Essays on Monetary and Fiscal Policy

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Lukas Hack

im Frühjahrs-/Sommersemester 2024

Abteilungssprecher	Prof. Klaus Adam, Ph.D.
Referent	Prof. Dr. Eckhard Janeba
Koreferent	Prof. Dr. Matthias Meier

Tag der Verteidigung	16.07.2024

Acknowledgements

First and foremost, I would like to thank my supervisors, Eckhard Janeba and Matthias Meier, for their invaluable guidance and support throughout my time as a graduate student and many insightful discussions about this thesis and my other research projects. All of my work greatly benefited from their feedback and experience.

This thesis is the result of many stimulating collaborations with my co-authors. In particular, I want to thank Klodiana Istrefi, Matthias Meier, Davud Rostam-Afschar, and Hannes Twieling for many inspiring conversations. In addition, I have benefited greatly from discussions with Klaus Adam, Antoine Camous, Antonio Ciccone, Husnu Dalgic, Andreas Gulyas, David Koll, Moritz Kuhn, Krzysztof Pytka, Jan Schymik, Arthur Seibold, Michèle Tertilt, Carlos Urrutia, Carsten Trenkler, and many others.

I want to thank the amazing people I have met during my Ph.D. studies, particularly, Ursula Berresheim, Thibault Cézanne, Lukas Diebold, Tommaso Gasparini, Marina Hoch, Bjarne Horst, Alexander Matusche, Yann Müller, Oliver Pfäuti, Timo Reinelt, Ulrich Roschitsch, and Hannes Twieling with whom I share the passion for macroeconomics. I learned a lot from all of you. I thank Daniel Overbeck and Hannes Twieling for countless discussions about our research and Felix Köhler, Karl Schulz, Sebastian Seitz, and Paul Steger for a great time at the chair. I thank Yulia Evsyukova and So Jin Lee and the entire Ph.D. cohort of 2019 for all I have learned from them. Finally, being a student at the Graduate School of Economic and Social Sciences at the University of Mannheim was instrumental in preparing this thesis and I want to thank all professors and administrators involved.

None of this would have been possible without the unconditional and lifelong support from my family, especially my parents Esther and Aron, as well as Nobbe and my grandmother Jutta. Most importantly, I am greatly indebted to my wife Katharina and my dog Pascha, who always supported me in preparing this thesis. This thesis would not have been written without Katharina's unconditional support, encouragement, and love.

Contents

List of Figures

List of Tables

Preface

Monetary and fiscal policies encompass a host of policy instruments. These instruments affect the economy in general, and business cycle fluctuations in particular. Understanding business cycle fluctuations and how they are shaped by monetary and fiscal policies is a classic topic in the theoretical and in the empirical literature (e.g., Bianchi, 2012; Leeper, 1991). Expanding our knowledge in this domain is important to understand business cycles from a positive perspective, and, ultimately, to improve economic policies that tame the business cycle in a welfare enhancing manner.

While most of the empirical literature studies either fiscal or monetary policy in isolation (e.g., Christiano, Eichenbaum, and Evans, 2005; Ramey, 2011; Romer and Romer, 2010), there is a myriad of interactions between both policies. For example, fiscal policymakers may increase government spending or lower taxes to boost economic activity. Such policies results in higher demand for goods and service, and, hence, encourages private firms to raise sales prices, which jacks up inflation. To counteract inflation, monetary policy may systematically respond by hiking interest rates. The overall effect on inflation and output depends on the strength and interaction of both policies.

Another example of the interaction across policies is tax bracket creep which may occur when nominal household income is taxed progressively. For example, monetary policy may lower interest rates to stimulate the economy leading to inflation. When inflation results in higher nominal household income, then tax rates may increase. This affects even taxpayers that do not see real income growth. These tax rate changes shape incentives that affect economic decisions, and ultimately feedback into the economy. The total effect of the monetary expansion depends on the fiscal side of the economy through the tax system.

In this thesis, I make progress towards our understanding of monetary and fiscal policies. In Chapter 1, I propose a new identification approach that I use to estimate how the effects of discretionary government spending depend on time-varying systematic monetary policy. Chapter 2 expounds that time-varying systematic monetary policy poses challenges for conventional empirical strategies that aim to isolate monetary policy shocks, i.e., exogenous interest rate changes. Lastly, in Chapter 3, I study tax bracket creep and how it affects the macroeconomy in general, and the propagation of monetary policy shocks specifically.

In Chapter 1, which is joint work with Klodiana Istrefi and Matthias Meier, I propose a novel identification design to estimate the effects of U.S. systematic monetary policy on the propagation of macroeconomic shocks. The design consists of three elements. First, a time-varying measure of systematic monetary policy based on the historical composition of hawks and doves in the Federal Open Market Committee (FOMC). The historical FOMC composition is measured using newspaper articles, which portray FOMC members as either a hawk that is more concerned about inflation, or as a dove that is more concerned about employment and growth. Second, to obtain exogenous variation in the FOMC composition, I propose an instrumental variables that leverages the FOMC rotation of voting rights. Finally, the identification design combines the measure of systematic monetary policy and the instrument in a state-dependent local projection that can be applied to any macroeconomic shock of interest.

In the empirical application, I study the interactions of discretionary fiscal policy with the systematic monetary policy response. Specifically, I ask how the transmission of government spending shocks is affected by systematic monetary policy. I find that a dovish FOMC supports the expansionary effects of higher spending by delaying interest rate hikes. This leads to a stronger output expansion for a given increase in government spending, i.e., to larger fiscal multipliers. Conversely, output does not expand when the FOMC is hawkish, but inflation expectations are contained. The latter suggests that inflationary pressure is successfully counteracted via a more aggressive systematic monetary policy response. An extensive sensitivity analysis and two case studies further corroborate the plausibility of the results. My estimates may be used to discipline analysis based on structural models of fiscal-monetary interactions (e.g., Bianchi and Ilut, 2017; Leeper, Traum, and Walker, 2017), and may directly inform policymakers.

Chapter 2, which is joint work with Klodiana Istrefi and Matthias Meier, studies how time-variation in systematic monetary policy poses a challenge for conventional approaches that aim to identify exogenous monetary policy shocks. This challenge arises because conventional approaches implicitly assume that systematic monetary policy is constant over time. In contrast, in an environment with time-varying systematic monetary policy, two problems arise. First, the resulting empirical monetary policy shock measures do not isolate exogenous interest rate changes. Instead, the empirical shock measures are contaminated by systematic monetary policy interacted with endogenous macroeconomic variables. This contamination renders the empirical shocks predictable. Second, the contamination further biases impulse response estimates away from the actual response to a truly exogenous monetary policy shock.

There are two empirical contributions in Chapter 2. First, I empirically confirm the theoretical result that fluctuations in systematic monetary policy predict empirical shocks that are estimated as in Romer and Romer (2004). Second, I propose a new monetary policy shock measure that is orthogonal to systematic monetary policy. Based on this new shock, I find that U.S. monetary policy has shorter lags and substantially stronger effects on inflation and output. I obtain similar results for additional macroeconomic outcomes that help to understand the transmission mechanism of monetary policy. The estimates are informative about the effectiveness of monetary policy and our new shocks may be used in future research to construct policy counterfactuals (e.g., Leeper and Zha, 2003; McKay and Wolf, 2023), study the optimality of monetary policy (e.g., Barnichon and Mesters, 2023), estimate structural macroeconomic equations (e.g., Barnichon and Mesters, 2020), or estimate DSGE model (e.g., Christiano, Eichenbaum, and Evans, 2005).

In Chapter 3, I investigate how inflation alters the tax rates of individual taxpayers through tax bracket creep and how this feeds back to the macroeconomy. To isolate bracket creep from other sources of tax rate changes, I propose a non-parametric decomposition of changes in tax rates. Applying the decomposition to German administrative tax records, I find sizeable bracket creep episodes. While the overall importance of bracket creep has decreased over time due to institutional changes, the post-Covid inflation surge led to a resurgence. To better understand how bracket creep feeds back to the macroeconomy, I analytically characterize the labor supply response to bracket creep and study monetary policy transmission in the presence of bracket creep in a New Keynesian model with incomplete markets. The model predicts that a given inflation reduction via monetary policy rate hikes leads to larger output costs when the tax code is indexed to inflation, revealing a potential caveat of such indexation schemes. My results may be informative for the policy debate about the potential implementation of automatic tax indexation schemes.

In summary, the thesis offers a quantitative assessment of monetary and fiscal policy and their interactions. These interactions are quantitatively relevant from a descriptive perspective. For a normative analysis, one may consult structural models to derive prescriptions for the optimal design of economic policies. My results can inform such structural models to obtain reliable policy recommendations that are immune to the Lucas (1976) critique.

Chapter 1

Identification of Systematic Monetary Policy

Joint with Klodiana Istrefi and Matthias Meier.

1.1 Introduction

Monetary policy is not random but a purposeful response to macroeconomic conditions. This response represents systematic monetary policy. Fundamentally, the systematic response reflects the preferences of the policymakers, e.g., concerning price stability and employment, which change over time as the policymakers change. As a consequence, the effects of macroeconomic shocks differ across time, depending on systematic monetary policy. In theory, systematic monetary policy is well-known to be important for the propagation of macroeconomic shocks. However, there is no direct evidence on the causal effects of systematic monetary policy in the U.S.[1]

The main contribution of Chapter 1 is an identification design to estimate the causal effects of the Federal Reserve's systematic monetary policy on the propagation of macroeconomic shocks. We use historical fluctuations in the composition of hawks and doves in the Federal Open Market Committee (FOMC) to measure time variation in systematic monetary

[1] A vast empirical literature estimates the effects of monetary policy shocks (e.g., the pioneering work by Bernanke and Blinder, 1992; Cochrane and Piazzesi, 2002; Romer and Romer, 1989). These shocks are commonly understood as deviations from a policy rule, whereas most policy variation is due to systematic monetary policy, i.e., the rule itself. While evidence on monetary policy shocks may be informative about the effects of systematic monetary policy under certain assumptions (e.g., McKay and Wolf, 2023), we propose to directly estimate the causal effects of systematic monetary policy.

policy. To address the concern that these fluctuations are endogenous to economic and political developments, we propose an instrument that exploits the mechanical rotation of voting rights in the FOMC. To the best of our knowledge, our FOMC rotation instrument is the first instrument for systematic monetary policy.

We then apply the identification design to address a classical question in macroeconomics: How do the effects of fiscal policy depend on the response of monetary policy? This question is deemed crucial in the policy (e.g., Blinder, 2022) and academic debate (e.g., Farhi and Werning, 2016; Woodford, 2011). However, the debate lacks causal evidence. Providing causal evidence is the second contribution of this chapter. We show that the Federal Reserve's systematic monetary policy has a significant effect on the GDP response to fiscal policy. When the FOMC is dovish, it delays tightening in response to an expansionary fiscal spending shock, which supports the expansion of GDP. Conversely, GDP does not expand, rather contracts, under a hawkish FOMC that tightens faster and more aggressively. Fiscal multipliers are between two and three when the FOMC is dovish and below zero when it is hawkish.

We measure time variation in systematic U.S. monetary policy building on the narrative classification of FOMC members by Istrefi (2019) which uses news archives to classify members of the FOMC as hawks and doves, for the period 1960 to 2023. Hawks are more concerned about inflation, while doves are more concerned about supporting employment and growth. Our measure of systematic monetary policy is the aggregate Hawk-Dove balance for each FOMC meeting.[2]

The Hawk-Dove balance is an appealing measure of systematic monetary policy because it parsimoniously summarizes the aggressiveness of the FOMC towards fulfilling one or the other leg of the dual mandate, without having to specify a policy reaction function or the policy tools.

