Monetary Policy, Markup Dispersion, and Aggregate TFP*

Matthias Meier† Timo Reinelt‡

January 25, 2020

Abstract

This paper studies a novel transmission mechanism of monetary policy. If firms with lower pass-through from marginal costs to prices charge higher markups, then contractionary monetary policy shocks increase the dispersion in markups and lower aggregate productivity. This mechanism is supported by US data: (a) firms with higher markups adjust prices less frequently, and contractionary monetary policy shocks (b) increase the relative markup of firms that adjust prices less frequently, (c) increase markup dispersion across firms, and (d) decrease aggregate productivity. We show analytically that one explanation for the negative correlation between firm-level pass-through and markup is heterogeneity in price-setting frictions. Firms with stickier prices optimally set higher markups for precautionary reasons. In a calibrated New Keynesian model with heterogeneous price rigidity, monetary policy shocks explain substantial fluctuations in aggregate productivity.

Keywords: Monetary policy, heterogeneous price rigidity, precautionary price setting, markup dispersion, aggregate productivity

*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 helpful comments and suggestions at an early stage of this project. We also wish to thank numerous discussants and 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. Timo Reinelt gratefully acknowledges financial support from Stiftung Geld & Währung. 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, Department of Economics and Center for Doctoral Studies in Economics (CDSE), Block L7, 3–5, 68161 Mannheim, Germany; E-mail: timo.reinelt@gess.uni-mannheim.de

1 Introduction

This paper revisits the monetary transmission channel, a matter of substantial importance for positive and normative questions in business cycle research. The workhorse theoretical framework to study monetary transmission, the New Keynesian model, is built on the assumption of rigid nominal prices. In fact, micro-level price data underpin this assumption.[1] Price data also reveal sizable heterogeneity in the price adjustment frequency across and within industries, which suggests heterogeneity in price-setting frictions. Compared to the baseline New Keynesian model, such heterogeneity has important implications. Monetary policy shocks have larger and more persistent effects and the divine coincidence disappears.[2] In this paper, we characterize a novel and overlooked theoretical implication of heterogeneous price stickiness, and show that it is both empirically supported and quantitatively important for the monetary transmission channel.

The mechanism is as follows. Suppose firms with more rigid prices charge higher markups. Contractionary monetary policy shocks further increase the markup of firms with more rigid prices relative to firms with more flexible prices. This widens the markup distribution across firms, which leads to lower allocative efficiency of inputs across firms, and hence lower measured aggregate total factor productivity (TFP).[3] Conversely, expansionary shocks lower markup dispersion and raise aggregate TFP. One explanation why firms with more rigid prices have higher markups is a precautionary price setting motive, which we analyze for various price setting frictions.[4,5]

We investigate the empirical relevance of this mechanism in US data. We combine high-frequency identified monetary policy shocks with quarterly firm-level balance-sheet data and price adjustment frequencies. To estimate firm-level markups, we follow the approach in Hall (1988) and De Loecker and Warzynski (2012). We derive four testable implications of the transmission mechanism, none of which is rejected by the data. First, firms with higher markups adjust prices less frequently, even when comparing firms within the same two or four-digit industries. Second, after contractionary monetary policy shocks, markups of firms with stickier prices increase by relatively more than markups of firms with more flexible prices. Third, within-industry markup dispersion increases after the same shock. The response is persistent and peaks about two years after the shock. Fourth, aggregate TFP falls after contractionary monetary policy shocks, even after adjusting for utilization.

[1] Evidence for the US is presented in Carlton (1986), Bils and Klenow (2004), Nakamura and Steinsson (2008), and Gorodnichenko and Weber (2016). Evidence for the Euro area is surveyed in Dhyne et al. (2006).

[2] For positive implications, see Carvalho (2006), Nakamura and Steinsson (2010), Carvalho and Schwartzman (2015), and Pasten et al. (2018). For normative implications, see Woodford (2010) and Eusepi et al. (2011).

[3] Markup (or price) dispersion (within sectors) is well known to lower aggregate TFP in the New Keynesian model, see, e.g., Galí (2015). Relatedly, the adverse effects of higher markup dispersion on aggregate productivity are also discussed in a macro development literature on factor misallocation, e.g., Hsieh and Klenow (2009).

[4] Because the profit function is asymmetric in a firm's price, firms with stronger price setting frictions optimally set higher markups. Firms respond similarly to higher uncertainty, see Fernandez-Villaverde et al. (2015).

[5] A competing explanation is that, in a menu cost environment, firms set lower markups and adjust prices more frequently if they face more elastic demand. However, all our empirical results hold irrespectively of differences in demand elasticities between narrowly-defined industries.

The estimated magnitudes are large. A monetary policy shock that lowers aggregate output by 1% two years after the shock lowers TFP by 0.8% and utilization-adjusted TFP by 0.4%. Through the lens of a model in the tradition of Hsieh and Klenow (2009), the increase in markup dispersion accounts for more than half of the utilization-adjusted aggregate TFP response at a two-year horizon. At more distant horizons, markup dispersion accounts for a decreasing fraction of the aggregate TFP response.

We study the quantitative relevance of our mechanism in a New Keynesian model with heterogenous price-setting frictions. In the calibrated model, firms with more rigid prices set higher markups. Hence, contractionary monetary policy shocks lead to higher markup dispersion and lower aggregate TFP. We calibrate heterogeneous price rigidities to the within-sector dispersion in price adjustment frequencies documented in Gorodnichenko and Weber (2016). We find that aggregate TFP falls by 0.34%, which is 60% of the empirical peak response of utilization-adjusted aggregate TFP.

According to the traditional view, fluctuations in aggregate productivity are driven by exogenous technology shocks. Whereas technology shocks change potential output, monetary policy shocks do not. With a Taylor rule that depends on the output gap, this can make a difference. Suppose the monetary authority calculates the output gap under the counterfactual belief that the aggregate TFP response to monetary policy shocks is driven by technology shocks. Then interest rates are adjusted less aggressively and monetary policy shocks have a roughly 20% larger effect on GDP.

The model solution technique plays an important role and is likely a reason why our mechanism has been overlooked in prior work. In the deterministic steady state of our model, all firms charge the same markup, irrespective of their price rigidity.[6] We compute the stochastic steady state, to which the economy converges in the presence of uncertainty but absent of shocks, using non-linear solution methods. This steady state captures precautionary price setting. Firms in the stickiest quintile charge a 10.9% higher markup than firms in the most flexible quintile.

Another interesting implication of the TFP channel of monetary policy is that markups can fall after contractionary monetary policy shocks. While the textbook New Keynesian model predicts that contractionary monetary policy raises markups, some empirical evidence points in the opposite direction, see Nekarda and Ramey (2019). We show that if the endogenous TFP response of monetary policy shocks is sufficiently strong, aggregate markup can decline after contractionary monetary policy shocks. Importantly, the argument extends to sector and even firm-level markups. If, for example, firms produce many goods of which prices are adjusted differently often, monetary policy has similar effects on firm-level TFP as on aggregate TFP.

This paper is closely related to three branches of literature. First, a growing literature studies the effects of heterogeneous price rigidity for monetary transmission, which we have referenced in the beginning of this paper. Our contribution is to show that such heterogeneity gives rise to TFP effects of monetary policy. Closely related is Baqaee and Farhi (2017), who show that negative money supply shocks lower aggregate TFP if sticky-price firms have higher ex-ante markups than flexible-

[6] In the presence of trend inflation, markup dispersion exists even in the deterministic steady state of a homogeneous firm model. However, with positive trend inflation, contractionary monetary policy shocks lower markup dispersion and increase aggregate TFP, see Ascari and Sbordone (2014).

price firms. Our results are less general than Baqaee and Farhi (2017) insofar as we abstract from input-output networks, but more general as we provide analytical results for dynamic environments. Importantly, we also show that a positive markup–rigidity correlation arises endogenously from heterogenous rigidity. In addition, we provide empirical evidence in support of the channel and a quantitative investigation.

Second, this paper confirms previous empirical findings that aggregate productivity responds to monetary policy shocks, e.g., Christiano et al. (2005), Moran and Queralto (2018), Garga and Singh (2019), and Jordà et al. (2020). The existing literature predominantly explains the aggregate productivity response as a response in average firm-level productivity. For example, variable utilization and fixed costs generate a productivity response in Christiano et al. (2005). However, they only explain a small fraction of the empirical productivity response. Moran and Queralto (2018) and Garga and Singh (2019) provide empirical evidence for an adverse R&D response to tighter monetary policy, which may ultimately affect productivity.

Third, this paper relates to a growing literature that studies allocative efficiency over the business cycle. Various business cycle shocks may transmit through misallocation: e.g., aggregate demand shocks in Basu (1995), aggregate productivity shocks in Khan and Thomas (2008), uncertainty shocks in Bloom (2009), financial shocks in Khan and Thomas (2013), supply chain disruptions in Meier (2018). Interestingly, short-run decreases in interest rates differ in sign from the effect of long-run decreases. Whereas we show that the former lowers misallocation, Gopinath et al. (2017) shows that the latter increases misallocation through size-dependent financial frictions. Relatedly, Oikawa and Ueda (2018) study the long-run effects of nominal growth through reallocation across heterogeneous firms. We further contribute to the literature by providing direct empirical evidence that measured misallocation responds to monetary policy shocks. Previously, Eisfeldt and Rampini (2006) documented that misallocation is countercyclical.

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

2 Analytical results

In this section, we characterize a novel transmission mechanism through which monetary policy shocks affect markup dispersion and aggregate TFP.

2.1 Fluctuations in markup dispersion, aggregate TFP, and aggregate markup

Let i index a firm and t time. A firm's markup is the firm's price P_{it} over aggregate price P_t and real marginal cost X_t, i.e. $\mu_{it} \equiv P_{it}/(P_t X_t)$. Define pass-through from marginal cost to price as

$$\varepsilon_{it} \equiv \frac{\partial \log P_{it}}{\partial \log X_t}. \tag{2.1}$$

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* $\mathrm{Corr}_t(\log \mu_{it}, \varepsilon_{it}) < 0$*, 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 $\mathrm{Corr}_t(\log \mu_{it}, \varepsilon_{it}) > 0$*.*

Proof: See Appendix A.1.

Contractionary monetary policy shocks that lower real marginal costs will raise the dispersion of markups if firms with higher markup have lower pass-through.

Fluctuations in markup dispersion imply fluctuations in aggregate TFP, because the distribution of markups determines the allocative efficiency of inputs across firms. To characterize the link between markup dispersion and aggregate TFP, we build on the seminal work by Hsieh and Klenow (2009) and Baqaee and Farhi (2019). Representing markup deviations as wedges in a static model and assuming monopolistic competition together with Dixit-Stiglitz aggregation, changes in aggregate TFP are

$$\Delta \log \mathtt{TFP}_t = -\frac{\eta}{2}\Delta \mathbb{V}_t(\log \mu_{it}) \; + \; \Big[\Delta \text{ exogenous productivity}\Big], \tag{2.2}$$

where η is the substitution elasticity between variety goods. Hence, an increase in the variance of log markups by 1% lowers aggregate TFP by $\frac{\eta}{2}$%. For a more detailed derivation, see Appendix A.2.

The monetary transmission mechanism is substantially altered if monetary policy shocks affect aggregate TFP. Interestingly, this can change the sign the aggregate markup response to monetary policy shocks. In a large class of New Keynesian models, markups increase after contractionary monetary policy shocks. However, recent evidence in Nekarda and Ramey (2019) points in the opposite direction. The New Keynesian model can be reconciled with this evidence if monetary policy shocks affect aggregate TFP sufficiently strongly. To make a simple example, consider an aggregate production function $Y_t = \mathtt{TFP}_t L_t$. Following Hall (1988), the aggregate markup is

$$\mu_t = \frac{P_t Y_t}{W_t L_t} = \frac{\mathtt{TFP}_t}{W_t / P_t}. \tag{2.3}$$

In standard New Keynesian models, tighter monetary policy reduces aggregate demand which lowers real wages and, hence, markups increase. However, the aggregate markup may increase if aggregate TFP falls 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. In Section 4, we revisit the effects of monetary policy on the aggregate markup in a quantitative New Keynesian model.

2.2 An explanation for the negative pass-through–markup correlation

We next show that firm-level heterogeneity in various price-setting frictions can explain a negative correlation between firm-level pass-through and markup. Consider a risk-neutral investor that sets prices in a monopolistically competitive environment with an isoelastic demand curve and subject to adjustment costs:

$$\max_{\{P_{it+j}\}_{j=0}^{\infty}} \mathbb{E}_t \sum_{j=0}^{\infty} \beta^t \left[\left(\frac{P_{it+j}}{P_{t+j}} - X_{t+j} \right) \left(\frac{P_{it+j}}{P_{t+j}} \right)^{-\eta} Y_{t+j} - \text{adjustment cost}_{it+j} \right] \tag{2.4}$$

Adjustment costs differ across firms and may be deterministic or stochastic. This formulation nests the Calvo (1983) friction, 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 is the source of 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 (Calvo/Taylor), or lowers the likelihood of costly price re-adjustments (Rotemberg/Barro).

Our goal is to characterize this effect. However, analytically solving the non-linear price-setting problem with adjustment costs and aggregate uncertainty in general equilibrium is not feasible. Instead, we study the problem in partial equilibrium. We assume that aggregate price, real marginal costs, and aggregate demand, denoted by (P_t, X_t, Y_t), follow an iid joint log-normal process. We denote the (co-)variances of innovations by 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

$$P_{it}^* = \frac{\eta}{\eta - 1} P_t X_t \frac{\mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j \theta_i^j \frac{X_{t+j}}{X_t} \left(\frac{P_{t+j}}{P_t} \right)^{\eta} \frac{Y_{t+j}}{Y_t} \right]}{\mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j \theta_i^j \left(\frac{P_{t+j}}{P_t} \right)^{\eta-1} \frac{Y_{t+j}}{Y_t} \right]}, \tag{2.5}$$

and the associated markup is μ_{it}^*. To isolate the role of uncertainty in price setting, we focus on the stochastic steady state, described by the unconditional means $\bar{P}$, $\bar{X}$, and $\bar{Y}$. The following proposition characterizes the upward price-setting bias as a function of the Calvo parameter and establishes a condition under which firms with lower pass-through optimally 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,*

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

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

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

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

Proof: See Appendix A.3.

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

Rotemberg friction. Consider the price setting problem subject to a Rotemberg (1982) quadratic price adjustment cost, 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

$$(1-\eta)\left(\frac{P_{it}}{P_t}\right)^{1-\eta} Y_t + \eta X_t \left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t = \phi_i \left(\frac{P_{it}}{P_{it-1}} - 1\right)\frac{P_{it}}{P_{it-1}} - \phi_i \mathbb{E}_t\left[\left(\frac{P_{it+1}}{P_{it}} - 1\right)\frac{P_{it+1}}{P_{it}}\right]. \quad (2.6)$$

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 a first-order approximation of* (2.6) *at $\phi_i = 0$ yields*

$$\mu_{it}^* \geq \frac{\eta}{\eta-1} \quad \text{and} \quad \frac{\partial \mu_{it}^*}{\partial \phi_i} \geq 0, \quad \text{with equality only if} \quad \phi_i = 0.$$

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 the Rotemberg parameter ϕ_i,

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

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

Proof: See Appendix A.4.

Menu costs. Consider Barro (1972) menu costs, but assume firms face different 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 B.

3 Empirical Evidence

In this section, we provide empirical evidence for the mechanism. We capture heterogeneity in pass-through by differences in the extensive margin of price adjustment. Lower price adjustment frequencies are tightly connected to lower pass-through against the backdrop of Calvo (1983), Taylor (1979), or Barro (1972) frictions. Our four main empirical findings are the following. First, firms that adjust prices less frequently charge higher markups. Second, contractionary monetary policy shocks increase the relative markup of firms that adjust prices less frequently. Third, contractionary monetary policy shocks increase markup dispersion across firms. Fourth, contractionary monetary policy shocks lower aggregate productivity.

3.1 Data, identification, and estimation

Firm-level markups. We use quarterly balance-sheet data of US firms from Compustat.[7] We estimate firm-level markups by adopting the approach of Hall (1988) and De Loecker and Warzynski (2012). If firms have a flexible factor of adjustment V_{it}, then cost minimization implies that the markup μ_{it} of firm i in quarter t can be computed as

$$\mu_{it} = \frac{\text{output elasticity of } V_{it}}{\text{revenue share of } V_{it}}. \tag{3.1}$$

The challenge is to estimate the output elasticity. We follow De Loecker et al. (forthcoming) in assuming a two-digit-industry and time specific Cobb–Douglas production function that combines a composite input of labor and materials, the flexible factor, with capital. In Compustat data, we use costs of goods sold to measure expenditures for labor and material, and property, plants, and equipment to measure the capital stock. We then estimate these production functions to obtain industry-time specific output elasticities of variable inputs. Details are provided in Appendix C.[8] The firm-time-specific revenue share is given by the ratio of costs of goods sold to sales.

