THE MISALLOCATION CHANNEL OF MONETARY POLICY

MATTHIAS MEIER† AND TIMO REINELT‡

April 9, 2019

Abstract

While the effects of monetary policy shocks on macroeconomic aggregates are well-documented, surprisingly little is known about the aggregate TFP response. We establish that monetary policy shocks lower aggregate TFP and increase the dispersion of markups across firms. To rationalize these empirical findings, we propose a heterogenous firm New Keynesian model, in which monetary policy is transmitted through factor misallocation. The key departure from the standard model is firm heterogeneity in price rigidity. This endogenously gives rise to markup heterogeneity through precautionary price setting. In the model, monetary policy shocks increase markup dispersion, increase misallocation, and lower aggregate TFP.

Keywords: Monetary Policy, TFP, Misallocation, Firm heterogeneity.

Contact information:
Timo Reinelt
University of Mannheim
Block L7, 3-5, Room 245
D-68161 Mannheim
E-mail: timo.reinelt@gess.uni-mannheim.de
Phone: +49 621 181-1802
Fax: +49 621 181-1807

† Universität Mannheim, Department of Economics, Block L7, 3-5, 68161 Mannheim, Germany; E-mail: m.meier@uni-mannheim.de.
‡ Universität Mannheim, Center for Doctoral Studies in Economics (CDSE), Block B6, 30-32, 68159 Mannheim, Germany; E-mail: timo.reinelt@gess.uni-mannheim.de.
1. INTRODUCTION

The nature of the monetary transmission channel is a classical question in macroeconomics. Understanding it is of central importance for positive questions of business cycle research and for normative aspects of monetary policy. This paper establishes a novel transmission channel, the *misallocation channel* of monetary policy. Monetary policy shocks affect the dispersion of markups across firms, which affects allocative efficiency and thereby aggregate TFP.

The motivating empirical finding of this paper is that monetary policy shocks lower aggregate TFP. While a vast empirical literature estimates the effects of monetary policy shocks on macroeconomic aggregates,\(^1\) little is known about the effects of monetary policy shocks on aggregate productivity.\(^2\) This is surprising when contrasted with the prominent role of productivity fluctuations in DSGE models.\(^3\) To identify US monetary policy shocks, we use high-frequency changes in federal funds futures around monetary policy announcements.\(^4\) We find that a contractionary monetary policy shock that raises the federal funds rate by 25 basis points depresses aggregate TFP by up to 0.6%.

Three possible explanations why monetary policy shocks may lower measured aggregate TFP are capacity utilization, aggregate markups, and R&D. First, if the shock lowers utilization and the TFP measure does not take this into account, the decline in TFP might be an artefact. Second, commonly used measures of TFP (Fernald, 2014) in the tradition of Solow (1957) use the labor income share as the Solow weight. TFP is then mismeasured in the presence of aggregate markups, see Hall (1986). Tighter monetary policy can lower such mismeasured TFP even when true TFP is unchanged. Empirically, adjusting for utilization and markups reduces the response of aggregate TFP by half. A contractionary monetary policy shock that raises the federal funds rate by 25 basis points depresses adjusted aggregate TFP by up to 0.3%. Third, tighter monetary policy may lower true TFP because

---

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

\(^2\)An exception is Evans (1992) who shows that TFP fluctuations are predictable from lags of the money stock and treasury bill rates and Evans and Santos (2002) who find that TFP declines after monetary policy shocks. For labor productivity, Christiano et al. (2005) provide empirical evidence that it falls in response to monetary policy shocks, which their model only partially explains through utilization and fixed costs.

\(^3\)Productivity fluctuations are not only a central driver of business cycles in Real Business Cycle models (Kydland and Prescott, 1982), they remain important in many estimated New Keynesian models (Smets and Wouters, 2007), and in business cycle accounting (Brinca et al., 2016).

\(^4\)This identification strategy follows a recent literature, see, e.g., Gertler and Karadi (2015), Gorodnichenko and Weber (2016), and Nakamura and Steinsson (2018).
firms reduce efforts to raise their productivity, e.g. R&D investment.\textsuperscript{5} Empirically, however, we find that R&D investment does not significantly respond to monetary policy shocks.

This paper argues that misallocation explains the remaining TFP response to monetary policy shocks. To provide empirical evidence in support of such a misallocation channel, we focus on markup dispersion. We use quarterly US firm-level data from Compustat and estimate firm-level markups following \textit{De Loecker and Warzynski} (2012) and \textit{De Loecker et al.} (2018). We focus on variation of markups within narrowly-defined sectors and within quarters. This measure is justified by theory and related to a literature that studies misallocation as a source of cross-country TFP differences.\textsuperscript{6} We establish that contractionary monetary policy shocks significantly raise markup dispersion. This response is driven by firms with high average markups, consistent with our model.

The model we propose is a heterogenous firm New Keynesian model. It only has a single departure from the basic New Keynesian model: firm heterogeneity in price rigidity. A key mechanism is that firms with more rigid prices have higher markups.\textsuperscript{7} This mechanism is related to \textit{Fernandez-Villaverde et al.} (2015), in which markups increase in response to uncertainty shocks. In our setup, firms with more rigid prices effectively face more uncertainty when resetting prices and thus charge a higher markup than firms with less rigid prices. Thus, heterogeneity in price rigidity endogenously generates markups dispersion and lowers TFP in the stochastic steady state. A contractionary monetary policy shock increases markup dispersion further: firms with a high markup tend to have more rigid prices, so their markups increase by more than for firms with low markups and less rigid prices. Hence, the shock lowers TFP. Quantitatively, we show that this misallocation channel of monetary policy can account for roughly half of our empirically estimated peak TFP response.