Identifying the causal effects of systematic monetary policy, independent of how it is measured, is challenging because of endogeneity. For example, systematic monetary policy may change in response to unemployment or inflation (Davig and Leeper, 2008). Similarly, the appointment of central bankers can depend on economic and political circumstances, e.g., as documented for the Nixon administration (Abrams, 2006; Abrams and Butkiewicz,

[2] Istrefi (2019) shows that these preferences match with narratives on monetary policy, preferred interest rates, dissents, and forecasts of FOMC members. Bordo and Istrefi (2023) study the origins of these preferences linking them to early-life experiences and education. Instead, we use the Hawk-Dove classification to study the effects of systematic monetary policy on the propagation of macroeconomic shocks. More specifically, we construct an instrument for the Hawk-Dove balance, propose a novel identification design, and apply it to study the effects of government spending shocks.

2012). We discuss this identification challenge through the lens of a New Keynesian model in which the coefficients of the monetary policy rule fluctuate in response to macroeconomic shocks. The model dynamics can be represented as a state-dependent local projection. The OLS estimates of the local projection will fail to identify the causal effects of systematic monetary policy because they are contaminated by unobserved shocks that change the monetary policy rule. Instead, we show that an instrument that captures exogenous variation in systematic monetary policy achieves identification.

We construct an instrument that levers exogenous variation in the Hawk-Dove balance arising from the FOMC rotation of voting rights. The rotation is an annual mechanical scheme that shuffles four out of twelve voting rights among eleven Federal Reserve Bank presidents.[3] We construct an FOMC rotation instrument that is the Hawk-Dove balance among the four FOMC member which the rotation assigns voting rights in a given year. Importantly, the mechanic nature of the rotation renders it orthogonal to economic and political developments. Moreover, the rotation is considered relevant by Fed watchers in the media, the correlation between rotation instrument and overall Hawk-Dove balance is 0.64, and the instrument passes multiple weak instrument tests.

Our identification design combines the measure of systematic monetary policy and the instrument in a state-dependent local projection that can be applied to any macroeconomic shock of interest. Specifically, we regress an outcome of interest on the shock, the shock interacted with the Hawk-Dove balance, the Hawk-Dove balance in levels, and possibly further controls. The instrument vector is given by the vector of regressors when replacing the Hawk-Dove balance with the FOMC rotation instrument. This local projection is in line with the dynamics of a New Keynesian model with time-varying systematic monetary policy. However, different from a New Keynesian model, our design identifies the effects of systematic monetary policy without imposing strong structural assumptions. Instead, we leverage historical variation in the composition of policy preferences among FOMC members. This allows us to study the effects of counterfactual Hawk-Dove balances.[4]

We apply our identification design to study the effects of government spending shocks in the U.S. We focus on the military spending shocks in Ramey (2011) and Ramey and Zubairy (2018) for the period 1960-2014.[5] We find that the real GDP response depends

[3]Relatedly, Ehrmann, Tietz, and Visser (2022) studies how voting rights affect the communication of Federal Reserve Bank presidents and the market reaction to this communication.

[4]This means we can study counterfactual interest rate responses that are associated with historical variation in the Hawk-Dove balance. In contrast, we cannot study counterfactual interest rate responses that did not occur in the data.

[5]In the post-Korean War sample, Ramey (2011) finds that these shocks have weak explanatory power for

significantly on systematic monetary policy. The GDP response to an expansionary shock increases in the share of dovish FOMC members, and decreases in the share of hawks. When the Hawk-Dove balance exceeds the sample average by two doves, quarterly GDP increases by up to 0.7% in response to a military spending shock, which is expected to raise cumulative military spending by 1% of GDP over the next five years. Conversely, quarterly GDP falls by up to 0.3% when the Hawk-Dove balance exceeds the sample average by two hawks.[6] In contrast to the IV estimates, OLS underestimates the dependence of the GDP response on systematic monetary policy at short horizons, but overestimates it at longer horizons.

A common metric to assess the effectiveness of fiscal spending is the spending multiplier, the dollar increase of real GDP per additional dollar of real government spending. We estimate the two- and four-year cumulative spending multipliers and find strong dependence on systematic monetary policy. While multipliers under a hawkish FOMC are typically insignificant with point estimates at or below 0, we find that dovish multipliers are between 2 and 3 and statistically significant. Moreover, the average multipliers are larger and much more precisely estimated when accounting for systematic monetary policy compared to a linear model that omits this state dependency. These results are robust to various modeling choices, as we show in an extensive sensitivity analysis.

We further inspect the mechanism behind the FOMC-dependent effects of spending shocks. We show that nominal interest rates rise under a hawkish FOMC. Under a dovish FOMC, nominal rates initially fall, and rise only with substantial delay. In more detail, when the Hawk-Dove balance exceeds the sample average by two hawks, the federal funds rate (FFR) starts to increase within one year after the shock, and increases by up to 50 basis points at a two-year horizon. Conversely, when the FOMC is dovish, the FFR falls and remains below the pre-shock level for more than two years after the shock, and then sharply rises toward a 50 basis point increase three years after the shock. The different interest rate responses are consistent with the fiscal multiplier estimates across hawkish and dovish FOMCs. Moreover, we find that hawkish policy is more successful in containing inflation (expectations) and that the monetary policy response primarily transmits to real GDP through private consumption.

Finally, we complement our quantitative analysis with narrative evidence from the histor-

contemporaneous government spending. In contrast, we show that the shocks have statistically significant dynamic effects on government spending when accounting for time-varying systematic monetary policy.

[6] For comparison, an increase of the Hawk-Dove balance by two doves or two hawks roughly corresponds to one standard deviation in the change of the Hawk-Dove balance.

ical records of the FOMC meetings. These records reveal that FOMC members and staff frequently discuss changes in (military) government spending, their potential impact on the economy and inflation, and the FOMC's policy response. We provide case studies of two important military spending buildup events in the 1960s, associated with the U.S. Space Program and the Vietnam War. We show that a hawkish FOMC indeed tightens faster after military buildups, whereas a dovish FOMC delays action.

Relation to literature. This chapter contributes to a literature that aims to identify the effects of systematic monetary policy on the propagation of macroeconomic shocks. Closely related are McKay and Wolf (2023) and Barnichon and Mesters (2023) who use multiple monetary policy (news) shocks to estimate the effects of counterfactual monetary policy rules.[7] Under the assumption that systematic monetary policy affects private agents only through changes in the policy instrument, their approach allows identifying the effects of a large set of counterfactual interest rate paths. Instead, our approach leverages historical variation in systematic monetary policy, which avoids potential problems related to the identification and size of monetary policy shocks. The identification of monetary policy shocks is subject to a long-running and ongoing debate (e.g., Bauer and Swanson, 2023b; Ramey, 2016). A key concern is that empirical monetary policy shocks may be contaminated by other business cycle shocks. In fact, one reason for contamination may be time variation in systematic monetary policy.[8] In addition, the effects of monetary policy shocks are typically small, particularly in more recent decades (Ramey, 2016). This may restrict the analysis to more modest policy counterfactuals to avoid extrapolation errors.
A closely related, earlier literature constructs monetary policy counterfactuals via monetary policy shocks (e.g., Bernanke, Gertler, and Watson, 1997; Kilian and Lewis, 2011).[9] Yet, this approach is subject to the Lucas critique (Sargent, 1979). Our identification design is not subject to the Lucas critique because we explicitly model and estimate how the dynamics depend on systematic monetary policy. Another closely related paper is Cloyne, Jordà, and Taylor (2021), which leverages time-invariant cross-country differences in the policy rate response to fiscal shocks to estimate the role of systematic monetary policy

[7] McKay and Wolf (2023) focus on constructing policy counterfactuals, whereas Barnichon and Mesters (2023) uses a similar approach to study optimal policy. Relatedly, Wolf (2023) uses the approach of McKay and Wolf (2023) to provide fiscal policy shock counterfactuals for a strict inflation-targeting central bank.

[8] For example, if the rule changes over time, Romer and Romer (2004) and high-frequency identified shocks may be contaminated, see Chapter 2 of this dissertation.

[9] A further related paper on the intersection of shocks and systematic policy is Arias, Caldara, and Rubio-Ramirez (2019) which identifies monetary policy shocks via sign restrictions on systematic monetary policy.

on the propagation of fiscal consolidation shocks. Whereas Cloyne, Jordà, and Taylor (2021) leverages cross-country differences, we leverage exogenous historical variation in U.S. systematic monetary policy.

An alternative approach to estimate the effects of time-varying systematic monetary policy uses non-linear VAR models (e.g., Primiceri, 2005; Sims and Zha, 2006). A key advantage of our approach is that it requires weaker identifying assumptions and addresses the potential endogeneity of systematic monetary policy. This chapter also relates to a literature studying macroeconomic models with exogenous changes in systematic monetary policy (e.g., Bianchi, 2013; Davig and Leeper, 2007; Leeper, Traum, and Walker, 2017) or endogenous changes (e.g., Barthélemy and Marx, 2017; Davig and Leeper, 2008). Our time series approach requires fewer structural assumptions and provides moments to discipline such models.

Finally, this chapter relates to a large empirical literature that estimates the government spending multiplier. Most empirical estimates find an average fiscal spending multiplier between 0.5 and 1.5 (e.g., Barro and Redlick, 2011; Blanchard and Perotti, 2002; Mountford and Uhlig, 2009; Ramey, 2011). Our findings show that the average fiscal spending multiplier may be downward biased and substantially less precisely estimated when not accounting for time-varying systematic monetary policy. Further closely related are recent papers that study the effects of government spending shocks at the zero lower bound (e.g., Miyamoto, Nguyen, and Sergeyev, 2018; Ramey and Zubairy, 2018). Zero lower bound episodes are endogenous to the business cycle which means the estimates may reflect monetary policy but also the shocks leading to it. Instead, we isolate the causal effects of monetary policy on the propagation of fiscal policy. Another related paper is Nakamura and Steinsson (2014), which estimates relative regional multipliers that difference out the response of monetary policy. This chapter also relates to recent papers that estimate state-dependencies of the multiplier, e.g., depending on the economy being in a recession (Auerbach and Gorodnichenko, 2012; Ghassibe and Zanetti, 2022; Jordà and Taylor, 2016; Ramey and Zubairy, 2018); sign of the shock (Barnichon, Debortoli, and Matthes, 2022; Ben Zeev, Ramey, and Zubairy, 2023); exchange-rate regime, trade openness, and public debt (Ilzetzki, Mendoza, and Végh, 2013); foreign holdings of debt (Broner, Clancy, Erce, and Martin, 2022); and tax progressivity (Ferriere and Navarro, 2024). Compared to this literature, our analysis tackles the endogeneity problem of the state variable. The state we consider captures the monetary policy reaction, and our results highlight the importance of fiscal-monetary interaction for macroeconomic stabilization and the role of who decides monetary policy.

1.2 Identification challenge

In this section, we present a stylized non-linear New Keynesian model in which systematic monetary policy may fluctuate endogenously. We use the model to expound the challenge of empirically identifying the effects of systematic monetary policy on the propagation of macroeconomic shocks.