[7] Compustat data has two central advantages over most other firm-level balance-sheet data: First, it is available at quarterly frequency instead of annual (e.g., ASM) or every five years (e.g., Census). This is important to precisely estimate responses to monetary policy shocks. Second, while Compustat only contains listed firms, it does cover all sectors.

[8] As robustness checks, we consider (i) a four-digit industry-specific translog production function and (ii) cost shares as direct estimates of the output elasticity, see Subsection 3.6.

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 only reported once in the associated year. We further drop observations if real sales are below 1 million USD or if quarterly sales growth is above 100% or below -67%. We finally drop the bottom and top 5% of the estimated markups. Appendix D.1 provides more details including summary statistics in Table 3. Our results are robust to dropping any of the data treatments, see Subsection 3.6.

Estimating output elasticities in the presence of infrequent price adjustment can be problematic. The approach of De Loecker and Warzynski (2012) is valid only if the relationship between input demand V_{it} and firm-level productivity is monotonic conditional on other determinants of demand for V_{it}. Under common assumptions, increases in productivity raise V_{it} when prices are adjusted, but lower V_{it} when prices are fixed. Hence, the relationship can only be monotonic when conditioning on price adjustment. Extending the approach to account for infrequent price adjustment is challenging. It would require firm-level prices and may be further complicated by the presence of multi-product firms and multiple distribution channels. Note that in our subsequent empirical analysis, the baseline specifications focus on deviations of firm-level log markups from their industry or industry-quarter specific mean. We do so primarily to control for industry characteristics such as their competitiveness or production technology. It follows that our main empirical results are not affected by potential biases in the estimated output elasticities.

Price rigidity. We use average industry-level price adjustment frequencies over 2005–2011 from Pasten et al. (2018) and based on PPI micro data.[9] The data is at the level of five-digit industries, which ensures variation in price rigidity between firms, even within more broadly-defined industries. We further use the Compustat segment files, which provides sales and five-digit industry code of business units within firms. We use this within-firm composition of sales to compute the sales-weighted average of industry-specific price adjustment frequencies. To construct a firm-level price adjustment frequency, we use the sales-weighted average, available for a quarter of firms, and otherwise the five-digit industry-specific adjustment frequency. More details are provided in Appendix D.2.[10] By constructing firm-level price rigidity from sectoral rigidities, we underestimate the true extent of heterogeneity across firms. We expect this to bias our regression results toward zero. Finally, we define the implied price duration as $-1/\log(1 - \text{adjustment frequency})$.

Monetary policy shocks. We identify monetary policy shocks using high-frequency futures prices based on the federal funds rate, which we acquired from the Chicago Mercentile Exchange. 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 is unchanged in that window and that no shock other than the monetary announcement occurs in or shortly before the time window. We denote the price of a future by f, and let τ denote the time of a

[9] We thank Michael Weber for generously sharing this data with us.

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

monetary announcement.[11] 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

$$\varepsilon_\tau^{\mathrm{MP}} = \mathrm{f}_{\tau+\Delta\tau^+} - \mathrm{f}_{\tau-\Delta\tau^-}. \tag{3.2}$$

We aggregate the shocks to quarterly frequency following the approach in Ottonello and Winberry (2019). 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 agents have to respond. Let t denote quarters, then the quarterly shocks are

$$\varepsilon_t^{\mathrm{MP}} = \sum_{\tau\in\mathcal{D}(t)} \phi(\tau)\varepsilon_\tau^{\mathrm{MP}} + \sum_{\tau\in\mathcal{D}(t-1)} (1-\phi(\tau))\varepsilon_\tau^{\mathrm{MP}}, \tag{3.3}$$

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.[12] Following Nakamura and Steinsson (2018), our baseline excludes the apex of the financial crisis from 2008Q3 to 2009Q2.[13] The monetary policy shock series covers 1995Q2 through 2018Q3. We discuss alternative monetary policy shock series at the end of this section. Table 4 in the Appendix reports summary statistics and Figure 7 shows the shock series.

3.2 Correlation between price rigidity and markup

We first study the correlation between price rigidity and markup. To compare markups with average price adjustment frequencies and implied price durations, we compute average firm-level markups over 2005–2011. There are various reasons why markups may differ over time and across industries. For example, industries differ in their competitiveness and production technologies. Throughout our empirical analysis, we therefore focus on the variation within industries. We regress average firm-level markups on average firm-level price adjustment frequencies, without fixed effects and with fixed effects at the two-digit or four-digit industry level. Table 1 summarizes the results. In all specifications, we find that firms with stickier prices set significantly higher markups.[14]

---

[11] 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).

[12] 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 central bank information about the state of the economy. Our results remain broadly robust when including these meetings, see Subsection 3.6.

[13] When estimating dynamic responses, we further do not regress post-Great Recession outcomes on pre-Great Recession shocks. Our results are robust to including the Great Recession, see Subsection 3.6.

[14] In addition, we find that firms with stickier prices have significantly higher markups even when controlling for firm size, leverage, and the share of liquid assets.

Table 1: Regressions of (log) markup on price stickiness at the firm level

	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.200	-0.230	-0.153			
	(0.0252)	(0.0289)	(0.0673)			
Implied price duration				0.0405	0.0405	0.0257
				(0.00381)	(0.00454)	(0.00994)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	3973	3973	3964	3973	3973	3964
Adjusted R^2	0.018	0.105	0.342	0.027	0.108	0.342

Notes: Regressions of (log) firm-level markup on firm-level price adjustment frequency and implied price duration, respectively. Heteroskedasticity-robust standard errors in parentheses.

3.3 Price stickiness and the markup response to monetary policy shocks

The previous subsection shows that on average high-markup firms adjust prices less frequently. Models of frictional price adjustment predict that the relative markups of firms with stickier prices should increase by more in response to contractionary monetary policy shocks. We next provide evidence to support this hypothesis.

We estimate panel local projections of firm-level log markups on the interaction between monetary policy shocks and firm-level price rigidity. As before, the measures of firm-level price rigidity are the price adjustment frequency and 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, and the ratio of liquid assets to total assets.[15] Our selection of controls is motivated by recent work in Ottonello and Winberry (2019) and Jeenas (2018), who study the transmission of monetary policy shocks through financial constraints. The panel regressions we estimate are

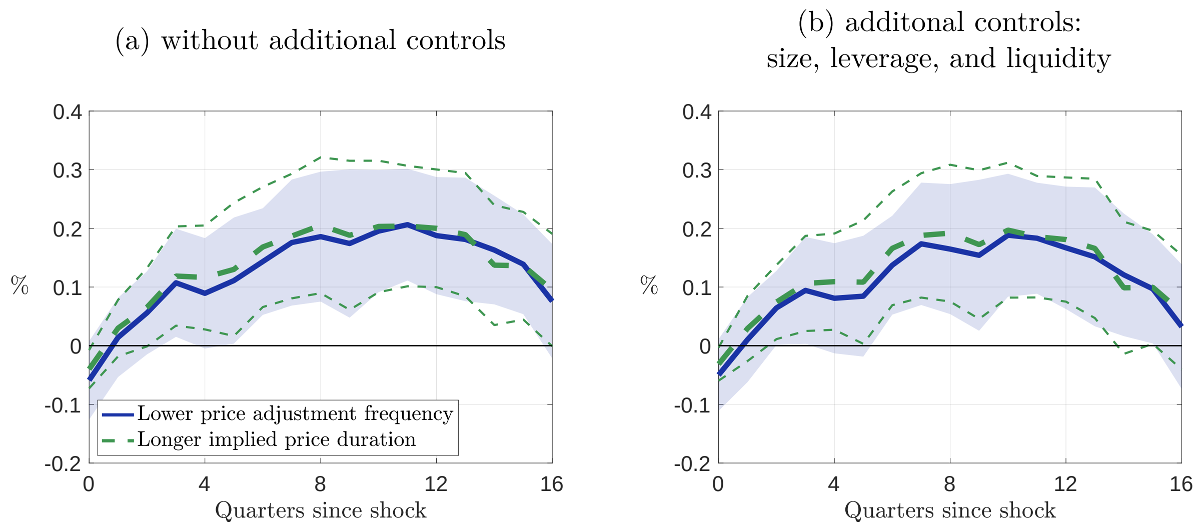
$$y_{it+h} - y_{it-1} = \alpha_i^h + \alpha_{st}^h + B^h Z_{it}\varepsilon_t^{\mathrm{MP}} + \Gamma^h Z_{it} + \gamma^h(y_{it-1} - y_{it-2}) + u_{it}^h \tag{3.4}$$

for $h = 0, \ldots, 16$, 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 industry mean from the measure of price rigidity. The main coefficients of interest are the elements of $\{B^h\}$ associated to price rigidity. These capture the relative markup increase for firms with stickier prices. Figure 1 shows the results. The markups of firms with stickier prices increase by significantly more after contractionary monetary policy shocks.[16] Importantly, we obtain almost identical estimates when including additional controls. To the extent that these

[15]We demean the additional firm-level controls by the firm-level mean to focus on within-firm variation.

[16]Note that we estimate a linear model, so the effects of an expansionary monetary policy shocks are simply reversed in sign. Throughout the paper, our (sometimes implicit) sign normalization is to plot and discuss the effects of contractionary monetary policy shocks.

Figure 1: Relative markup response of firms with stickier prices to monetary policy shocks



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 (or with an implied price duration one standard deviation above) the two-digit-industry-quarter mean. That is, we plot the appropriately scaled coefficients in B^h that are associated to price rigidity in the panel local projections (3.4). In panel (a), Z_{it} contains only price stickiness. In panel (b), Z_{it} also contains lagged log assets, leverage, and liquidity and their interactions with a monetary policy shock. Leverage and liquidity are measured relative to the firm-level means, in order to isolate within-firm variation. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

controls capture firm-level differences in financial constraints, the price rigidity coefficients in panel (a) do not pick up financial constraints.[17]

3.4 Effects of monetary policy on within-industry markup dispersion

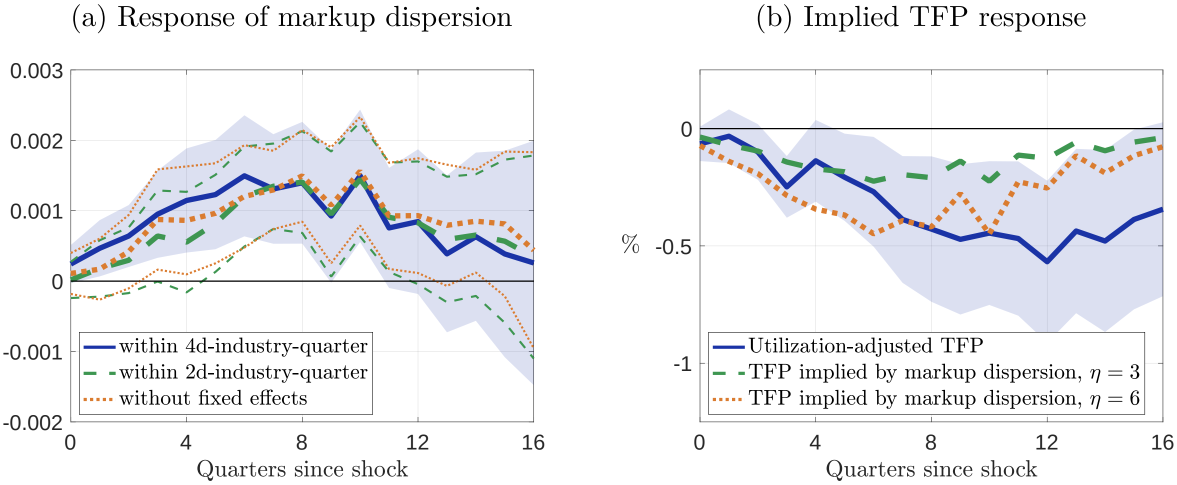
We next estimate the response of markup dispersion to monetary policy shocks. If firm-level differences in the markup response are mainly driven by price rigidity, an increase in markup dispersion will be trivial against the backdrop of our previous empirical findings. In general, however, markups may respond differently for many reasons, including different financial constraints. We therefore estimate the markup dispersion response directly.

To compute markup dispersion within industry and time, we substract from firm-level markups the two or four-digit industry s and quarter t specific means. Our baseline measure of markup dispersion is the cross-sectional variance $\mathbb{V}_t(\log \mu_{it} - \overline{\log \mu}_{st})$. Figure 7 in the Appendix plots this measure of markup dispersion over time.[18] To estimate the effects of monetary policy shocks on

[17]Figures 8 in the Appendix shows the interaction with the other three firm characteristics included in specification (ii). The interaction with firm size is insignificant. Firms with higher pre-shock leverage tend to increase their markups by more. The interaction with liquidity is mostly insignificant.

[18]Similar to De Loecker et al. (forthcoming), we find an increasing trend in markup dispersion.

Figure 2: Responses of markup dispersion to monetary policy shocks



Notes: Panel (a) shows the responses of markup dispersion to a one-standard deviation contractionary monetary policy shock. That is, we plot the coefficients in β^h in (3.5). Markup dispersion is measured within two-digit and four-digit industry-quarters as well as without fixed effects, respectively. Panel (b) shows the imputed response of TFP, implied by the response of markup dispersion within four-digit industry-quarters, according to $\Delta \log \mathtt{TFP}_t = -\frac{\eta}{2}\Delta\mathbb{V}_t(\log \mu_{it})$, see equation (2.2), and using $\eta = 3$ and $\eta = 6$, respectively. Alongside, it shows the empirical response of utilization-adjusted TFP from Figure 3. The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

markup dispersion, we estimate

$$y_{t+h} - y_{t-1} = \alpha^h + \beta^h \varepsilon_t^{\mathrm{MP}} + \gamma_0^h \varepsilon_{t-1}^{\mathrm{MP}} + \gamma_1^h (y_{t-1} - y_{t-2}) + u_t^h, \tag{3.5}$$

for $h = 0, \ldots, 16$ and where y_t is markup dispersion. Figure 2(a) shows the response of markup dispersion, captured by the coefficients $\{\beta^h\}$. The key finding is that markup dispersion increases significantly and persistently, both within two-digit and four-digit industry-quarters.

3.5 Effects of monetary policy on aggregate productivity

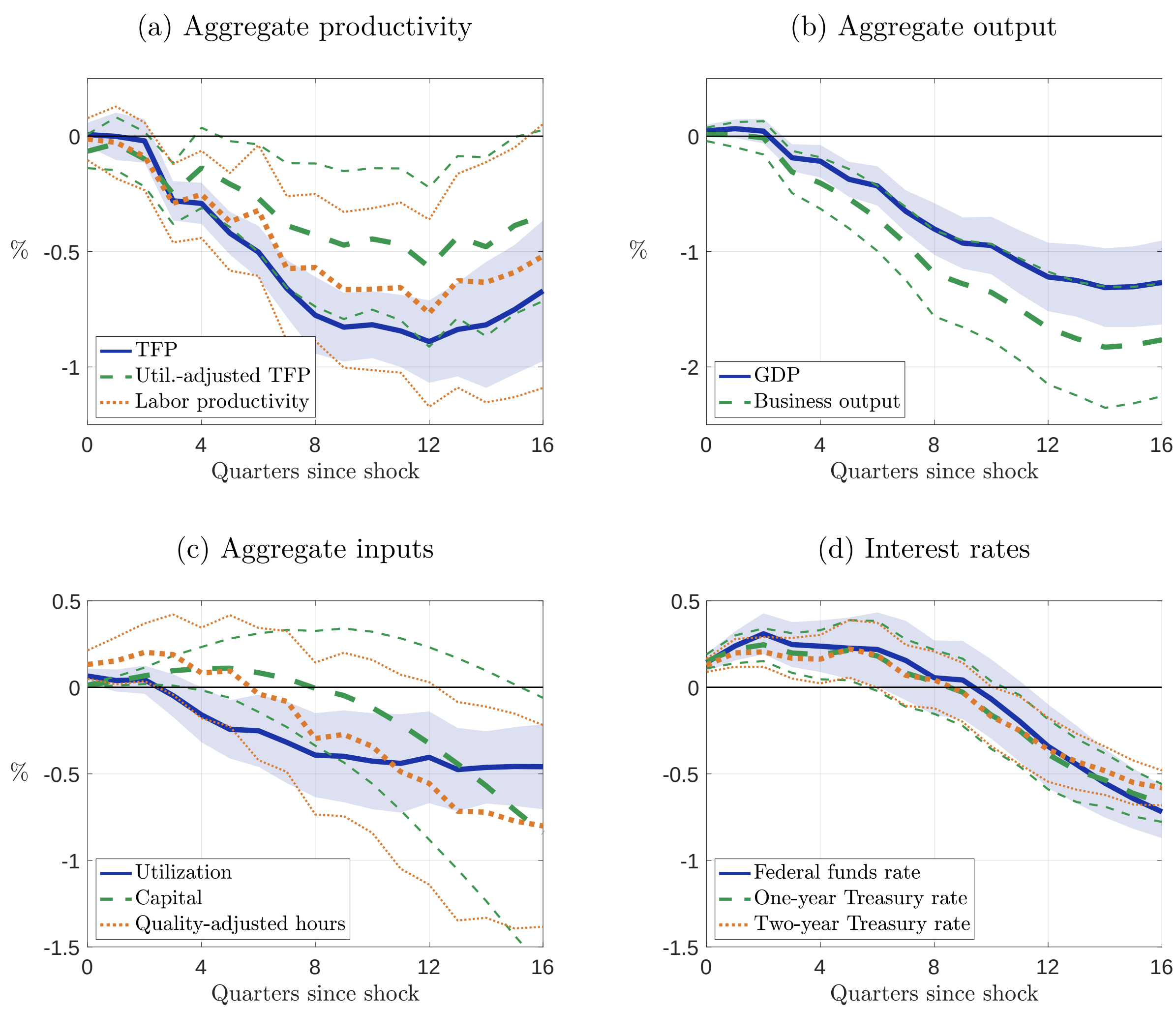
We next show that monetary policy lowers aggregate productivity. We consider aggregate TFP and utilization-adjusted aggregate TFP from Fernald (2014), as well as labor productivity.[19] In addition, we estimate the response of aggregate output, aggregate inputs, and interest rates.[20]

Figure 3 shows the estimated responses to one-standard deviation contractionary monetary policy shocks. A monetary policy shock of this magnitude raises the federal funds rate by up to 30 basis points, see panel (d). The response of the federal funds rate is comparable to the responses of

[19] Aggregate TFP is $\Delta \log \mathtt{TFP} = \Delta y - w_k \Delta k - (1 - w_k)\Delta \ell$, with Δy real business output growth, w_k the capital income share, Δk real capital growth (based on separate perpetual inventory methods for 15 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 7(c) in the Appendix plots the different aggregate productivity series.