This paper relates to a growing empirical literature that studies the heterogeneous impact of monetary policy shocks on firms as a means to better understand the relevant monetary transmission channels. An early contribution is the seminal \textit{Gertler and Gilchrist} (1994), which finds that small firms are more responsive to

\textsuperscript{5}E.g., \textit{Comin and Gertler} (2006) develop a DSGE model in which firms choose R&D investment.\textsuperscript{6}To name but a few, see \textit{Hsieh and Klenow} (2009), \textit{Asker et al.} (2014), \textit{Midrigan and Xu} (2014).\textsuperscript{7}New Keynesian models with heterogeneous price rigidity have been studied before, e.g., \textit{Carvalho and Schwartzman} (2015) and \textit{Carvalho and Nechio} (2016). Differently from previous work, we study the misallocation channel that this heterogeneity gives rise to. Importantly, capturing this misallocation channel requires either global or higher-order local solution techniques.
monetary policy shocks than large firms. More recently, Ippolito et al. (2018) show that firms with unhedged floating rate loans are more responsive. Ottonello and Winberry (2018) find that firms with low leverage are more responsive, while Jeenas (2018) shows that firms with little liquid assets are more responsive. In contrast, our paper studies responses in the cross-sectional distribution of firms, which map into misallocation. Eisfeldt and Rampini (2006) provide evidence of countercyclical fluctuations in capital misallocation. Instead, we study fluctuations in misallocation, measured as markup dispersion, conditional on monetary policy shocks. A further related paper is Gopinath et al. (2017), which argues the productivity slowdown in South Europe is the result of increased misallocation, in turn a consequence of low interest rates and size-dependent financial frictions. Interestingly, our findings go in the opposite direction. We find that unexpectedly higher interest rates increase misallocation.

In addition, this paper relates to a number of papers that study misallocation as a transmission mechanism for business cycle shocks: for example, aggregate productivity shocks in Khan and Thomas (2008), financial shocks in Khan and Thomas (2013), uncertainty shocks in Bloom (2009), and supply chain disruptions in Meier (2018). Closely related, Pasten et al. (2018) and the Appendix of Baqae and Farhi (2018) propose models to study the effects of monetary policy in the presence of heterogenous price rigidity and an input-output structure. Different from these papers, we focus our model on (endogeneous) markup dispersion and misallocation, and provide empirical evidence in support of the transmission channel.

The remainder of this paper is organized as follows. Section 2 discusses theories of fluctuations in measured TFP. Section 3 provides new empirical evidence on the effects of monetary policy shocks. Section 4 presents our model, and Section 5 discusses quantitative results. Section 6 concludes and an Appendix follows.

2. TFP FLUCTUATIONS

Measured aggregate TFP can fluctuate for several reasons. The traditional view is that productivity fluctuates exogenously. In this section we briefly review the measurement of TFP. We show how variable factor utilization and the presence of aggregate markups generate TFP movements in response to demand shocks. Lastly, we show how factor misallocation through markup dispersion lowers aggregate TFP.
2.1. The Solow residual

Measuring aggregate TFP goes back to the seminal Solow (1957), which we briefly review here. Consider a constant returns to scale aggregate production function

\[(2.1) \quad Y = A \cdot F(K, L),\]

where \(A\) is productivity, \(K\) and \(L\) denote aggregate capital and labor, respectively. Log-differentiating with respect to time yields

\[(2.2) \quad \Delta y = \Delta a + \frac{\partial y}{\partial k} \Delta k + \frac{\partial y}{\partial \ell} \Delta \ell,\]

where lowercase letters denote the natural logarithms of capitalized variables. We denote nominal wages and nominal rents by \(W\) and \(R\), respectively, and prices by \(P\). The labor and capital shares are then

\[(2.3) \quad w_{\ell} = \frac{W L}{P Y} \quad \text{and} \quad w_k = \frac{R K}{P Y}.\]

Under perfect competition on factor and output markets, factor prices equal their marginal products. Hence \(w_k = \frac{\partial y}{\partial k}\) and \(w_{\ell} = \frac{\partial y}{\partial \ell}\). This yields the Solow residual,

\[(2.4) \quad \Delta \text{TFP} = (\Delta y - \Delta k) - w_{\ell}(\Delta \ell - \Delta k).\]

Under the above assumptions, \(\Delta \text{TFP}\) equals actual productivity growth \(\Delta a\). For the remainder of this paper, we refer to \(\Delta \text{TFP}\) as measured aggregate TFP according to equation (2.4), which is consistent with the baseline TFP series in Fernald (2014).

The first potential reason why measured aggregate TFP may change is variable capacity utilization. If utilization is adjusted proportionally for capital and labor by \(\Delta u\), we can compute utilization-adjusted TFP as

\[(2.5) \quad \Delta \text{TFP}_{\text{util}} = \Delta \text{TFP} - \Delta u,\]

which corresponds to the utilization-adjusted TFP series in Fernald (2014).

2.2. Aggregate markups and TFP

Consider now an aggregate price markup, denoted \(\mu\), over aggregate marginal costs, denoted \(X\), such that \(P = \mu X\). Under the constant returns to scale assumption,
tion, cost minimization implies \( w_\ell = \frac{1}{\mu} \frac{W_L}{XY} = \frac{1}{\mu} \frac{\partial y}{\partial \ell} \). Now suppose the economy is hit by a monetary policy shock (or any non-productivity shock) which lowers \( \Delta y - \Delta k \) and leaves exogenous productivity unchanged, \( \Delta a = 0 \). If markups are strictly positive, \( \mu > 1 \), then such a shock changes aggregate TFP as measured by Equation (2.4) by

\[
(2.6) \quad \Delta \text{TFP} = \frac{\mu - 1}{\mu} (\Delta y - \Delta k).
\]

Thus, market power may explain why measured, aggregate TFP falls after contractionary monetary policy shocks. Importantly, the TFP decline is amplified if the price markup increases after the shock.

Fundamentally, in the presence of markups, the correct Solow weight is not the labor income share, but the labor expenditure share. The insight that markups distort the original Solow residual goes back to Hall (1986), who suggests a markup-adjusted TFP measure,

\[
(2.7) \quad \Delta \text{TFP}_{\text{Hall}} = (\Delta y - \Delta k) - \mu w_\ell (\Delta \ell - \Delta k),
\]

which uses the labor expenditure share as effective Solow weight.

