markups may differ over time and across industries, some of which may have little to do with monetary policy. For example, industries differ in production technologies (e.g., high vs. low fixed costs) or in how much market power the average firm in an industry has. Across time, technological progress or shifts in demand across sectors may shift markups. Throughout this paper, we therefore restrict our attention to **markup dispersion within industry and time**. We examine the cross-sectional variance, $\mathbb{V}_t(\tilde{\mu}_{it})$, where $\tilde{\mu}_{it} = \log \mu_{it} - \overline{\log \mu}_{st}$ is the difference of log markup of firm i from the mean log markup in the industry s that firm i operates in.

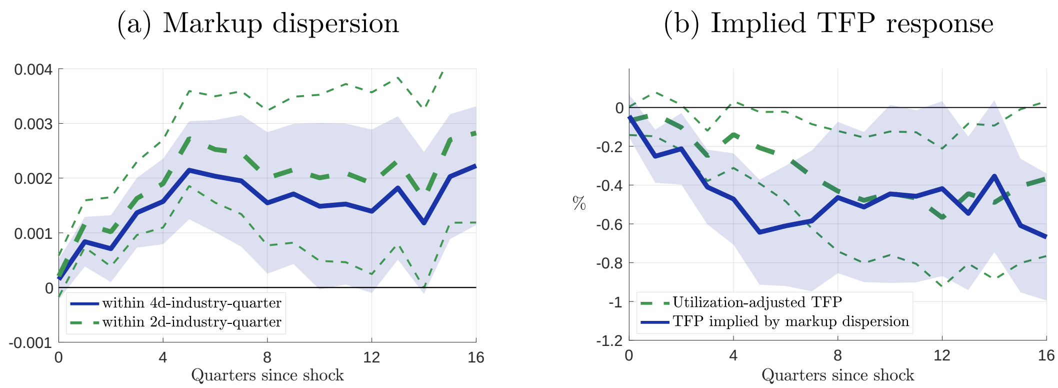
Following De Loecker et al. (2018), we assume firms in a given industry and quarter have a common output elasticity. This assumption is satisfied, e.g., if firms operate the same Cobb-Douglas technology in a given industry-quarter. Again following De Loecker et al. (2018), we consider as X_{it} the composite input of labor and materials. This is well observed in Compustat through costs of goods sold (`cogsq`) and allows us to compute the revenue share in (2.5) by dividing through sales (`saleq`).[13]

We consider all industries except public administration, finance, insurance, real estate, and utilities. An industry is either a two-digit or a four-digit SIC code. We drop firm-quarter observations if the revenue share of labor and material is in the top or bottom 1% of its quarter-specific distribution. We drop firm-quarter observations if the firm reports only once in the respective year. To avoid imprecisely estimated average markups per industry-quarter, we only consider industry-quarters if we observe at least five firms. Figure 5(d) in the Appendix shows the evolution of the markup dispersion series.

To estimate the response of markup dispersion to a one-standard deviation contractionary monetary policy shock, we use again equation (2.3). Figure 2(a) shows the response of markup dispersion. The key finding is that markup dispersion significantly increases. The response is persistent, with a shape that resembles the aggregate productivity responses. Whether an industry is defined at the two-digit or four-digit level makes little difference for our result.

[13] As in De Loecker and Warzynski (2012) and De Loecker et al. (2018), we use a non-parametric projection to purge the sales of measurement error.

Figure 2: Responses of markup dispersion to a monetary policy shock



Notes: The plots show the responses of markup dispersion to a one-standard deviation contractionary monetary policy shock. Panel (a) shows markup dispersion within two-digit and four-digit industry-quarters, respectively. Panel (b) shows the imputed response of TFP, implied by the response of markup disperesion within four-digit industry-quarters, according to Equation (2.4). For comparison, Panel (b) also shows the empirical response of utilization-adjusted TFP as shown in Figure 1. Inference is as in Figure 1.

Relevance

To examine the relevance of the estimated markup dispersion responses, we make use of equation (2.4). We impute the (percentage) TFP response by multiply the response of markup dispersion by $100 \cdot \eta/2$. In Figure 2(b), we compare the estimated utilization-adjusted TFP response with the imputed TFP response. When setting $\eta = 6$, which is the estimate in Christiano et al. (2005), the imputed TFP response comes strikingly close to the estimated one. This simple exercise shows that the estimated increase in markup dispersion may not only qualitatively, but also quantitatively be of key importance to understand the effects of monetary policy shocks on aggregate productivity and aggregate output.

2.4. ***Firm-level heterogeneity in the markup response***

If contractionary monetary policy shocks increases the relative markups of firms with relatively high markups before the shock, then markup dispersion across firms increases. We provide evidence which supports this account.

Empirically, we estimate panel local projections on the interaction of monetary policy shocks and pre-shock firm characteristics. Our variable of interest is the relative log markup of a firm. Let $y_{it} = \tilde{\mu}_{it}$, and denote denote by X_{it-1} a vector of firm-specific characteristics observed in $t-1$, which we interact with the monetary policy shock and add as control. We estimate the following panel local projection

$$y_{it+h} - y_{it-1} = \alpha_{th} + B^h X_{it-1}\varepsilon_t^{\mathrm{MP}} + \Gamma^h X_{it-1} + u_{it}^h. \tag{2.6}$$

We consider two specifications for X_{it-1}: (i) only the firm-level markup $\tilde{\mu}_{it}$, (ii) additionally including firm-level size (`atq`), leverage ((`dlcq`+`dlttq`)/`atq`), and the liquid asset ratio (`cheq`/`atq`). All variable are expressed in logs and demeaned for each industry-quarter. We drop industry-quarters with less than five observations, and the top/bottom 1% of all variables in X_{it-1} and $y_{it+h} - y_{it-1}$. Our selection of controls is motivated by recent work in Ottonello and Winberry (2018) and Jeenas (2018), who study the heterogeneous effects of monetary policy shocks on capital.

We are primarily interested in the coefficient in B^h for $h = 0, \ldots, 16$ associated with the firm's past markup. This coefficient captures the relative markup increase of firms with a relatively higher pre-shock markup. For specification (i), the estimated coefficients are shown in Figure 3(a). A firm with a markup one standard deviation above the industry mean in quarter $t-1$, responds to a quarter t contractionary monetary policy shock by increasing its markup by up to 0.5% more than other firms in the industry. Importantly, this result is also consistent with a negative correlation between firm-level markups and price rigidity. We obtain almost identical estimates for specification (ii), see Figure 3(b), which suggests that our results are not driven by financial constraints.[14]

2.5. ***Pass-through and markups***

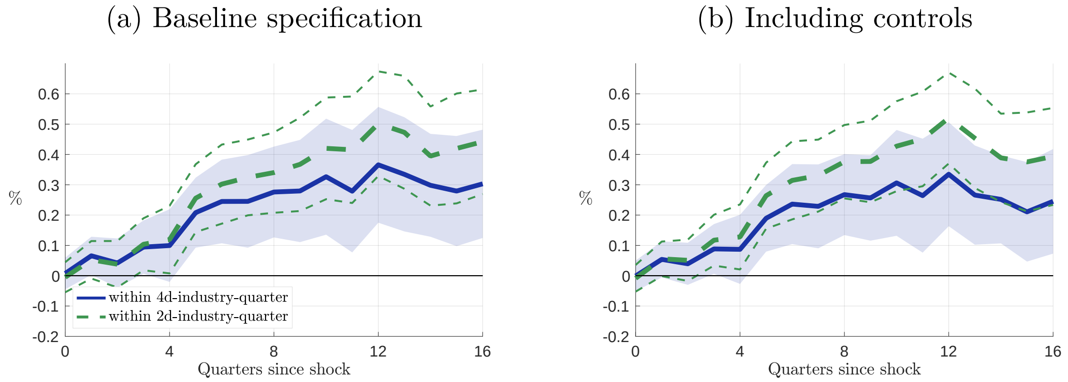
If firms with higher markups have lower pass-through, then their markups increase by more after monetary policy shocks, and, given their higher *initial* markup, this raises markup dispersion. We show that the data supports this account: industries with stickier prices have higher markups.

We use price adjustment frequencies at the level of narrowly-defined industries based on the PPI micro data.[15] We compute various industry-level markup measures

[14]Figure 11 in the Appendix shows the interaction with the other three firm characteristics included in specification (ii). The interaction with firm size is mostly insignificant. Firms with higher pre-shock leverage tend to increase their markups by more. Finally, firms with a higher pre-shock share of liquid assets increase their markups by more.

[15]The adjustment frequencies are an average over 2005-2011 based on Pasten et al. (2018), and were generously shared with us by Michael Weber.

Figure 3: Firm-level heterogeneity in the markup response to monetary policy shocks



Notes: Response of firm-level markup to a monetary policy shock interacted with pre-shock firm-level markup. Inference is based on two-way clustered standard errors by firms and quarters. The shaded area is a one-standard error band.

based on Nekarda and Ramey (2011) and Nekarda and Ramey (2013) and using the NBER-CES manufacturing industry database. The first markup measure is the *price-cost margin*

$$\mu_{jt}^{PCM} = \frac{S_{jt} + \Delta I_{jt} - \text{payroll}_{jt} - \text{materical cost}_{jt}}{S_{jt} + \Delta I_{jt}}$$

for industry j and year-quarter t, and where S denotes sales and ΔI is the change of total inventories. The second measure is the *average markup*, computed as

$$\mu_{jt}^{A} = \frac{\text{Cost share of labor}_{jt}}{\text{Revenue share of labor}_{jt}},$$

where the revenue share is payroll over sales and the labor cost share is payroll over the sum of payroll, material cost, and capital cost. Capital cost is the product of real capital stock and user cost. The user cost is the real interest rate on one-year treasuries, plus the credit spread from Gilchrist and Zakrajsek (2012), plus the industry-specific depreciation rate from the BEA. The third measure is the *marginal markup*, μ_{jt}^{M}. This is similar to μ_{jt}^{A} with the difference that we replace the average real wages (in the payroll) with an estimate of the marginal wage from Nekarda and Ramey (2013). This adjustment may correct for a procyclical bias in markups arising from higher wage costs due to overtime hours. In addition, we distinguish between markups based on the (marginal) payroll of all employees or based on production workers only.

To compare markups with price adjustment frequencies, we first average the industry-level markups for 2005-2011. Table I shows the results when regressing the various industry-level markup measures on the price adjustment frequency. In all specifications, we obtain significantly negative coefficients. This is consistent with the notion that firms with stickier prices set higher markups.

TABLE I

Regressions of markup on price stickiness at the industry level

	Price-cost margin	Markup	Markup (Production)
		Average markup	
Price adjustment frequency	-0.3377***	-0.1363**	-0.2791***
	(0.1171)	(0.0558)	(0.0571)
		Marginal markup	
		-0.1144**	-0.2661***
		(0.0562)	(0.0571)

Notes: The number of observations for all regressions is 177. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.6. ***Robustness***

Alternative monetary policy shocks

Our baseline monetary policy shock uses the three-month ahead federal funds future around scheduled FOMC meetings. For robustness, we consider the current month federal funds futures, and a policy indicator, similar to Nakamura and Steinsson (2018), computed as first principal component of the current/three-month federal funds futures and the 2/3/4-quarters ahead Eurodollar futures. We also consider a shock series including unscheduled meetings and conference calls.

High-frequency future price changes may not only reflect conventional monetary policy shocks, but also the release of private central bank information about the state of the economy, see Jarocinski and Karadi (2018). We apply two alternative strategies to control for such information shocks. 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 (2018), we discard daily monetary policy shocks if the associated high-frequency change in the S&P500 moves in the same direction. A different concern may be that unconventional monetary policy may drive our result. We address this by setting daily monetary policy shocks at Quantiative Easing (QE) announcements to zero.

Figures 6 shows the responses of aggregate productivity. Across the seven monetary policy shock series, the responses are comparable in magnitude and the vast majority are statistically significant. Figure 7 shows the response of markup dispersion for all monetary policy shock series. The baseline results again hold broadly robust and we estimate responses in the same order of magnitude. Finally, Figure 12 shows that markups of firms with higher pre-shock markups robustly increase.