A New Keynesian model. The model is a textbook New Keynesian model (e.g., Gali, 2015) except for a monetary policy rule with time-varying coefficients. Households choose consumption, labor and bond holdings to maximize $E_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - N_t^{1+\varphi} \right)$ subject to budget constraints. Intermediate good firms produce variety goods using $Y_{it} = x_t^a N_{it}$ where x_t^a is exogenous productivity. The price of a variety good can be reset with a constant probability $1-\theta$. Final good firms produce the final good $Y_t = \left(\int_0^1 Y_{it}^{(\epsilon-1)/\epsilon} di \right)^{\epsilon/(\epsilon-1)}$. A fiscal policy authority finances government spending $G_t = \gamma Y x_t^s$ with lump-sum taxes where $\gamma \in [0,1)$, Y is steady-state output, and x_t^s denotes exogenous changes in fiscal spending. Goods market clearing requires $Y_t = C_t + G_t$. The exogenous variables follow stable AR(1) processes $\log x_t^k = \rho_k \log x_{t-1}^k + \varepsilon_t^k$ with $\varepsilon_t^k \sim (0, \sigma_k^2)$ for $k = a, s$ respectively. A monetary policy rule closes the model. Letting lowercase letters denote (log) deviations from the steady state, the monetary authority sets nominal interest rates i_t according to

$$i_t = \widetilde{\phi}_t \pi_t, \tag{1.1}$$

where $\widetilde{\phi}_t \in (1, \infty)$ is *systematic monetary policy* which fluctuates according to a stable AR(1)

$$\phi_t = \rho_\phi \phi_{t-1} + \zeta^s \varepsilon_t^s + \zeta^a \varepsilon_t^a + \eta_t, \tag{1.2}$$

where $\widetilde{\phi}_t = \phi + \phi_t$ and ϕ denotes the unconditional mean of $\widetilde{\phi}_t$. Importantly, we allow systematic monetary policy to be endogenous, as ϕ_t may respond to macroeconomic shocks $(\varepsilon_t^s, \varepsilon_t^a)$.[10] Such endogeneity creates an empirical identification challenge as we discuss

[10] For DSGE models with exogenous changes in the Taylor rule coefficients see Davig and Leeper (2007) and Bianchi (2013), for endogenous changes see Davig and Leeper (2008) and Barthélemy and Marx (2017).

toward the end of this section. In addition, we allow for exogenous changes in systematic monetary policy, captured by the exogenous policy shifter η_t. We assume that ε_t^s, ε_t^a, and η_t are mutually independent and identically distributed over time. Accounting for the effects of systematic monetary policy ϕ_t, the approximate equilibrium dynamics of GDP are given by

$$y_t = a + b_s x_t^s + b_a x_t^a + c_s x_t^s \phi_t + c_a x_t^a \phi_t + d\phi_t, \tag{1.3}$$

where a, b_s, b_a, c_s, c_a, d are coefficients that depend on the deep structural parameters of the model. Appendix 1.A.1 provides details on the derivation.

Identification challenge. We next discuss the challenge of identifying the effects of systematic monetary policy from a regression when y_t is generated by (1.3). Without loss of generality, we focus our discussion on the fiscal spending shock. Consider an econometrician who observes $\{y_t, \varepsilon_t^s, \phi_t\}$, and estimates the state-dependent local projection

$$y_{t+h} = \alpha^h + \beta^h \varepsilon_t^s + \gamma^h \varepsilon_t^s \phi_t + \delta^h \phi_t + v_{t+h}^h, \tag{1.4}$$

for $h = 0, \dots, H$ forecast horizons. For $h = 0$, the residual v_{t+h}^h contains lagged spending shocks, contemporaneous and lagged technology shocks, and the interaction of these shocks with ϕ_t. For $h > 0$, the residual further contains shocks (ε_t^s, ε_t^a) and policy shifter (η_t) occuring between t and $t+h$. The estimands in (1.4) are

$$\beta^h = b_s(\rho_s)^h \ , \qquad \gamma^h = c_s(\rho_s \rho_\phi)^h \ , \qquad \delta^h = d(\rho_\phi)^h \ . \tag{1.5}$$

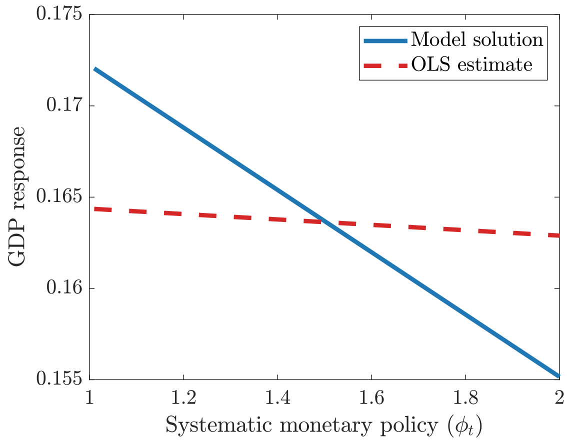
Both β^h, the average effect of the spending shock, and γ^h, the differential effect associated with ϕ_t, diminish in the forecast horizon h.

We next ask whether the OLS estimates of $(\beta^h, \gamma^h, \delta^h)$ are consistent, i.e., whether they asymptotically recover the estimands in (1.5).[11] Consistency holds under the strong exogeneity assumption $\zeta^s = \zeta^a = 0$, that is if ϕ_t is independent of the macroeconomic shocks. In contrast, if ϕ_t correlates with at least one of the shocks, the OLS estimates do *not* consistently estimate $(\beta^h, \gamma^h, \delta^h)$.[12] If, for example, ϕ_t responds to a spending shock, the

[11] We explicitly include δ^h in the vector of coefficients because including the (endogenous) control variable ϕ_t in the regression is important for identification, as ϕ_t is correlated with ε_t^s and $\varepsilon_t^s \phi_t$ in general.

[12] If the econometrician observes and includes *all* shocks and corresponding interaction terms in the regression according to equation (1.3), then the OLS estimates will be consistent without the exogeneity assumption. In practice, this is infeasible as many shocks are (partially) unobserved.

Figure 1.1: GDP response and systematic monetary policy



Notes: The solid line shows the model solution for the GDP response to a spending shock as a function of systematic monetary policy (ϕ_t), i.e., $b_s + c_s\phi_t$, with b_s and c_s given by (1.6) and the parametrization: $\beta = 0.99$, $\theta = 0.75$, $\epsilon = 9$, $\varphi = 2$, $\gamma = 0.2$, $\bar{\phi} = 1.5$, $\zeta^s = 1$, $\zeta^a = 0.25$, $\sigma_s = \sigma_a = 1$. The dashed line shows the OLS estimate $\hat{\beta}^0 + \hat{\gamma}^0\phi_t$ based on a regression of (1.3) when the terms in u_t are unobserved. The estimands are $\beta^0 = b_s = 0.164$ and $\gamma^0 = c_s = -0.017$, and the large-sample OLS estimates are $\hat{\beta}^0 = 0.164$ and $\hat{\gamma}^0 = -0.002$.

OLS estimator will be contaminated by the response of GDP to the spending shock.
Now suppose the econometrician observes an instrument ϕ_t^{IV} that is correlated with ϕ_t (relevance), but uncorrelated with all past, present, and future macroeconomic shocks ε_t^s and ε_t^a and that is uncorrelated with all past and future policy shifters η_t (exogeneity). Consider the IV estimates of $(\beta^h, \gamma^h, \delta^h)$ when using $(\varepsilon_t^s, \varepsilon_t^s\phi_t^{IV}, \phi_t^{IV})$ as instrument vector for the regressors $(\varepsilon_t^s, \varepsilon_t^s\phi_t, \phi_t)$. The IV estimator consistently estimates $(\beta^h, \gamma^h, \delta^h)$, even when ϕ_t fluctuates endogenously in response to macroeconomic shocks ($\zeta^a, \zeta^s \neq 0$). For further details, see Appendix 1.A.2. This result guides the remainder of this chapter in which we propose an instrument for systematic monetary policy and use it to estimate the causal effects of systematic monetary policy.

Illustration. To illustrate the effects of systematic monetary policy and the identification challenge, we focus on a special case of our economy in which $\rho_s = \rho_a = \rho_\phi = 0$. To understand how ϕ_t affects the GDP response to the fiscal spending shock ε_t^s, we need to know

$$b_s = \gamma\,(1 + \lambda\phi)\,\omega^{-1}\ , \qquad c_s = -\gamma(1-\gamma)\lambda\varphi\omega^{-2}\ , \tag{1.6}$$

where $\omega = 1 + \lambda\,(\varphi(1-\gamma) + 1)\,\phi$, $\lambda = (1-\theta)(1-\beta\theta)/\theta$. Since $b_s > 0$ and $c_s < 0$ (under standard parameter restrictions), the GDP response falls in the strength of the monetary policy reaction to inflation. This is the monetary offset (e.g., Christiano, Eichenbaum, and Rebelo, 2011; Woodford, 2011).

The solid line in Figure 1.1 illustrates the monetary offset. The dashed line illustrates the OLS bias in the estimated GDP response to the spending shock. In our example, the OLS estimate strongly understates the role of systematic monetary policy.

1.3 Identification design

In this section, we propose an identification design to study how systematic monetary policy in the U.S. shapes the propagation of macroeconomic shocks. Our identification design relies on three crucial elements: (i) a measure of systematic monetary policy, (ii) an instrument for systematic monetary policy, and (iii) a state-dependent local projection regression that combines (i) and (ii) to tackle the identification challenge discussed in the preceding section.

1.3.1 Hawk-Dove balance in the FOMC

In the following, we build on the classification of Federal Open Market Committee (FOMC) members into hawks and doves by Istrefi (2019) and argue that the Hawk-Dove balance captures well variation in systematic monetary policy over time.

The FOMC. The FOMC is the committee of the Federal Reserve that sets U.S. monetary policy. The FOMC consists of 12 members: the seven members of the Board of Governors of the Federal Reserve System, including the Federal Reserve Chair, the president of the Federal Reserve Bank (FRB) of New York, and four of the remaining 11 FRB presidents, who serve one-year terms on a rotating basis.[13]

Individual policy preferences. To measure the policy preferences of FOMC members we use the Istrefi (2019) classification of FOMC members as hawks and doves, for the period 1960-2023.[14] Underlying this classification are more than 20,000 real-time media

[13] While non-voting FRB presidents attend the FOMC meetings and participate in the discussions, we focus on the voting FOMC, the decision-making body, in line with the literature that studies central bank decision making by committees (e.g., Belden, 1989; Blinder, 2007; Bordo and Istrefi, 2023; Riboni and Ruge-Murcia, 2023; Riboni and Ruge-Murcia, 2010).

[14] The data in Istrefi (2019) covers 1960 through 2014. The data is currently extended up to the first meeting of 2023. Thus, our sample covers all 634 (scheduled) FOMC meetings between 1960 and 2023.

articles from over 30 newspapers and business reports of Fed watchers (available in news archives like ProQuest Historical Newspapers and Factiva) mentioning individual FOMC members. Istrefi (2019) uses these articles to categorize individual FOMC members as hawks or doves for each FOMC meeting based on the news information available up until the meeting. So, the Hawk-Dove classification is a panel that tracks FOMC members over time, at FOMC meeting frequency. Hawks are perceived to be more concerned with inflation, while doves are more concerned with employment and growth.[15] Through the lens of our model in Section 1.2, we can think about hawks as preferring a larger inflation coefficient ϕ_t than doves. However, the Hawk-Dove classification we use is not tied to assuming a specific policy rule.

Overall, 129 of the 147 FOMC members between 1960 and 2023 are classified as hawk or dove. The news coverage for the remaining 18 members does not allow classification (as hawk or dove) for any meeting, as some served in the early 1960s with sparse media coverage and others are very recent appointments in the FOMC. The majority (95) of the classified FOMC members are consistently hawks or doves over time while the rest switches camps at least once. Swings are equally split in either direction and quite uniformly distributed over time. On average, the 34 swinging FOMC members switch camps at only 1.8% of the member-meeting pairs.

While true policy preferences are unobserved, Istrefi (2019) shows that perceived preferences match well with policy tendencies that are unknown in real-time to the public, as expressed by preferred interest rates, with forecasting patterns of individual FOMC members, and with dissents. In addition, Bordo and Istrefi (2023) show that the FOMC members' educational background, e.g., whether they graduated from a university related to the Chicago school of economics, and early life experience, i.e., whether they grew up during the Great Depression, predicts the Hawk-Dove classification. The long lasting effect of the early life experience in the formation of policy preferences is consistent with the very few swings in our sample.