[20] We use specification (3.5) in all cases except for interest rates, for which we use a specification in levels, $y_{t+h} = \alpha^h + \beta^h \varepsilon_t^{\mathrm{MP}} + \gamma_0^h \varepsilon_{t-1}^{\mathrm{MP}} + \gamma_1^h y_{t-1} + u_t^h$.

Figure 3: Macroeconomic responses to monetary policy shocks



Notes: The plots show the responses to a one-standard deviation contractionary monetary policy shock. The local projections in Panel (d) are estimated in levels rather than differences. The shaded and bordered areas indicate one standard error bands based on the Newey–West estimator.

the one-year and two-year Treasury rates. Importantly, tighter monetary policy lowers aggregate productivity. The responses of all three aggregate productivity measures are statistically significant. At a two-year horizon, aggregate TFP falls by 0.8%, labor productivity by 0.6% and utilization-adjusted aggregate TFP by 0.4%, see panel (a). Aggregate TFP accounts for about 50–80% of the output response at a two-year horizon. While the response of utilization is significant and accounts for 30–50% of the aggregate output response, the responses of capital and quality-adjusted hours are insignificant for the most part. This highlights the role of the aggregate TFP response for understanding the real effects of monetary policy.

We next we ask how much of the utilization-adjusted aggregate TFP response the response of markup dispersion can account for. Using equation (2.2), we compute the implied TFP response by multiplying the response of markup dispersion by $\eta/2$. 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 peaks at about 0.5%. This 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.

In Figure 10 in the Appendix we reproduce the result in Nekarda and Ramey (2019) that the aggregate markup falls after monetary policy shocks. The same is true on average for firm-level markups. As explained in Subsection 2.1, an explanation may lie in the TFP reponse.

3.6 Robustness

Markup estimation. Estimation of firm-level markups using the method of De Loecker and Warzynski (2012) requires an assumption on the production technology of firms. As a baseline specification, we have assumed a Cobb–Douglas production function with industry-time-specific parameters. Based on De Loecker et al. (forthcoming), we consider two alternatives. First, a translog production function with industry-specific coefficients. This gives rise to *firm- and time-specific* output elasticities. Second, we estimate the output elasticity directly through the cost share, costs of goods sold divided by total costs. This is a valid estimator of the output elasticity under flexible adjustment of *all* input factors. See Appendix C for details on both alternative approaches.

All our results are robust to computing markups based on a translog production function or cost shares. Table 5 in the Appendix shows the correlation between average markup and price rigidity. Figure 12 shows the relative markup response to monetary policy shocks. Figure 13 shows that markup dispersion robustly increases.

Firm-level data treatment and delisting. Balance-sheet data may contain erroneous and economically implausible entries. We examine the robustness of our results when tightening or relaxing our baseline data treatment. First, we keep small firms with real quarterly sales below 1 million 2010 USD. Second, we keep firms with real sales growth above 200% or below -66%. 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.

Table 6 in the Appendix shows that the correlation between markups and price rigidity is robust across data treatments. Figure 14 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. Figure 15 shows that markup dispersion robustly increases after contractionary monetary policy shocks.

A well-known recent trend is the delisting of public firms. The number of firms in our baseline sample drops over time from about 5,000 to 2,500 firms. Thus, a valid concern is that this may affect our results. We address this concern 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, see above. Second, we estimate whether the number of firms in Compustat responds to monetary policy shocks. Figure 9(b) shows that the response is small and insignificant.

Alternative monetary policy shocks. For robustness, we consider a policy indicator, similar to Nakamura and Steinsson (2018), computed as the first principal component of the current/three-month federal funds futures and the 2/3/4-quarters ahead Eurodollar futures. We further 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. 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 (forthcoming), 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.

Figure 16 in the Appendix shows the robustness of the interaction of firm-level price rigidity with the monetary policy shock for all monetary policy shock series. Figure 17 in the Appendix shows the response of markup dispersion for all monetary policy shock series. Figure 18 in the Appendix shows the responses of aggregate productivity for all monetary policy shock series. The baseline results hold robust and we estimate responses in the same order of magnitude.

LP-IV. We reproduce our main results with the LP-IV method (Stock and Watson, 2018). More precisely, we replace the MP shocks $\varepsilon_t^{\mathrm{MP}}$ by the quarterly change in the one-year treasury rate and use $\varepsilon_t^{\mathrm{MP}}$ as an instrument. Figure 20 in the Appendix shows that our results are robust to using this method.

Treatment of the 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) in Figure 14 and Figure 15 in the Appendix show that the firm-level heterogeneity in the markup response and the increase in markup dispersion, respectively, after contractionary monetary policy shocks is robust in all samples. Figure 19 shows the robustness of the decline in aggregate productivity.

TFP (mis)measurement. Hall (1986) shows that the Solow residual is misspecified in the presence of market power. He shows that instead of the capital income share w_{kt} as the Solow weight for capital and $1 - w_{kt}$ for labor, the correct weights are $\mu_t w_{kt}$ and $1 - \mu_t w_{kt}$, where μ_t 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 et al. (forthcoming) to compute Hall's weights. Figure 11(c) in the Appendix shows that this barely affects our results.

We further investigate whether measurement error in quarterly TFP data is responsible for the effect 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 11(d) 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 11(a) and (b). Notably, the response of investment-specific TFP is significantly stronger than that of consumption-specific TFP.

4 New Keynesian Model with heterogeneous price stickiness

In this section, we study a standard New Keynesian model with heterogeneity in price rigidity. In this setup, a negative correlation between price rigidity and markup arises endogenously in the stochastic steady state. The calibrated model explains almost two thirds of the peak response in utilization-adjusted TFP to monetary policy shocks.

4.1 Model setup

Our model setup builds on Carvalho (2006). 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. 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. In addition, it seems more plausible to assume equal demand elasticities within than across sectors.

There is a continuum of monopolistically competitive intermediate goods firms that produce with a linear technology in labor. Firms may reset their prices with a firm-specific probability $1 - \theta_i$ in any given period. They set prices to maximize the value of the firm to the households. 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, that would prevail under flexible prices. Monetary policy follows a Taylor rule with interest rate smoothing, that is subject to monetary policy shocks, $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$.

4.2 Quantitative results

In the model, a 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. We assume that there are five equally large groups of firms, indexed by $k \in \{1, \ldots, 5\}$, which differ in their price adjustment frequencies $1 - \theta_k$. We calibrate $\{\theta_k\}$ to match the empirical distribution of *within-industry* price adjustment frequencies, as documented in Gorodnichenko and Weber (2016). They report means and standard deviations of monthly price adjustment frequencies within five sectors. We first compute a 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. Our empirical results in Figure 3 show that the hours response to monetary policy shocks is small on impact, but larger at longer horizons. 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. The calibrated Frisch elasticity is $\varphi = 0.1175$, which is low compared to the macroeconomics literature, but which falls 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

Table 2: Calibration

Externally calibrated parameters			
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)
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
Internally calibrated parameters			
Parameter		Value	Target
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%

Notes: This table shows calibrated parameters for the New Keynesian model with heterogeneous price rigidity. The distribution of price adjustment frequencies is chosen to match the within-sector distribution reported in Gorodnichenko and Weber (2016).

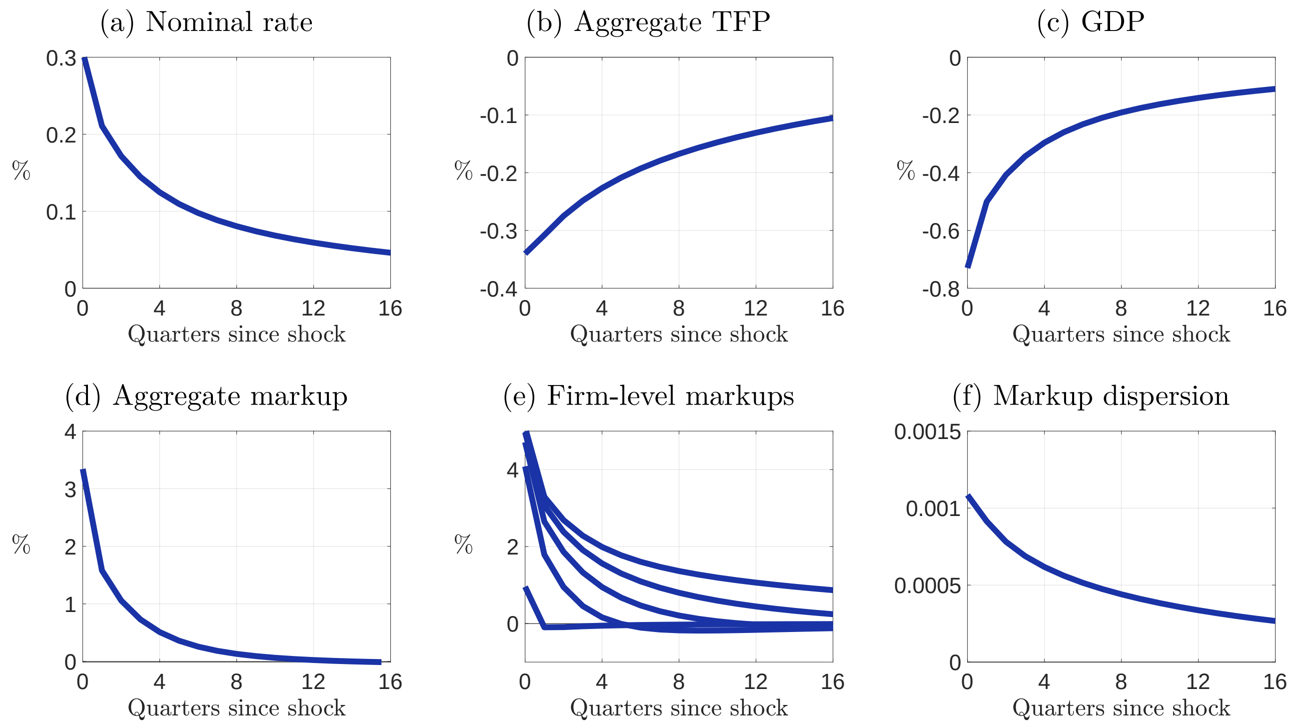
deviation monetary policy shock of 30bp, see Figure 3. This yields $\sigma_\nu = 0.00415$.

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.[21] In the stochastic steady state, markups of the firms with the largest price rigidity are 10.9% higher than markups of firms with the most flexible prices. The negative correlation between markups and pass-through implies that contractionary monetary policy shocks increase markup dispersion and lower aggregate TFP, see Proposition 1.

Figure 4 shows the responses to a one standard deviation contractionary monetary policy shock. The shock depresses aggregate demand and real marginal costs. In response firms want to lower

[21] 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 more general setup of heterogeneous price rigidities.

Figure 4: Model responses to monetary policy shocks



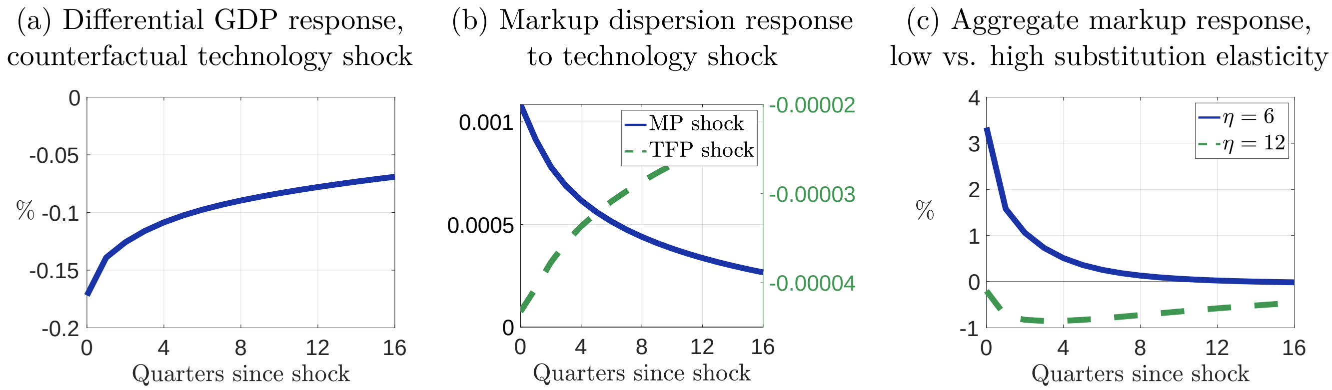
Notes: This figure shows responses to a one standard deviation contractionary monetary policy shock.

their prices. For firms with stickier prices, however, pass-through is lower and on average their markup increases 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, see Figure 3. The results further 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, which generates additional persistence in the responses.

We argue in this paper that an important aspect of the monetary transmission channel is the response of aggregate TFP. In contrast, traditional business cycle models assume that the only source of fluctuations in aggregate TFP are 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 observed aggregate TFP movements after monetary policy shocks to originate from technology shocks. In particular, we compute natural output counterfactually supposing the TFP response to monetary policy shocks is driven by technology shocks.[22] Panel (a) in Figure 5 shows the difference between the GDP responses under the counterfactual technology shock and the baseline

[22]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 5: Additional model results and counterfactuals



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 with a technology shock. 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.

response.[23] Output drops by up to 0.17 percentage points more if the monetary authority attributes aggregate TFP fluctuations to technology shocks, and the response is markedly more persistent.

This finding highlights the importance for the monetary authority to assess the sources of observed aggregate TFP fluctuations. Panel (b) in Figure 5 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.[24] 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.

Panel (c) in Figure 5 highlights the important role of the elasticity of substitution η. A larger value increases the misallocation costs of markup dispersion and thus the TFP loss after a monetary policy shock. Interestingly, a more pronounced TFP decline can give rise to *decreasing* aggregate markups $\mu_t = \frac{P_t}{W_t}\mathtt{TFP}_t$, as discussed in Section 2. 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. This example shows that aggregate markups may fall after contractionary monetary policy shocks. Figure 23 in the Appendix shows further impulse responses to monetary policy shocks under $\eta = 14$.

In the Appendix, we analyze the robustness of our results in two dimensions. First, in Figure 24, we vary the Frisch elasticity of labor supply φ. The markup dispersion and TFP responses are higher for less elastic labor supply and dampened for more elastic labor supply. Second, in Figure 25, we raise the lowest price adjustment frequency, $1-\theta_1$ to the level of the second quintile group $1-\theta_2 =$ 6.78%. This dampens the increase in markup dispersion, and an aggregate TFP response of -0.27%.

[23]Figure 21 in the Appendix provides further impulse responses for this counterfactual scenario.
[24]Figure 22 in the Appendix provides further impulse responses for the technology shock.

5 Conclusion

This paper studies an overlooked monetary transmission mechanism. If firms with relatively high markups have relatively low pass-through from marginal costs to prices, then contractionary monetary policy shocks widen the distribution of markups across firms. This implies a change in the allocation of inputs across firms, which lowers measured aggregate TFP. We use aggregate and firm-level data to test several implications of this mechanism. We find that firms with higher markups adjust prices less frequently, and that contractionary monetary policy shocks increase the relative markup of these firms, increase markup dispersion across firms, and lower aggregate productivity. The estimated magnitudes suggest that the mechanism we propose is quantitatively important to understand the effects of monetary policy on real economic activity. We show that an explanation for the negative correlation between markup and price stickiness is the stickiness itself. Firms with stickier prices optimally set higher markups for precautionary reasons. In a calibrated New Keynesian model, introducing heterogeneous stickiness allows us to explain a large share of the empirically estimated TFP response to monetary policy shocks.