TFP measurement and market power

Hall (1986) shows that the Solow residual is misspecified in the presence of market power. Hall shows that instead of using the capital income share, w_k, as Solow weight for capital, and $1 - w_k$ for labor, the correct weights are μw_k and $1 - \mu w_k$, where μ is the aggregate markup. We therefore examine the response of markup-corrected (utilization-adjusted) aggregate TFP to monetary policy shocks. We consider the series of average markups from De Loecker and Eeckhout (2018) to compute Hall's weights. Figure 8(c) in the Appendix shows that this barely affects our results.

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 8(a) and (b). Notably, the response of investment-specific TFP is significantly stronger than that of consumption-specific TFP.

Firm-level data treatment and delisting

Balance-sheet data may contain erroneous and economically implausible entries. We examine the robustness of our results for three alternative data treatments, all of which drop additional observations compared to our baseline date treatment. First, instead of dropping the top/bottom 1% of the revenue share distribution per quarter, we drop the top/bottom 5%. Second, we drop firm-quarter observations with real quarterly sales below 1 million 2010 USD. Third, we drop all firm-quarter observations if real quarterly sales growth is above 200% or below -66%.

Figure 9(a)-(c) show the response of markup dispersion for these data treatments. The exclusion of observations with large absolute sales growth makes no difference to the results. Under the other two alternatives, we see some shift in the response function toward zero. However, the responses remain positive and significant.

A well-known recent trend is the delisting of formerly public firms. Figure 10(a) plots the number of firms in our sample after data treatment, which drops from about 6,000 to 3,000 firms. Thus, a valid concern is that this may affect our results. We address this concern in two ways. First, we estimate whether the number of firms in Compustat responds to monetary policy shocks. Figure 10(b) shows that the response is small and insignificant. Second, we condition our estimation on firms that are in the sample for at least 16 quarters, corresponding to our maximal forecast horizon. Figure 9(d) shows that this restriction only marginally affects our results.

2.7. *Alternative explanations of lower TFP*

This paper focuses on higher markup dispersion as explanation for lower aggregate TFP in response to contractionary monetary policy shocks. An alternative explanation for lower aggregate productivity is lower *average* firm-level TFP.

Contractionary monetary policy shocks may lower average firm-level TFP because (a) the productivity of existing firms falls (relative to trend), or (b) the pool of operating firms changes. The productivity of existing firms may fall because firms invest less resources into developing new products and processes or into improving existing ones. Resources spent on such activities will be partially captured by measured R&D investment. In fact, panel (d) of Figure 8 shows that aggregate R&D expenditures fall after contractionary monetary policy shocks. This reconfirms the findings in Moran and Queralto (2018) and Garga and Singh (2019). Hence, the R&D channel constitutes a complementary explanation for the drop in aggregate productivity.

Conversely, even if the productivity of existing firms remains unchanged by the monetary policy shock, aggregate productivity may fall if the selection of firms worsens in the sense that the share of low-productivity firms increases. This seems rather unlikely from a theoretical perspective. In firm models with endogeneous entry and exit a la Hopenhayn (1992), contractionary shocks improve the selection of firms, which increases average firm-level TFP.

3. THEORETICAL RESULTS

In this section, we provide an analytical characterization of our mechanism for two common types of price setting frictions. When firms have differently strong Calvo or Rotemberg frictions, firms with stronger frictions have lower pass-through from marginal cost to price and set higher markups. If the markup-pass-through correlation is indeed negative, then markup dispersion increases after contractionary monetary policy shocks.

3.1. ***Model setup and definitions***

The basic economic environment is common across the two models. Time is discrete, lasts forever, and is indexed by t. Firms, indexed by i, are monopolistically competitve with isoelastic demand function $Y_{it} = \left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t$, where P_{it} is the firm's price, P_t the aggregate price level, Y_t aggregate demand, and $\eta > 1$ is the own-price elasticity of demand. Firms operate a linear technology, $Y_{it} = L_{it}$, and hire labor L_{it} at the real wage W_t. We consider a risk-neutral investor with discount factor $\beta \in (0, 1)$, who sets prices to maximize the net present value of profits subject to firm-specific price setting frictions. Real period t profit is given by

$$\text{profit}_{it} = \left(\frac{P_{it}}{P_t} - W_t\right)\left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t. \tag{3.1}$$

This profit function is asymmetric in the price P_{it}. Profits are (statically) maximized at $\tilde{P}_{it} = \frac{\eta}{\eta-1} P_t W_t$ and fall more rapidly below $\tilde{P}_{it}$ than above it, see Fernandez-Villaverde et al. (2015). When price setting is frictional and subject to uncertainty, then this asymmetry gives rise to a *precautionary upward bias* in price setting.

Firms take aggregate prices and aggregate demand as given. To allow for analytical tractability, we assume a multivariate log-normal distribution for P_t, W_t, and Y_t,

$$\log \begin{pmatrix} P_t/\bar{P} \\ W_t/\bar{W} \\ Y_t/\bar{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} -\frac{\sigma_p^2}{2} \\ -\frac{\sigma_w^2}{2} \\ -\frac{\sigma_y^2}{2} \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & & \\ \sigma_{pw} & \sigma_w^2 & \\ \sigma_{py} & \sigma_{wy} & \sigma_y^2 \end{bmatrix} \right). \tag{3.2}$$

The *markup* is defined as the ratio of price to marginal cost

$$\mu_{it} = \frac{P_{it}}{P_t W_t}. \tag{3.3}$$

The *pass-through* from real marginal costs to price is the elasticity

$$\varepsilon_{it} = \frac{\mathrm{d} \log P_{it}}{\mathrm{d} \log W_t}. \tag{3.4}$$

3.2. ***Precautionary price setting***

CALVO FRICTION

Consider a Calvo (1983) friction, parametrized by a *firm-specific* price adjustment probability $1 - \theta_i \in (0, 1)$. The price setting problem of the firm is

$$\max_{P_{it}} \mathbb{E}_t \sum_{j=0}^{\infty} \theta_i^j \beta^j \left(\frac{P_{it}}{P_{t+j}} - W_{t+j} \right) \left(\frac{P_{it}}{P_{t+j}} \right)^{-\eta} Y_{t+j}. \tag{3.5}$$

The first-order condition for P_{it} is satisfied by the optimal reset price

$$P_{it}^* = \frac{\eta}{\eta - 1} P_t W_t \frac{\mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \theta_i^j \frac{W_{t+j}}{W_t} \left(\frac{P_{t+j}}{P_t} \right)^{\eta} \frac{Y_{t+j}}{Y_t}}{\mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \theta_i^j \left(\frac{P_{t+j}}{P_t} \right)^{\eta - 1} \frac{Y_{t+j}}{Y_t}}. \tag{3.6}$$

We denote the associated optimal reset markup by μ_{it}^*. To isolate the role of uncertainty in price setting, we focus on $P_t = \bar{P}$ and $W_t = \bar{W}$. The following proposition characterizes precautionary price setting.

PROPOSITION 1 (Calvo: precautionary price setting) *If $P_t = \bar{P}$, $W_t = \bar{W}$, and $(\eta - 1)\sigma_p^2 + \sigma_{py} + \eta\sigma_{pw} + \sigma_{wy} > 0$, then the firm optimally sets a higher markup than statically optimal, and the markup further increases in the Calvo parameter θ_i,*

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

Proof: See Appendix D.1.

Under the Calvo friction, pass-through ε_{it} is stochastic: $\varepsilon_{it} = 0$ with probability θ_i and $\varepsilon_{it} > 0$ else. We next characterize how *expected* pass-through depends on θ_i.

PROPOSITION 2 (Calvo: pass-through) *Let $\bar{\varepsilon}_{it}$ denote the expected pass-through ε_{it}. If $P_t = \bar{P}$, then the expected pass-through from*
(i) a transitory change in W_t (around $W_t = \bar{W}$), or
(ii) a permanent change in W_t,
falls monotonically in the Calvo parameter θ_i,

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

Proof: See Appendix D.2.

Proposition 1 and 2 imply that firms with lower pass-through optimally set higher markups, and that markups and pass-through correlate negatively across firms.

Corollary 1 *Under the conditions of Proposition 1 and 2, low-pass through firms optimally set higher markups* $\frac{\partial \mu_{it}^*}{\partial \bar{\varepsilon}_{it}} < 0$. *If firms differ in the Calvo parameter* θ_i, *then the correlation between markups and pass-through is negative* $\operatorname{Corr}(\mu_{it}^*, \varepsilon_{it}) < 0$.

Rotemberg friction

Consider a Rotemberg (1982) friction, parametrized by a *firm-specific* cost shifter $\phi_i \geq 0$. The firm's price-setting problem is

$$\max_{\{P_{i,t+j}\}_{j=0}^{\infty}} \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \left(\frac{P_{i,t+j}}{P_{t+j}} - W_{t+j} \right) \left(\frac{P_{i,t+j}}{P_{t+j}} \right)^{-\eta} Y_{t+j} - \frac{\phi_i}{2} \left(\frac{P_{i,t+j}}{P_{i,t+j-1}} - 1 \right)^2 . \tag{3.7}$$

The first-order condition for P_{it} is

$$\begin{aligned}(1-\eta)\left(\frac{P_{it}}{P_t}\right)^{1-\eta} Y_t + \eta W_t \left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t \\ - \phi_i \left[\left(\frac{P_{it}}{P_{i,t-1}} - 1\right)\frac{P_{it}}{P_{i,t-1}}\right] + \phi_i \mathbb{E}_t \left[\left(\frac{P_{i,t+1}}{P_{it}} - 1\right)\frac{P_{i,t+1}}{P_{it}}\right] = 0\end{aligned} \tag{3.8}$$

The following proposition characterizes precautionary price setting.

Proposition 3 (Rotemberg: precautionary price setting) *If* $P_{t-1} = P_t = \bar{P}$, $W_{t-1} = W_t = \bar{W}$, *and* $\sigma_p^2 + \sigma_w^2 + 3\sigma_{pw} > 0$, *then a first-order approximation of* (3.8) *at* $\phi_i = 0$ *yields*

$$\mu_{it}^* \geq \frac{\eta}{\eta - 1}, \quad \textit{and} \quad \frac{\partial \mu_{it}^*}{\partial \phi_i} \geq 0.$$

The two inequalities are strict if the Rotemberg parameter $\phi_i > 0$.

Proof: See Appendix D.3.

We next characterize how pass-through depends on the Rotemberg parameter ϕ_i.

Proposition 4 (Rotemberg: pass-through) *There exists an $\tilde{\eta} \gg 1$ such that if $\eta \in (1, \tilde{\eta})$ and $P_t = \bar{P}$, then a first-order approximation of (3.8) at $\phi_i = 0$ has the following implications. The pass-through from*
(i) a transitory change in W_t (around $W_t = \bar{W}$), or
(ii) a permanent change in W_t,
falls monotonically in the Rotemberg parameter ϕ_i

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

Proof: See Appendix D.4.

If we express the price adjustment costs in terms of profits at ϕ_i, then pass-through falls in ϕ_i for all $\eta > 1$. In Proposition 4, the upper bound is $\tilde{\eta} \approx 1 + (\sigma_p^2 + \sigma_w^2 + 3\sigma_{pw})^{-1}$, which is above common estimates of η for common parametrizations. For relatively large variances $\sigma_p = \sigma_w = 0.05$ and under $\sigma_{pw} = \sigma_p \sigma_w$, we obtain $\tilde{\eta} \approx 80$. Proposition 3 and 4 imply a negative correlation between firm-level markups and pass-through.

Corollary 2 *Under the conditions of Proposition 3 and 4, low-pass through firms set higher markups $\frac{\partial \mu_{it}}{\partial \varepsilon_{it}} < 0$. If firms differ in Rotemberg parameter ϕ_i, then the correlation between markups and pass-through is negative,* $\mathrm{Corr}(\mu_{it}, \varepsilon_{it}) < 0$.

3.3. *Response of markup dispersion to shocks*

We finally show that the correlation between firm-level markup and firm-level pass-through is a critical moment for the response of markup dispersion to shocks.