### 2.3. Misallocation and TFP

Even if the TFP measure correctly accounts for utilization and aggregate markups, aggregate TFP can fluctuate absent aggregate productivity shocks. Factor misallocation across firms lowers aggregate TFP. A stylized model of heterogeneous firms allows to quantify the TFP losses from misallocation.

Consider a mass of firms, indexed by \( i \), that operate a Cobb-Douglas technology, \( Y_i = K_i^\alpha L_i^{1-\alpha} \), combining labor \( L_i \) and capital \( K_i \), and earn profits \( \pi_i = (1 - \tau_i)P_Y Y_i - WL_i - RK_i \), where the demand curve is given by \( Y_i = (P_i/P)^{-\eta}Y \). The wedge \( \tau_i \) is a shortcut that captures nominal frictions that prevent a firm from setting its static optimal price markup. A model-consistent measure of aggregate TFP can be computed as

\[
(2.8) \quad \text{TFP}_{\text{model}} = \log Y - \alpha \log K - (1 - \alpha) \log L.
\]

where aggregate output is determined by the CES aggregator \( Y = \left( \int Y_i^{\frac{1}{\eta}} \, di \right)^{\frac{\eta}{\eta-1}} \).
and aggregate capital and labor are given by \( K = \int K_i \, di \) and \( L = \int L_i \, di \), respectively. As shown by Hsieh and Klenow (2009), if the pricing friction \( \tau_i \) is log-normally distributed, TFP is given by

\[
\text{TFP}_{\text{model}} = -\frac{\eta}{2} \mathbb{V} \log(1 - \tau_i).
\]

The TFP losses from misallocation are captured by markup dispersion. Because of imperfect substitutability of goods, output is maximized if firms are of equal size. Markup dispersion induces size differences, which impairs allocative efficiency and reduces aggregate output. This result allows us to empirically study the misallocation channel of monetary policy by estimating the response of markup dispersion to monetary policy shocks.

3. EMPIRICAL EVIDENCE

This section provides novel empirical evidence. We first show that contractionary monetary policy shocks lower aggregate TFP. Only about half of the TFP response can be explained by mismeasurement or utilization adjustment. Toward explaining the remaining drop in TFP, we show that contractionary monetary policy shocks raise markup dispersion, and vice versa for expansionary monetary policy shocks. We further find that the response in markup dispersion is mostly driven by markup dispersion across firms with high average markups.

3.1. Identification of monetary policy (MP) shocks

We identify monetary policy shocks using high-frequency futures data, which capture market expectations about the federal funds rate. The identifying restriction is that during a narrow time window around FOMC announcements, no shock other than the monetary policy shock affects the price of futures. We denote the price of a future by \( f_\tau \), where \( \tau \) is the time of the monetary announcement. A monetary policy
shock in FOMC meeting period $\tau$ is defined as

\[
\varepsilon^{MP}_{\tau} = \omega(\tau) (f_{\tau+\Delta \tau^+} - f_{\tau-\Delta \tau^-}),
\]

We specify a thirty minute window around FOMC announcements, setting $\Delta \tau^- = 10$ minutes and $\Delta \tau^+ = 20$ minutes, as in Gorodnichenko and Weber (2016).

Because both the micro data and the macro data we use is at quarterly frequency, we need to aggregate daily shocks. We assign daily shocks fully to the current quarter if they occur on the first day of the quarter. If they occur within the quarter, we partially assign the shock to the subsequent quarter. In this way, we weight shocks across quarters corresponding to the amount of time firms have had to respond. Formally, let $t$ denote quarters, then we compute quarterly shocks as

\[
\varepsilon^{MP}_t = \sum_{\tau \in D(t)} \phi(\tau) \varepsilon^{MP}_{\tau} + \sum_{\tau \in D(t-1)} (1 - \phi(\tau)) \varepsilon^{MP}_{\tau},
\]

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

Our baseline monetary shock is based on the three-month ahead federal funds future as in Gertler and Karadi (2015). For robustness, we also construct surprises in the current month federal funds futures, and use the policy news shock of Nakamura and Steinsson (2018), who extract the first principal component of the current and next month federal funds futures and the 2/3/4-quarters ahead Eurodollar futures. Panel (a) of Figure 6 in the Appendix plots the three shock series. Table I in the Appendix reports summary statistics.

**3.2. Macro evidence: MP shocks lower aggregate TFP**

We first document that monetary policy shocks lower aggregate TFP. We estimate the dynamic responses of aggregate productivity using Jordà (2005)’s local

---

8Specifically the current-month federal funds futures settle on the month’s average effective overnight federal funds rate. Shocks happen at different days of the month. To make them comparable, we use an adjustment $\omega(\tau)$. We have $\omega(\tau) = (\text{total number of days in announcement month}) / (\text{remaining number of days in announcement month after meeting in } \tau)$. If $\tau$ is within the last seven days of the month, we use the unadjusted change in the next-month federal funds future. Any other federal funds future we consider, as well as the Eurodollar futures, have their respective reference periods in the future. Then the adjustment collapses to $\omega(\tau) = 1$. 

projections. Our baseline specification is

\[(3.3) \quad x_{t+h} - x_{t-1} = \alpha^h + \beta^h e^{\text{MP}}_t + \gamma^0 e^{\text{MP}}_{t-1} + \gamma^1 (x_{t-1} - x_{t-2}) + u^h_t,\]

where \(x_t\) denotes log aggregate productivity in quarter \(t\), and \(h = 0, \ldots, 16\) the horizon. The coefficient \(\beta^h\) is the cumulative response of productivity growth \(h\) periods after a monetary policy shock. We include lagged productivity growth to control for a potentially changing conduct of monetary policy along the path of productivity growth. We include one lag of the monetary policy shock to control for the serial correlation that arises from time aggregation in (3.2).