A promising topic for future research is to study why firms with stickier prices charge higher markups. The explanation we propose, precautionary price setting, takes price stickiness as given. Instead, the underlying reasons for differences in stickiness may also explain markup differences. For example, firms may commit to stable prices to render a product more attractive, which allows the firm to charge a higher markup.

References

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.

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

——— (2019): "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.

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

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

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

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

Carlton, D. W. (1986): "The Rigidity of Prices," *The American Economic Review*, 76, 637–658.

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.

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 (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.

De Loecker, J., J. Eeckhout, and G. Unger (forthcoming): "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.

Dhyne, E., L. J. Alvarez, H. Le Bihan, G. Veronese, D. Dias, J. Hoffmann, N. Jonker, P. Lunnemann, F. Rumler, and J. Vilmunen (2006): "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, 20, 171–192.

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.

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.

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.

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 (forthcoming): "Deconstructing monetary policy surprises: the role of information shocks," *American Economic Journal: Macroeconomics.*

Jeenas, P. (2018): "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.

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

Meier, M. (2018): "Time To Build and the Business Cycle," *mimeo U Mannheim*.

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

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

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.

——— (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.

Nekarda, C. J. and V. A. Ramey (2019): "The Cyclical Behavior of the Price-Cost Markup," mimeo, University of California San Diego.

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

Ottonello, P. and T. Winberry (2019): "Financial Heterogeneity and the Investment Channel of Monetary Policy," *mimeo*.

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

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

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.

Woodford, M. (2010): "Chapter 14 - Optimal Monetary Stabilization Policy," Elsevier, vol. 3 of *Handbook of Monetary Economics*, 723 – 828.

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 Proofs

A.1 Proof of Proposition 1

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

$$\mathbb{V}_t[\log \mu_{it}] = \int (\log P_{it} - \log P_t - \log X_t)^2 di - \left[\int (\log P_{it} - \log P_t - \log X_t) di\right]^2. \tag{A.1}$$

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

$$\frac{\partial \mathbb{V}_t[\log \mu_{it}]}{\partial \log X_t} = 2\int \log(\mu_{it})\varepsilon_{it} di - 2\int \log(\mu_{it}) di \int \varepsilon_{it} di = 2\text{Cov}_t[\log \mu_{it}, \varepsilon_{it}]. \tag{A.2}$$

Hence, the markup variance falls in $\log X_t$ if $\text{Corr}_t[\log \mu_{it}, \epsilon_{it}] < 0$. □

A.2 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

$$\max_{P_{it}} (\tau_{it} P_{it} - X_{it}) Y_{it} \qquad \text{s.t.} \quad Y_{it} = (P_{it}/P_t)^{-\eta} Y_t, \tag{A.3}$$

where τ_{it} is a *markup wedge* in the spirit of Hsieh and Klenow (2009) and Baqaee and Farhi (2019). 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

$$\log \texttt{TFP}_t = \log \left(\int Y_{it}^{(\eta-1)/\eta} di\right)^{\eta/(\eta-1)} - \log \int \frac{Y_{it}}{A_{it}} di. \tag{A.4}$$

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 $\log \texttt{TFP}_t$ 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

$$\log \texttt{TFP}_t = -\frac{\eta}{2}\mathbb{V}_t(\log \mu_{it}) + \mathbb{E}_t(\log A_{it}) + \frac{\eta-1}{2}\mathbb{V}_t(\log A_{it}). \tag{A.5}$$

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.

A.3 Proof of Proposition 2

We assume that

$$\log \begin{pmatrix} P_t/\bar{P} \\ X_t/\bar{X} \\ Y_t/\bar{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} -\frac{\sigma_p^2}{2} \\ -\frac{\sigma_x^2}{2} \\ -\frac{\sigma_y^2}{2} \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & & \\ \sigma_{px} & \sigma_x^2 & \\ \sigma_{py} & \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right). \tag{A.6}$$

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

$$C_{it} \equiv E_t \left[\frac{X_{t+1}}{X_t} \left(\frac{P_{t+1}}{P_t} \right)^{\eta} \frac{Y_{t+1}}{Y_t} \right], \tag{A.7}$$

$$D_{it} \equiv E_t \left[\left(\frac{P_{t+1}}{P_t} \right)^{\eta-1} \frac{Y_{t+1}}{Y_t} \right], \tag{A.8}$$

$$\Psi_{it} \equiv \frac{1+\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}}, \tag{A.9}$$

which allows us to rewrite the first-order condition in (2.5) as

$$P_{it}^* = \frac{\eta}{\eta-1} P_t X_t \Psi_{it}. \tag{A.10}$$

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

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

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

$$(\eta-1)\sigma_p^2 + \sigma_{py} + \eta\sigma_{px} + \sigma_{xy} > 0. \tag{A.11}$$

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

$$\frac{\partial \Psi_{it}}{\partial \tilde{\theta}_i} = \frac{C_{it} - D_{it}}{(1+\tilde{\theta}_i D_{it})^2} > 0, \quad \text{and hence} \quad \frac{\partial \Psi_{it}}{\partial \theta_i} > 0. \tag{A.12}$$

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

$$\bar{\varepsilon}_{it} = (1-\theta_i)\frac{\partial \log P_{it}}{\partial \log X_t} = (1-\theta_i)\left(1+\Phi_{it}\right), \quad \text{where} \quad \Phi_{it} = \frac{\partial \log \Psi_{it}}{\partial \log X_t} \tag{A.13}$$

and

$$\Phi_{it} = \frac{\tilde{\theta}_i \frac{\partial C_{it}}{\partial \log X_t}(1+\tilde{\theta}_i D_{it}) - (1+\tilde{\theta}_i C_{it})\tilde{\theta}_i \frac{\partial D_{it}}{\partial \log X_t}}{(1+\tilde{\theta}_i D_{it})^2} \Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i D_{it}} \Psi_{it}^{-1} = -\frac{\tilde{\theta}_i C_{it}}{1+\tilde{\theta}_i C_{it}} < 0. \tag{A.14}$$

Hence pass-through becomes

$$\bar{\varepsilon}_{it} = \frac{1-\theta_i}{1+\tilde{\theta}_i C_{it}} \in (0,1). \tag{A.15}$$

In addition, the pass-through falls in θ_i,

$$\frac{\partial \bar{\varepsilon}_{it}}{\partial \theta_i} = -(1+\Phi_{it}) + (1-\theta_i)\frac{\partial \Phi_{it}}{\partial \theta_i} < 0. \tag{A.16}$$

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 $X_t = \bar{X}$,

$$\frac{\partial \log P_{it}^*}{\partial \log \bar{X}} = 1. \tag{A.17}$$

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

A.4 Proof of Proposition 3

Let us first define

$$C_{it} = \left(\frac{P_{it}}{P_{i,t-1}} - 1\right)\frac{P_{it}}{P_{i,t-1}}, \tag{A.18}$$

$$D_{it} = \mathbb{E}_t\left[\left(\frac{P_{i,t+1}}{P_{it}} - 1\right)\frac{P_{i,t+1}}{P_{it}}\right], \tag{A.19}$$

such that we can re-write the first-order condition in equation (2.6) more compactly as

$$(1-\eta)\left(\frac{P_{it}}{P_t}\right)^{1-\eta} Y_t + \eta X_t \left(\frac{P_{it}}{P_t}\right)^{-\eta} Y_t = \phi_i (C_{it} - D_{it}). \tag{A.20}$$

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,

$$\bar{C}_{it} = \left(\frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} - 1\right)\frac{\bar{P}_{it}}{\bar{P}_{i,t-1}} = (\Pi_{pt}\Pi_{xt})^2 - \Pi_{pt}\Pi_{xt}, \tag{A.21}$$

$$\bar{D}_{it} = E_t\left[\left(\frac{\bar{P}_{i,t+1}}{\bar{P}_{it}} - 1\right)\frac{\bar{P}_{i,t+1}}{\bar{P}_{it}}\right] = \frac{\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\}}{(\Pi_{pt}\Pi_{xt})^2} - \frac{\exp\left\{\sigma_{pw}\right\}}{\Pi_{pt}\Pi_{xt}}. \tag{A.22}$$

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 $d\phi_i = \phi_i$, we obtain

$$(1-\eta)^2 \left(\frac{P_{it}}{\bar{P}_t}\right)^{1-\eta} Y_t \mathrm{dlog} P_{it} - \eta^2 X_t \left(\frac{\bar{P}_{it}}{P_t}\right)^{-\eta} Y_t \mathrm{dlog} P_{it} = (\bar{C}_{it} - \bar{D}_{it}) d\phi_i. \tag{A.23}$$

This yields

$$\Psi_{it} \equiv \frac{\mathrm{dlog} P_{it}}{d\phi_i} = \frac{\bar{D}_{it} - \bar{C}_{it}}{(\eta-1)^\eta \eta^{1-\eta} X_t^{1-\eta} Y_t}, \tag{A.24}$$

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

$$\sigma_p^2 + \sigma_x^2 + 2\sigma_{px} > 0, \tag{A.25}$$

for which a sufficient condition is that the correlation

$$\rho_{px} \equiv \frac{\sigma_{px}}{\sigma_p \sigma_x} > -1. \tag{A.26}$$

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

$$\varepsilon_{it} = 1 + \frac{\partial \Psi_i}{\partial \log X_t} d\phi_i. \tag{A.27}$$

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

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0, \tag{A.28}$$

which is equivalent to examining the conditions for

$$\frac{\partial \bar{D}_{it}}{\partial \log X_t} - \frac{\partial \bar{C}_{it}}{\partial \log X_t} + (\eta - 1)(\bar{D}_{it} - \bar{C}_{it}) < 0. \tag{A.29}$$

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

$$\frac{\partial \bar{C}_{it}}{\partial \log X_t} = 2(\Pi_{pt}\Pi_{xt})^2 - \Pi_{pt}\Pi_{xt}, \tag{A.30}$$

$$\frac{\partial \bar{D}_{it}}{\partial \log X_t} = -2(\Pi_{pt}\Pi_{xt})^{-2} \exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} + (\Pi_{pt}\Pi_{xt})^{-1} \exp\left\{\sigma_{px}\right\}. \tag{A.31}$$

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

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0 \qquad \text{if} \quad \eta < \tilde{\eta}^{\text{transitory}} = 2 + \frac{1 + \exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\}}{\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} - \exp\left\{\sigma_{px}\right\}} \tag{A.32}$$

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

$$\frac{\partial \bar{C}_{it}}{\partial \log X_t} = 2(\Pi_{pt}\Pi_{wt})^2 - \Pi_{pt}\Pi_{wt}, \quad \frac{\partial \bar{D}_{it}}{\partial \log X_t} = 0. \tag{A.33}$$

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

$$\frac{\partial \Psi_i}{\partial \log X_t} < 0 \qquad \text{if} \quad \eta < \tilde{\eta}^{\text{permanent}} = 1 + \frac{1}{\exp\left\{\frac{3}{2}\sigma_p^2 + \frac{3}{2}\sigma_x^2 + 4\sigma_{px}\right\} - \exp\left\{\sigma_{px}\right\}} \tag{A.34}$$

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

B 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

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

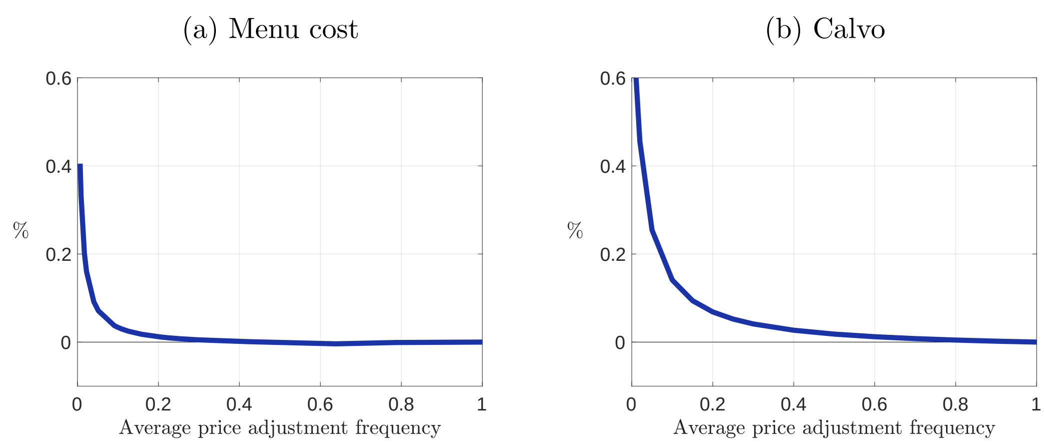
where p is the price a firm sets and Z denote a vector of the aggregate state variables price level (P), aggregate demand (Y), and marginal costs (X). The firm chooses to adjust prices in the presence of the menu cost ξ.

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.

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 6 plots the price setting policy p^* at the unconditional mean of Z for different average price adjustment frequencies. We compare menu costs in panel

Figure 6: Precautionary price setting under menu costs and Calvo



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

(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.

C Markup estimation

We follow De Loecker and Warzynski (2012) and De Loecker et al. (forthcoming) in estimating firm-level markups from firm-level balance sheet data. If a firm can frictionlessly adjust its variable inputs V, then it follows from cost minimization that the markup can be written as the product of the output elasticity θ_{it}^V and the inverse cost share of V:

$$\mu_{it} = \theta_{it}^V \frac{Q_{it}}{V_{it}} \tag{C.1}$$

Letting lower cases denote logs, we assume that sales are given by

$$q_{it} = \phi_{st}(v_{it}, k_{it}) + \omega_{it} + \varepsilon_{it}, \tag{C.2}$$

where ϕ_{st} denotes the industry- and potentially time-specific production function, ω_{it} is firm-level revenue productivity, and ε_{it} is an unanticipated shocks to production, that is observed by the firm only after having made input choices, or classical measurement error. We assume that revenue productivity follows an AR(1) process.

Our baseline markup measure is based on a time- and two-digit-industry-specific Cobb–Douglas production function of a composite input of labor and materials (measured by costs of goods sold) and capital:

$$\phi_{st}(v_{it}, k_{it}; \theta_{st}^V, \theta_{st}^K) = \theta_{st}^V v_{it} + \theta_{st}^K k_{it} \tag{C.3}$$

The estimation proceeds in two stages. First, we project sales q_{it} on a third-order polynomial in v_{it} and k_{it} to obtain an estimate of $\phi_{st}(v_{it}, k_{it}) + \omega_{it}$ denoted by $\hat{q}_{it}$. This projection captures unobserved productivity ω_{it} under the condition that the demand for inputs v_{it} strictly increases in productivity. In the second stage, we use GMM to find the parameters of the production function: For a given set parameters, revenue productivity can be recovered as $\omega_{it}(\theta_{st}^V, \theta_{st}^K) = \hat{q}_{it} - \phi_{st}(v_{it}, k_{it}; \theta_{st}^V, \theta_{st}^K)$. In turn, estimated productivity innovations ξ_{it} can be obtained as the residuals of a regression of ω_{it} on ω_{it-1}. The moment conditions are based on the notion that predetermined input choices are orthogonal to the productivity innovation:[25]

$$\mathbb{E}\left[\xi_{it}(\theta_{st}^V, \theta_{st}^K)\begin{pmatrix} v_{it-4} \\ k_{it} \end{pmatrix}\right] = 0 \tag{C.4}$$

We find $(\theta_{st}^V, \theta_{st}^K)$ for all industries s and quarters t by minimizing the associated GMM objective functions.[26] Based on equation (C.1), we compute the firm-level markup by multiplying the estimated output elasticity with the ratio of sales and costs of goods sold. We drop the bottom and top 5% of markups. As a robustness

[25] Similar to De Loecker et al. (forthcoming), who use the first lag of cost of goods sold as an instrument in annual data, we use the fourth lag in quarterly data. This choice improves the stability of our estimated output elasticities within industries over time.

[26] For any given quarter t we use the preceding five years of data to estimate the output elasticities. This follows De Loecker et al. (forthcoming). While they use a centered five-year window, we use a backward-looking one because we are interested in dynamic effects after shocks.

check, we only drop the bottom and top 1%.