Aggregate Hawk-Dove balance. To measure variation in systematic monetary policy over time, we aggregate the cross-section of individual FOMC member preferences into an aggregate Hawk-Dove balance for each meeting (cf. Istrefi, 2019). We do so because

[15] A typical example of a newspaper quote used to categorize a hawk reads: *Volcker leans toward tight-money policies and high interest rates to retard inflation*, New York Times, 2 May 1975. For a dove: *The weakness of Treasury prices and higher yields was seen reflecting the view that Bernanke will be 'pro-growth' and perhaps less hawkish on inflation, said John Roberts, managing director at Barclays Capital in New York*, Dow Jones Capital Markets Report, 24 October 2005.

the nature of monetary policy-making by committee involves the aggregation of diverse individual policy preferences in a collective decision.[16]

We adopt a symmetric numerical scale for the qualitative Hawk-Dove classification in order to aggregate the preferences. We define $Hawk_{i\tau}$ as the policy preference of FOMC member i at FOMC meeting τ:

$$Hawk_{i\tau} = \begin{cases} +1 & \textit{Consistent hawk} \\ +\frac{1}{2} & \textit{Swinging hawk} \\ 0 & \textit{Preference unknown} \\ -\frac{1}{2} & \textit{Swinging dove} \\ -1 & \textit{Consistent dove} \end{cases} \tag{1.7}$$

A consistent hawk is an FOMC member that has not been categorized as a dove in the past. In contrast, a swinging hawk has been a dove at some point in the past. The definition of a consistent dove and a swinging dove is analogous. We assign a lower weight to swingers as they are often perceived as 'middle-of-the-roaders' with more moderate leanings to the hawkish or dovish side (Istrefi, 2019).[17] Finally, we assign $Hawk_{i\tau} = 0$ when the policy preference of the FOMC member is (yet) unknown.

We next aggregate the individual policy preferences in (1.7). We compute the aggregate Hawk-Dove balance by

$$Hawk_{\tau} = \frac{1}{|\mathcal{M}_{\tau}|} \sum_{i \in \mathcal{M}_{\tau}} Hawk_{i\tau} \tag{1.8}$$

where $\mathcal{M}_{\tau}$ denotes the set of FOMC members at meeting τ. A full FOMC consists of $|\mathcal{M}_{\tau}| = 12$ members but $|\mathcal{M}_{\tau}|$ is occasionally below 12 because of absent members or vacant positions.[18] The Hawk-Dove balance in (1.8) is the arithmetic average across individual preferences. This is our baseline aggregation of the Hawk-Dove balance in the FOMC and

[16] Relatedly, Blinder (1999) writes: *While serving on the FOMC, I was vividly reminded of a few things all of us probably know about committees: that they laboriously aggregate individual preferences; that they need to be led; that they tend to adopt compromise positions on difficult questions; and–perhaps because of all of the above–that they tend to be inertial.*

[17] Our empirical findings are robust to not distinguishing between consistent and swinging preferences, see Section 1.4.5.

[18] When a substitute temporarily replaces an absent FOMC member, we assume the substitute acts in the interest of the original FOMC member and assign the same policy preference, see Appendix 1.B for details. This assumption affects less than one percent of all observations and is not important for our results.

conforms well with the consensual mode in which the FOMC typically operates.[19][20] In Section 1.4.5, we show that our empirical findings are robust to alternatively using the median of preferences or putting a higher weight on the Fed Chair's preference. Finally, we aggregate $Hawk_\tau$ from meeting frequency to quarterly frequency. We compute the Hawk-Dove balance $Hawk_t$ for quarter t as the average balance in the first month of the quarter. If the first month is without a meeting, we use the first preceding month with a meeting.

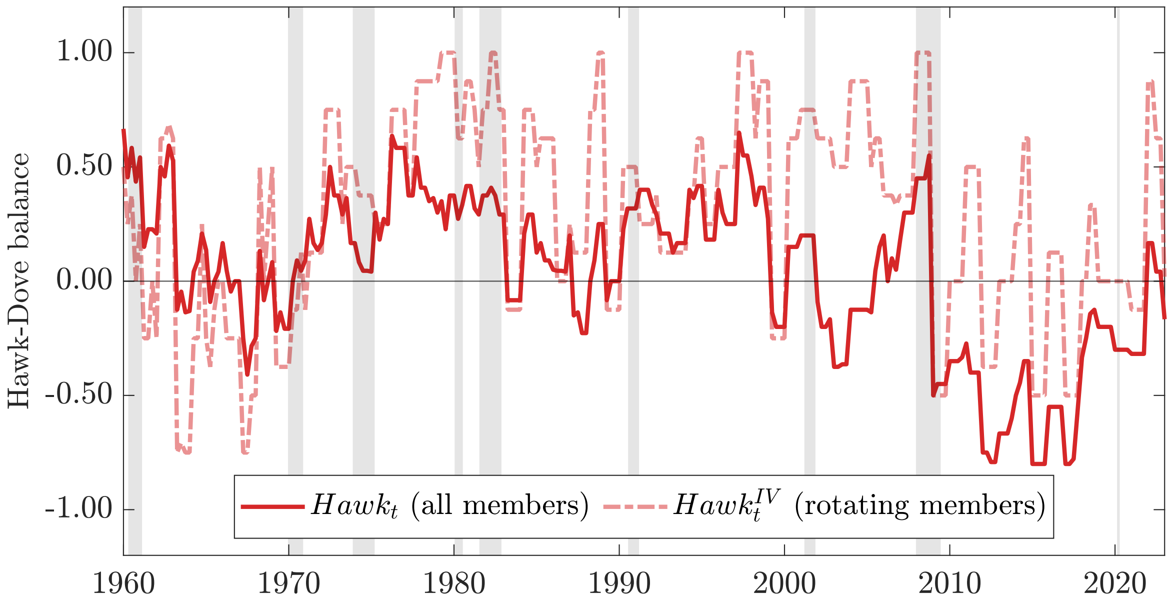
We present the evolution of the Hawk-Dove balance from 1960 to 2023 as the solid line in Figure 1.2. There is considerable variation in this balance, featuring both hawkish and dovish majorities. The variation reflects the turnover of rotating FOMC members, the turnover of non-rotating FOMC members, and changes in policy preferences of incumbent FOMC members. We discuss the importance of these components for $Hawk_t$ fluctuations in Subsection 1.3.2.

Systematic monetary policy. The aggregate Hawk-Dove balance $Hawk_t$ represents our measure of systematic U.S. monetary policy. It accounts for the diversity of views within the FOMC on how policy should be adjusted to promote both, price stability and maximum employment. This diversity is usually expressed in FOMC meetings through different forecasts of individual members, through dissents, and in public through speeches. While the Fed's response to macroeconomic shocks is sophisticated and depends on various economic factors, we argue that our Hawk-Dove balance matches well with narratives of monetary policy in the U.S. (Istrefi, 2019). For example, the dovish leaning of $Hawk_t$ in the mid-1960s coincides with a period of delays and hesitation from the FOMC to take anti-inflationary action (Meltzer, 2005). The hawkish majorities in the 1970s might be surprising given the high inflation rates in this period. Yet it is consistent with monetary policy being misguided by an underestimated natural rate of unemployment (DeLong, 1997; Romer and Romer, 2002) and persistence of inflation (Primiceri, 2006). In particular, Orphanides (2004) argues that for the periods before and after Paul Volcker's appointment

[19] Riboni and Ruge-Murcia (2010) argue that a consensus model fits actual policy decisions of the Federal Reserve. In addition, Riboni and Ruge-Murcia (2023) provide evidence suggesting that policy proposals of the Fed Chair are the result of a compromise, reflecting a balance of power within the FOMC.

[20] Cieslak, Hansen, McMahon, and Xiao (2023) construct a Hawk-Dove score based on the language in FOMC meeting transcripts. In contrast to our measure which captures FOMC members preferences about monetary policy, their measure captures (a hawkish or dovish) sentiment on current direction of policy changes. Furthermore, Ferguson, Kornejew, Schmelzing, and Schularick (2023) classify central bank governors in 80 countries as hawks and doves, with respect to financial sector support, for the periods preceding banking crises.

Figure 1.2: Hawk-Dove balance in the FOMC



Notes: The solid red line shows the quarterly time series of the aggregate Hawk-Dove balance of the FOMC ($Hawk_t$) from 1960 until 2023. The dashed red line shows the aggregate Hawk-Dove balance of the subgroup of rotating FRB presidents with voting right in period t, the FOMC rotation instrument ($Hawk_t^{IV}$). Grey bars indicate NBER dated recessions.

in 1979, policy was broadly similar and consistent with a strong reaction to Greenbook inflation forecasts.[21] During the 1980s, the perception of a less hawkish FOMC reflects nominations of dovish Board members by President Reagan. In addition, it is consistent with the imperfect credibility of hawkish policy during the Volcker disinflation, as observed in persistently elevated long-term interest rates (indicative of inflation expectations) in this period (Goodfriend and King, 2005). Overall, this suggests that the Hawk-Dove balance captures important aspects of the Fed's systematic policy-making.

Our approach of measuring systematic policy via $Hawk_t$ has several advantages to alternative approaches such as calibrating or estimating policy rules (e.g., Bauer, Pflueger, and Sunderam, 2022; Clarida, Gali, and Gertler, 2000). Importantly, we do not have to specify a particular reaction function, nor do we need to restrict the analysis to specific policy instruments or communication strategies.[22] We further avoid the well-known identification

[21] Moreover, Orphanides (2003) shows that a dovish Taylor rule with a sufficiently large weight on the output gap would have resulted in substantially higher inflation.

[22] For a summary of alternative policy rules that the FOMC consults, see here: https://www.federalreserve.gov/monetarypolicy/policy-rules-and-how-policymakers-use-them.htm. Policy instruments have been changing over our sample, from targeting monetary aggregates to targeting

issues that plague the estimation of monetary policy rules (Carvalho, Nechio, and Tristão, 2021; Cochrane, 2011). Independently of the policy tool or policy rule, our measure reflects the aggressiveness of the FOMC towards fulfilling one or the other leg of the dual mandate. In addition, the Hawk-Dove balance reflects public beliefs, in real-time, about monetary policymakers. In contrast, ex-post estimates of systematic monetary policy may inadvertently use ex-post information not available at the time of the policy decision, potentially giving rise to misleading conclusions (Orphanides, 2003).

Comparability over time. A potential concern with the classification of FOMC members into hawks and doves is that the meaning of being a hawk or dove might have changed over time. We argue this is likely no major concern. First, Istrefi (2019) has classified each member as a hawk or dove based on a common and time-invariant definition, that is the policy leaning with regard to the dual mandate of the Fed: maximum employment and stable prices. Second, given that preferences tend to be stable, we would expect many swings after large changes in the meaning of hawks and doves. However, swings in measured preferences are rare suggesting that the meaning of being a hawk or dove is relatively stable over time. Third, the fact that we observe large and persistent fluctuations in $Hawk_t$ is incompatible with the Hawk-Dove classification being a relative ranking, according to which hawks are those FOMC members which are more hawkish than the contemporaneous average policy preference among FOMC members, and analogously for doves. Finally, in a robustness exercise in Section 1.4, we show that our results are robust to using an alternative Hawk-Dove balance which accounts for potential trends in the meaning of hawks and doves.

Relation to monetary policy shocks. Empirically estimated monetary policy shocks are often considered to reflect changes in central bank preferences (Christiano, Eichenbaum, and Evans, 1999; Ramey, 2016). Hence, they may be related to the Hawk-Dove balance, our measure of systematic monetary policy. In Chapter 2 of this dissertation, we characterize this relationship. Because conventional identification strategies assumes a time-invariant policy rule, the empirical monetary policy shocks may indeed capture time variation in systematic monetary policy. However, the relationship between empirical monetary policy shocks and systematic monetary policy is non-linear and also depends on the state of the economy (e.g., the inflation rate). Instead, our Hawk-Dove balance provides a cleaner

the Fed Funds rate, conducting balance sheets policy, and through forward guidance communication.

measure of systematic monetary policy.