Proposition 5 (Markup dispersion) *If and only if the correlation between markups and pass-through exists and is strictly negative,*

$$\mathrm{Corr}(\log \mu_{it}, \bar{\varepsilon}_{it}) < 0,$$

then the cross-sectional dispersion in markups decreases in real marginal costs

$$\frac{\partial\, V(\log \mu_{it})}{\partial \log W_t} < 0.$$

Proof: See Appendix D.5.

If contractionary monetary policy shocks lower real marginal costs, then markup dispersion increases if the correlation between (log) markup and pass-through is strictly negative. As shown before, this correlation is typically negative in models with (heterogeneous) Calvo or Rotemberg price setting frictions.

4. NEW KEYNESIAN MODEL WITH HETEROGENEOUS PRICE STICKINESS

In this section we integrate heterogeneity in price rigidity into standard a New Keynesian model. The endogenous correlation of price rigitidy with markup differences in the stochastic steady state leads to TFP losses after contractionary monetary policy shocks.

4.1. *Model setup*

Households

We assume a representative infinitely-lived household seeking to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(C_t, N_t), \tag{4.1}$$

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. 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 represents employment and W_t the nominal wage. D_t are aggregate dividends. We impose the solvency constraint $\lim_{T\to\infty} \mathbb{E}_t[\Lambda_{t,T} \frac{B_T}{P_T}] \geq 0$ for all t, where $\Lambda_{t,T} = \beta^{T-t} U_{c,T}/U_{c,t}$ is the stochastic discount factor.

Final good firms

A representative firm produces final goods Y_t using a Dixit-Stiglitz aggregator

$$Y_t = \left(\int_0^1 Y_{i,t}^{\frac{\eta-1}{\eta}} \, \mathrm{d}i \right)^{\frac{\eta}{\eta-1}}, \tag{4.2}$$

where η is the elasticity of substitution between differentiated goods, $Y_{i,t}$.

Intermediate good firms

The economy is populated by a continuum of intermediate good firms, indexed by $i \in [0,1]$. Each intermediate good firm produces a differentiated good using the Cobb-Douglas production technology $Y_{i,t} = A_t N_{i,t}$. A_t is average firm-level productivity, which follows $\log A_t = \rho_a \log A_{t-1} + \varepsilon_{a,t}$. The innovations $\varepsilon_{a,t} \sim \mathcal{N}(0, \sigma_a^2)$ are classical productivity shocks. Final good aggregation generates an isoelastic demand schedule for intermediate goods given by $Y_{i,t} = (P_{i,t}/P_t)^{-\eta} Y_t$, where $P_t = (\int_0^1 P_{i,t}^{1-\eta} di)^{1/(1-\eta)}$ denotes the price index and $P_{i,t}$ the firm-level price. Following Calvo (1983), firms may reset their prices $P_{i,t}$ with firm-specific probability $1-\theta_i$ in any given period. The price setting policy maximizes the value of the firm to its shareholder,

$$\max_{P_{i,t}} \sum_{j=0}^{\infty} \theta_i^j \mathbb{E}_t \left[\frac{\Lambda_{t,t+j}}{P_{t+j}} \left(\frac{P_{it}}{P_{t+j}} - W_{t+j} \right) \left(\frac{P_{it}}{P_{t+j}} \right)^{-\eta} Y_{t+j} \right]. \tag{4.3}$$

Monetary authority

The monetary authority aims to stabilize inflation (P_t/P_{t-1}) and fluctuations in output, Y_t, around its natural level, denoted $\tilde{Y}_t$, by following the Taylor-type rule, subject to monetary policy shocks ν_t,

$$R_t = R_{t-1}^{\rho_r} \left[\frac{1}{\beta} \left(\frac{P_t}{P_{t-1}} \right)^{\phi_\pi} \left(\frac{Y_t}{\tilde{Y}_t} \right)^{\phi_y} \right]^{1-\rho_r} \exp\{\nu_t\}, \quad \nu_t \sim \mathcal{N}(0, \sigma_\nu^2). \tag{4.4}$$

Market clearing

The final goods market clears if $Y_t = C_t$. The labor market clears if $N_t = \int_0^1 N_{i,t} \mathrm{d}i$.

Equilibrium

The competitive equilibrium is defined by firm-level allocations $\{Y_{i,t}, N_{i,t}\}_{t=0}^{\infty}$ and prices $\{P_{i,t}\}_{t=0}^{\infty}$ for all i, and aggregate allocations and prices $\{Y_t, N_t, P_t, R_t\}_{t=0}^{\infty}$ such that households and firms maximize their objective functions, the monetary authority follows the policy rule, and both markets clear in every period t.

4.2. *Calibration and solution*

In the model a period is a quarter. We choose the discount factor β to match an annual real interest rate of 3%. Preferences are separable $U(C_t, N_t) = \log(C_t) - N_t^{1+\varphi}/(1+\varphi)$, where φ is the inverse Frisch elasticity. We set $\eta = 6$, which is the estimate in Christiano et al. (2005). Note that a low η tends to diminish the quantitative relevance of our mechanism: It weakens the incentive for precautionary price setting and lowers the aggregate TFP losses from markup dispersion. For robustness, we also consider $\eta = 21$ as in Fernandez-Villaverde et al. (2015). For the Taylor rule, we follow Christiano et al. (2016) and set $\rho_r = 0.85$, $\varphi_\pi = 1.5$, and $\varphi_y = 0.05$. Compared to textbook New Keynesian models, the central departure of our model is heterogeneity in price rigidity. For simplicity, we assume that half of the intermediate goods firms reset prices every period, that is $\theta_i = 0$ for $i \in [0, 0.5]$, while the other half set prices infrequently, on average every two years, $\theta_i = 7/8$ for $i \in (0.5, 1]$. Hence, the average price duration is one year.

We next calibrate critically important model parameters. First, our empirical results in Figure 14 show that the hours response is relatively small (and insignificant) compared to the GDP response. We calibrate the Frisch elasticity by targeting the relative hours response two years after the shock. Importantly, this does not restrict the magnitude of the aggregate TFP response in the model. The calibrated Frisch elasticity of labor supply is $\varphi = 1/0.125$, which is at the lower end of estimates in the literature, see Chetty et al. (2011).

Another critical model feature is the extent of aggregate uncertainty. Precautionary price setting behavior is the firms' endogenous reaction to aggregate uncertainty. The higher aggregate uncertainty, the larger the markup differences between firms with sticky prices and firm with flexible prices. For simplicity, monetary policy shocks are the only source of uncertainty, i.e., we abstract from productivity shocks. We calibrate the variance of the monetary policy shocks such that a one standard deviation shock increases the interest rate by 30 basis points, consistent with our empirical estimate in Figure 14. This results in $\sigma_\nu = 0.58\%$.

To capture pre-shock markup differences that rise 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, in which markup differences across firms exist. We apply the method developed by Meyer-Gohde (2014), which uses a third-order perturbation around the deterministic steady state to approximate the stochastic steady state as well as a first-order approximation of the model dynamics around the stochastic steady state.

4.3. ***Quantitative results***

Precautionary price setting

In the stochastic steady state, markups of sticky-price firms are 4.76% higher than markups of flexible-price firms. Hence, we have a negative correlation between markups and pass-through, which implies that contractionary monetary policy shocks increase markup dispersion and lower TFP, see Proposition 5.

Effects of monetary policy shocks

Figure 4 shows the quantitative effects of a one-standard deviation contractionary monetary policy shock. The shock depresses aggregate demand, in response to which flexible-price firms revise their prices downward to keep their markup unchanged. Sticky-price firms, however, have lower pass-through and on average their markups increase. Since the sticky-price firms have higher initial markups, this unambiguously raises markup dispersion. This worsens the allocation of factors across firms and thereby depresses aggregate TFP.[16]

The proposed mechanism is quantitatively important and in line with our empirical evidence. The increase in markup dispersion is about half the size of the peak empirical response, see Section 2. The model explains half of the estimated drop in utilization-adjusted TFP. Moreover, the model matches empirical observation that firms with high pre-shock markups increase their markups further after contractionary monetary policy shocks.

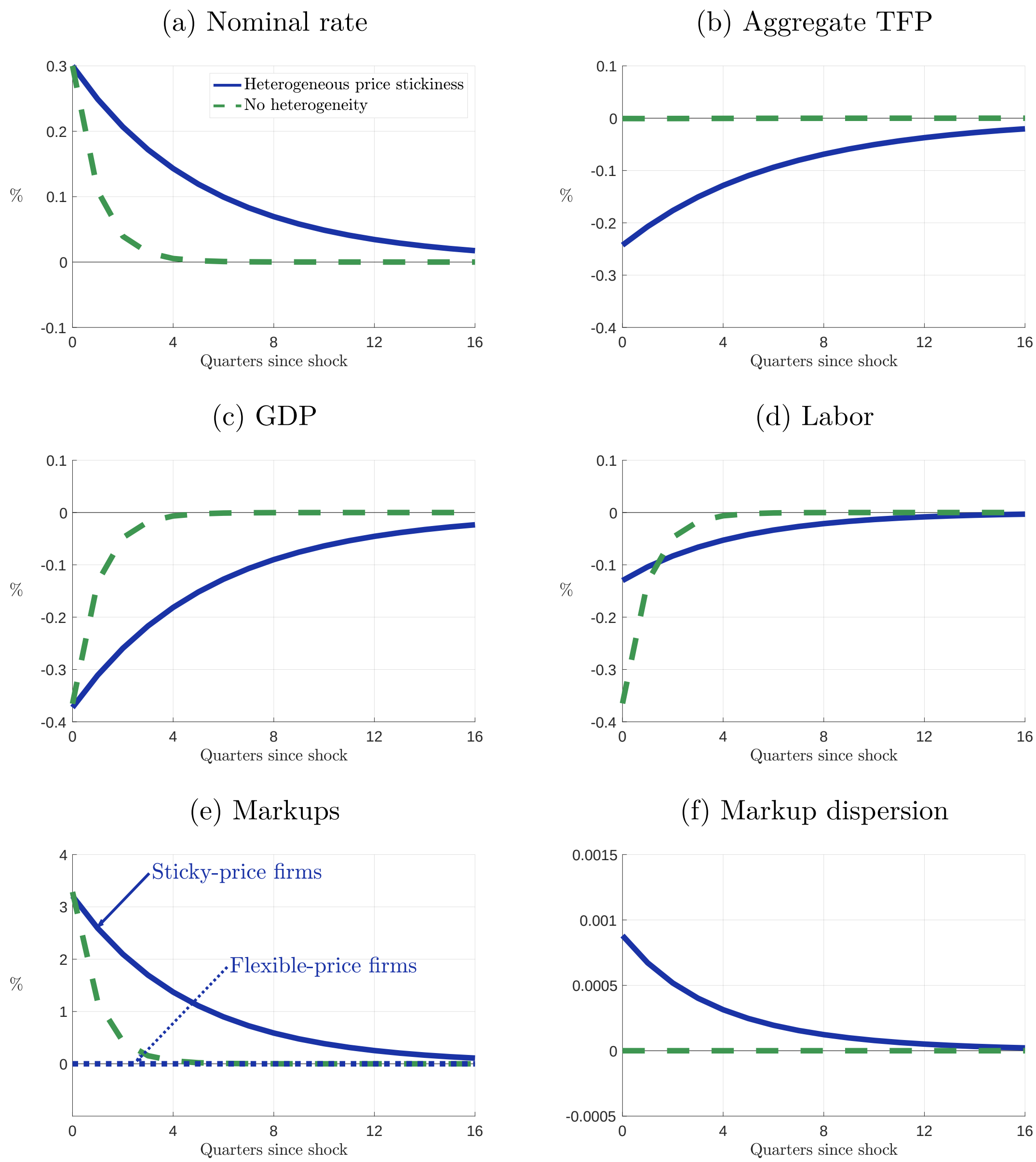
To highlight that the model's transmission mechanism critically builds on heterogeneous price rigidity, we shut down all heterogeneity and set $\theta_i = 3/4$ all $i \in [0, 1]$, while keeping all other parameters unchanged. The dashed lines in Figure 14 show the responses under this counterfactual scenario. While the output response is of comparable magnitude, we observe no visible aggregate TFP response.