Alternatively, we estimate a translog production function with four-digit-industry-specific parameters:

$$\phi_s(v_{it}, k_{it}) = \beta_s^V v_{it} + \beta_s^K k_{it} + \beta_s^{VV} v_{it}^2 + \beta_s^{KK} k_{it}^2 \tag{C.5}$$

Estimation is analogous to above using the moment conditions[27]

$$\mathbb{E}\left[\xi_{it}(\theta_{st}^V, \theta_{st}^K)\begin{pmatrix} v_{it-4} \\ v_{it-4}^2 \\ k_{it} \\ k_{it}^2 \end{pmatrix}\right] = 0. \tag{C.6}$$

This production function gives rise to firm-time-specific output elasticities given by

$$\theta_{it}^V = \beta_s^V + 2\beta_s^{VV} v_{it}. \tag{C.7}$$

Yet another way of estimating the output elasticity of variable inputs at the four-digit-industry-quarter-level is to use the cost share (Hall, 1988). This requires an estimate of the user cost of capital. Following De Loecker et al. (forthcoming), we set $R_t = \texttt{FFR}_t/4 + \pi_t + 3\%$ where $\texttt{FFR}_t$ is the federal funds rate, π_t is measured as the quarterly growth rate in the GDP deflator, and the annual depreciation rate and risk premium is assumed to be 12%. The cost share of an individual firm is then given by the ratio of costs of goods sold to total costs (costs of goods sold plus $R_t K_t$). For every four-digit-industry-quarter, we estimate the output elasticity of variable inputs as the median cost share across firms.

D Data construction and descriptive statistics

D.1 Firm-level balance sheet data

We use quarterly firm-level balance sheet data of U.S.-incorporated firms for the period 1995Q1 to 2017Q2 from the Compustat North America. In case of duplicate firm-quarter observations, we keep only the industry format and delete all remaining duplicates. Our industry classification is based on the North American Industry Classification System (NAICS). We exclude firms in utilities (2-digit NAICS code 22), finance, insurance, and real estate (52 and 53), and public administration (99). We disregard observations of sales (`saleq`), costs of goods sold (`cogsq`), and property, plant, and equipment (net PPE, `ppentq`, and gross PPE, `ppegtq`), that are non-positive. We fill one-quarter gaps in 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 $K_{i0} = \texttt{ppegtq}_{i0}$ and recursively compute $K_{it} = K_{it-1} + (\texttt{ppentq}_{it} - \texttt{ppentq}_{it-1})$. We drop single firm-quarter observations in terms of sales, costs of goods sold, and the capital stock for a given year. Finally, we delete observations if sales are below 1 million US$ or quarterly sales growth is above 100% or below -67%. Table 3 shows descriptive statistics for our baseline sample.

[27] For consistency, we also use fourth lags of cost of goods sold as instruments as we do above. However, the choice here is inconsequential for the estimated output elasticities.

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$.

D.2 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. (2018) and was generously shared with us by Michael Weber. We define the implied price duration as $-1/\log(1 - \text{adjustment frequency})$.

We obtain firms' yearly industry sales 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 industry code are not reported. If only some segment industry codes are missing, we assign the firm-level industry code to those.

We then compute every firm's yearly 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.

D.3 Monetary policy shocks

We compute monetary policy shocks by high-frequency identification as described in Subsection 3.1. Table 4 reports summary statistics for all monetary policy shocks and Figure 7 shows the shock series.

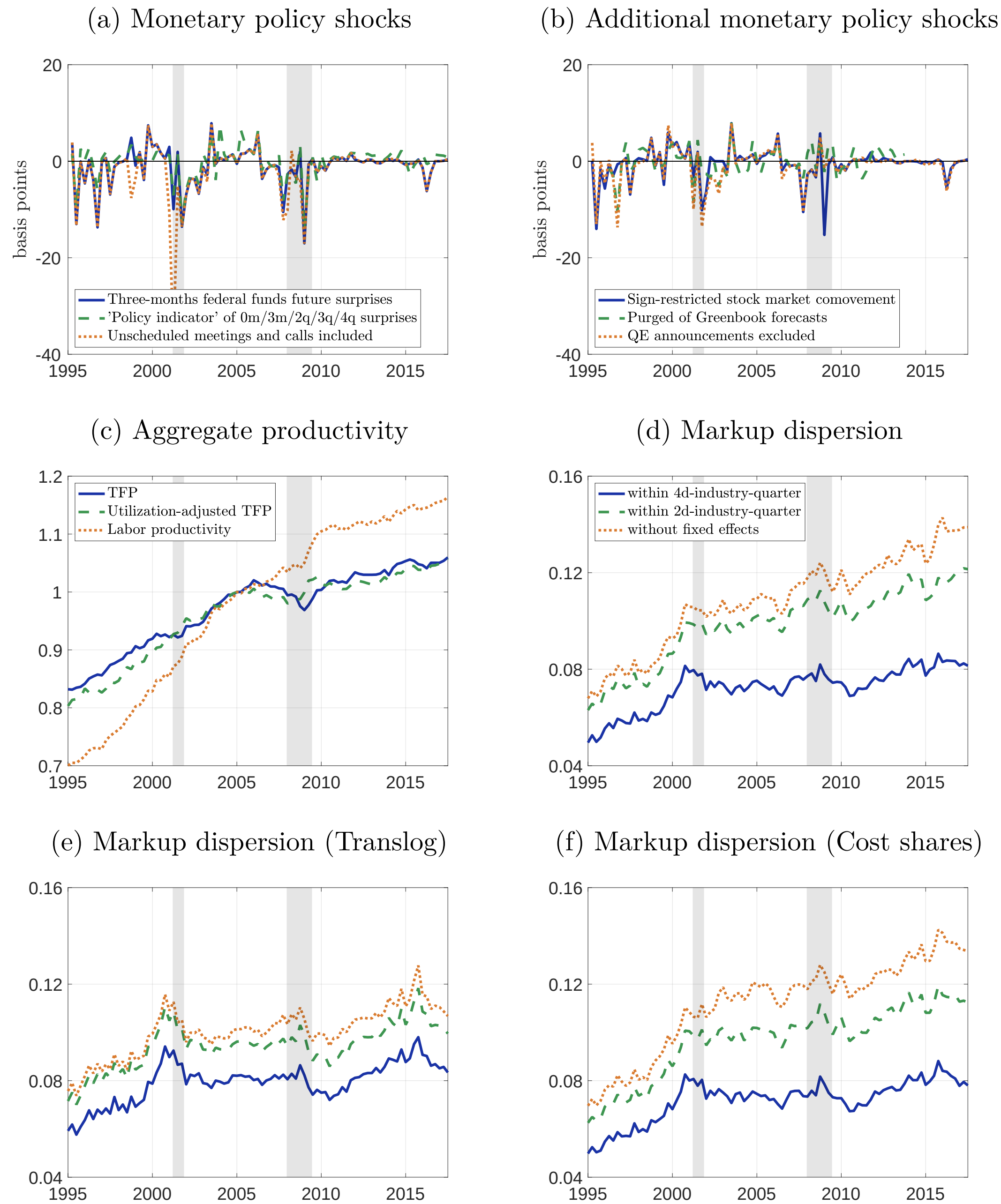
Table 4: Summary statistics of monetary policy shocks

	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.

D.4 Time series plots of monetary policy shocks, markup dispersion, and aggregate productivity

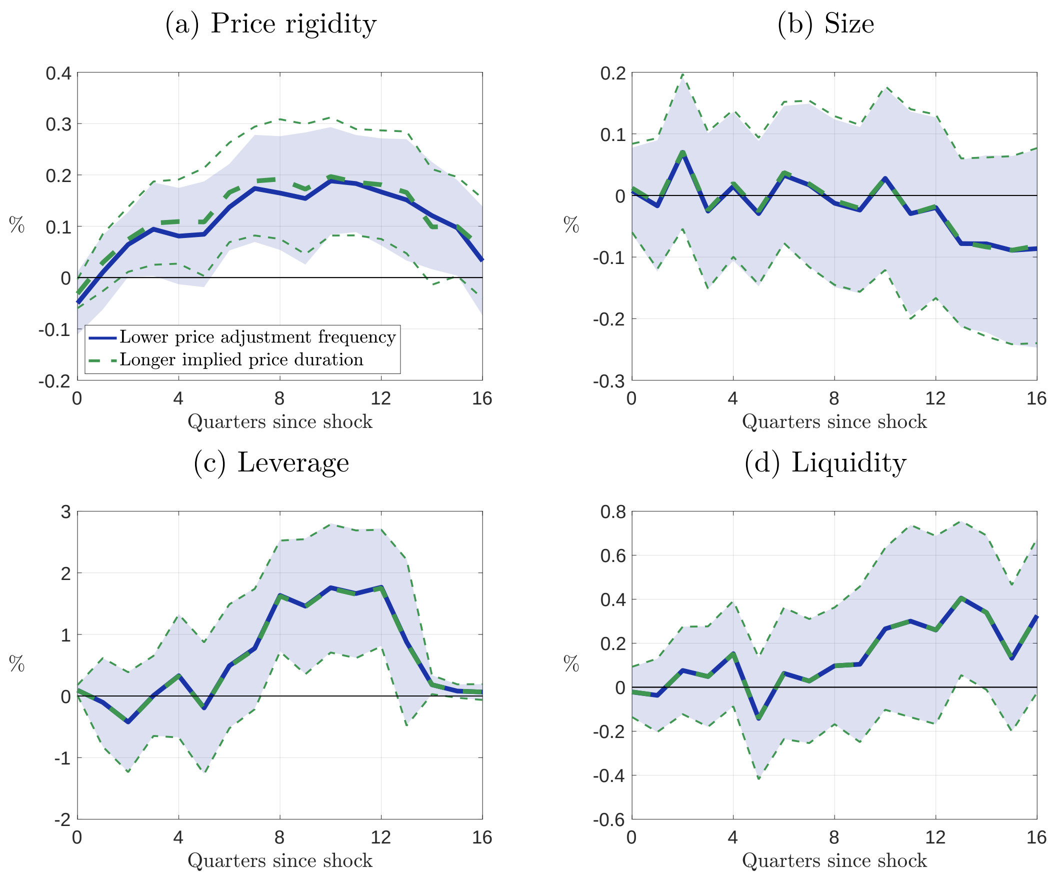
Figure 7: Monetary policy shocks, aggregate productivity, and markup dispersion



Notes: Aggregate productivity (in logs), markup dispersion, and monetary policy shocks are at quarterly frequency. Aggregate (utilization-adjusted) TFP is from Fernald (2014). Labor productivity is from FRED. Markup dispersion is computed from Compustat balance sheet data. Shaded gray areas indicate NBER recession dates.

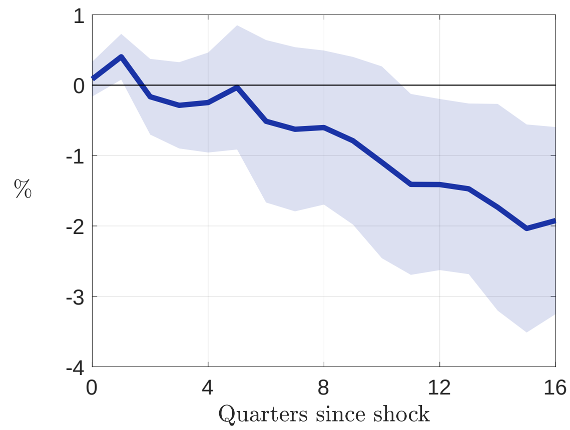
E Additional empirical results

Figure 8: Relative markup response of firms with stickier prices to monetary policy shocks: other interaction variables



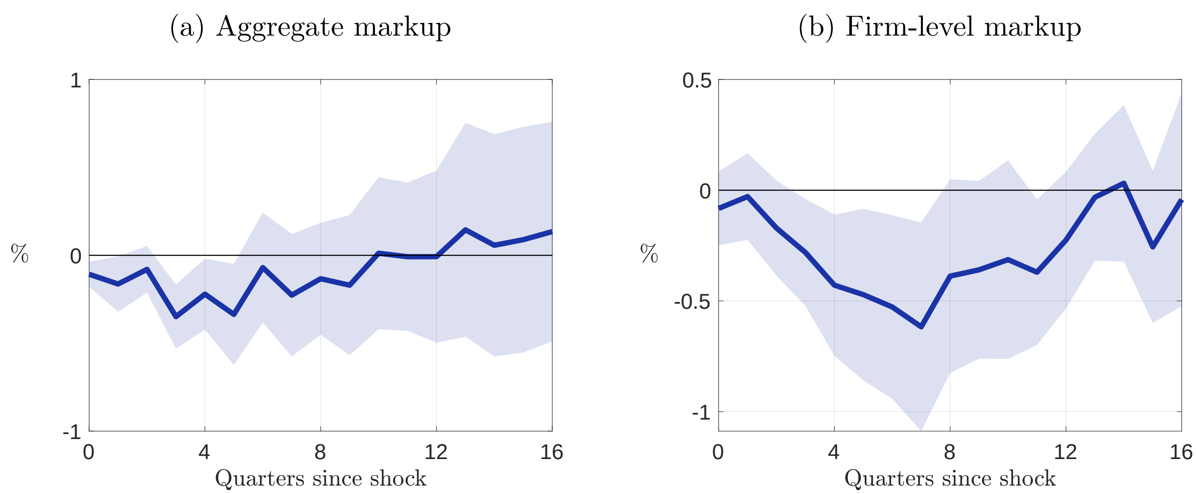
Notes: This figure shows the interaction coefficients between the monetary policy shock and price adjustment frequency, size, leverage, and liquidity, respectively, from the panel local projection in equation (3.4). Panel (a) shows the relative response of firms with a price adjustment frequency one standard deviation below (or with an implied price duration one standard deviation above) the two-digit-industry-quarter mean and replicates Figure 1(b). Panels (b)–(d) show the response of firms with lagged size, leverage, and liquidity, respectively, one standard deviation above their firm-level mean. The blue solid line (green dashed line) corresponds to the regression with price adjustment frequency (implied price duration) as the measure of price rigidity. The shaded area is a 90% error band clustered by firms and quarters.

Figure 9: Response of firm-level observations after monetary policy shocks



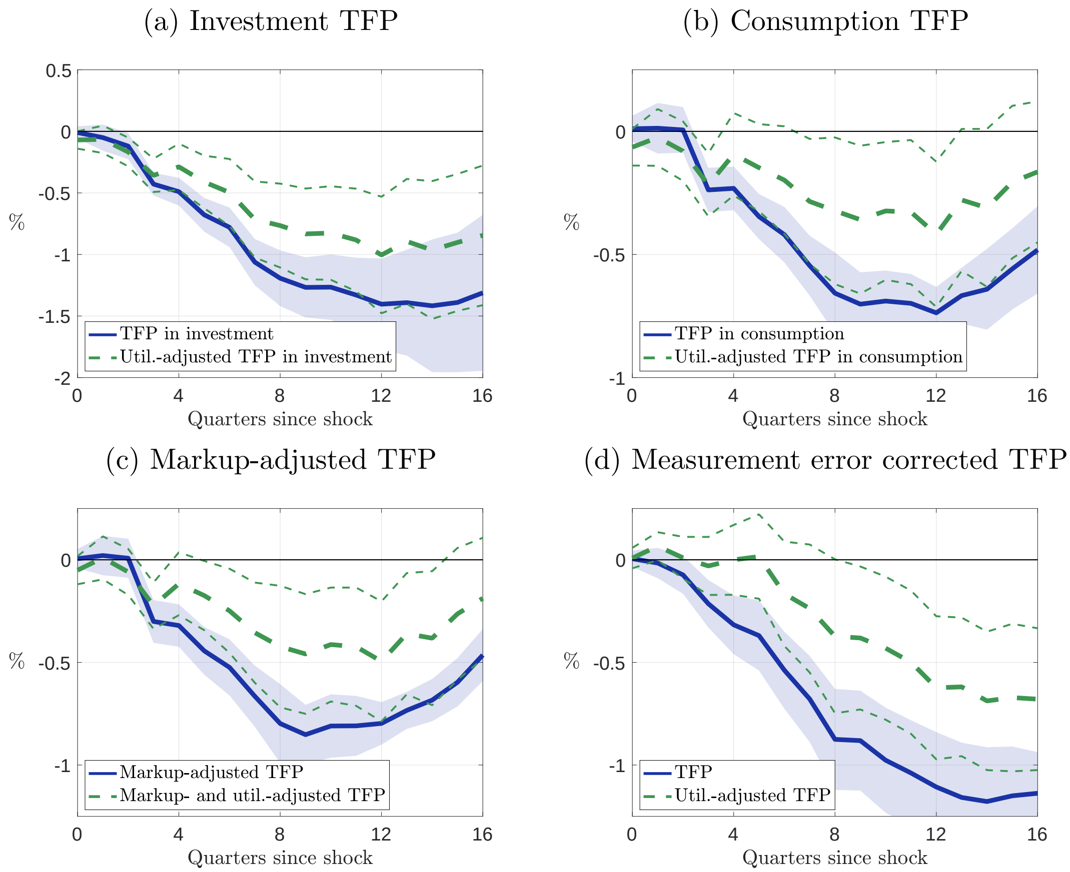
Notes: This figure shows the response of the number of firm-level observations in our sample to monetary policy shocks obtained from local projections as in equation (3.5). The shaded area is a one standard error band based on Newey–West.