1.3.2 FOMC Rotation Instrument

We next propose and discuss a novel FOMC rotation instrument that allows us to identify the effects of systematic monetary policy, even if monetary policy is endogenous to the state of the economy (cf. Section 1.2).

Potential endogeneity. Systematic monetary policy may change depending on the state of the economy. For example, the Federal Reserve may become more dovish in response to high unemployment, or more hawkish in response to high inflation (cf. Davig and Leeper, 2008). Empirically, Chang, Maih, and Tan (2021) find that the parameters of the monetary policy rule respond to macroeconomic shocks. Changes in systematic monetary policy may also be driven by political pressure. For example, Abrams (2006) and Abrams and Butkiewicz (2012) document the influence of the Nixon administration on the FOMC.[23] Political pressure may lead to endogenous fluctuations in the Hawk-Dove balance through swings of incumbent FOMC members and through new appointments. In this context, note that members of the Board of Governors and the Fed Chair require a nomination from the U.S. President for their first and any subsequent term.

FOMC rotation instrument. To address the endogeneity of the Hawk-Dove balance we propose an instrument which leverages exogenous variation in $Hawk_t$ that arises from the annual FOMC rotation. Each year, four FOMC memberships rotate among eleven FRB presidents following a mechanical scheme that has been in place since the early 1940s. According to the scheme, some FRB presidents become FOMC members every second year (Cleveland and Chicago) and others every third year (Philadelphia, Richmond, Boston, Dallas, Atlanta, St. Louis, Minneapolis, San Francisco and Kansas City). As the rotation of voting rights is independent of the state of the economy, it induces exogenous variation in $Hawk_t$. To leverage the variation from the FOMC rotation we propose a novel instrument, which we refer to as FOMC rotation instrument. Formally, the instrument is

[23] More recently, Bianchi, Gómez-Cram, Kind, and Kung (2023) and Camous and Matveev (2021) document that President Trump exerted pressure on the Fed and Drechsel (2024) identifies the effects of political pressure on the Fed.

given by

$$Hawk_{\tau}^{IV} = \frac{1}{|\mathcal{R}_{\tau}|} \sum_{i \in \mathcal{R}_{\tau}} Hawk_{i\tau}, \tag{1.9}$$

where $\mathcal{R}_{\tau}$ denotes the set of rotating FOMC members at FOMC meeting τ. A full set of rotating members consists of $|\mathcal{R}_{\tau}| = 4$ members.[24] We aggregate the FOMC rotation instrument to quarterly frequency analogously to the Hawk-Dove balance.

In Figure 1.2, the dashed line presents the FOMC rotation instrument over time. On average, the rotating presidents are more hawkish than the overall FOMC Hawk-Dove balance, reflecting the fact that FRB presidents tend to be more hawkish than governors (Bordo and Istrefi, 2023; Chappell, McGregor, and Vermilyea, 2005; Istrefi, 2019). Both series display sizable variation over time, but fluctuations in the instrument $Hawk_t^{IV}$ are more short-lived, with a year-over-year autocorrelation of 0.19 compared to 0.66 for $Hawk_t$, see Table 1.1.

Table 1.1: Summary statistics

	Mean	Median	SD	Autocorr	Corr	Min	Max	T
$Hawk_t$	0.04	0.09	0.35	0.66	-	-0.80	0.67	253
$Hawk_t^{IV}$	0.28	0.33	0.45	0.19	0.64	-0.75	1.00	253

Notes: This table shows summary statistics for the quarterly time series from 1960 until 2023. $Hawk_t$ is the average Hawk-Dove balance of the FOMC. $Hawk_t^{IV}$ is the FOMC rotation instrument. "Autocorr" refers to the year-over-year autocorrelation. "Corr" refers to the correlation with $Hawk_t$.

Relevance of instrument. Our instrument $Hawk_t^{IV}$ aggregates the policy preferences of one-third of the FOMC members, capturing a significant part of the variation in the overall Hawk-Dove balance $Hawk_t$. In fact, the correlation between $Hawk_t$ and $Hawk_t^{IV}$ is 0.64. We further study the explanatory power of the FOMC rotation instrument via a stylized first-stage regression by projecting $Hawk_t$ on $Hawk_t^{IV}$ and a constant. Applying the weak instrument test from Montiel Olea and Pflueger (2013) yields an effective F-statistic of 46.13 which is above 37.42, the critical value for rejecting a relative weak

[24] In our sample, $|\mathcal{R}_{\tau}| = 4$ for 625 out of 634 FOMC meetings and $|\mathcal{R}_{\tau}| = 3$ for the remaining nine meetings because of an absent member.

instrument bias exceeding 5%.[25] This suggests that the instrument satisfies the relevance condition. A more thorough assessment of instrument strength for our main results is delegated to Section 1.4.4.

We further provide a decomposition of $Hawk_t$ into intensive margin changes of incumbent FOMC members' policy preferences and extensive margin changes in the composition of the FOMC due to entry and exit, see Appendix 1.C for details. We find that extensive margin changes in the FOMC composition due to the rotation account for 53% of the variance in yearly changes of $Hawk_t$. The turnover of non-rotating FOMC members accounts for almost another quarter of the variance, and the remainder is due to preference changes of incumbent FOMC members and various covariance terms. Both the first-stage regression and the variance decomposition strongly suggest that our instrument is relevant for $Hawk_t$.

Finally, the rotation is considered important by Fed watchers in the media. Each year before the rotation, they discuss its implications for monetary policy. A typical media discussion, here an article in The New York Times from January 1, 2011, reads as follows:

> *As the Federal Reserve debates whether to scale back, continue or expand its $600 billion effort to nurse the economic recovery, four men will have a newly prominent role in influencing the central bank's path. The four men are presidents of regional Fed banks, and under an arcane system that dates to the Depression, they will become voting members in 2011 on the Federal Open Market Committee, [...] the change in voting composition is likely to give the committee a somewhat more hawkish cast. This could amplify anxieties about unforeseen effects of Bernanke's policies [...]. Two of the four new voters are viewed as hawkish on inflation, meaning that they tend to be more worried about unleashing future inflation than they are about reducing unemployment in the short run.*

Exogeneity of instrument. We next argue that variation in $Hawk_t^{IV}$ is quasi-exogenous. First, the rotation scheme is mechanical and time-invariant and therefore unrelated to the state of the economy. Second, new appointments of FRB presidents are relatively infrequent and unlikely to be influenced by the federal government. FRB presidents are appointed by the Board of Directors of the respective Federal Reserve district. The direc-

[25] We also reject the null of the weak instrument bias exceeding 5% when adding four lags of $Hawk_t$ and $Hawk_t^{IV}$ to control for serial correlation in both variables. In either case, we use Newey-West standard errors with automatic bandwidth selection.

tors are to represent the financial institutions and the broader public in the district.[26] In contrast, members of the Board of Governors (including the Fed Chair) are nominated by the U.S. president and confirmed by the Senate. Furthermore, the average tenure of an FRB president is eleven years but only seven years for a governor in our sample. Relatedly, Bordo and Istrefi (2023) show that different from governors, there is no correlation between the preferences of the FRB presidents and the U.S. president's party at the time of their appointment. In addition, some regional FRBs have persistent leanings toward either the dovish or the hawkish camp. For example, the Cleveland FRB president is typically a hawk whereas the president of the San Francisco FRB is typically a dove.

Third, of potential concern are swings of FRB presidents between being a hawk or dove. If swings are driven by macroeconomic shocks this will introduce endogeneity in the FOMC rotation instrument. Yet, we argue that swings are a negligible threat to the exogeneity of our instrument. For rotating FOMC members, swings occur only in 1.3% of member-meetings pairs.[27] In addition, we find that swings account for a negligible fraction of the variance of the rotation instrument. In particular, we decompose $Hawk_t^{IV}$ into intensive margin changes of preferences (swings) and extensive margin changes of the composition of rotating FOMC members due to either the rotation or appointments, see Appendix 1.C for details. The rotation accounts for 93% of the variance in yearly changes of $Hawk_t^{IV}$, appointments for 7% and swings for less than 1%. In addition, among the few swings that did happen, some do not appear linked to the state of the economy.[28] To address residual concerns about swings, our sensitivity analysis considers an alternative Hawk-Dove balance which mutes the effects of swings. Our results are robust to these alternatives, see Section 1.4.5.

Fourth, $Hawk_t^{IV}$ displays relatively short-lived time series fluctuations that are unlikely to be correlated with slow-moving macroeconomic trends, such as increasing market power, female labor force participation, and various technological innovations. Similarly, $Hawk_t^{IV}$ is uncorrelated with business cycle fluctuations. For example, the correlation between $Hawk_t^{IV}$ and yearly real GDP growth is -0.02 and statistically insignificant. In contrast,

[26] Formally, the Board of Governors approves the appointments of FRB presidents. In the words of former Governor Kevin Warsh *it would be reasonably unprecedented in modern times, for the Reserve Bank's preferred choice not to ultimately be accepted by the Board of Governors* (Bordo, 2016).

[27] Specifically, in 2533 member-meeting observation, we observe only 34 swings.

[28] Bordo and Istrefi (2023) discuss three major swing waves in the FOMC during 1960-2014. The first wave is a hawkish wave influenced by inflation dynamics in the late 1960s to early 1970s. The second wave is a hawkish swing in the early 1990s, related to the discussion on inflation targeting inspired by the announcements of the Reserve Bank of New Zealand and Bank of Canada. Finally, the third swing wave is a dovish one in the late 1990s, following a new understanding of the economy.

the correlation between $Hawk_t$ and GDP growth is 0.15 and significant at the 5% level. Overall, the above arguments support the validity of our FOMC rotation instrument for identifying the causal effects of systematic monetary policy. To the best of our knowledge, we are the first that propose an instrument for systematic monetary policy. We believe this is a substantial contribution to the literature which opens up myriad research questions.

A validation exercise for $Hawk_t$ and $Hawk_t^{IV}$. Given our definition of hawkish policy makers and conventional wisdom about hawkish monetary policy, we should expect a hawkish FOMC to respond more aggressively to inflation. As validation exercise, we empirically test this correlation via a dynamic Taylor rule regression. We use $Hawk_t^{IV}$ as instrument in a local projection of the federal funds rate on the Greenbook inflation forecast interacted with $Hawk_t$. We find that a hawkish FOMC indeed raises the federal funds rate significantly more aggressively in the presence of higher inflation forecasts.[29] For more details on the exercise, the results, and a weak instrument test, see Appendix 1.D. Overall, this exercise suggests that $Hawk_t$ and $Hawk_t^{IV}$ capture important variation in systematic monetary policy.

1.3.3 Local projection framework

Finally, we propose to combine $Hawk_t$ and $Hawk_t^{IV}$ in a state-dependent local projection framework that permits causal identification of how systematic monetary policy shapes the propagation of various macroeconomic shocks. The setup of the local projection is consistent with the New Keynesian model discussed in Section 1.2.