Monetary policy and (mis)perceived aggregate TFP fluctuations

We argue in this paper that an important aspect of the monetary transmission channel is the response of aggregate TFP. In contrast, in traditional business cycle models, the only source of fluctuations in aggregate TFP are exogeneous aggregate productivity shocks. This motivates us to examine the success of a given Taylor

[16] An expansionary monetary policy shock leads to mirrored impulse-responses and, in particular, an increase in TFP. The finding is not driven by price dispersion within the group of sticky-price firms, which leads to TFP losses conditional on monetary policy shocks of any sign.

Figure 4: Model responses to a monetary policy shock



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

rule in stabilizing output, if the monetary authority in our model (mis)perceives observed aggregate TFP movements after monetary policy shocks to be the response to exogeneous aggregate productivity shocks. In particular, we compute natural output $\tilde{Y}_t$ counterfactually as the flexible price output series after an aggregate productivity shocks that replicates the aggregate TFP response to a monetary policy shock. For comparability with our baseline model, we recalibrate σ_ν to ensure the same interest rate response to a one-standard deviation monetary policy shock, but keep all other parameters unchanged. Figure 13 in the Appendix compares the baseline (solid) to the (mis)perceived Taylor rule (dotted). We find that output drops by 7% more and aggregate TFP by 8% more if the monetary authority wrongly attributes aggregate TFP fluctuations to productivity shocks.

This finding highlights the importance for monetary authorities to assess the sources of observed aggregate TFP fluctuations. Figure 14 in the Appendix shows the response to an exogenous contractionary productivity shock with the size and persistence that matches the endogenous response of TFP to a monetary policy shock. Compared to the monetary policy shock, the output response is about one third smaller. In particular, the behavior of labor and markup dispersion helps to discriminate between productivity and monetary policy shocks. A productivity shock increases labor and lowers markup dispersion, while the monetary policy shock has the opposite effects.

The role of the substitution elasticity and the Frisch elasticity

Two parameters are critical for the transmission of monetary policy shocks to aggregate TFP. First, the Frisch elasticity of labor supply determines the response of marginal costs to shocks. The more volatile marginal costs, the stronger our precautionary price setting mechanism. Second, the elasticity of substitution between differentiated goods specifies how costly it is for consumers to substitute away from high-markup goods. Thereby it determines the TFP loss from markup dispersion. We examine the quantitative implications of changes in these two parameter values.

Figure 15 in the Appendix shows the response to a contractionary monetary policy shock when the Frisch elasticity of labor supply is set to $\varphi = 1$. The effect of monetary policy on aggregate TFP almost vanishes and the drop in GDP is almost completely generated by a decline in labor. Because marginal costs are less volatile, sticky-price firms exert less precautionary price setting; they charge only 0.65% higher markups. The TFP losses around the stochastic steady state are therefore smaller. The drop in GDP is of the same size as in the baseline calibration. However, this drop is then counterfactually generated solely by a drop in labor.

Figure 16 in the Appendix shows the response to a contractionary monetary policy shock when the elasticity of substitution between differentiated goods is set to $\eta = 21$, as in Fernandez-Villaverde et al. (2015). A larger elasticity of substitution increases the costs of substituting away from high-markup goods. Hence, any given dispersion in markups leads to a stronger decline in aggregate TFP. Compared to the baseline model, the markup differences in the stochastic steady state are comparable in magnitude, while the equilibrium responses of markups to monetary policy shocks are dampened. However, their effects on aggregate TFP are comparable in magnitude to the baseline. We consider two cases: (i) we keep $\varphi = 1/0.125$, and (ii) we recalibrate φ to match again the relative response of labor. Interestingly, under $\eta = 21$ the calibrated Frisch elasticity increases to $\varphi = 1/0.4$.

5. CONCLUSION

Despite the prominent role of productivity fluctuations in modern business cycle theory, surprisingly few papers have studied the nexus between monetary policy and aggregate productivity. Traditionally, monetary policy shocks have been considered as aggregate demand shocks, and productivity shocks as aggregate supply shocks. Our paper blurs this distinction by emphasizing the aggregate productivity effects of monetary policy shocks.

We argue that heterogeneity in firm-level price-setting frictions play a central role to understand the response of aggregate productivity to monetary policy shocks. Firms with stickier prices set higher markups because of precautionary price setting motives. Contractionary monetary policy shocks increase the relative markup of firms with stickier prices and thereby raise the dispersion in markups across firms. This is associated with a reallocation of resources across firms that lowers aggregate productivity. In this paper, we have characterized this mechanism, provided empirical evidence in its support, and proposed a quantitative New Keynesian model to show the quantitative relevance of the mechanism.

REFERENCES

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

Baqaee, D. R. and E. Farhi (2018): "Productivity and Misallocation in General Equilibrium," *NBER Working Paper Series.*

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

Bayer, C., A. M. Mecikovsky, and M. Meier (2018): "Misallocation, Markups, and Technology," *mimeo U Mannheim.*

Bils, M. and P. J. Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.

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.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review*, 101, 471–75.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?" *Handbook of Macroeconomics*, 1, 65–148.

——— (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.

Clayton, C., X. Jaravel, and A. Schaab (2018): "Heterogeneous Price Rigidities and Monetary Policy," mimeo, London School of Economics.

De Loecker, J. and J. Eeckhout (2018): "Global Market Power," *mimeo.*

De Loecker, J., J. Eeckhout, and G. Unger (2018): "The Rise of Market Power and the Macroeconomic Implications," *mimeo.*

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

Evans, C. L. (1992): "Productivity shocks and real business cycles," *Journal of Monetary*

Economics, 29, 191–208.

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.

Galí, J. (2015): *Monetary Policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and its Applications*.

Garga, V. and S. R. Singh (2019): "Output Hysteresis and Optimal Monetary Policy," Tech. rep., University of California Davis.

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

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

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

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

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

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 (2018): "Deconstructing monetary policy surprises: the role of information shocks," .

Jeenas, P. (2018): "Firm Balance Sheet Liquidity, Monetary Policy Shocks, and Investment Dynamics," *mimeo*.

Jordà, Ò. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

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

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

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

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 held at the Stern School of Business at New York University.

Nakamura, E. and J. Steinsson (2008): "Five Facts about Prices: A Reevaluation of Menu Cost Models," *The Quarterly Journal of Economics*, 123, 1415–1464.

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

Nekarda, C. J. and V. A. Ramey (2011): "Industry evidence on the effects of government spending," *American Economic Journal: Macroeconomics*, 3, 36–59.

——— (2013): "The Cyclical Behavior of the Price-Cost Markup," Working Paper 19099, National Bureau of Economic Research.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 703–708.

Ottonello, P. and T. Winberry (2018): "Financial Heterogeneity and the Investment Channel of Monetary Policy," .

Pasten, E., R. Schoenle, and M. Weber (2018): "The Propagation of Monetary Policy Shocks in a Heterogeneous Production Economy," *mimeo*.

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, Elsevier B.V., vol. 2, 71–162, 1 ed.

Romer, C. D. and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

Appendix

APPENDIX A: DESCRIPTIVE STATISTICS

TABLE II

Summary statistics for monetary policy shocks

(a) Daily shocks

	mean	sd	min	max	count
Three-month Fed funds future surprises	-0.48	3.71	-18.50	11.50	189
... unscheduled meetings and conference calls included	-0.77	4.59	-37.00	11.50	219
... purged of Greenbook forecasts	0.00	3.52	-11.77	11.99	152
... sign-restricted stock market comovement	-0.26	3.39	-18.50	11.50	189
... QE announcements excluded	-0.41	3.53	-16.00	11.50	182
Current-month Fed funds future surprises	-0.46	4.05	-20.62	13.00	189
'Policy indicator' surprise	0.00	3.27	-15.84	10.26	180

(b) Quarterly 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
Current-month Fed funds future surprises	-0.98	4.21	-17.78	13.57	94
'Policy indicator' surprise	-0.05	3.43	-14.13	7.45	94

Notes: Summary statistics for monetary policy shocks in basis points. Daily shocks are computed from Equation (2.1). Quarterly shocks are aggregated from daily frequency using Equation (2.2). All shocks are in basis points.

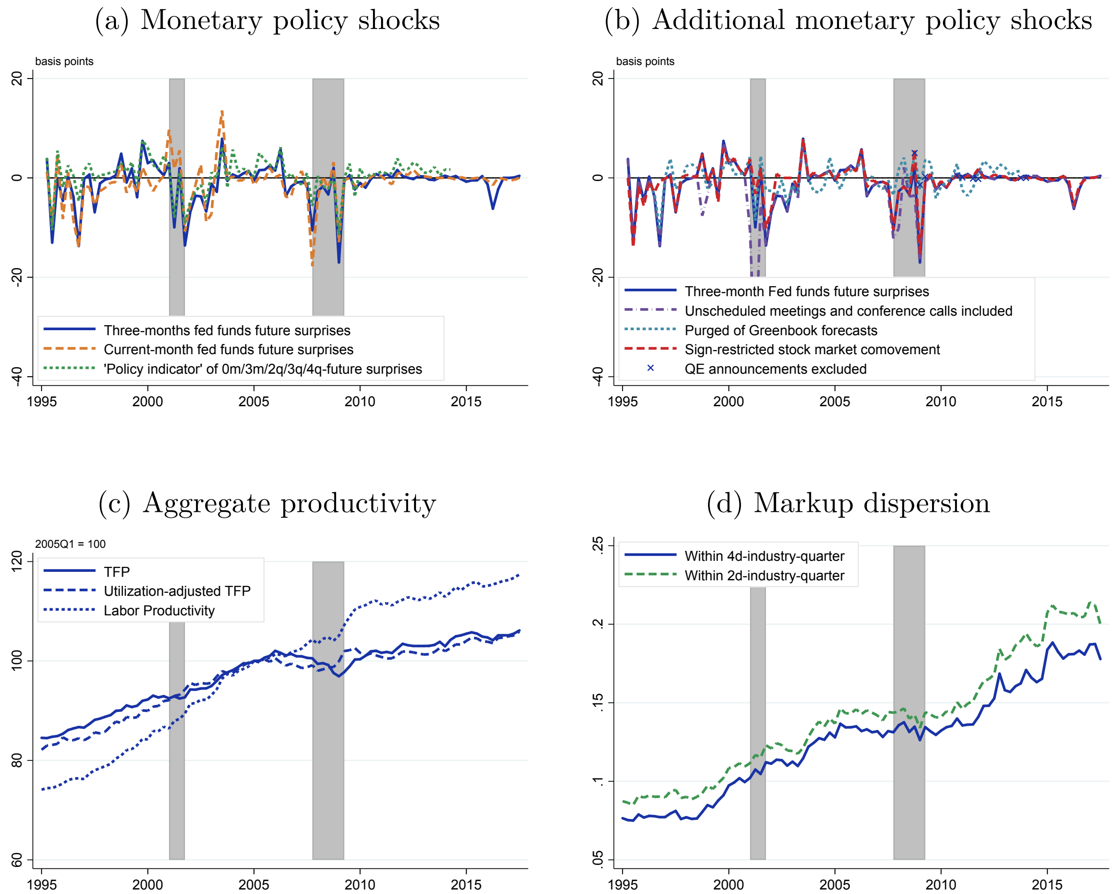
TABLE III

Summary statistics for Compustat data

	mean	sd	min	max	count
Sales	517.23	2804.95	0.00	155150.27	413727
Property, Plant	954.55	5633.15	0.00	269010.41	413727
Variable costs	365.62	2106.25	0.00	119295.70	413727
Assets	2436.84	15525.16	0.00	860586.34	411873

Notes: Summary statistics for Compustat data. All variables are in millions of 2010Q1 US$.