Figure 10: Response of aggregate and firm-level markups to monetary policy shocks



Notes: Panel (a) shows the response of the aggregate markup measured as the inverse aggregate labor share to monetary policy shocks (see equation (2.3)). The shared area is a one standard error band based on Newey–West. Panel (b) shows the average response of firm-level markups to monetary policy shocks, i.e., coefficients $\{\beta^h\}$ in the panel local projections $y_{it+h} - y_{it-1} = \alpha_i^h + \beta^h \varepsilon_t^{\mathrm{MP}} + \Gamma^h Z_{it-1} + u_{it}^h$. Z_{it-1} includes log assets, leverage, and liquidity. The shaded area is a 90% error band clustered by firms and quarters.

Figure 11: Further productivity responses



Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). 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. (forthcoming). 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.

F Robustness of main findings

F.1 Results for additional markup series

Table 5: Regressions of markup on price stickiness at the firm level: additional markup series

(a) Markups based on translog production function

	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.116 (0.0230)	-0.171 (0.0277)	-0.161 (0.0741)			
Implied price duration				0.0231 (0.00350)	0.0249 (0.00446)	0.0222 (0.0112)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	3923	3923	3923	3923	3923	3923
Adjusted R^2	0.007	0.072	0.147	0.010	0.071	0.147

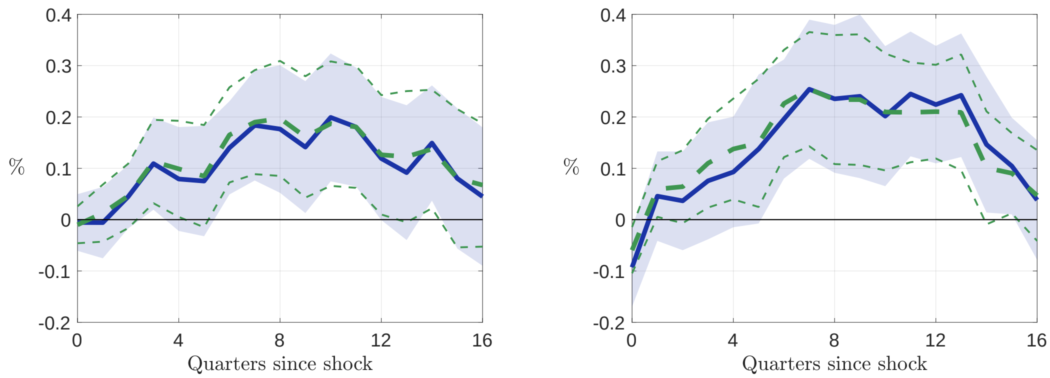
(b) Markups based on cost shares

	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.294 (0.0235)	-0.314 (0.0257)	-0.160 (0.0670)			
Implied price duration				0.0507 (0.00374)	0.0488 (0.00438)	0.0269 (0.00998)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	3963	3963	3963	3963	3963	3963
Adjusted R^2	0.037	0.176	0.365	0.041	0.175	0.366

Notes: Regression of firm-level markup (averaged over 2005–2011) on firm-level price adjustment frequency and implied price duration, respectively. Panel (a) uses markups based on an industry-specific translog production function, which gives rise firm-quarter-specific output elasticities. Panel (b) uses markups based on output elasticities estimated as the industry-quarter-specific median cost share. Heteroskedasticity-robust standard errors in parentheses.

Figure 12: Relative markup response of firms with stickier prices to monetary policy shocks: alternative markup measures

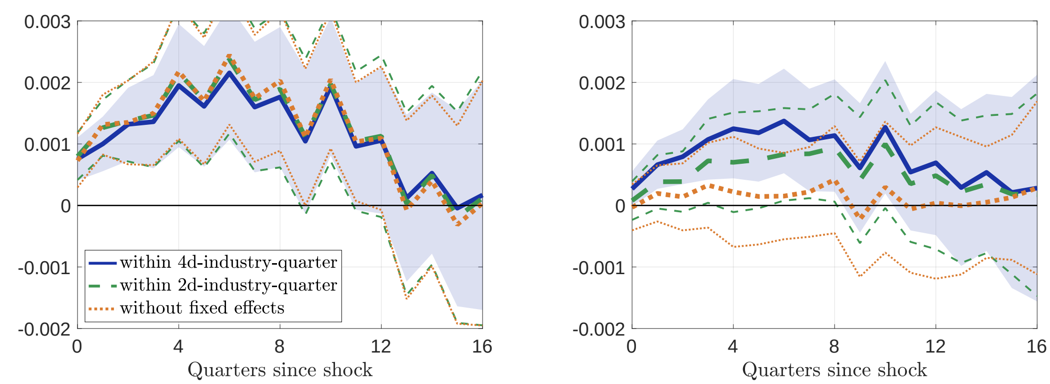
(a) Translog (b) Cost shares



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 (or with an implied price duration one standard deviation above) from panel local projections as in equation (3.4). The regressions include interactions with lagged log assets, leverage, and liquidity and their interactions with a monetary policy shock. Panel (a) uses markups based on an industry-specific translog production function, which gives rise firm-quarter-specific output elasticities. Panel (b) uses markups based on output elasticities estimated as the industry-quarter-specific median cost share. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

Figure 13: Responses of markup dispersion to monetary policy shocks: alternative markup measures

(a) Translog (b) Cost shares



Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). Panel (a) uses markups based on an industry-specific translog production function, which gives rise firm-quarter-specific output elasticities. Panel (b) uses markups based on output elasticities estimated as the industry-quarter-specific median cost share. Markup dispersion is measured within two-digit and four-digit industry-quarters as well as without fixed effects, respectively. The shaded and bordered areas indicate one standard error bands based on Newey–West.

F.2 Results for alternative data treatments

Table 6: Regressions of markup on price stickiness at the firm level: alternative data treatments

(a) Keep small firms

	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.155	-0.175	-0.0641			
	(0.0292)	(0.0314)	(0.0736)			
Implied price duration				0.0407	0.0308	0.0130
				(0.00441)	(0.00517)	(0.0108)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	4560	4560	4551	4560	4560	4551
Adjusted R^2	0.006	0.149	0.208	0.014	0.150	0.208

(b) Keep firms with excessive growth

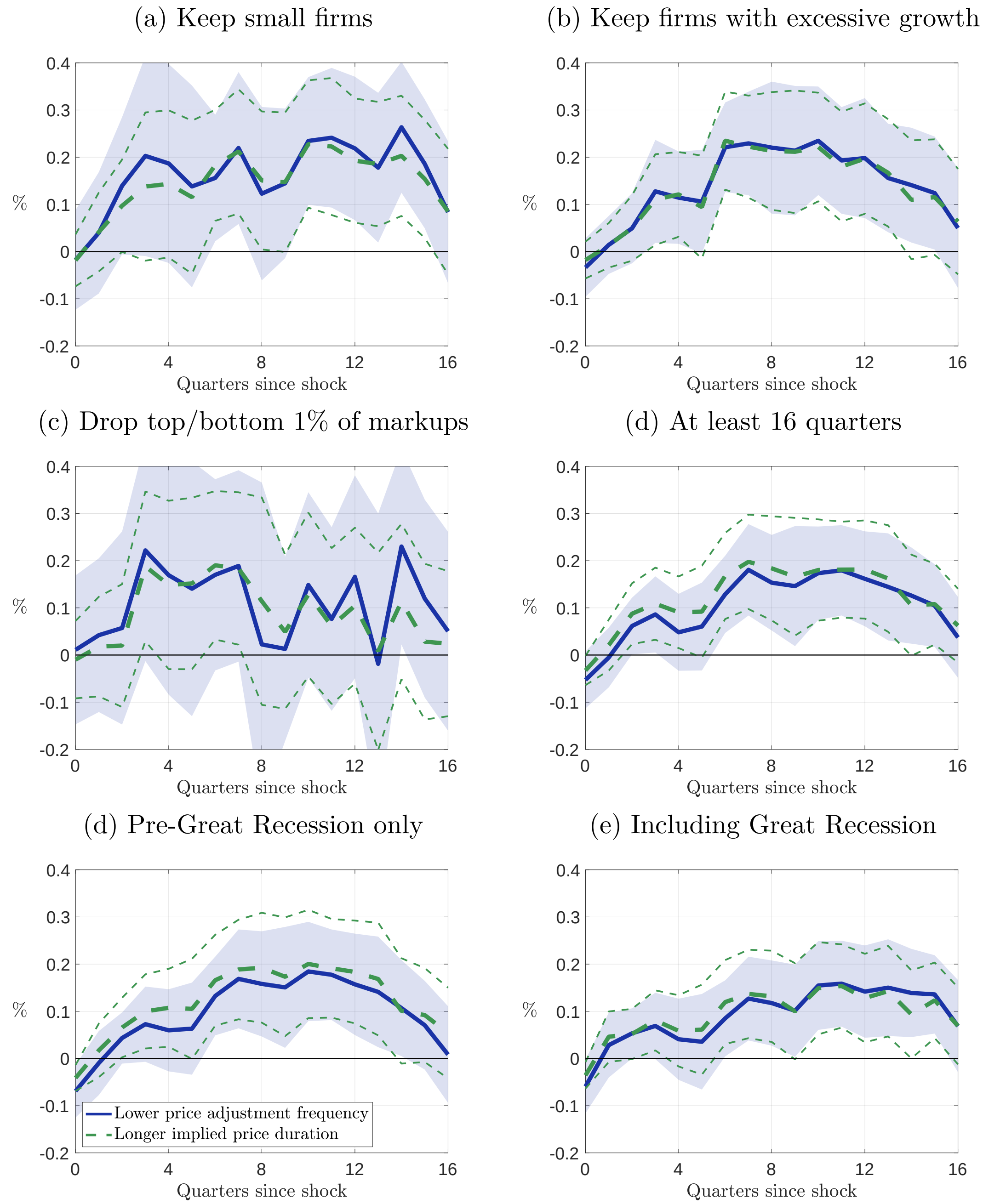
	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.209	-0.242	-0.166			
	(0.0249)	(0.0282)	(0.0697)			
Implied price duration				0.0414	0.0407	0.0245
				(0.00376)	(0.00447)	(0.0101)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	4354	4354	4344	4354	4354	4344
Adjusted R^2	0.018	0.094	0.276	0.026	0.096	0.276

(c) Drop top/bottom 1% of markups

	(1)	(2)	(3)	(4)	(5)	(6)
Price adjustment frequency	-0.217	-0.162	-0.111			
	(0.0347)	(0.0459)	(0.0880)			
Implied price duration				0.0517	0.0309	0.0222
				(0.00534)	(0.00707)	(0.0159)
Industry FE	–	2-digit	4-digit	–	2-digit	4-digit
Observations	4235	4235	4226	4235	4235	4226
Adjusted R^2	0.006	0.135	0.206	0.014	0.136	0.206

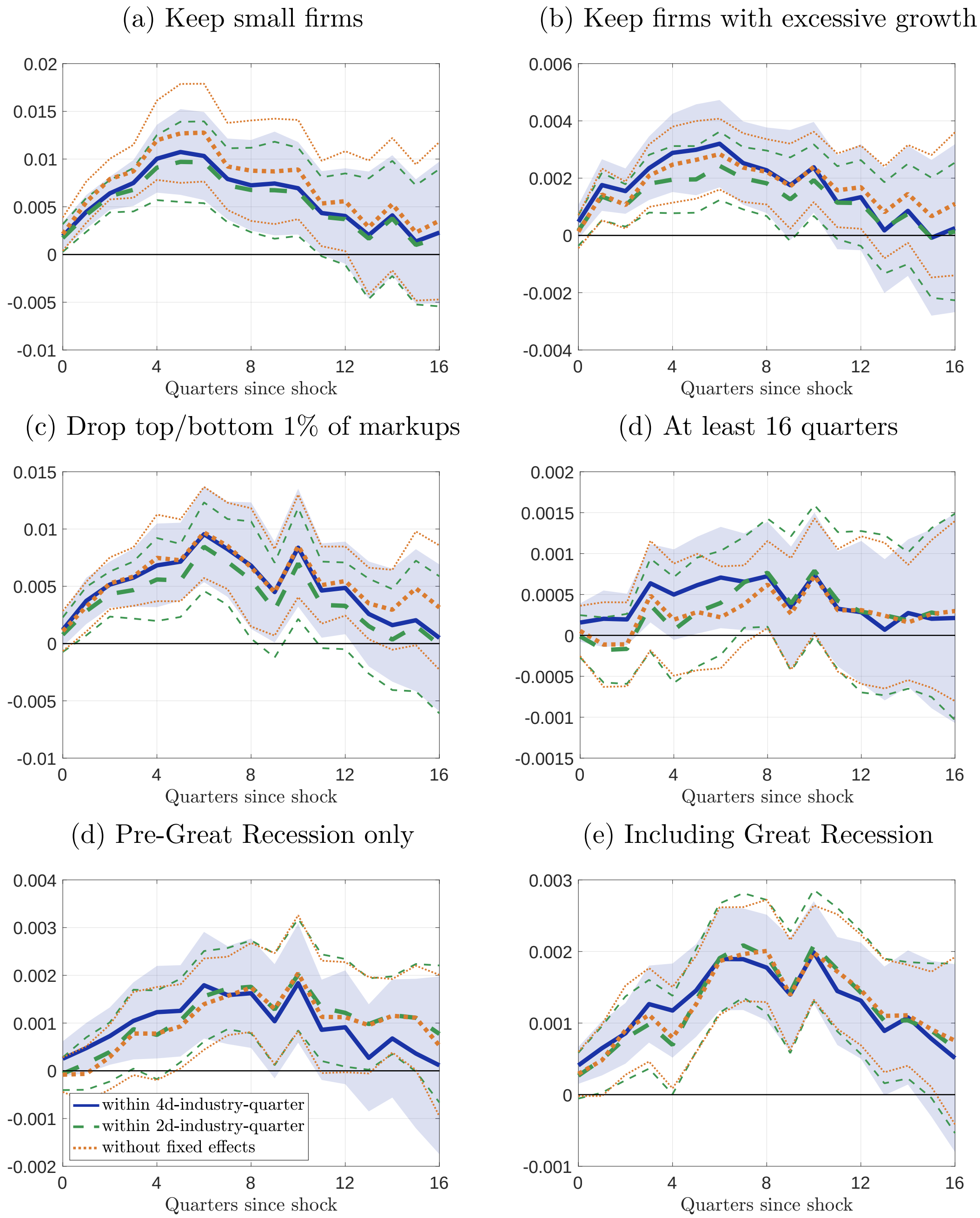
Notes: Regression of firm-level markup (averaged over 2005–2011) on firm-level price adjustment frequency and implied price duration, respectively. *Keep small firms* does not drop firms with sales below 1 million (in 2012 US$). *Keep firms with excessive growth* does not drop firms with growth above 100% or below -67%. *Drop top/bottom 1%* drops markups in the top/bottom 1% in the quarter instead of 5%. Heteroskedasticity-robust standard errors in parentheses.

Figure 14: Relative markup response of firms with stickier prices to monetary policy shocks: alternative data treatments



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 (or with an implied price duration one standard deviation above) from panel local projections as in equation (3.4). The regressions include interactions with lagged log assets, leverage, and liquidity and their interactions with a monetary policy shock. *Keep small firms* does not drop firms with sales below 1 million (in 2012 US$). *Keep firms with excessive growth* does not drop firms with growth above 100% or below -67%. *Drop top/bottom 1%* drops markups in the top/bottom 1% in the quarter instead of 5%. *At least 16 quarters* restricts the sample to firms with at least 16 quarters of consecutive observations. *Pre-Great Recession only* considers only observations before 2008Q3. *Including Great Recession* does not drop the period 2008Q3–2009Q2 from the sample. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

Figure 15: Responses of markup dispersion to monetary policy shocks: alternative data treatments

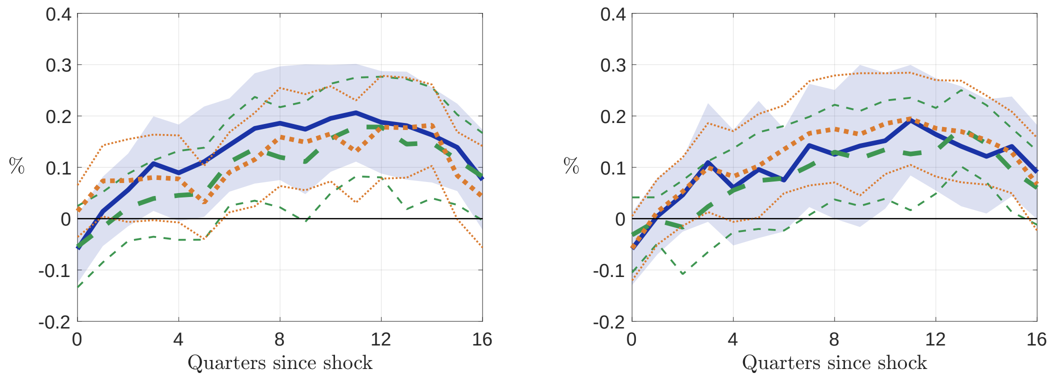


Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). See the notes to Figure 14 for details on the data treatments. The shaded and bordered areas indicate one standard error bands based on Newey–West.