We regress an outcome variable of interest, x_{t+h}, on a macroeconomic shock of interest, ε_t^s, the interaction of the shock with the Hawk-Dove balance $Hawk_t$, as well as $Hawk_t$ in levels and a vector of additional control variables Z_{t-1}. Formally,

$$x_{t+h} = \alpha^h + \beta^h \varepsilon_t^s + \gamma^h \varepsilon_t^s (Hawk_t - \overline{Hawk}) + \delta^h (Hawk_t - \overline{Hawk}) + \zeta^h Z_{t-1} + v_{t+h}^h, \quad (1.10)$$

for $h = 0, \ldots, H$ forecast horizons. $\overline{Hawk}$ denotes the arithmetic sample mean of $Hawk_t$. To address the potential endogeneity of $Hawk_t$, we use the instrument vector

$$q_t = \left[\ 1,\ \varepsilon_t^s,\ \varepsilon_t^s \left(Hawk_t^{IV} - \overline{Hawk}^{IV}\right),\ \left(Hawk_t^{IV} - \overline{Hawk}^{IV}\right),\ Z_{t-1}\ \right] \quad (1.11)$$

[29] This is in line with the findings in Bordo and Istrefi (2023) who provide OLS estimates of a Taylor rule regression augmented by the Hawk-Dove balance.

for the regressors in (1.10). The two key coefficients in (1.10) are β^h and γ^h, which capture the average response, when the Hawk-Dove balance equals its sample average, and the differential response, when the FOMC is more or less hawkish than the sample average.[30]

Based on Section 1.2, the IV estimator is consistent if the instrument $Hawk_t^{IV}$ is orthogonal to all macroeconomic shocks (both observed shocks ε_t^s and other unobserved shocks) at all lags and leads. In the next section, we discuss whether the identifying assumptions are satisfied in the context of a government spending shock.

In general, this framework can be used to study the propagation of any shock through systematic U.S. monetary policy. Our framework permits revisiting a range of important empirical questions, such as the role of systematic monetary policy for the effects of oil-related shocks (e.g., Bernanke, Gertler, and Watson, 1997; Kilian and Lewis, 2011), technology shocks (e.g., Galí, López-Salido, and Vallés, 2003), news shocks (e.g., Barsky and Sims, 2011), fiscal spending shocks (e.g., Ramey and Zubairy, 2018), and tax shocks (e.g., Romer and Romer, 2010). Moreover, our framework allows the estimation of a new set of moments that can be used to discipline structural models with time variation in systematic monetary policy, such as regime-switching models (e.g., Bianchi, 2013; Bianchi and Ilut, 2017; Davig and Leeper, 2007).

1.4 Government spending and monetary policy

In this section, we use our identification design to estimate how the effects of U.S. government spending shocks depend on systematic monetary policy. We find that a hawkish FOMC significantly dampens the expansionary effects of increased government spending on GDP, while a dovish FOMC supports it. Relatedly, we find sizable differences in the fiscal multiplier depending on the hawkishness or dovishness of the FOMC. We further provide evidence on the strength of our instrument, and perform an extensive sensitivity analysis.

[30] Formally, we define state-dependent impulse responses as

$$\mathbb{E}\left[x_{t+h}|\varepsilon_t^s=\varepsilon, Hawk_t=Hawk\right]-\mathbb{E}\left[x_{t+h}|\varepsilon_t^s=0, Hawk_t=Hawk\right]=\left[\beta^h+\gamma^h\left(Hawk-\overline{Hawk}\right)\right]\varepsilon,$$

where both expectations additionally condition on the control vector Z_{t-1}.

1.4.1 Data and identifying assumptions

We next discuss the data (in addition to $Hawk_t$ and $Hawk_t^{IV}$) and the identifying assumptions for our analysis of government spending shocks.

Variables. We first specify the local projection framework (1.10)-(1.11). Our baseline shock of interest, ε_t^s in (1.10), is the military spending shock constructed by Ramey (2011) and Ramey and Zubairy (2018), based on a narrative approach to identify surprise build-ups (or build-downs) in U.S. military spending. The shock is constructed as the present value of expected changes in real defense spending over the next years, typically up to a horizon of five years, and expressed relative to real potential GDP. The two outcome variables of interest, x_{t+h} in (1.10), are real GDP and real government spending, both expressed relative to real potential GDP.[31] Finally, the vector of control variables, Z_{t-1} in (1.10), includes four lags of real GDP and real government spending, both relative to potential output and four lags of the fiscal spending shock. If we restrict $\gamma^h = \delta^h = 0$, our specification of (1.10) corresponds to equation (1) of Ramey and Zubairy (2018). This facilitates the comparability of our results with the literature.[32]

Sample. Our baseline sample covers the period from 1960Q1 to 2014Q4, which is the longest possible sample for which the Hawk-Dove balance and the fiscal spending shocks are available. Our sample includes important military spending shocks, e.g., the Vietnam War, the Carter-Reagan military buildup, and 9/11. On the other hand, our sample excludes WWII and the Korean War which are important events in Ramey (2011) and Ramey and Zubairy (2018).[33] In the context of studying the response of monetary policy to fiscal spending shocks, however, it may be desirable to exclude these events because monetary policy was less autonomous from fiscal policy prior to the Treasury-Fed Accord in 1951. Between 1942 and 1951, the Fed was constrained to support government bond prices by

[31] Detrending by potential GDP is the so-called Gordon and Krenn (2010) transformation. Compared to using log variables, this avoids using an ex-post multiplication with the GDP/G ratio, which substantially varies over time, to obtain the fiscal spending multiplier.

[32] In Section 1.4.5 we present various sensitivity checks, including additional control variables such as lags of $Hawk_t$ or interactions of $Hawk_t$ with the control vector.

[33] Ramey (2011) shows that excluding the Korean War renders military spending shocks a weak instrument for contemporaneous government spending. In general, it is not surprising that military spending shocks are a weak instrument for contemporaneous government spending because the shocks largely pertain to future spending. Therefore, we do not use military spending shocks as an instrument but as shocks in our local projection framework (1.10) and find a significant dynamic government spending response, see Section 1.4.2.

pegging short-term interest rates.

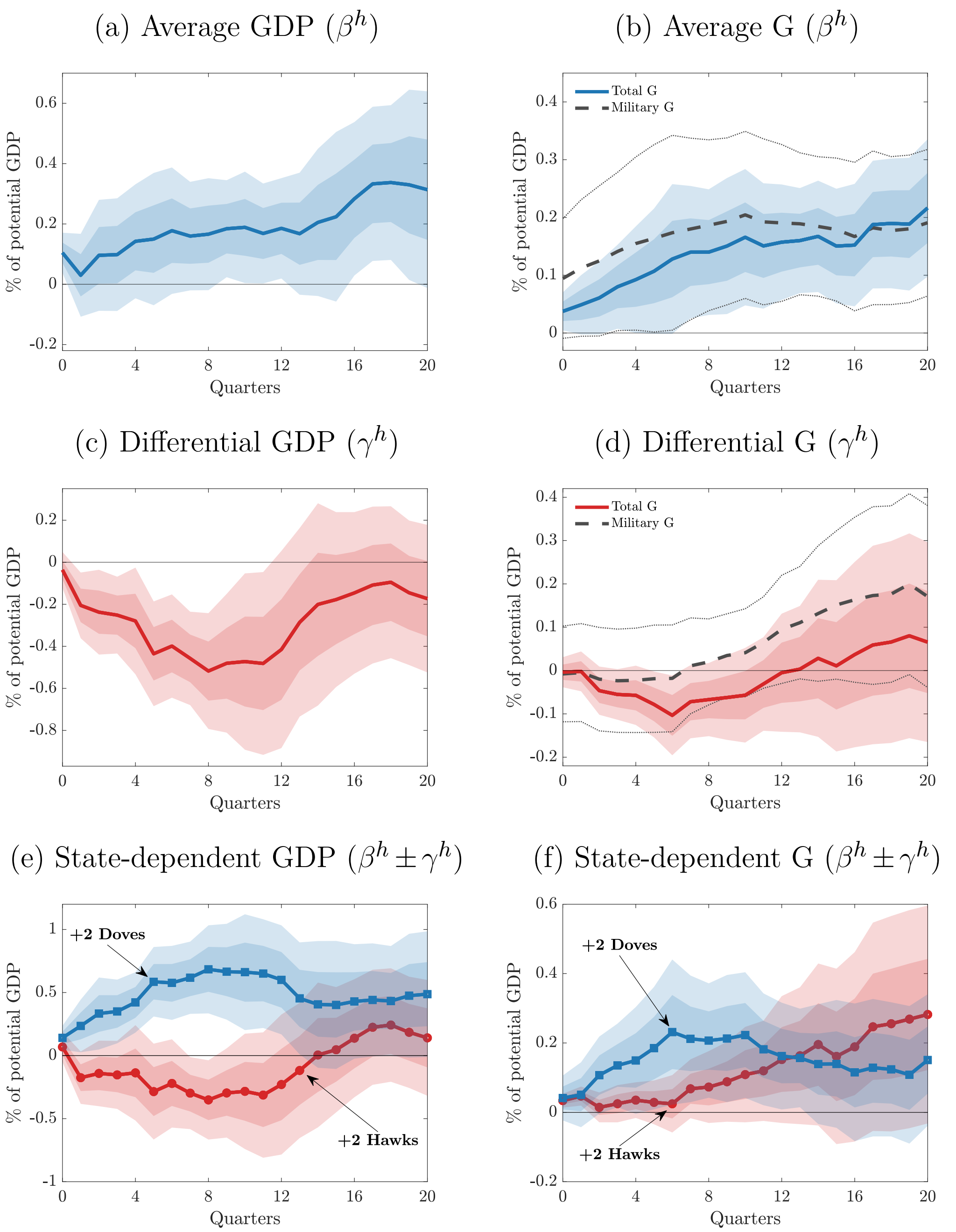
Identifying assumptions. Two key identifying assumptions are necessary for the causal interpretation of the estimates of β^h and γ^h in (1.10). The first assumption is that the FOMC rotation instrument is orthogonal to all macroeconomic shocks at all leads and lags. This is plausible for various reasons as discussed in Section 1.3.2. More specifically, given that fluctuations in $Hawk_t^{IV}$ are relatively short-lived and uncorrelated with real GDP growth, it is unlikely that our estimates capture differences in the response across booms and busts (e.g., Auerbach and Gorodnichenko, 2012; Ramey and Zubairy, 2018).
The second assumption is that military spending shocks are random shocks. In particular, the distribution of military spending shocks may not depend on systematic monetary policy. According to Ramey and Shapiro (1998) and Ramey and Zubairy (2018), military spending shocks are unanticipated changes in spending plans triggered by geopolitical events and are therefore exogenous to the economy. This argument similarly applies when conditioning on systematic monetary policy. We provide three additional arguments as to why the military spending shocks are independent of systematic monetary policy: (i) the response of military spending to the shock does not depend on systematic monetary policy; (ii) the news quotes used to construct military spending shocks as described in the supplementary appendix to Ramey and Zubairy (2018) do not mention monetary policy, the Federal Reserve, or the FOMC for our sample; and (iii) the Hawk-Dove balance does not predict spending shocks. The specific concern the last point addresses is that military spending shocks might be timed to episodes with a more dovish FOMC. To test this concern we regress future military spending shocks on $Hawk_t$ and use $Hawk_t^{IV}$ as an instrument. We find no significant effects of the Hawk-Dove balance on contemporaneous or future military spending shocks, see Figure 1.E.1 in Appendix 1.H.

1.4.2 GDP and government spending

We next present our empirical estimates of the causal effects of systematic monetary policy on the responses of real GDP and real government spending to fiscal spending shocks. We find that expansionary spending shocks raise GDP more strongly when the FOMC is dovish.