As measures of aggregate productivity, we consider TFP and utilization-adjusted TFP from Fernald (2014), which correspond to equations (2.4) and (2.5). Additionally we compute a markup-corrected measure of aggregate TFP, see equation (2.7). This requires price markups and we use the estimated markup series in De Loecker et al. (2018). We further consider labor productivity, which obviates the need to specify an aggregate production function. Panel (c) of Figure 6 in the Appendix plots the five productivity series. Our sample runs from 1995Q1 to 2018Q3. We exclude the apex of the financial crisis in 2008Q3 to 2009Q2.

Figure 1 shows the estimated responses of aggregate productivity to a contractionary monetary policy shock. The shock is scaled to raise the federal funds rate by 25 basis points at peak. The figure has three main takeaways. First, tighter monetary policy lowers aggregate productivity. This decline is both statistically and economically significant. A 25 basis point unexpected increase in the federal funds rate lowers aggregate productivity by up to 0.6% within the first three years. Second, the response of aggregate productivity builds up gradually and is highly persistent. Third, the differences across productivity measures are relatively small. Utilization adjustment accounts for about one third of the baseline TFP response.

Panels (c) and (d) of Figure 1 show that both TFP responses are only marginally affected when using the markup-adjusted TFP series instead of the baseline series.

---

9Fernald (2014) computes \(\Delta y\) as real business output growth, \(\Delta k\) as real capital growth (after applying the perpetual inventory method to 15 types of NIPA investment categories), \(\Delta \ell\) as the growth of hours worked plus growth in labor composition/quality, and growth in the utilization rate as growth in hours per worker.

10Labor productivity is real output per hour in the nonfarm business sector, in FRED: OPHNFB.

11We want to avoid that our results are predominantly driven by extraordinarily large macroeconomic shocks around the Great Recession. Including this period strengthens our findings.

12At a six-year horizon, the responses of TFP and utilization-adjusted TFP regress toward zero and become less significant, see panel (c) of Figure 9 in the Appendix.
Figure 1: Responses of aggregate productivity to a 25 basis points contractionary monetary policy shock

(a) TFP

(b) Utilization-adjusted TFP

(c) Markup-adjusted TFP

(d) Markup- and utilization-adjusted TFP

(e) Labor productivity

Notes: The plots show the responses of aggregate productivity to a contractionary monetary policy shock, i.e., coefficients $\beta_h$ in equation (3.3). TFP and utilization-adjusted TFP is from Fernald (2014). Markup-adjustment follows equation (2.7) using estimated markups from De Loecker et al. (2018). The shaded area is a one standard error band based on the Newey-West estimator.
Given the prominent role of aggregate markups in the monetary transmission of New Keynesian models, we consider this result important.

Finally, we investigate the robustness of our results in various directions. First, we use alternative monetary policy shocks, the high-frequency changes in the current month federal funds future and the one in Nakamura and Steinsson (2018). Figure 7 shows that the responses of baseline TFP, utilization-adjusted TFP, and labor productivity are broadly robust. Second, to investigate whether some outliers drive the empirical results, panel (a) of Figure 10 provides a scatterplot of the local projection. While some events play a larger role than others, the results are clearly not driven by a few outliers. Third, the aggregate TFP response is not exclusively the response of TFP in either the investment good or the consumption good sector, see panels (a) and (b) of Figure 9.

Another concern may be that our results are specific to unconventional monetary policy during the sample period. To address this concern, we drop all monetary policy shocks during QE announcements. Figure 8 shows that this leaves our results practically unchanged.

Yet another concern may be the informational content of monetary policy announcement. If the Fed has private information about the future state of the economy, policy announcements will be signals of such information, see Nakamura and Steinsson (2018) and Jarocinski and Karadi (2018). In this case, monetary policy shocks are not fully exogenous. To address this concern, we adopt the ‘poor-man sign restriction’ in Jarocinski and Karadi (2018), which excludes any monetary policy shock that coincide with stock market price movements in the same direction. While this excludes half of the initial daily shocks, the quarterly shock series is similar to the original one, see panel (b) of Figure 6. Importantly, the productivity responses are, if anything, somewhat stronger than the baseline results, see Figure 8.

### 3.3. Micro evidence: MP shocks increase markup dispersion

Motivated by the empirical evidence that aggregate TFP strongly falls, we next ask whether markup dispersion across firms can account the TFP decline. Using firm-level data, we show that markup dispersion indeed increases in response to contractionary monetary policy shocks. In addition, we show that firms with high markups tend to increase their markups even more after contractionary monetary policy shocks.

To study markup dispersion, we use balance sheet data from Compustat. Compu-
MEIER & REINELT

stat has two advantages. First, it provides data for a large number of US firms at quarterly frequency. With annual firm-level data (e.g., IRS data), we would need to aggregate monetary policy shocks over a full year, which would dilute the informativeness of the shocks. Second, while Compustat only contains listed firms, it does cover all sectors. In contrast, there is excellent establishment level data, which, however, only covers the manufacturing sector (e.g., AMS data).

Recall from Section 2 that in a stylized economy, the first-best allocation is achieved if markups are equalized across firms. Against this benchmark, a higher cross-sectional dispersion in markups means misallocation. In quarterly Compustat data, we consider all industries except finance, insurance, real estate, and public administration. To estimate firm-level markups, we follow De Loecker and Warzynski (2012) and De Loecker et al. (2018) and compute the markup as the ratio of sales \( \text{SALEQ} \), purged of measurement error using a non-parametric projection, to costs of variable inputs \( \text{COGSQ} \). This inverse cost share estimates the markup up to an industry-specific output elasticity of variable inputs. Since we focus on within-industry markup dispersion, we can ignore this constant. We eliminate observations from the sample if the cost share is in the top or bottom 1% in the quarter.

To estimate the effects of monetary policy shocks on markup dispersion, we re-use the regression model in equation (3.3), where \( x_t \) now becomes the cross-sectional variance of log markups, within industries and periods. The baseline monetary policy shock is, again, based on high-frequency changes in the three-month federal fund futures around FOMC announcements. Panel (a) of Figure 2 shows the response of the quarterly cross-sectional markup dispersion to a contractionary monetary policy shock. The key finding is that markup dispersion significantly increases. Further, the response of markup dispersion is quite persistent, albeit less than the response of aggregate productivity.