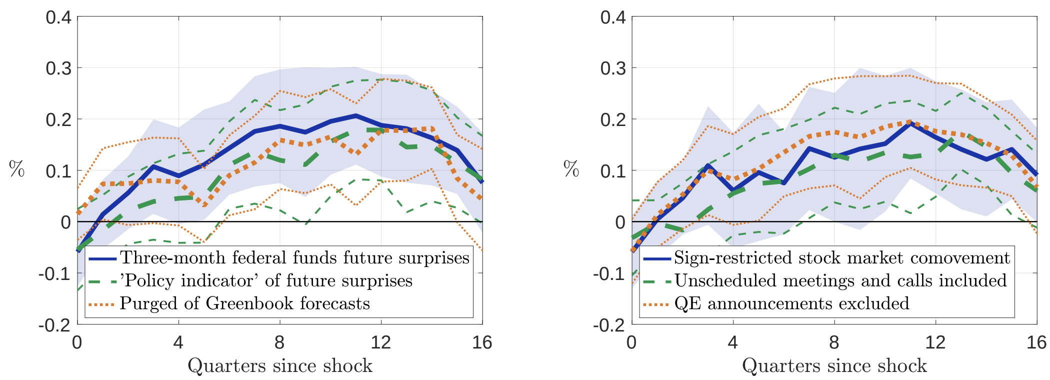
F.3 Results for additional monetary policy shocks

Figure 16: Relative markup response of firms with stickier prices to monetary policy shocks for additional monetary policy shocks

(a) Price adjustment frequency



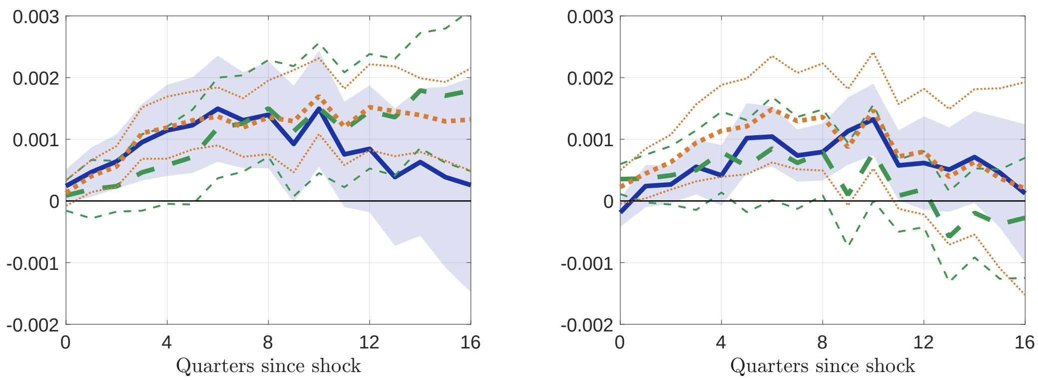
(b) Implied price duration



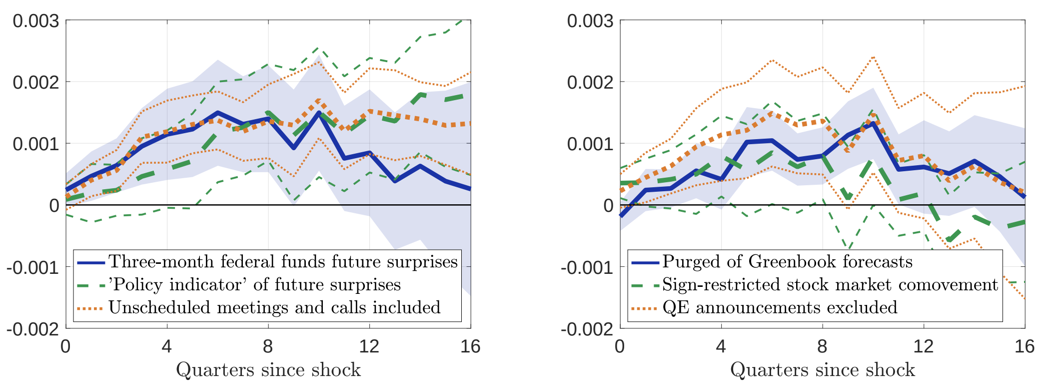
Notes: The figures show the response to a one standard deviation contractionary monetary policy shock of the firm-level markup of firms with a price adjustment frequency one standard deviation below (or with an implied price duration one standard deviation above) from panel local projections as in equation (3.4). The regressions include interactions with lagged log assets, leverage, and liquidity and their interactions with a monetary policy shock. The shaded and bordered areas indicate 90% error bands clustered by firms and quarters.

Figure 17: Markup dispersion IRFs for additional monetary policy shocks

(a) within 2d-industry-quarter



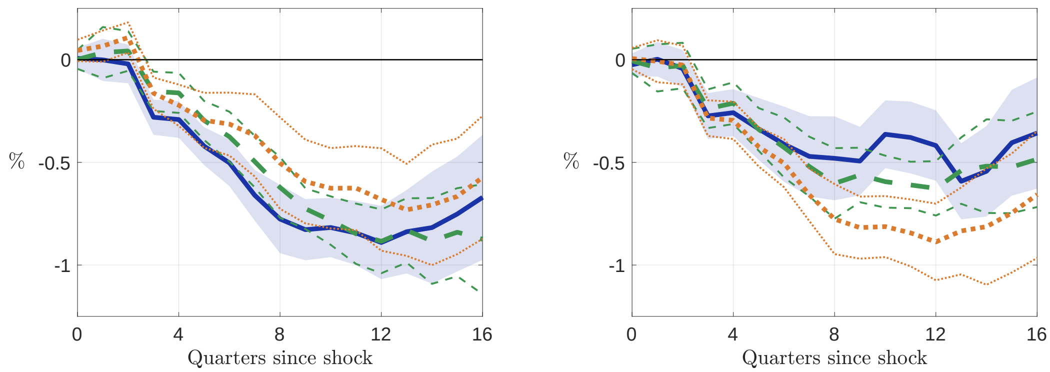
(b) within 4d-industry-quarter



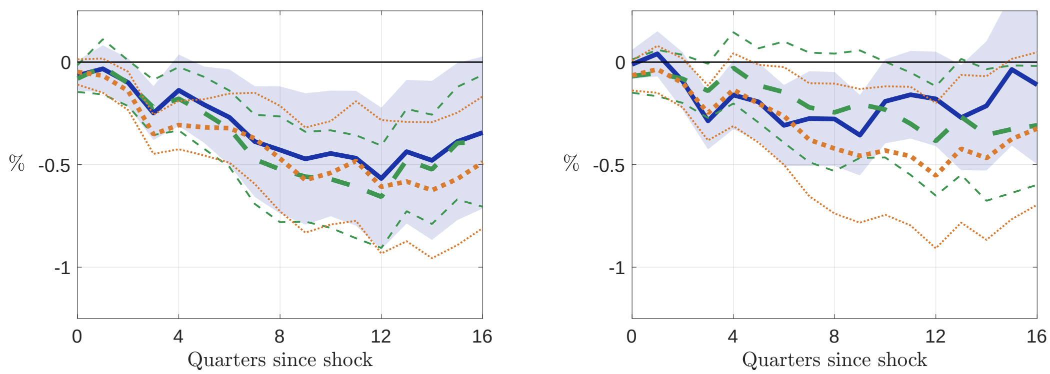
Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 18: Productivity IRFs for additional monetary policy shocks

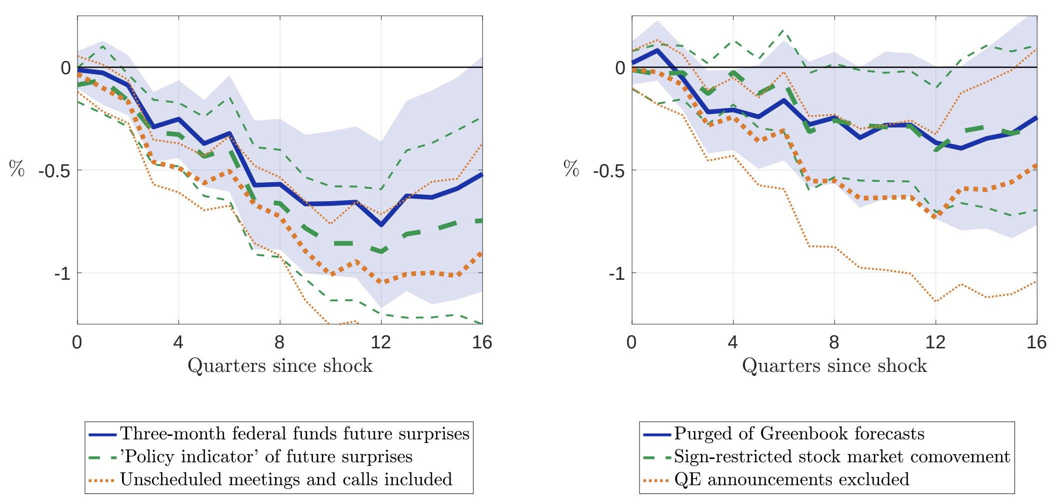
(a) TFP



(b) Utilization-adjusted TFP



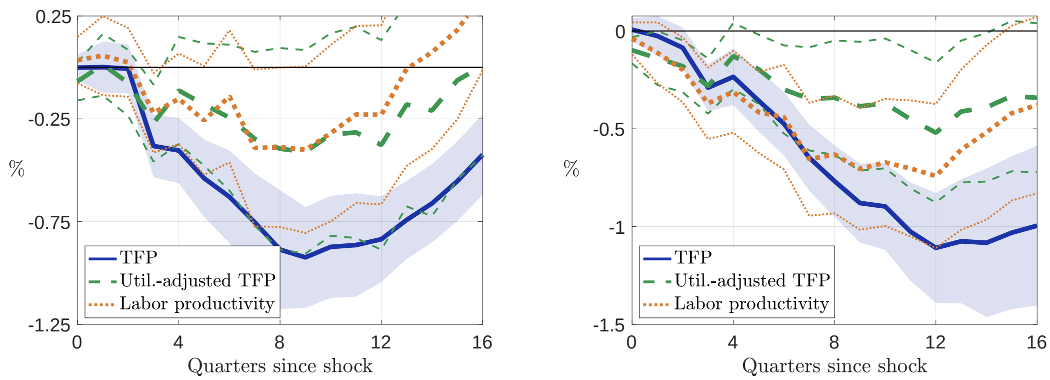
(c) Labor productivity



Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). The shaded and bordered areas indicate one standard error bands based on Newey–West.

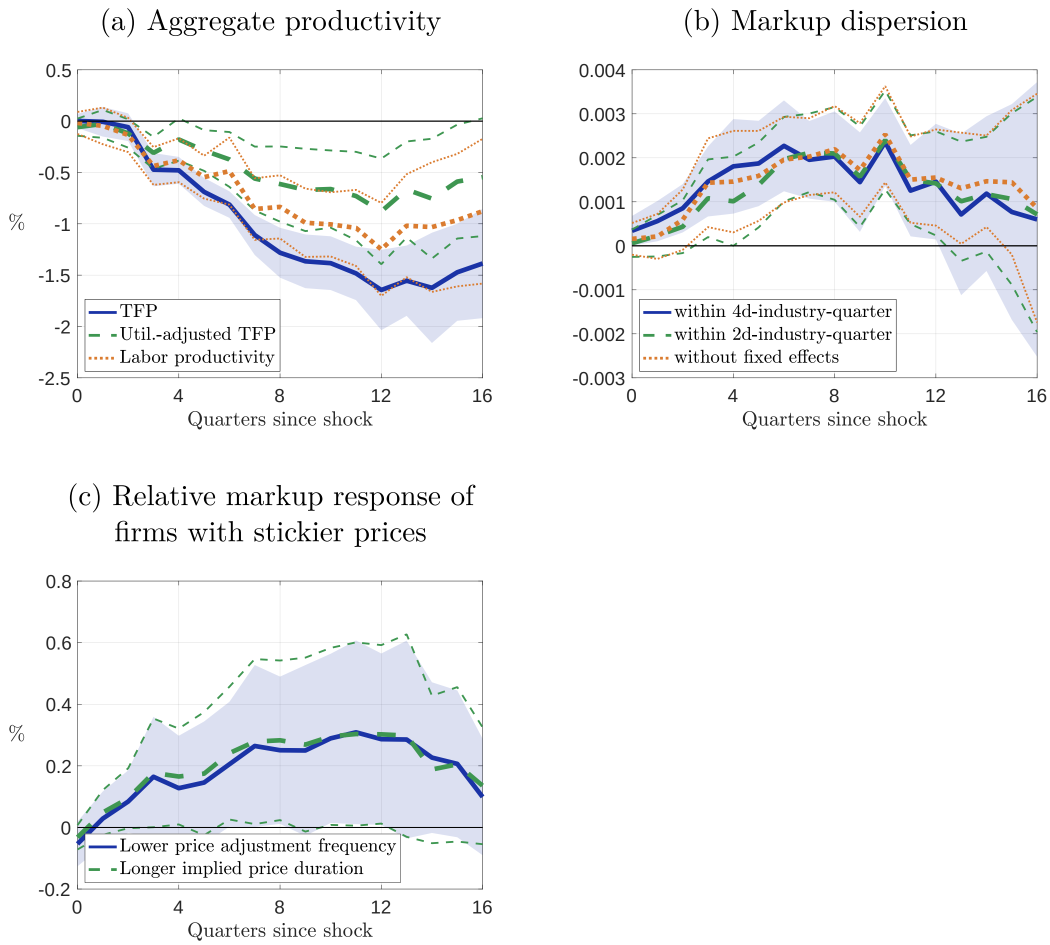
Figure 19: Responses of aggregate productivity for the Pre-Great Recession Period and including the apex of the Great Recession

(a) Pre-Great Recession (b) Include apex of Great Recession



Notes: Responses to monetary policy shocks obtained from local projections as in equation (3.5). *Pre-Great Recession only* considers observations until 2008Q2. *Including Great Recession* does not drop the period 2008Q3–2009Q2 from the sample. The shaded and bordered areas indicate one standard error bands based on Newey–West.

Figure 20: Main results using LP-IV



Notes: Responses to monetary policy shocks obtained from local projections with instrumental variables (LP-IV), $y_{t+h} - y_{t-1} = \alpha^h + \beta^h \Delta R_t + \gamma_1^h (y_{t-1} - y_{t-2}) + u_t^h$, and analogues of the panel local projections, where the changes in the one-year Treasury rate, ΔR_t, are instrumented with the monetary policy shocks $\varepsilon_t^{\mathrm{MP}}$. 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 clustered by firms and quarters.