Figure 5: Time series plots



(a) Monetary policy shocks

(b) Additional monetary policy shocks

(c) Aggregate productivity

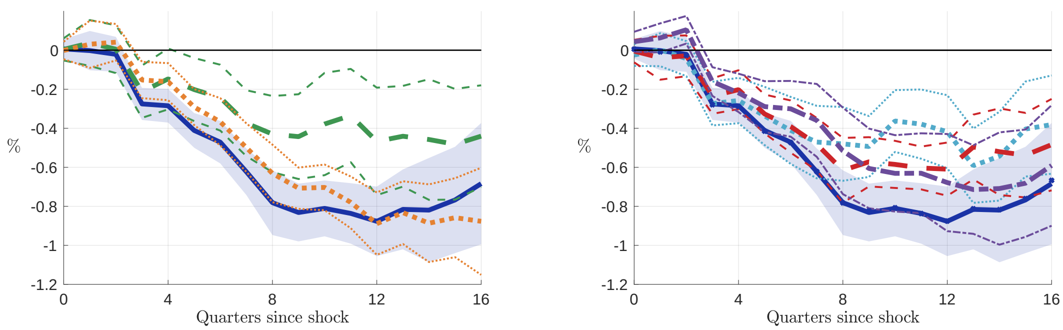
(d) Markup dispersion

Notes: Aggregate productivity, markup dispersion, and monetary policy shocks are at quarterly frequency. Shaded gray areas indicate NBER recession dates.

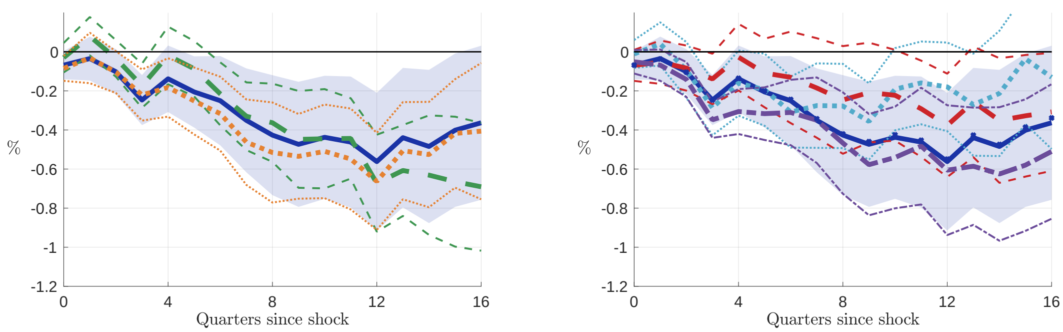
APPENDIX B: ROBUSTNESS OF EMPIRICAL FINDINGS

Figure 6: Productivity IRFs for additional monetary policy shocks

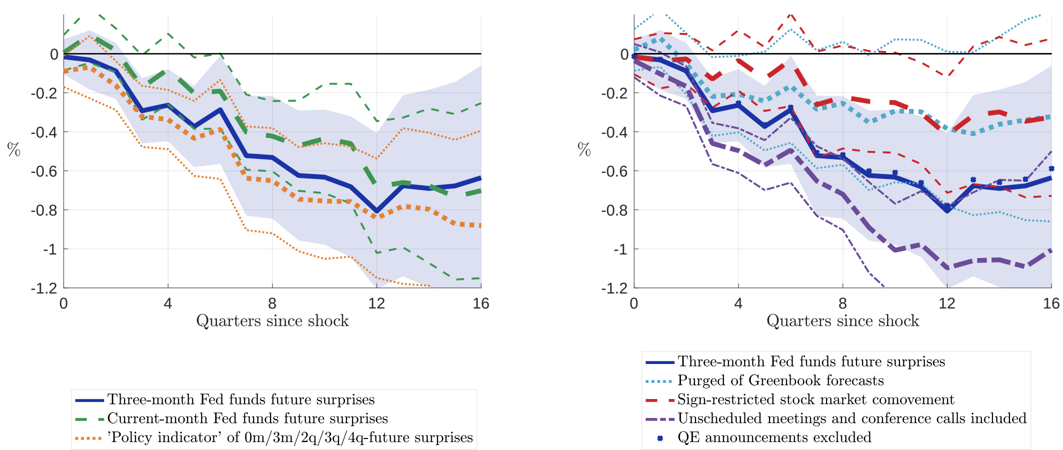
(a) TFP



(b) Utilization-adjusted TFP



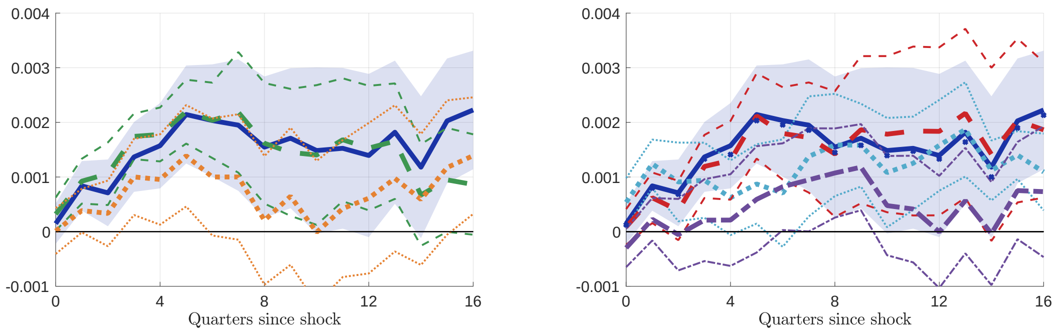
(c) Labor productivity



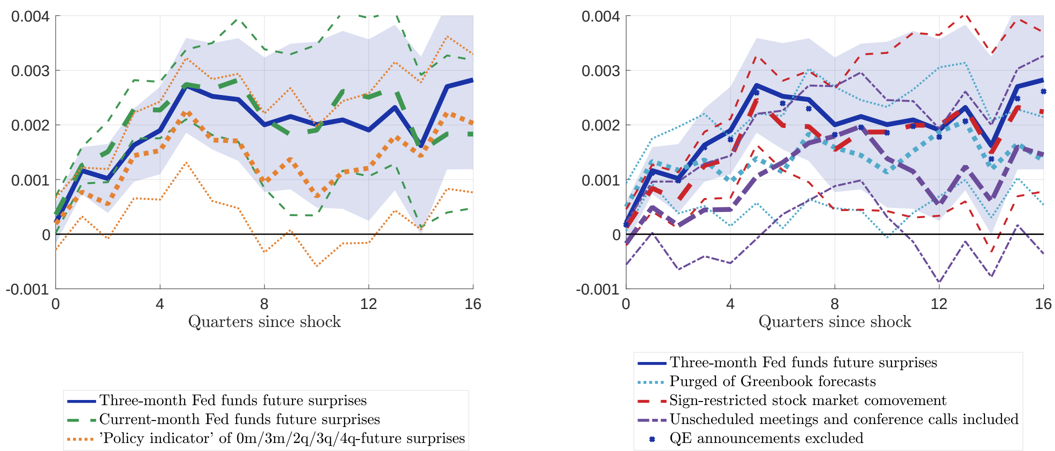
Notes: The shaded and bordered areas indicate a one standard error band based on the Newey-West estimator.

Figure 7: Markup dispersion IRFs for additional monetary policy shocks

(a) Within four-digit industry-quarter



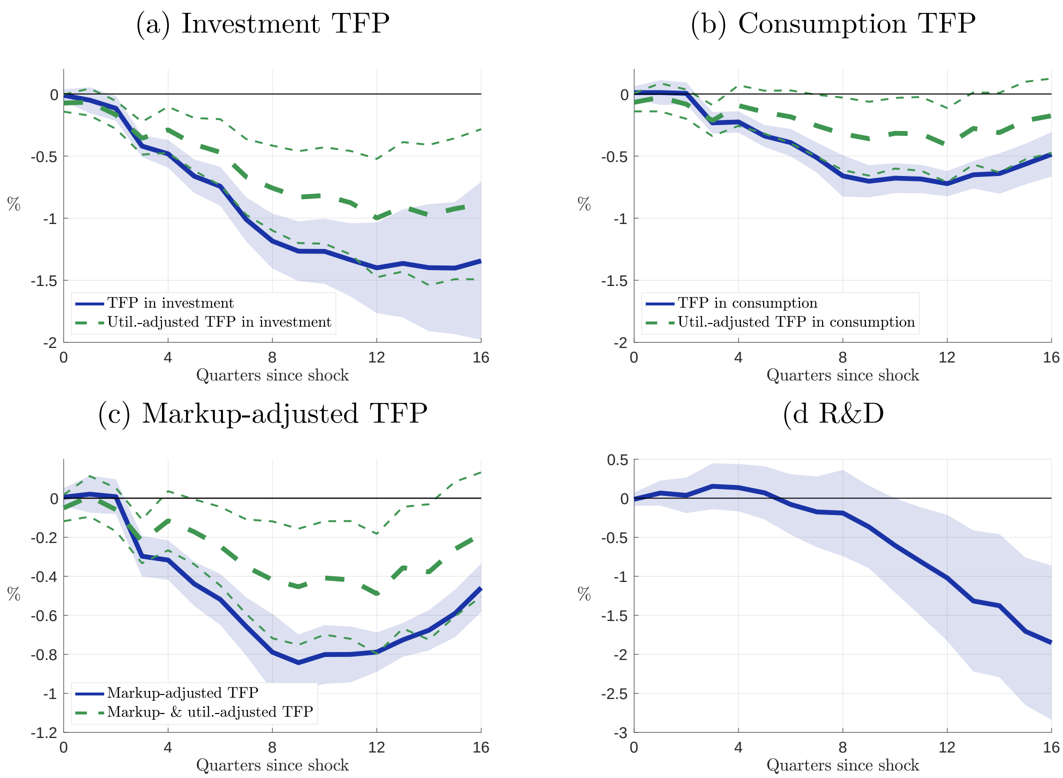
(b) Within two-digit industry-quarter



Notes: The shaded and bordered areas indicate a one standard error band based on the Newey-West estimator.

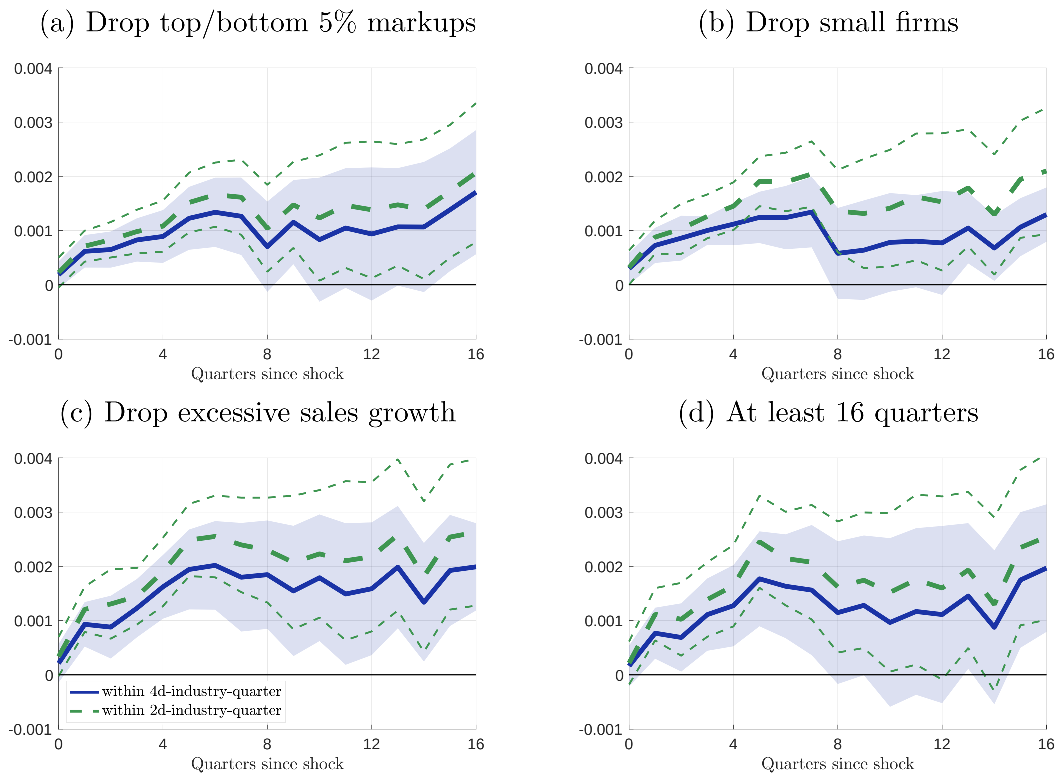
APPENDIX C: ADDITIONAL EMPIRICAL RESULTS

Figure 8: Further productivity responses and the response of R&D



Notes: Investment-TFP and Consumption-TFP are from Fernald (2014). Inference is based on Newey-West standard errors. The shaded and bordered areas show a one standard error band for the response of TFP.