In addition to correcting for measurement error in sales, we examine alternative additional data treatments. First, we show the results when trimming the top and bottom 2.5% and 5% of cost shares, respectively. Second, the information provided by small firms might be more erroneous, so we drop all observations with real sales below 1 mln (in 2010 US$). Third, we drop observations with real sales growth above 200% or below -66%. Fourth, we look at the combination of all data treatments. Panel (b) of Figure 2 shows that these alternative, and, arguably, more aggressive, data treatments diminish the magnitudes of the response only slightly.

Furthermore, the response of markup dispersion holds up to the same scrutiny
Figure 2: Responses of markup dispersion to a 25 basis points
contractionary monetary policy shock

(a) Within 2d-industry-quarter

(b) Within 4d-industry-quarter

(c) Within 2d-industry-quarter: Data treatments

(d) Within 4d-industry-quarter: Data treatments

Notes: The plot shows the response of markup dispersion within two-digit and four-digit
industry-quarters, respectively. The baseline specification drops the top and bottom 1% of
cost shares. 2.5% trimming and 5% trimming drops the corresponding tails. Exclude
small firms drops observations with real sales below 1 mln (in 2010 US$). Exclude exces-
sive growth drops observations with real sales growth above +200% or below -66%. All
treatment combines these restrictions. Inference as in Figure 1.

that we applied to the TFP response. Panel (a) and (b) of Figure 11 shows that our
results are broadly robust to alternative monetary policy shock series, one based on
the current-month federal funds future and one following Nakamura and Steinsson
(2018). Panel (c) and (d) of Figure 11 show that higher markup dispersion in
response to contractionary MP shocks is neither driven by QE-related shocks, nor
by the information component of monetary announcement. To investigate whether
peculiar outliers may drive the result, panel (b) of Figure 10 provides a scatterplot
to introspect what drives our $\beta^h$ coefficient.

Next, we ask whether the increase in markup dispersion can quantitatively account for the estimated negative TFP response. To map markup dispersion into TFP, we use the accounting model in Section 2. Denoting the monetary policy shock by $\varepsilon^{\text{MP}}$, it follows that

$$\frac{\partial \text{TFP}}{\partial \varepsilon^{\text{MP}}} = -\frac{\eta}{2} \frac{\partial \mathbb{V} \log \tau_i}{\partial \varepsilon^{\text{MP}}}.$$  

(3.4)

Assuming a demand elasticity of $\eta = 21$ (Fernandez-Villaverde et al., 2015), the estimated increase in $\mathbb{V} \log \tau_i$ of 0.001 (over-)explains a decline in aggregate TFP by 1.05%. To summarize, the presented evidence suggests that the misallocation channel of monetary policy shocks is of quantitative importance.

At the steady state of the basic New Keynesian model, both positive and negative demand shocks raise markup dispersion. We therefore separately estimate the responses to positive and negative monetary policy shocks on markup dispersion. We consider the extended regression model

$$x_{t+h} - x_{t-1} = \alpha^h + \beta^h_{(+)} \varepsilon^t_{t} \mathbb{I}_{\{\varepsilon^t_{t} \geq 0\}} + \beta^h_{(-)} \varepsilon^t_{t} \mathbb{I}_{\{\varepsilon^t_{t} < 0\}}$$

$$+ \gamma^h_0 \varepsilon^{t-1}_{t-1} \mathbb{I}_{\{\varepsilon^{t-1}_{t-1} \geq 0\}} + \gamma^h_0 \varepsilon^{t-1}_{t-1} \mathbb{I}_{\{\varepsilon^{t-1}_{t-1} < 0\}}$$

$$+ \gamma^h_1 (x_{t-1} - x_{t-2}) + u^h_t,$$

(3.5)

which includes the expansionary and contractionary monetary policy shock as separate regressors. By normalizing expansionary shocks to positive sign, both $\beta^h$ coefficients can be interpreted as the response of markup dispersion to monetary policy shocks in absolute value. As Figure 3 shows, markup dispersion increases after contractionary shocks, and decreases after expansionary shocks.

To learn more about firm heterogeneity in markup responses, we estimate the firm-level response of markups to monetary policy shocks conditional on a firm’s long-run markup. We measure a firm’s long-run markup by the average markup four years before the shock and denote it by $\bar{x}_{it-1}$. Letting $x_{it}$ denote the markup of firm $i$ in period $t$, we estimate the markup response using the regression equation

$$x_{it+h} - x_{it-1} = \alpha_{ht} + \beta^h_{t} \varepsilon^t_{t} \bar{x}_{it-1} + \Gamma' Z_{it-1} + u^h_{it},$$

(3.6)

which includes an industry-quarter fixed effect for horizon $h$. We also include the
lagged monetary policy shock, the long-run markup, lagged firm size, and its interaction with monetary policy in the regression. Figure 4 shows that a firm with a 1 percentage point higher long-run markup increases its markup by up to 4 percentage points relative to the industry-quarter average, in response to a 25 basis points monetary policy shock.

In Section 4 we show that these patterns are consistent with heterogeneity in price stickiness. Firms with stickier prices set higher markups for precautionary reasons and their markups increase even more after contractionary monetary policy shocks. Therefore, markup differences are amplified after contractionary monetary policy shocks, which increases misallocation that results in a TFP loss.