Baseline IV estimates. Figure 1.3 shows the responses of real GDP and real government spending (G) to a military spending shock conditional on systematic monetary policy ($Hawk_t$). The estimates are based on the local projection framework (1.10)-(1.11)

Figure 1.3: Responses to spending shocks conditional on monetary policy



Notes: The figure shows responses of real GDP and real government spending (G), separately for total G and military G, to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. The dotted lines indicate 95% confidence bands for military G.

as specified in Section 1.4.1. The solid lines show the point estimates and the shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.[34] All estimates of β^h and γ^h are normalized to correspond to an expansionary shock that raises the expected present discounted value of future military spending by one percent of GDP.[35]

Panels (a) and (b) show the IV estimates of β^h for GDP and G, which capture the responses when $Hawk_t$ equals its sample average. The average responses of both GDP and G are positive and significantly different from zero at most horizons beyond the first year. Both responses build up gradually and exceed 0.15% for GDP and 0.11% for total G after one year. The response of military G (dashed line) resembles total G, meaning the expansion of total G primarily reflects higher military G.[36]

Panels (c) and (d) show the estimates of γ^h, which capture the differential responses of GDP and G when the FOMC exceeds the average Hawk-Dove balance by two hawks. Specifically, γ^h is scaled to capture an increase in $(Hawk_t - \overline{Hawk})$ of 2/12. This means, for example, that two FOMC members with unknown preferences are replaced by two consistent hawks, or that two FOMC members swing from dovish to hawkish. An increase in $Hawk_t$ by 2/12 slightly exceeds one standard deviation of the change in $Hawk_t$ which is 0.15. Importantly, the GDP response is lower after a fiscal expansion when the FOMC is more hawkish. This effect is statistically significant at the 5% level until three years after the shock. The estimated magnitudes are sizable. Between two and three years after the shock the GDP response is more than 0.4% lower under a more hawkish FOMC. Conversely, the GDP response is 0.4% higher when there are two more doves in the FOMC. The differential response of government spending (G) is also negative at horizons until three years after the shock, albeit smaller in absolute terms and less significant.

The differential response of military G is insignificant at all horizons. This results supports our identifying assumption that the military spending shock does not depend on the Hawk-Dove balance. In contrast, the negative γ^h for total G means non-military G falls in response to more hawkish monetary policy. This fiscal policy response is unsurprising in an environment of tighter monetary policy and constitutes a part of the transmission of systematic monetary policy.

Panels (e) and (f) of Figure 1.3 show $\beta^h \pm \gamma^h$, the state-dependent responses when $Hawk_t$

[34] For the Newey-West standard errors, we set the bandwidth to $h+1$, where h is the horizon in (1.10). A truncation parameter rule (Lazarus, Lewis, Stock, and Watson, 2018) or automatic bandwidth selection leads to similar results.

[35] Normalizing the responses to a shock size of 1% of GDP approximately normalizes to one standard deviation of the shock series, which is 1.17% of GDP.

[36] Military G is defined relative to real potential GDP, analogous to total G, see Appendix 1.B for details.

exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The GDP response strongly varies between the dovish and the hawkish FOMC. The dovish FOMC supports the GDP expansion while the hawkish FOMC undoes the GDP expansion. Quantitatively, GDP increases by up to 0.68% under the dovish FOMC, but falls by up to 0.35% under the hawkish FOMC. The former response is highly statistically significant, whereas the latter response is less precisely estimated.
Overall, our evidence suggests that monetary offset of fiscal spending shocks is not a constant feature of monetary policy but varies strongly with the Hawk-Dove balance in the FOMC. In contrast to the GDP response, government spending displays smaller and less significant differences in the state-dependent responses.

Comparison with OLS. We compare our IV estimates presented above with the OLS counterparts that do not use the FOMC rotation instrument. Figure 1.4 shows the response of GDP as a function of the FOMC's Hawk-Dove balance, in the first and fourth year after the shock. In the first year, the OLS estimates substantially understate the dependence of the GDP response on the Hawk-Dove balance. In contrast, the OLS estimates overstate this dependence in the fourth year.[37] This comparison suggests that ignoring the endogeneity of $Hawk_t$ leads to biased conclusions about the role of systematic monetary policy for fiscal spending shocks.

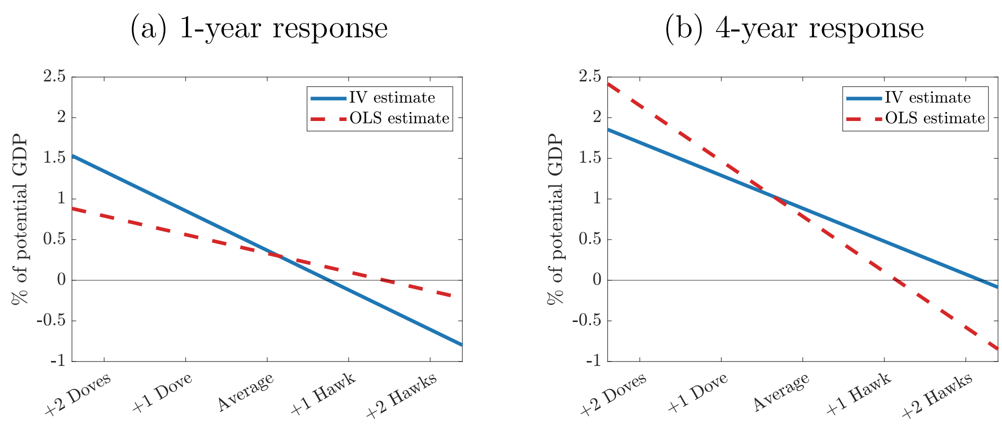
1.4.3 Fiscal spending multiplier

A key object for the design and evaluation of fiscal policies is the fiscal spending multiplier. We use our framework to estimate how the fiscal spending multiplier depends on the hawkishness of the FOMC. We find that a dovish FOMC leads to substantially larger multipliers, relative to an average or a more hawkish FOMC composition.

Definition and estimation. The multiplier is defined as the dollar amount by which GDP increases per dollar increase in fiscal spending (both in real terms). A common procedure is to compute the multiplier as the cumulative response of GDP to a spending shock divided by the cumulative response of government spending to the same shock over some horizons of interest (e.g., Mountford and Uhlig, 2009; Ramey and Zubairy, 2018). To study how systematic monetary policy shapes the fiscal multiplier, we define the monetary

[37] Figure 1.E.3 in the Appendix presents the same responses for two and three years after the shock. Figure 1.E.2 presents the OLS estimates of β^h and γ^h.

Figure 1.4: GDP responses for OLS and IV



Notes: The figure shows the yearly real GDP response to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV and OLS estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The displayed estimates are computed as $\sum_{h=H-3}^{H}[\beta^h + \gamma^h(Hawk_t - \overline{Hawk}_t)]$ for H =4 quarters in Panel (a) and $H = 16$ quarters in Panel (b).

policy-dependent fiscal multiplier as

$$FM^H(\chi) = \frac{\sum_{h=0}^{H}\left(\beta^h_{\text{GDP}} + \gamma^h_{\text{GDP}}\chi\right)}{\sum_{h=0}^{H}\left(\beta^h_{\text{G}} + \gamma^h_{\text{G}}\chi\right)} \tag{1.12}$$

where H is the forecast horizon, β_i^h and γ_i^h are the average and differential responses of outcome $i \in \{\text{GDP}, \text{G}\}$ to a spending shock, and χ indicates some level of the Hawk-Dove balance in deviation from the sample mean $(Hawk_t - \overline{Hawk})$.[38] We estimate the responses for cumulative GDP and government spending jointly by seemingly unrelated regressions, see Appendix 1.F.2. This allows us to compute standard errors that account for serial correlation and the cross-correlation between the numerator and denominator of (1.12).[39]

Results. Table 1.2 presents the IV estimates of the fiscal spending multipliers $FM^H(\chi)$ for both a two-year and a four-year horizon. For an average Hawk-Dove balance, $\chi = 0$, the cumulative spending multiplier is 1.3 at both horizons, and significantly different from zero at the 10% level. Analogous to Figure 1.3, we consider a range of χ from $-2/12$ to $+2/12$. As the FOMC becomes more dovish than average, the multiplier increases from 1.3

[38] Alternatively, one could discount future horizons in (1.12). For common discount rates, this will have a minor impact on our estimated fiscal multipliers.

[39] Our baseline inference procedure for the fiscal multiplier uses the Delta method in conjunction with Driscoll-Kraay standard errors. We further provide Anderson-Rubin type confidence sets that are robust to weak instruments and to the denominator of the multiplier being close to zero, see Section 1.4.4.

to 2.3 for one additional dove ($\chi = -1/12$), and to 3 for two additional doves ($\chi = -2/12$). The difference between the average and the dovish multipliers are similar across the two horizons. Moreover, the difference is statistically significant at the 5% level for the four-year horizon, see Table 1.E.1 in Appendix 1.E. Conversely, as the FOMC becomes more hawkish, the multiplier $FM^H(\chi)$ drops to zero or below and is insignificantly different from zero. The differences in $FM^H(\chi)$ across χ are mainly driven by differences in the cumulative GDP response rather than the G response. The differences in the GDP response across χ are larger in magnitude and more significant, see Table 1.E.1. This result is analogous to the findings in Figure 1.3.

Table 1.2: Government spending multipliers and monetary policy

	Baseline model					Linear
Outcome	+2 Hawks	+1 Hawk	Average	+1 Dove	+2 Doves	model
			Two-year horizon			
Multiplier	-4.825	-0.476	1.348	2.351	2.986	0.860
	(5.229)	(1.418)	(0.708)	(0.934)	(1.239)	(1.427)
GDP (*cum*)	-1.689	-0.282	1.124	2.531	3.937	0.616
	(0.989)	(0.768)	(0.649)	(0.689)	(0.865)	(1.057)
G (*cum*)	0.350	0.592	0.834	1.076	1.319	0.716
	(0.250)	(0.300)	(0.395)	(0.510)	(0.634)	(0.338)
			Four-year horizon			
Multiplier	-1.790	-0.001	1.308	2.307	3.095	0.838
	(2.637)	(0.862)	(0.475)	(0.808)	(1.162)	(1.449)
GDP (*cum*)	-2.735	-0.002	2.731	5.465	8.198	1.494
	(2.498)	(1.557)	(0.842)	(1.045)	(1.892)	(2.747)
G (*cum*)	1.528	1.808	2.088	2.368	2.649	1.782
	(1.010)	(0.804)	(0.734)	(0.848)	(1.079)	(0.689)

Notes: The table shows IV estimates of the cumulative fiscal spending multipliers $FM^H(\chi)$ in equation (1.12) for $H = 8$ (top panel) and $H = 16$ quarters (bottom panel), as well as the cumulative GDP response (numerator of $FM^H(\chi)$) and the cumulative G response (denominator of $FM^H(\chi)$). The coefficients are estimated using a cumulative version of the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. For our baseline model, the columns present different states of the Hawk-Dove balance between "+2 Hawks" ($\chi = +2/12$), "Average" ($\chi = 0$), and "+2 Doves" ($\chi = -2/12$). The linear model in the last column presents the estimates when we restrict $\gamma^h = \delta^h = 0$ in the local projection (1.10). Driscoll-Kraay standard errors are in parenthesis, see Appendix 1.F for details.

Comparison with linear model. We explicitly estimate how the fiscal spending multiplier depends on systematic monetary policy, whereas much of the related literature has estimated a single 'average' fiscal spending multiplier (e.g., Blanchard and Perotti, 2002; Ramey, 2016). To compare our results with this tradition in the literature, we estimate an average fiscal spending multiplier in a linear version of our framework when restricting $\gamma^h = \delta^h = 0$. The resulting fiscal multiplier is given by $\widetilde{FM}^H = (\sum_{h=0}^{H} \beta^h_{\text{GDP}})/(\sum_{h=0}^{H} \beta^h_{\text{G}})$ and the estimates are presented in the last column of Table 1.2. We find average multipliers of about 0.85 at both horizons. While this estimate is relatively close to the multiplier estimates in Ramey and Zubairy (2018) which range from 0.66 to 0.71 (see their Table 1), it is substantially below the multiplier of 1.3 for an average FOMC composition ($FM^H(0)$) in our baseline model. In addition, the standard errors for the multiplier in the linear model are substantially larger than the standard errors of $FM^H(0)$. This comparison suggests that accounting for systematic monetary policy is important for the magnitude and precision of multiplier estimates. Moreover, one potential reason for the broad range of multiplier estimates in the literature is not accounting for time variation in systematic monetary policy.