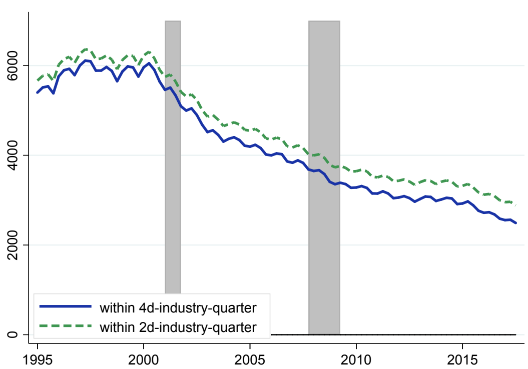
Figure 9: Responses of markup dispersion to a monetary policy shock: various data treatments



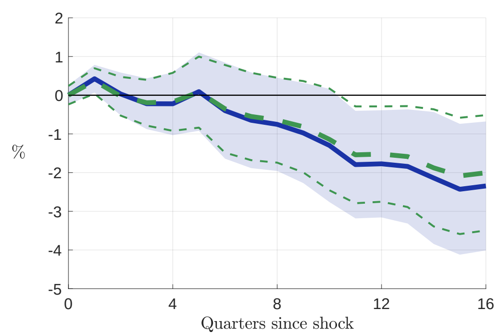
Notes: The plots show the responses of markup dispersion to a one-standard deviation contractionary monetary policy shock. Markup dispersion is within two-digit and four-digit industry-quarters, respectively. *5% trimming* drops the corresponding tails within quarters. *Drop small firms* drops observations with real sales below 1 mln (in 2010 US$). *Drop excessive growth* drops observations with real sales growth above +200% or below -66%. *At least 16 quarters* restricts the sample to firms with at least 16 quarters of observations. Inference as in Figure 1.

Figure 10: Number of firm-level observations over time and after shocks

(a) Time trend

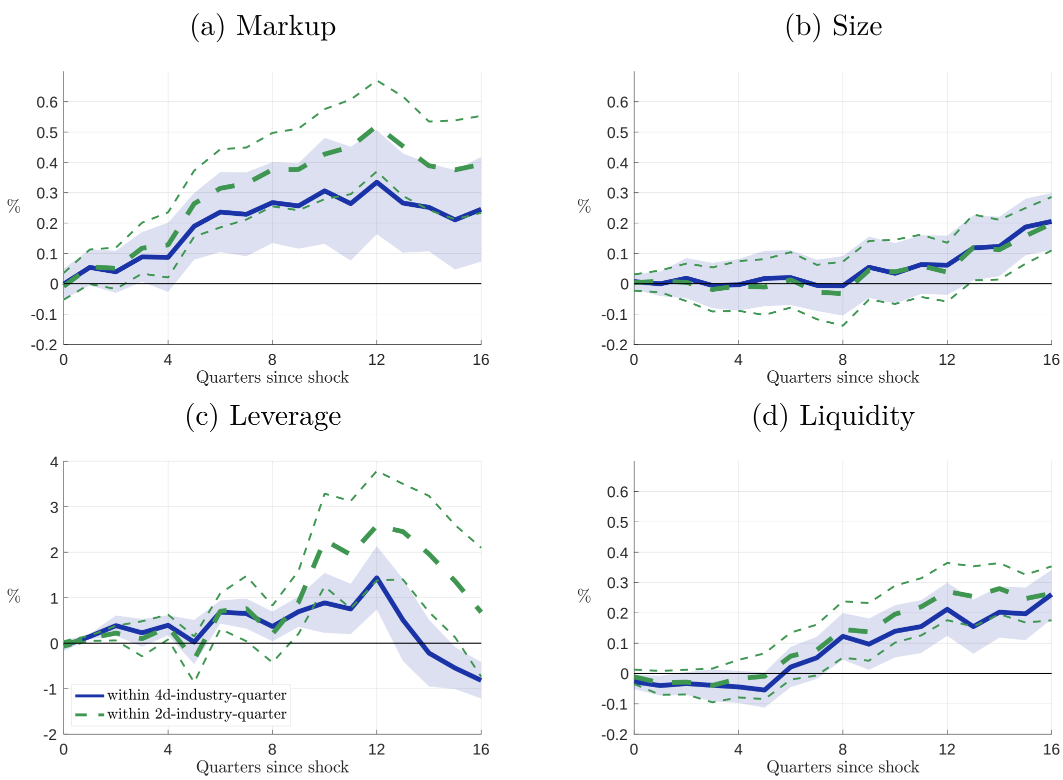


(b) Response to shock



Notes:

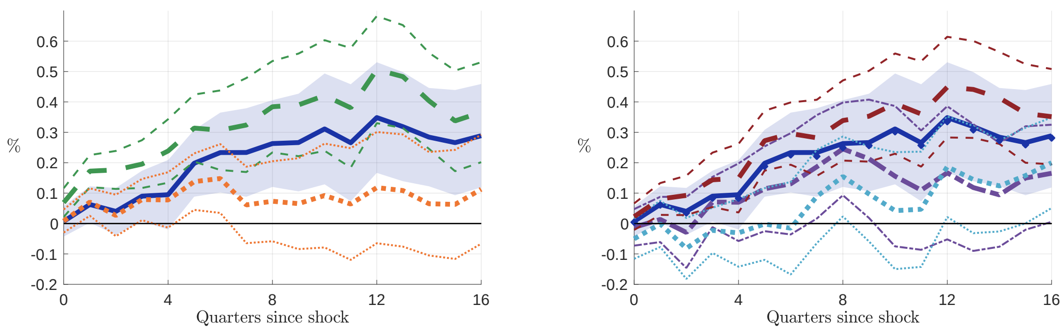
Figure 11: Firm-level heterogeneity in the markup response to monetary policy shocks: other interaction variables (including controls)



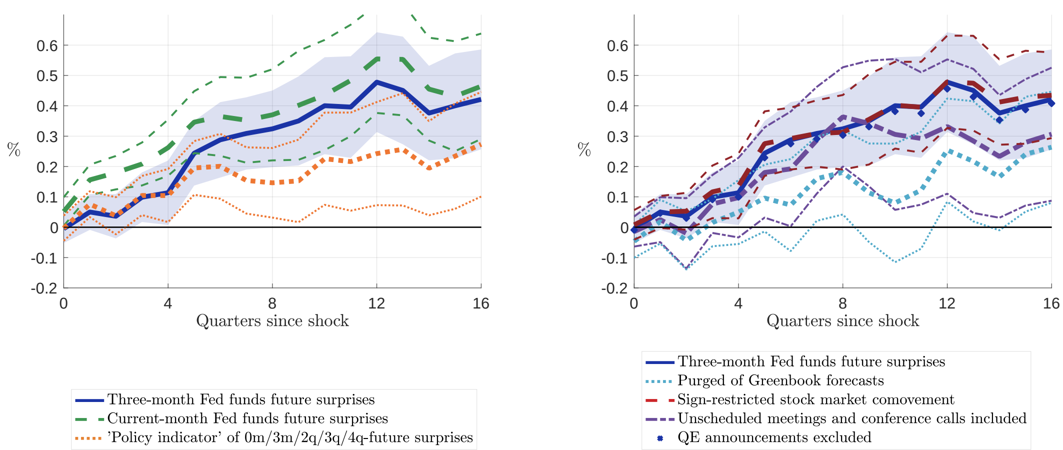
Notes: Inference is based on two-way clustered standard errors by firms and quarters. The shaded area is a one-s.e. band.

Figure 12: Firm-heterogeneity in markup response for additional monetary policy shocks (including controls)

(a) Within four-digit industry-quarter



(b) Within two-digit industry-quarter



Notes: The shaded and bordered areas indicate a one standard error band clustered by firms and quarters.

APPENDIX D: PROOFS

D.1. *Proof of Proposition 1*

First, we rewrite the first-order condition

$$P_{it}^* = = \frac{\eta}{\eta-1} P_t W_t \frac{1+\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}} \equiv \frac{\eta}{\eta-1} P_t W_t \Psi_{it}, \tag{D.1}$$

where

$$\tilde{\theta}_i \equiv \frac{\beta\theta_i}{1-\beta\theta_i}, \tag{D.2}$$

$$C_{it} \equiv E_t \frac{W_{t+1}}{W_t} \left(\frac{P_{t+1}}{P_t}\right)^{\eta} \frac{Y_{t+1}}{Y_t}, \tag{D.3}$$

$$D_{it} \equiv E_t \left(\frac{P_{t+1}}{P_t}\right)^{\eta-1} \frac{Y_{t+1}}{Y_t}. \tag{D.4}$$

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

$$C_{it} = \frac{\bar{W}\bar{P}^{\eta}\bar{Y}}{W_t P_t^{\eta} Y_t} \exp\left\{\eta(\eta-1)\frac{\sigma_p^2}{2} + \eta\sigma_{pw} + \eta\sigma_{py}\right\},$$

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

Since $\tilde{\theta}_i \in (0,1)$, we have $\Psi_{it} > 1$ if $C_{it} > D_{it}$. For $P_t = \bar{P}$ and $W_t = \bar{W}$, we have $C_{it} > D_{it}$ if

$$(\eta-1)\sigma_p^2 + \sigma_{py} + \eta\sigma_{pw} + \sigma_{wy} > 0. \tag{D.5}$$

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

$$\frac{\partial \Psi_{it}}{\partial \tilde{\theta}_i} = \frac{C_{it} - D_{it}}{(1+\tilde{\theta}_i D_{it})^2} > 0, \quad \text{and hence} \quad \frac{\partial \Psi_{it}}{\partial \theta_i} > 0. \tag{D.6}$$

D.2. *Proof of Proposition 2*

We first examine a **transitory** change in W_t away from $\bar{W}$. The expected pass-through is

$$\bar{\varepsilon}_{it} = (1-\theta_i)\frac{\partial \log P_{it}}{\partial \log W_t} = (1-\theta_i)\left(1 + \frac{\partial \log \Psi_{it}}{\partial \log W_t}\right) \equiv (1-\theta_i)\left(1+\Phi_{it}\right), \tag{D.7}$$

and

$$\Phi_{it} = \frac{\tilde{\theta}_i \frac{\partial C_{it}}{\partial \log W_t}(1+\tilde{\theta}_i D_{it}) - (1+\tilde{\theta}_i C_{it})\tilde{\theta}_i \frac{\partial D_{it}}{\partial \log W_t}}{(1+\tilde{\theta}_i D_{it})^2} \Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}} \Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i C_{it}} < 0. \tag{D.8}$$

Hence pass-through becomes

$$\bar{\varepsilon}_{it} = \frac{1-\theta_i}{1+\tilde{\theta}_i C_{it}} \in (0,1). \tag{D.9}$$

In addition, the pass-through falls in θ_i,

$$\text{(D.10)} \qquad \frac{\partial \bar{\varepsilon}_{it}}{\partial \theta_i} = -(1+\Phi_{it}) + (1-\theta_i)\frac{\partial \Phi_{it}}{\partial \theta_i} < 0$$

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

$$\text{(D.11)} \qquad \frac{\partial \log P^*_{it}}{\partial \log \bar{W}} = 1$$

Expected pass-through is then

$$\text{(D.12)} \qquad \bar{\varepsilon}_{it} = 1 - \theta_i,$$

which is falling in θ_i.