3.4. Alternative explanations of lower TFP

An alternative reason for lower TFP is that firms reduce efforts to raise their productivity after tighter monetary policy. Aggregate TFP then falls because firm-level productivity falls. One way in which firms can raise their own productivity is R&D investment. For instance, Comin and Gertler (2006) develops a DSGE model in which firms choose R&D investment. However, panel (c) of Figure 12 in the

Figure 3: Response of markup dispersion to contractionary and expansionary monetary policy shocks

Notes: Response of markup dispersion within four-digit industry-quarters to contractionary and expansionary monetary policy shocks. The shocks are normalized such that the shocks can be interpreted in absolute value. The responses are given by estimates of $\beta_{(+)}$ and $\beta_{(-)}$ in equation (3.5). Inference as in Figure 1.
Figure 4: Firm-level heterogeneity in the markup response to monetary policy shocks

Notes: Response of firm-level markup to a monetary policy shock interacted with firm-level long-run markup. The response is given by estimates of $\beta_h$ in equation (3.6). Inference is based on two-way clustered standard errors by firms and quarters. The shaded area is a one-s.e. band.

Appendix shows that R&D investments does not consistently and significantly fall after contractionary monetary policy shocks.

Higher uncertainty is an alternative explanation of higher misallocation and thus lower TFP. If tighter monetary policy raises uncertainty, this may increase capital misallocation, e.g., through the real option channel, see Bloom (2009). We consider four measures of uncertainty: macro uncertainty in Jurado et al. (2015), financial uncertainty in Ludvigson et al. (2018), the VIX used in Bloom (2009), and policy uncertainty in Baker et al. (2016). Figure 13 in the Appendix shows that uncertainty does not consistently increase across monetary policy shock measures. Whereas our baseline three-month federal funds future shock leads to somewhat higher uncertainty, the estimated response is negative when using current-month federal funds futures. In contrast, the TFP and markup dispersion responses are consistent across all monetary policy shock measures, see Figures 7 and 11. We conclude that uncertainty is not the main driver of misallocation in response to monetary policy shocks.

4. NEW KEYNESIAN MODEL WITH FIRM HETEROGENEITY

This section proposes a New Keynesian model, in which firm heterogeneity is the result of aggregate risk in combination with firm-level differences in price rigidity and precautionary price setting. The endogenous correlation of price rigidity with
markup differences in the stochastic steady state leads to TFP losses after contractionary monetary policy shocks.

4.1. Model setup

Households

We assume a representative infinitely-lived household seeking to maximize

\[
E_0 \sum_{t=0}^{\infty} \beta^t \exp\{Z_t\} U(C_t, N_t),
\]

subject to the period budget constraint \(P_tC_t + R_{t-1}B_t \leq B_{t-1} + W_tN_t + D_t\) 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 \(R_{t-1}\), \(N_t\) represents employment with \(W_t\) the nominal wage, and \(D_t\) are aggregate dividends. The discount factor shock process follows

\[
\log Z_t = \rho z \log Z_{t-1} + \varepsilon_{z,t}, \quad \varepsilon_{z,t} \sim N(0, \sigma_z^2).
\]

We impose the solvency constraint \(\lim_{T \to \infty} E_t[\Lambda_{t,T}B_T] \geq 0\) for all \(t\), where \(\Lambda_{t,T} = \beta^{T-t}U_{c,T}/U_{c,t}\) is the stochastic discount factor.

Final good firms

A representative firm produces final goods \(Y_t\) using a CES aggregate of differentiated goods \(Y_{i,t}\),

\[
Y_t = \left( \int_0^1 Z_i^\eta Y_{i,t}^{\eta-1} \, di \right)^{\frac{\eta}{\eta-1}},
\]

where \(\eta\) is the elasticity of substitution between differentiated goods and \(Z_i\) denotes a firm-specific demand shifter.
Intermediate good firms

Final good firms purchase differentiated goods from intermediate good firms. The economy is populated by a continuum of intermediate good firms indexed by \( i \in [0, 1] \). Each intermediate good firm produces a differentiated good using the Cobb-Douglas production technology

\[
Y_{i,t} = A_t N_{i,t}^{1-\gamma},
\]

where \( A_t \) is average firm-level productivity, which follows an exogeneous stochastic process given by

\[
\log A_t = \rho_a \log A_{t-1} + \varepsilon_{a,t}, \quad \varepsilon_{a,t} \sim \mathcal{N}(0, \sigma_a^2).
\]

where classical productivity shocks are represented by \( \varepsilon_{a,t} \). As a result of the final good firm’s technology, intermediate good firms face an isoelastic demand schedule given by

\[
Y_{i,t}(P_{i,t}) = Z_{i,t} \left( \frac{P_{i,t}}{P_t} \right)^{-\eta} Y_t,
\]

where \( P_t = \left( \int_0^1 Z_{i,t} P_{i,t}^{1-\eta} di \right)^{1/(1-\eta)} \) denotes the price index and \( P_{i,t} \) the firm-level price. A firm’s period dividends are given by

\[
D_{i,t}(P_{i,t}) = P_{i,t} Y_{i,t}(P_{i,t}) - W_t \left( \frac{Y_{i,t}(P_{i,t})}{A_t X_t} \right)^{1-\gamma}.
\]

Following Calvo (1983), firms may reset their prices \( P_{i,t} \) with probability \( 1 - \theta_i \) in any given period. In deviation from the standard model setup, the price stickiness parameter \( \theta_i \) may be different across firms. The price setting policy maximizes the value of the firm to its shareholder,

\[
\max_{P_{i,t}} \sum_{k=0}^\infty \theta_i^k \mathbb{E}_t \left[ \Lambda_{t,t+k} \frac{1}{P_{t+k}} D_{i,t+k}(P_{i,t}) \right].
\]

Market clearing

The market for the final good clears if \( Y_t = C_t \). The labor market clears if \( N_t = \int_0^1 N_{i,t} di \).
Monetary authority

The monetary authority aims to stabilize inflation \( \frac{P_t}{P_{t-1}} \) and fluctuations in output, \( Y_t \), around its steady state, denoted \( \bar{Y} \), by following the Taylor-type rule

\[
R_t = R^{\rho_r}_{t-1} \left[ \frac{1}{\beta} \left( \frac{P_t}{P_{t-1}} \right)^{\phi_s} \left( \frac{Y_t}{\bar{Y}} \right)^{\phi_y} \right]^{1-\rho_r} \exp\{\nu_t\}
\]

(4.9)

Monetary policy shocks are represented as exogenous fluctuations in \( \nu_t \).