G Details on the Quantitative New Keynesian Model

This section presents details on the quantitative New Keynesian model in Section 4. We assume a representative infinitely-lived household who maximizes

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

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, 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 $\lim_{T\to\infty} \mathbb{E}_t[\Lambda_{t,T} \frac{B_T}{P_T}] \geq 0$ 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

$$Y_t = \left(\int_0^1 Y_{it}^{\frac{\eta-1}{\eta}} \mathrm{d}i \right)^{\frac{\eta}{\eta-1}}, \tag{G.2}$$

where η is the elasticity of substitution between differentiated goods $\{Y_{it}\}$. Intermediate goods are with the 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. Final good aggregation implies an isoelastic demand schedule for intermediate goods given by $Y_{it} = (P_{it}/P_t)^{-\eta} Y_t$, where $P_t = (\int_0^1 P_{it}^{1-\eta} di)^{1/(1-\eta)}$ denotes the aggregate price index and P_{it} the firm-level price. Firms may reset their prices P_{it} with firm-specific probability $1-\theta_i$. The price setting policy maximizes the value of the firm to its shareholder,

$$\max_{P_{it}} \sum_{j=0}^{\infty} \theta_i^j \mathbb{E}_t \left[\frac{\Lambda_{t,t+j}}{P_{t+j}} \left(\frac{P_{it}}{P_{t+j}} - W_{t+j} \right) \left(\frac{P_{it}}{P_{t+j}} \right)^{-\eta} Y_{t+j} \right]. \tag{G.3}$$

The firm type k-specific price index is

$$P_{kt} = \left[(1-\theta_k) \tilde{P}_{kt}^{1-\eta} + \theta_k P_{kt-1}^{1-\eta} \right]^{\frac{1}{1-\eta}} \tag{G.4}$$

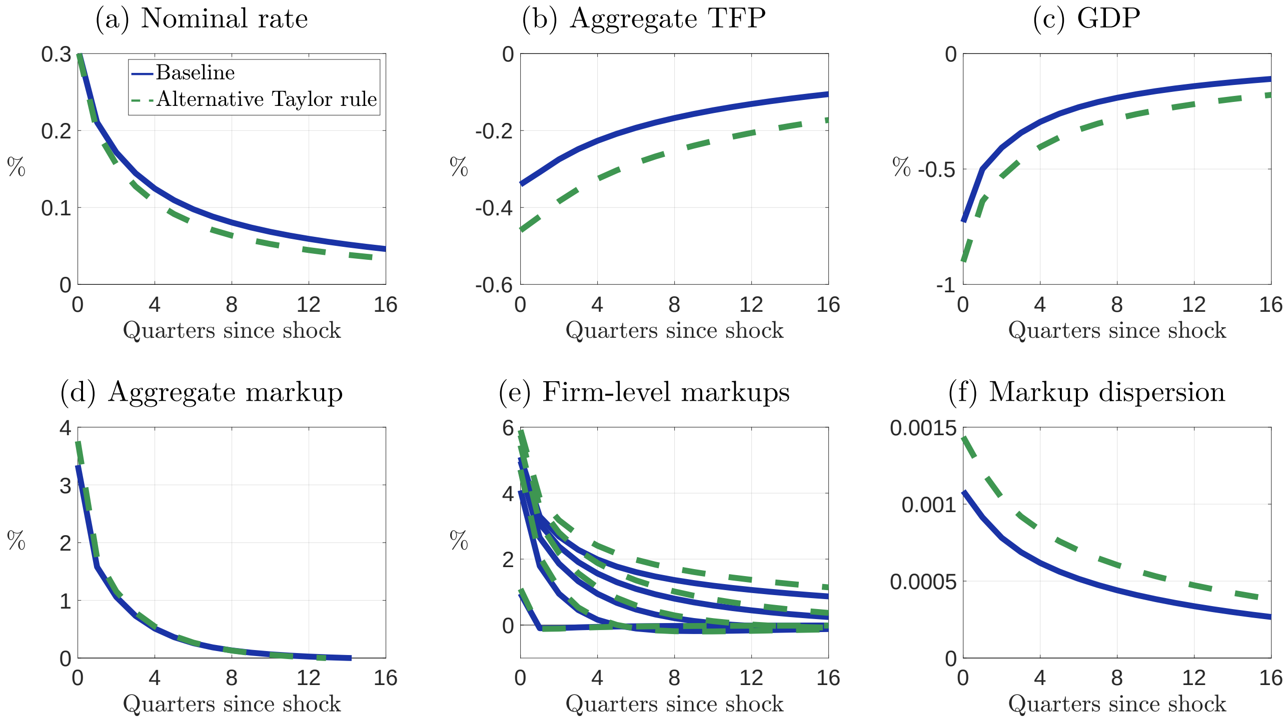
where $\tilde{P}_{kt}$ is the optimal reset price of a firm of type k. The monetary authority aims to stabilize inflation (P_t/P_{t-1}) and fluctuations in output, Y_t, around its natural level, denoted $\tilde{Y}_t$, by following the Taylor-type rule, subject to monetary policy shocks ν_t,

$$R_t = R_{t-1}^{\rho_r} \left[\frac{1}{\beta} \left(\frac{P_t}{P_{t-1}} \right)^{\phi_\pi} \left(\frac{Y_t}{\tilde{Y}_t} \right)^{\phi_y} \right]^{1-\rho_r} \exp\{\nu_t\}, \quad \nu_t \sim \mathcal{N}(0, \sigma_\nu^2). \tag{G.5}$$

The competitive equilibrium is defined by firm-level allocations $\{Y_{it}, N_{it}\}_{t=0}^{\infty}$ and prices $\{P_{it}\}_{t=0}^{\infty}$ 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. The final goods market clears, $Y_t = C_t$, and the labor market clears, $N_t = \int_0^1 N_{it} \mathrm{d}i$, in every period t.

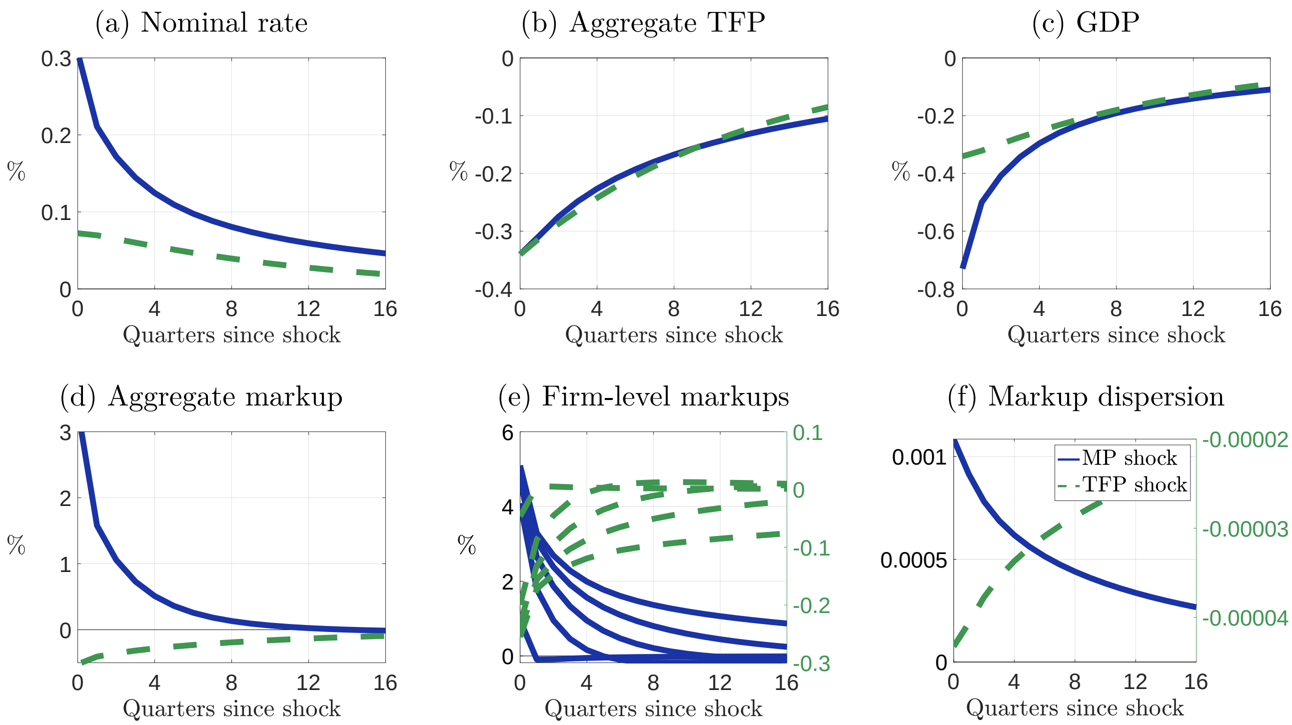
H Additional Model Results

Figure 21: Model responses to monetary policy shocks: Alternative Taylor rule



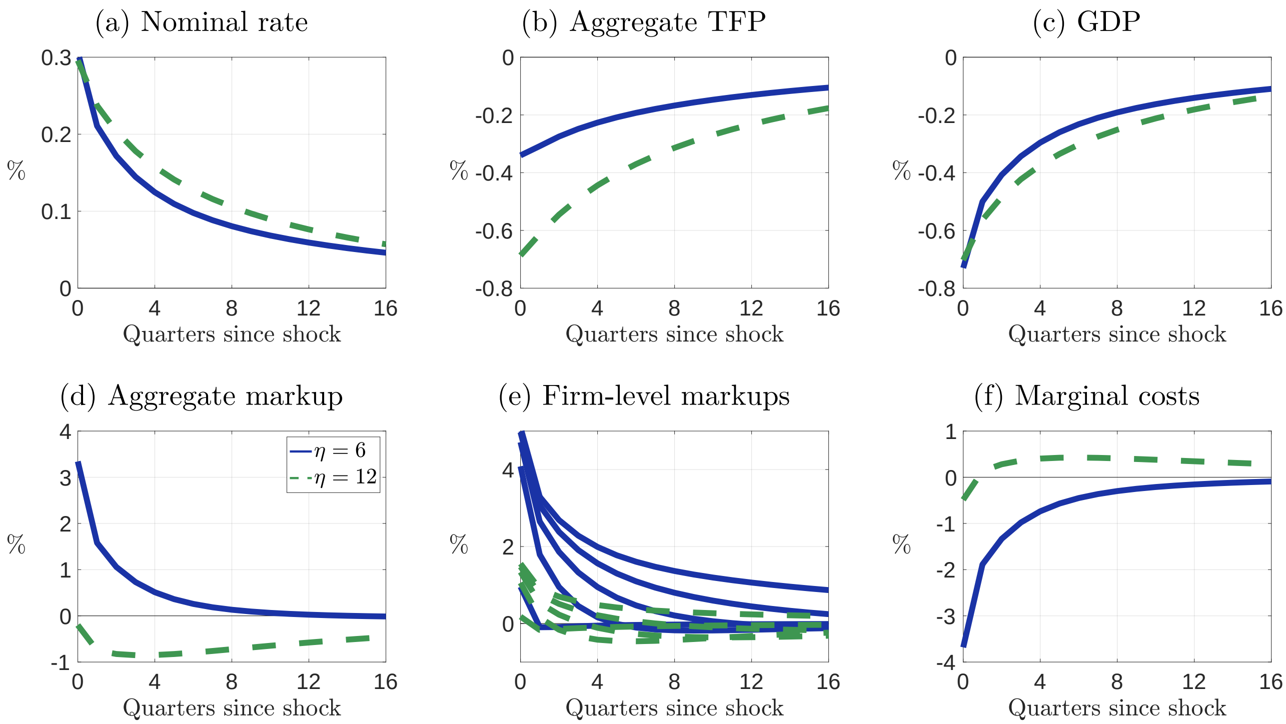
Notes: This figure shows impulse responses to a one standard deviation monetary policy shock. *Baseline* corresponds to the model in the main text. In particular, the central bank follows the Taylor rule in equation (G.5), which reacts to fluctuation in the output gap. The gap is defined relative to natural output (the level prevailing under flexible prices), which is unchanged after monetary policy shocks. *Alternative Taylor rule* corresponds to a setup in which the central bank computes natural output based on the observed movements in aggregate TFP. Conseqently, natural output is perceived to react to monetary policy shocks, which leads to a different policy response. The standard deviation of monetary policy shocks σ_ν is re-calibrated to match the response of the nominal rate of 30bp.

Figure 22: Model responses to TFP shocks



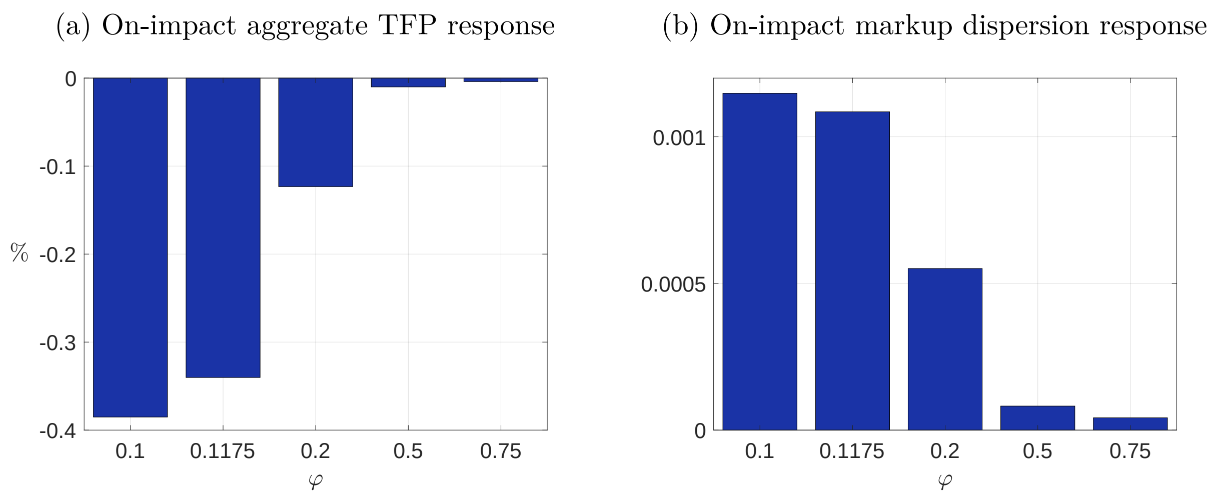
Notes: This figure compares the impulse responses to a one standard deviation monetary policy shock to those to a TFP shock. The persistence of TFP and the TFP shock size are calibrated to match the shape of the TFP response to monetary policy shocks.

Figure 23: Model responses to monetary policy shocks: Varying the Elasticity of substitution



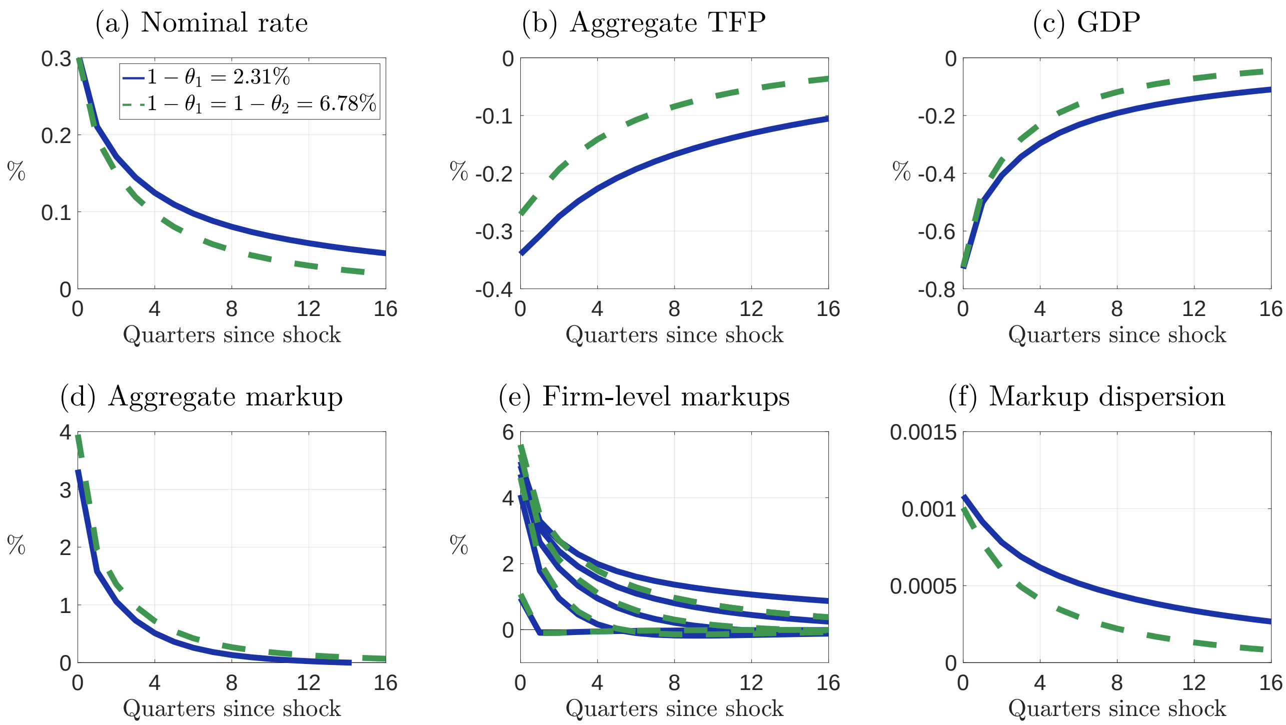
Notes: This figure shows impulse responses to a one standard deviation monetary policy shock for two values of the elasticity of substitution between variety goods η. The value 6 corresponds to our baseline calibration and the value 12 corresponds to an intermediate value of elasticities considered in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

Figure 24: Model responses to monetary policy shocks: Varying the Frisch elasticity of labor supply



Notes: This figure plots the size of on-impact impulse responses of TFP and markup dispersion to a one standard deviation monetary policy shock under different calibrations of the Frisch elasticity of labor supply φ. The value 0.1 corresponds to the lower end of short-run elasticity estimates surveyed by Ashenfelter et al. (2010). The value 0.1175 corresponds to our baseline calibration, which matches the contribution of TFP to the GDP response and is in the range of Ashenfelter et al. (2010). The value 0.5 corresponds to a lower end and an intermediate value (which is suspected to be potentially upward-biased) in Chetty et al. (2011). The standard deviation of monetary policy shocks σ_ν is re-calibrated to match the response of the nominal rate of 30bp.

Figure 25: Model responses to monetary policy shocks: Re-calibrating the largest price rigidity



Notes: This figure compares impulse responses to a one standard deviation monetary policy shock in the baseline calibration to the setting in which we reduce the largest price rigidity, θ_1, to the second-largest price rigidity, θ_2. The standard deviation of monetary policy shocks σ_ν is re-calibrated to match the response of the nominal rate of 30bp.