D.3. ***Proof of Proposition 3***

in order to write the first-order condition, equation (3.8), more compactly, we first define

$$\text{(D.13)} \qquad C_{it} = \left[\left(\frac{P_{it}}{P_{i,t-1}} - 1\right)\frac{P_{it}}{P_{i,t-1}}\right],$$

$$\text{(D.14)} \qquad D_{it} = \mathbb{E}_t\left[\left(\frac{P_{i,t+1}}{P_{it}} - 1\right)\frac{P_{i,t+1}}{P_{it}}\right].$$

We then 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

$$\text{(D.15)} \qquad \bar{P}_{it} = \frac{\eta}{\eta-1}P_t W_t.$$

In addition, we obtain

$$\text{(D.16)} \qquad \bar{C}_{it} = \left(\frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} - 1\right)\frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} = (\Pi_{pt}\Pi_{wt})^2 - \Pi_{pt}\Pi_{wt},$$

$$\text{(D.17)} \qquad \bar{D}_{it} = E_t\left[\left(\frac{\bar{P}_{i,t+1}}{\bar{P}_{it}} - 1\right)\frac{\bar{P}_{i,t+1}}{\bar{P}_{it}}\right] = \frac{\exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\}}{(\Pi_{pt}\Pi_{wt})^2} - \frac{\exp\left\{\sigma_{pw}\right\}}{\Pi_{pt}\Pi_{wt}}.$$

We next use a first-order approximation of the first-order condition in (3.8) at $\bar{\phi}_i$ and with respect to ϕ_i, $\log P_{it}$. Denoting deviations of those variables from their respective values at $\bar{\phi}_i$ by $\text{dlog}P_{it} = \log P_{it} - \log \bar{P}_{it}$ and $d\theta_i = \theta_i$, we obtain

$$\text{(D.18)} \qquad (1-\eta)^2\left(\frac{P_{it}}{\bar{P}_t}\right)^{1-\eta} Y_t \text{dlog}P_{it} - \eta^2 W_t\left(\frac{\bar{P}_{it}}{P_t}\right)^{-\eta} Y_t \text{dlog}P_{it} - \bar{C}_{it}d\theta_i + \bar{D}_{it}d\theta_i = 0.$$

After some reformulation this yields

$$\text{(D.19)} \qquad \frac{\text{dlog}P_{it}}{d\theta_i} = \frac{\bar{D}_{it} - \bar{C}_{it}}{(\eta-1)\left(\frac{\eta}{\eta-1}W_t\right)^{1-\eta} Y_t} \equiv \Psi_{it},$$

and hence

$$\log P_{it} \approx \log \bar{P}_{it} + \Psi_{it} d\theta_i. \tag{D.20}$$

The dynamic optimal markup is above the static optimal one if $P_{it} \leq \bar{P}_{it}$, which is satisfied if

$$\Psi_{it} < 0, \quad \text{and, which is satisfied if} \quad \bar{D}_{it} > \bar{C}_{it} \tag{D.21}$$

If we assume that $P_t = \bar{P}$ and $W_t = \bar{W}$, then $\Psi_{it} < 0$ if

$$\sigma_p^2 + \sigma_w^2 + 3\sigma_{pw} > 0. \tag{D.22}$$

Under the same condition, the price P_{it} decreases in θ_i.

D.4. ***Proof of Proposition 4***

We start from the approximated optimal price setting policy in equation (D.20). The pass-through from changes in the real wage W_t to price is

$$\varepsilon_{it} = 1 + \frac{\partial \Psi_i}{\partial \log W_t} d\theta_i. \tag{D.23}$$

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

$$\frac{\partial \Psi_i}{\partial \log W_t} < 0, \tag{D.24}$$

which is equivalent to examining the conditions for

$$\frac{\partial \bar{D}_{it}}{\partial \log W_t} - \frac{\partial \bar{C}_{it}}{\partial \log W_t} + (\eta - 1)(\bar{D}_{it} - \bar{C}_{it}) < 0. \tag{D.25}$$

We distinguish between a permanent and a transitory change in $\log W_t$. For a **transitory** change we have

$$\frac{\partial \bar{C}_{it}}{\partial \log W_t} = 2(\Pi_{pt}\Pi_{wt})^2 - \Pi_{pt}\Pi_{wt}, \tag{D.26}$$

$$\frac{\partial \bar{D}_{it}}{\partial \log W_t} = -2(\Pi_{pt}\Pi_{wt})^{-2} \exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\} + (\Pi_{pt}\Pi_{wt})^{-1} \exp\left\{\sigma_{pw}\right\}. \tag{D.27}$$

Evaluating these derivatives at $P_t = \bar{P}$ and $W_t = \bar{W}$, we obtain

$$\frac{\partial \bar{C}_{it}}{\partial \log W_t} = 1, \tag{D.28}$$

$$\frac{\partial \bar{D}_{it}}{\partial \log W_t} = -2\exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\} + \exp\left\{\sigma_{pw}\right\}. \tag{D.29}$$

Then $\frac{\partial \Psi_i}{\partial \log W_t} < 0$ if

$$\eta < \tilde{\eta}^{\text{transitory}} = 2 + \frac{1 + \exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\}}{\exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\} - \exp\left\{\sigma_{pw}\right\}} \tag{D.30}$$

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

$$\text{(D.31)} \qquad \frac{\partial \bar{C}_{it}}{\partial \log W_t} = 2(\Pi_{pt}\Pi_{wt})^2 - \Pi_{pt}\Pi_{wt},$$

$$\text{(D.32)} \qquad \frac{\partial \bar{D}_{it}}{\partial \log W_t} = (2-2)(\Pi_{pt}\Pi_{wt})^{-2}\exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\} + (1-1)(\Pi_{pt}\Pi_{wt})^{-1}\exp\left\{\sigma_{pw}\right\} = 0.$$

Evaluated at $P_t = \bar{P}$ and $W_t = \bar{W}$, we obtain

$$\text{(D.33)} \qquad \frac{\partial \bar{C}_{it}}{\partial \log W_t} = 1,$$

$$\text{(D.34)} \qquad \frac{\partial \bar{D}_{it}}{\partial \log W_t} = 0.$$

Then $\frac{\partial \Psi_i}{\partial \log W_t} < 0$ if

$$\text{(D.35)} \qquad \eta < \tilde{\eta}^{\text{permanent}} = 1 + \frac{1}{\exp\left\{\sigma_p^2 + \sigma_w^2 + 4\sigma_{pw}\right\} - \exp\left\{\sigma_{pw}\right\}}$$

Clearly, $\eta^{\text{permanent}} < \eta^{\text{transitory}}$ and we define

$$\text{(D.36)} \qquad \tilde{\eta} \equiv \eta^{\text{permanent}} \approx 1 + \frac{1}{\sigma_p^2 + \sigma_w^2 + 3\sigma_{pw}}.$$

D.5. ***Proof of Proposition 5***

The log markup is

$$\text{(D.37)} \qquad \log \mu_{it} = \log P_{it} - \log P_t - \log W_t.$$

Taking the total differential yields

$$\text{(D.38)} \qquad \text{dlog}(\mu_{it}) = \text{dlog}(P_{it}) - \text{dlog}(P_t) - \text{dlog}(W_t) = -\text{dlog}(P_t) - (1-\varepsilon_{it})\text{dlog}(W_t).$$

Denote by $x_t = \text{dlog}(W_t)$ a small change in real marginal costs. The markup excluding the change x_t is denoted μ_{it} and including x_t we write μ'_{it}. Then we

$$\text{(D.39)} \qquad \log \mu'_{it} = \log \mu_{it} - \text{dlog}(P_t) - (1-\varepsilon_{it})x_t = \log \mu_{it} - (1 + a_t - \varepsilon_{it})x_t,$$

where $a_t = \frac{\text{dlog}(P_t)}{\text{dlog}(W_t)}$ captures the co-movement of aggregate price and real marginal cost. The change x_t is small so that $x_t^2 \approx 0$. Then the variance of the markup after x_t is given by

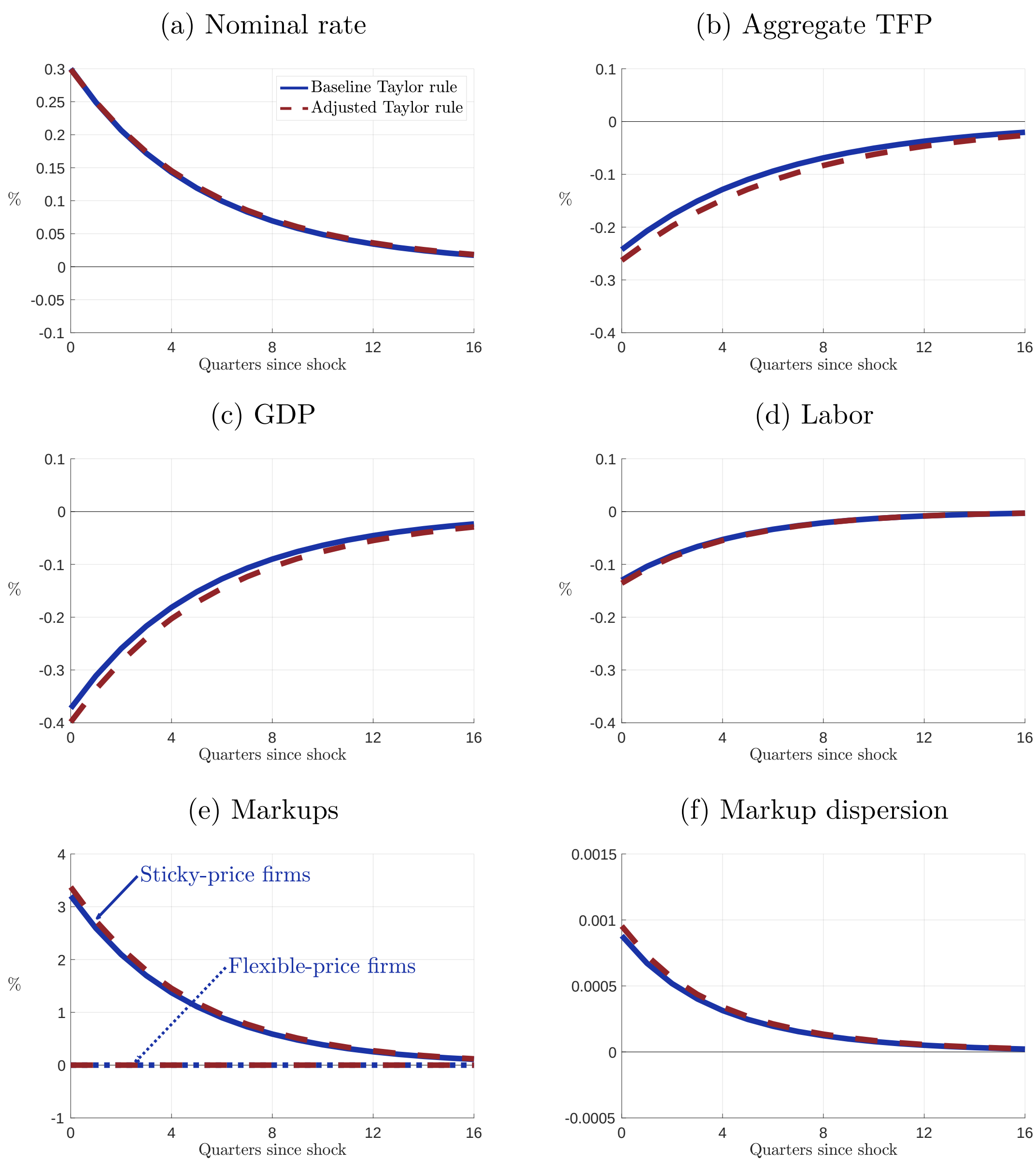
$$\text{(D.40)} \qquad V[\log \mu'_{it}] = V[\log \mu_{it}] + 2Cov(\log \mu_{it}, \epsilon_{it})x_t$$