4.2. Key features of the model

We aim to provide intuition for three critically important features of our model. First, because of price rigidity, firms set prices for several periods in advance. That introduces uncertainty in the price setting problem. In particular, firms are uncertain about future aggregate price levels, which move in response to productivity shocks. In addition, profits in (4.7) fall by more if the firm-specific price is low relative to the aggregate price than if it is high. In combination, price rigidity induces an upward bias in price setting, or precautionary price setting.

Second, within-sector markup dispersion impairs allocative efficiency. Absent differences in price rigidity, firms are (ex-ante) identical, so the first-best allocation requires all firms within a sector to produce an identical quantity of output. Conversely, markup dispersion implies output dispersion which lowers the aggregate output that can be produced with a given quantity of inputs. Therefore, markup dispersion leads to misallocation of resources and depresses aggregate TFP.\(^{13}\)

Third, monetary policy shocks that raise the interest rate lower demand and thereby decrease marginal costs. Hence, markups increase for firms which do not re-adjust their prices. On average, the markup increases by more for firms with more rigid prices compared to firms with less sticky prices. Consequently, the markups of firms with a higher markup before the shock tend to increase by more than the markups of firms with a lower markup before the shock. Hence, the monetary policy shock increases markup dispersion and depresses TFP endogenously. This is the key and novel mechanism proposed in this paper. It arises in the otherwise standard basic New Keynesian model, only by additionally assuming differences in price rigidity across firms. The next section will explore this mechanism quantitatively.

\(^{13}\)The insight that markup dispersion results in lower TFP draws on Hsieh and Klenow (2009).
5. QUANTITATIVE RESULTS

In this section we first calibrate the model. Using the calibrated model, we then investigate the effects of monetary policy shocks, with a particular focus on endogenous TFP responses.

5.1. Calibration

In the model a period is a quarter. We choose the discount factor $\beta$ to match an annual real interest rate of 4%. We assume separable preferences $U(C_t, N_t) = \frac{C_t^{1-\sigma} - 1}{1-\sigma} - \frac{N_t^{1+1/\varphi}}{1+1/\varphi}$ and assume $\sigma = 1$. We set the Frisch elasticity of labor supply to $\varphi = 1$ following Chetty et al. (2011). We follow Fernandez-Villaverde et al. (2015) and assume a baseline elasticity of substitution between differentiated goods of $\eta = 21$. We assume constant returns to scale, $\gamma = 1$.

The central feature of the model, whereby monetary policy shocks affect misallocation and TFP, is heterogeneity in price rigidity. We calibrate such heterogeneity in a stylized way by assuming that half of the intermediate goods firms set prices every period, $\theta_i = 0$ for $i \in [0, 1/2]$, while the remaining half of the firms set prices infrequently, on average every two years, $\theta_i = 7/8$ for $i \in (1/2, 1]$. Across all firms, the average price duration is one year. Without prior knowledge whether sticky firms are larger or smaller than flexible firms, we assume they are equally large on average.

We calibrate the demand shifter $Z_t$ to ensure equal size in the presence of different markups.

Precautionary price setting behavior is the firms’ endogenous reaction to aggregate uncertainty in aggregate productivity, monetary policy, and the discount factor. We use estimated standard deviations for the corresponding exogenous shocks from Smets and Wouters (2007) and set $\sigma_{a,t} = 0.45\%$, $\sigma_{z,t} = 0.24\%$, and $\sigma_{\nu,t} = 0.25\%$. This leads to a 1.6% markup difference in the stochastic steady state.

Our Taylor rule calibration follows Christiano et al. (2016) by setting $\rho_r = 0.85$, $\varphi_\pi = 1.5$, $\varphi_y = 0.05$, $\rho_\nu = 0$. For comparability with the empirically estimated IRFs, our baseline model IRFs show the responses to a monetary policy shock that raises the nominal interest rate by 25 basis points at peak.

5.2. Effects of monetary policy shocks

Figure 5 shows the effects of a contractionary monetary policy shock that increases the nominal interest rates by 25 basis points. The shock depresses aggregate demand,
Figure 5: Baseline responses to a monetary policy shock

(a) Nominal rate

(b) TFP

(c) GDP

(d) Labor

(e) Markups

(f) Inflation

Notes: This figure shows impulse responses to a monetary policy shock that increases the nominal rate by 25 basis points.
in response to which flexible-price firms revise their prices downward keeping their markup unchanged. Sticky-price firms, however, respond only sluggishly and on average their markups increase. Firms that differ in their price rigidity but that are otherwise identical, will produce different quantities corresponding to their different markups. This worsens the allocation of factors across firms and thereby depresses aggregate TFP.\footnote{An expansionary monetary policy shock leads to mirrored impulse-responses and, in particular, an increase in TFP. The results are \textit{not} driven by unconditional price dispersion, which leads to TFP losses conditional on monetary policy shocks of any sign.}

The proposed mechanism is quantitatively important and in line with our empirical evidence. The decline in TFP of 0.14\% is about half of the size estimated in Section 3. Moreover, the model explains the empirically estimated increase in markup dispersion and the empirical observation that firms with high average markups increase their markups further after contractionary monetary policy shocks.

To further highlight the model’s properties, we shut down all firm heterogeneity by setting $\theta_i = 3/4$ all $i \in [0, 1]$, and examine the effects of a monetary policy shock. The dashed lines in Figure 5 reproduce the well-known result that a New Keynesian model cannot generate first-order TFP effects. While the output response is of comparable magnitude, it is generated solely through a decrease in labor. Through the lense of the suggested model above, this is a misleading prediction.

The classical source of comovement in output and TFP are exogenous aggregate productivity shocks. In terms of understanding the business cycle, our model therefore suggests a more important role for monetary policy shocks, being able to \textit{jointly} capture sizeable output and TFP fluctuations. A further interesting implication of our model is the contemporaneous comovement of TFP and labor. Whereas the adverse TFP shock in the standard NK model increases labor, the endogeneous TFP response of a contractionary monetary policy shock in our model is associated with a decrease in labor.