Hence, if $Corr(\log \mu_i, \epsilon_i) < 0$, then the markup variance falls in x_t

$$\frac{\partial V[\log \mu'_{it}]}{\partial x_t} < 0.$$

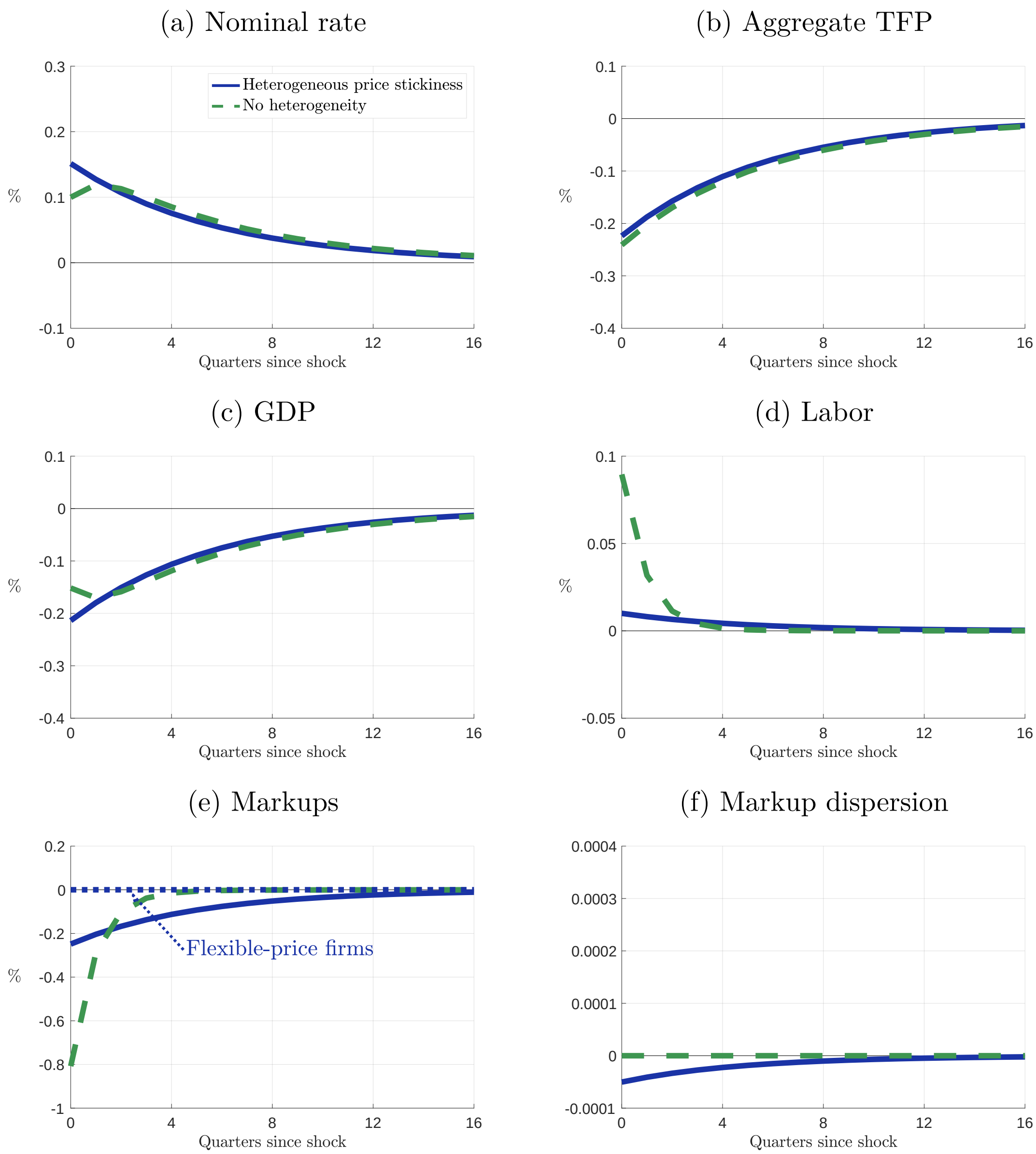
APPENDIX E: ADDITIONAL MODEL RESULTS

Figure 13: Model responses to a monetary policy shock: Alternative Taylor rule



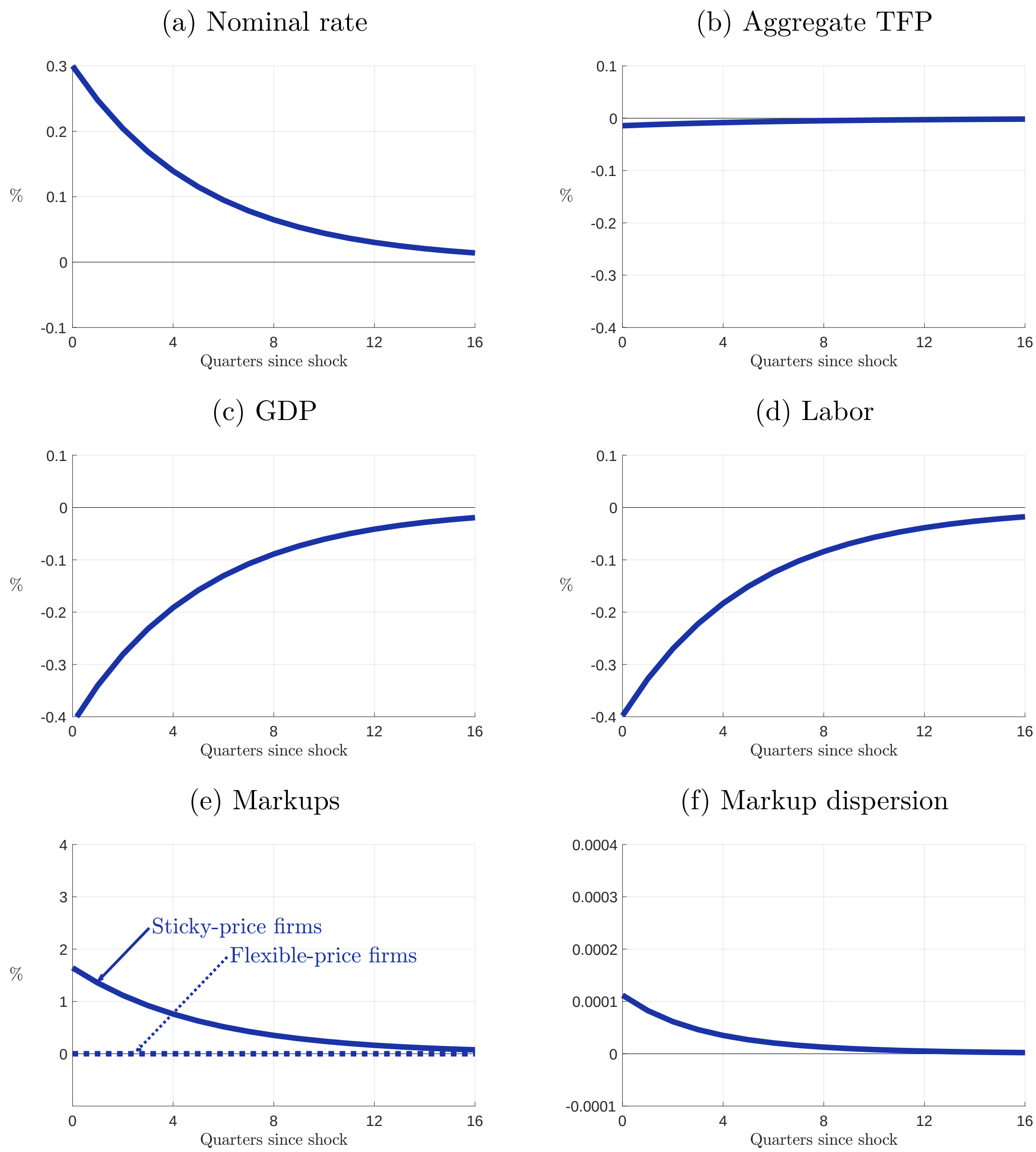
Notes: This figure shows impulse responses to a one standard deviation monetary policy shock.

Figure 14: Model responses to a TFP shock



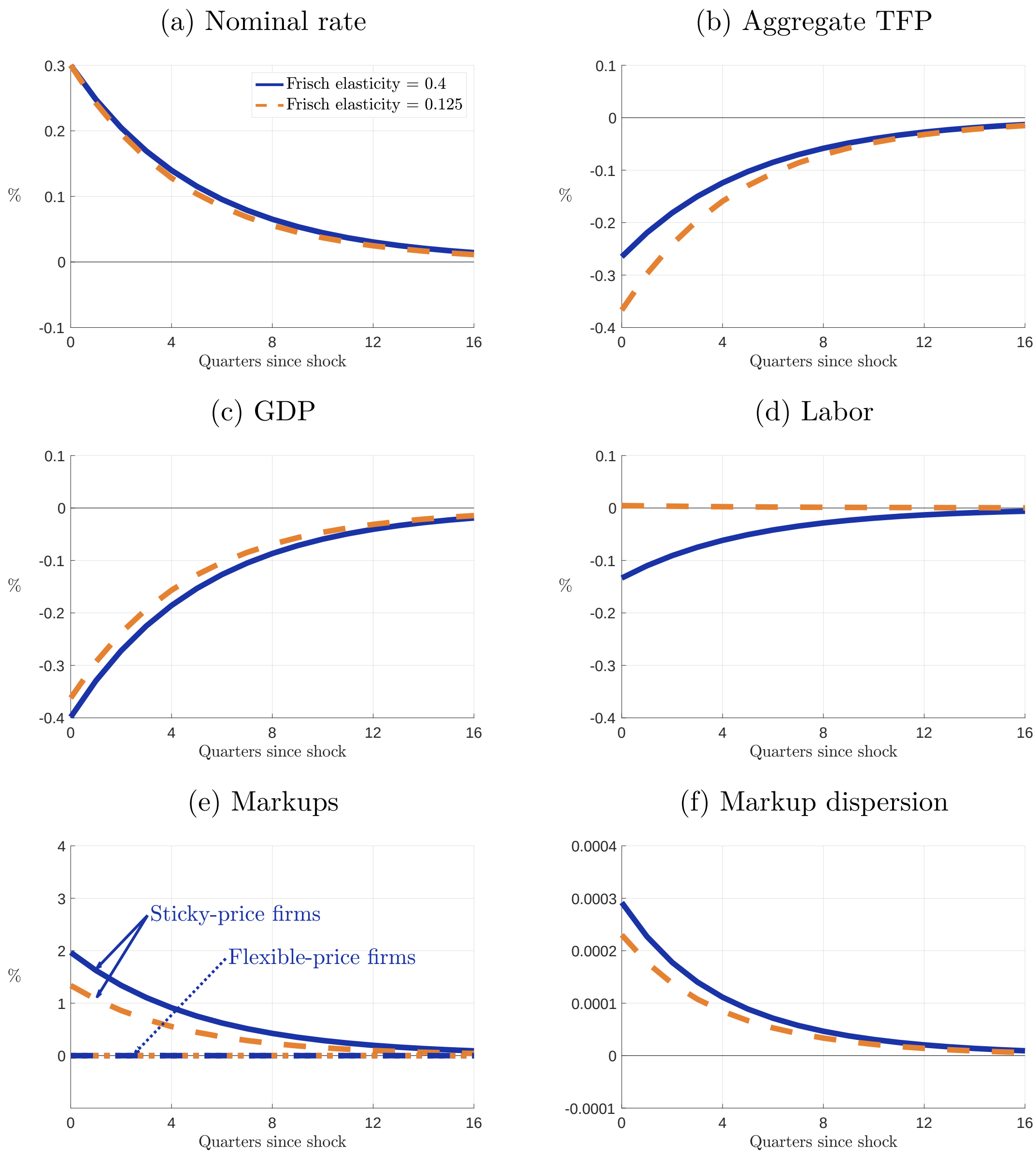
Notes: This figure shows impulse responses to an exogenous TFP shock that is calibrated to match the endogenous response of aggregate TFP to a monetary policy shock (see Figure 4). This results in $\sigma_a = 0.24\%$ and $\rho_a = 0.84$.

Figure 15: Model responses to a monetary policy shock:
high Frisch elasticity of labor supply ($\varphi = 1$)



Notes: This figure shows impulse responses to a one standard deviation monetary policy shock when the Frisch elasticity of labor supply is set to $\varphi = 1$.

Figure 16: Model responses to a monetary policy shock: high elasticity of substitution ($\eta = 21$)



Notes: This figure shows impulse responses to a one standard deviation monetary policy shock when the elasticity of substitution between differentiated goods is set to $\eta = 21$. We show the responses for the same Frisch elasticity as in the baseline parametrization, as well as the responses for a Frisch elasticity of 0.4, which matches the empirical contributions of aggregate TFP and labor to the drop in GDP.