6. CONCLUSION

This paper presents new empirical evidence, which shows that contractionary monetary policy shocks lower TFP and raise markup dispersion. We show that a heterogenous firm New Keynesian model rationalizes these empirical observations. The key departure from the standard model is to introduce heterogeneity in price
rigidity. This endogenously gives rise to markup heterogeneity. In the model, contractionary monetary policy shocks increase markup dispersion and lower aggregate TFP, and vice versa for expansionary shocks.
REFERENCES


THE MISALLOCATION CHANNEL OF MONETARY POLICY

Monetary Economics, 53, 369–399.


APPENDIX
### APPENDIX A: DESCRIPTIVE STATISTICS

#### TABLE I

**Summary statistics for monetary policy shocks**

(a) Daily shocks

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-month FF future</td>
<td>-0.0038</td>
<td>0.0346</td>
<td>-0.1850</td>
<td>0.1150</td>
<td>189</td>
</tr>
<tr>
<td>Sign-restricted three-month FF future</td>
<td>-0.0036</td>
<td>0.0272</td>
<td>-0.1850</td>
<td>0.1000</td>
<td>189</td>
</tr>
<tr>
<td>Current-month FF future</td>
<td>-0.0044</td>
<td>0.0403</td>
<td>-0.2062</td>
<td>0.1300</td>
<td>189</td>
</tr>
</tbody>
</table>

(b) Quarterly shocks

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-month FF future</td>
<td>-0.0077</td>
<td>0.0381</td>
<td>-0.1701</td>
<td>0.0787</td>
<td>94</td>
</tr>
<tr>
<td>Sign-restricted three-month FF future</td>
<td>-0.0073</td>
<td>0.0315</td>
<td>-0.1562</td>
<td>0.0787</td>
<td>94</td>
</tr>
<tr>
<td>Current-month FF future</td>
<td>-0.0093</td>
<td>0.0413</td>
<td>-0.1778</td>
<td>0.1357</td>
<td>94</td>
</tr>
<tr>
<td>Policy news shock</td>
<td>0.0056</td>
<td>0.0355</td>
<td>-0.1148</td>
<td>0.0754</td>
<td>77</td>
</tr>
</tbody>
</table>

Notes: Summary statistics for monetary policy shocks. Daily shocks are computed from Equation (3.1). Quarterly shocks are aggregated from daily frequency using Equation (3.2). The policy news shock is taken from Nakamura and Steinsson (2018).

#### TABLE II

**Summary statistics for Compustat data**

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>507.48</td>
<td>2668.21</td>
<td>0.00</td>
<td>181061.19</td>
<td>459295</td>
</tr>
<tr>
<td>Property, Plant</td>
<td>1199.41</td>
<td>5788.80</td>
<td>0.00</td>
<td>269010.41</td>
<td>459295</td>
</tr>
<tr>
<td>Variable costs</td>
<td>365.91</td>
<td>2051.16</td>
<td>0.00</td>
<td>127884.36</td>
<td>459295</td>
</tr>
<tr>
<td>Assets</td>
<td>2655.41</td>
<td>15055.80</td>
<td>0.00</td>
<td>860586.34</td>
<td>457353</td>
</tr>
</tbody>
</table>

Notes: Summary statistics for Compustat data. All variables are in millions of real 2010Q1 US$. 
Figure 6: Time series plots

(a) Monetary policy shocks
(b) Alterations of baseline shocks

(c) Aggregate productivity
(d) Markup dispersion

Notes: Aggregate productivity, markup dispersion, and monetary policy shocks are at quarterly frequency. Shaded gray areas indicate NBER recession dates.
Figure 7: Productivity IRFs for alternative monetary policy shocks

(a) TFP

(b) Utilization-adjusted TFP

(c) Markup-adjusted TFP

(d) Markup- and utilization-adjusted TFP

(e) Labor productivity

Notes: The shaded and bordered areas indicate a one standard error band based on the Newey-West estimator.
Figure 8: Productivity IRFs for alterations of baseline monetary policy shock

(a) TFP

(b) Utilization-adjusted TFP

(c) Markup-adjusted TFP

(d) Markup- and utilization-adjusted TFP

(e) Labor productivity

Notes: The shaded and bordered areas indicate a one standard error band based on the Newey-West estimator.
Notes: Investment-TFP and Consumption-TFP are from Fernald (2014). Estimating the productivity responses at a six-year horizon effectively changes the sample that we consider. Inference is based on Newey-West standard errors. The shaded and bordered areas show a one standard error band for the response of TFP.
Figure 10: Scatter plots on the responses of TFP and markup dispersion

(a) TFP

(b) Markup dispersion

Note: This figure plots residuals from a Frisch-Waugh-Lowell procedure that isolates the effective regressors and regressands that produce our coefficient of interest, $\beta_h$, which is the impulse response function, at $h = 8$. 
Figure 11: Markup dispersion IRFs for alternative and altered baseline monetary policy shocks

(a) Alternative monetary policy shocks

(i) Within 2d-industry-quarter

(ii) Within 4d-industry-quarter

(b) Sign-restricted shocks and Exclusion of QE dates

(i) Within 2d-industry-quarter

(ii) Within 4d-industry-quarter

Notes: The shaded and bordered areas show a one standard error band based on the Newey-West estimator.
Figure 12: Responses of GDP, Federal funds rate, R&D investment

(a) GDP  
(b) Federals funds rate  
(c) R&D

Notes: Real GDP, NIPA investment in R&D, and the quarterly average Federal funds rate are from FRED. The shaded area shows a one standard error band based on the Newey-West estimator.
Figure 13: Responses of uncertainty measures to a 25 basis points contractionary monetary policy shock

Notes: The shaded and bordered areas show a one standard error band based on the Newey-West estimator. Macro and financial uncertainty is from Ludvigson et al. (2018). Policy uncertainty is from Baker et al. (2016).