1.4.4 Weak instruments and robust inference

A common concern with IV estimates is the strength of the instrument. We provide evidence supporting the strength of our instruments, including weak instrument tests, reinforcing the contribution of our identification design. Finally, we provide robust inference for the estimated responses and fiscal multipliers.

First-stage results. Our local projection framework (1.10) contains two endogenous regressors, $\varepsilon^s_t(Hawk_t - \overline{Hawk})$ and $(Hawk_t - \overline{Hawk})$. The estimates of the two associated first-stage regressions are shown in Table 1.F.1 in the Appendix. We find that the instrumental variable $\varepsilon^s_t(Hawk^{IV}_t - \overline{Hawk}^{IV})$ has a positive effect on the endogenous variable $\varepsilon^s_t(Hawk_t - \overline{Hawk})$ that is significant at the one percent level. Similarly, $(Hawk^{IV}_t - \overline{Hawk}^{IV})$ has a positive and highly significant effect on $(Hawk_t - \overline{Hawk})$. In both regressions, the R^2 increases by about 0.4 when including the instruments as regressors. Taken together, these results suggest that our instruments are strong (Bound, Jaeger, and Baker, 1995).

Weak instrument tests. We use three statistical tests to assess the strength of our instrument more formally. First, we use the Montiel Olea and Pflueger (2013) test of weak instruments, which is popular in time series settings because it is robust to autocorrelation and heteroskedasticity. Formally, we test whether the relative weak instrument bias for the IV estimates of γ^h exceeds 10%, 20%, or 30%.[40] Panel (a) of Figure 1.5 shows the p-values of the weak instrument tests for the differential GDP response. At all horizons, even a relatively small 10% bias ($\tau = 0.1$) can be rejected at significance levels below 2%. The second weak instrument test we apply was recently developed by Lewis and Mertens (2022) and generalizes Montiel Olea and Pflueger (2013) to allow for multiple endogenous regressors. We apply this test to jointly evaluate whether the average relative bias across γ^h and δ^h exceeds some threshold τ and report the results in Panel (b) of Figure 1.5. A small average bias of 10% can be rejected at significance levels below 10% for most horizons. Moreover, we can reject a bias of 20% at the two percent level for all horizons. For government spending, both tests lead to the same conclusion, see Figure 1.F.1 in Appendix 1.F.

Lastly, we test for weak instruments via the reduced form of our regression framework. Following Chernozhukov and Hansen (2008), the hypothesis test of the reduced form estimates of γ^h against zero is equivalent to testing whether the instrument has zero relevance. Figure 1.F.2 in the Appendix shows that the reduced-form estimates for γ^h are significant, as in Figure 1.3. To summarize, all three tests indicate that our instruments are not weak.

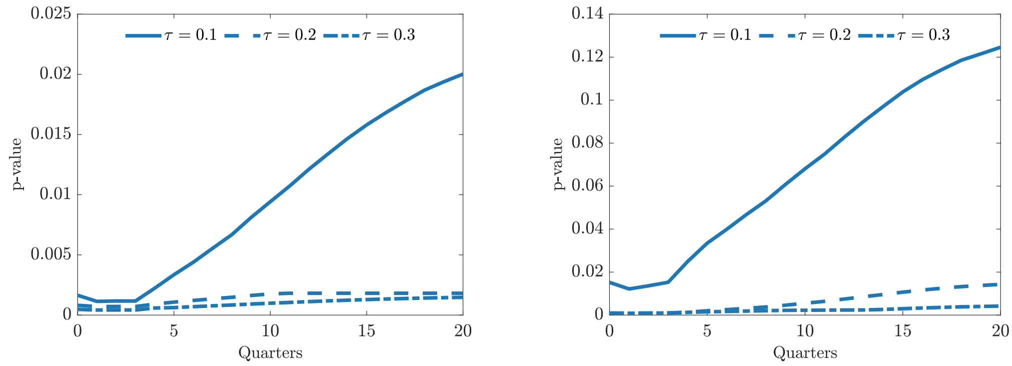
Robust inference for impulse responses. To address residual concerns about instrument strength, we further provide inference that is robust to weak instruments and allows for multiple endogenous regressors based on Andrews (2018). We find robust confidence sets for the differential GDP and G responses similar to our baseline intervals, see Figure 1.F.3 in the Appendix. This provides additional support for the strength of our instruments.

Robust inference for fiscal multipliers. We provide Anderson and Rubin (1949) type inference for the fiscal multiplier, following Andrews, Stock, and Sun (2019). Importantly, the procedure is based on a test statistic with a limiting distribution that does not depend

[40] We apply the test to γ^h because it is our main coefficient of interest (together with β^h), and because the Montiel Olea and Pflueger (2013) test can only be applied to a single endogenous regressor. For the other endogenous regressor, $(Hawk_t - \overline{Hawk})$ in levels, we estimate the first stage separately and plug in the fitted values in the second stage used to test the interaction term. If we alternatively replace the $Hawk_t$ level term by $Hawk_t^{IV}$ we obtain very similar results.

Figure 1.5: Weak instrument tests

(a) Montiel Olea and Pflueger (2013) (b) Lewis and Mertens (2022)



Notes: The figure shows p-values for rejecting the null of weak instruments for the responses of real GDP, based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The Montiel Olea and Pflueger (2013) test evaluates the null of the bias in γ^h exceeding a threshold τ. Similarly, the Lewis and Mertens (2022) test evaluates the null of the ℓ^2 norm of the bias in γ^h and δ^h exceeding a threshold τ. For the former, the endogenous regressor $Hawk_t$ is not tested but directly replaced by its first stage fitted value. The critical values and associated p-values are based on Newey-West standard errors.

on the strength of the instruments and that does not depend on the denominator of the fiscal multiplier being non-zero. We provide a detailed description of the implementation in Appendix 1.F.2. The robust confidence sets are presented in Figure 1.F.4. They leave our conclusions about fiscal multipliers in Table 1.2 broadly unchanged. In particular, we estimate dovish fiscal multipliers with p-values of 0.08 and 0.12 for +1 Dove and +2 Doves, respectively. The hawkish multipliers are highly insignificant. Finally, the average multiplier is significant with a p-value of 0.06, whereas the estimate of the multiplier in the liner model remains highly insignificant.

1.4.5 Sensitivity analysis

In this section, we provide an extensive sensitivity analysis to assess the robustness of our baseline results. We investigate alternative Hawk-Dove balances, an alternative spending shock, varying sample periods, and the inclusion of additional control variables.

Alternative Hawk-Dove balances. We address potential concerns regarding the aggregation of individual policy preferences and the comparability of preferences over time. While our baseline $Hawk_t$ aggregates individual preferences by an unweighted arithmetic average, we consider four alternative aggregation schemes. First, we use the median policy preference across FOMC members. Second, we use an arithmetic average but double the weight of the Fed Chair. Third, we use the arithmetic average but do not distinguish between consistent and swinging FOMC members when defining $Hawk_{i\tau}$ in (1.7). We estimate average and differential responses similar to the baseline, albeit smaller ones for the median aggregation, see Figure 1.G.1. Across the three alternative aggregations, we find multipliers similar to the baseline, see Table 1.G.1. In a fourth alternative aggregation, we consider the role of strong majorities in the FOMC. We construct an alternative Hawk-Dove balance which equals -1 if $Hawk_t$ falls below the first quartile or tertile of the distribution of $Hawk_t$ over time, +1 above the highest quartile or tertile, and zero otherwise. The estimated average and differential effects remain quite similar in terms of the shapes and significance of the results, see Figure 1.G.2. Both specifications also roughly align with the baseline multipliers, see Table 1.G.2.

We also address potential endogeneity concerns due to preference swings of policymakers by alternative rotation instruments. We either allow swings in the instrument only with a time lag of 8 or 16 quarters or impose that preferences equal the average preference of an FRB president, rendering them time-invariant. The results in Figure 1.G.3 are similar to our baseline, suggesting that swings in the instrument are not driving our results. The implied state-dependence of the fiscal multipliers in Table 1.G.1 is slightly muted compared to the baseline.[41]

Another potential concern is that the meaning of being a hawk or dove might have changed over time, see the discussion in Section 1.3.1. To account for trends in the Hawk-Dove balance, we consider an alternative Hawk-Dove balance which subtracts from the baseline $Hawk_t$ its backward-looking 5, 10, or 15-year moving average. The estimated average and differential responses are very similar to our baseline estimates, see Figure 1.G.4 in the Appendix. In addition, the average and dovish multipliers have similar magnitudes as the baseline while the hawkish multiplier is similarly imprecise, see Table 1.G.1 in the Appendix. Overall, our results reinforce the arguments in Section 1.3.1 that the classification of hawks and doves is indeed comparable over time.

Alternative spending shock. Our baseline shock is specific to military spending. We investigate the external validity of our results by using an alternative fiscal spending shock, which is identified from a timing restriction on total government spending as suggested by Blanchard and Perotti (2002), henceforth BP. They assume that only government spending shocks can affect government spending contemporaneously.

[41] The finding of muted state-dependence in the multipliers does not necessarily imply that our results are partly driven by endogenous swings. The alternative rotation instrument also takes out variation from swings which are exogenous, see the discussion in Section 1.3.2.

We find that GDP and G respond more swiftly compared to our baseline, see Figure 1.G.5. This is in line with the nature of the BP shock. More importantly, we find that a hawkish FOMC significantly dampens the expansionary effect on GDP. The average fiscal multiplier is around 1.4 for the four-year horizon, see Table 1.G.1, which is remarkably similar to our baseline multiplier. The fiscal multiplier ranges from 0.88 to 1.74 between the hawkish and dovish FOMC ($\chi = \pm 2/12$). While the variation in the multiplier is more compressed compared to the baseline, it is similarly significant.[42]

Great Recession and ZLB. Our baseline results are estimated using the sample from 1960Q1 to 2014Q4 which includes the Great Recession (GR) and the subsequent ZLB period. We investigate the sensitivity of our results on a sample that ends either in 2007Q4 to exclude the GR and ZLB period or in 2008Q4 to exclude the ZLB period. For both of these subsamples, our estimates are highly similar to the baseline, see Figure 1.G.6 for average and differential responses and the corresponding multiplier in Table 1.G.1.

Additional (non-linear) control variables. Finally, we investigate the sensitivity of our results to adding potentially important co-variates to the baseline specification of our local projection framework. The additional control variables are short-term and long-term interest rates, inflation, and the primary surplus. While the estimates are similar to the baseline, we naturally give up some statistical power, see Figure 1.G.7 and Table 1.G.1. Nevertheless, we estimate dovish multipliers around 2 which substantially exceeds the average multiplier, consistent with our baseline results. We further add lags of $Hawk_t$, or we consider non-linear controls by including interactions of $Hawk_t$ with the control variables. The results are remarkably close to the baseline, see Figure 1.G.8 and Table 1.G.1.

1.5 Inspecting the mechanism

In this section, we inspect the mechanism behind our findings in the previous section. We show that in response to an expansionary spending shock, nominal and real interest rates rise, and inflation is dampened under a hawkish FOMC. Conversely, interest rates initially

[42] The compressed variation in the multiplier appears consistent with the interest rate responses to BP shocks. Initially, interest rates significantly rise under a more hawkish FOMC, but the magnitude is smaller than for the baseline spending shocks. Starting two years after the shock, the differential interest rate response flips sign and interest rates are lower under a more hawkish FOMC. An important reason for the different interest rate responses across spending shocks may be the fact that G rises only temporarily in response to BP shocks, but rises persistently after military spending shocks.

fall and rise only with substantial delay under a dovish FOMC, supporting a crowd in of consumption and investment.

1.5.1 Additional responses

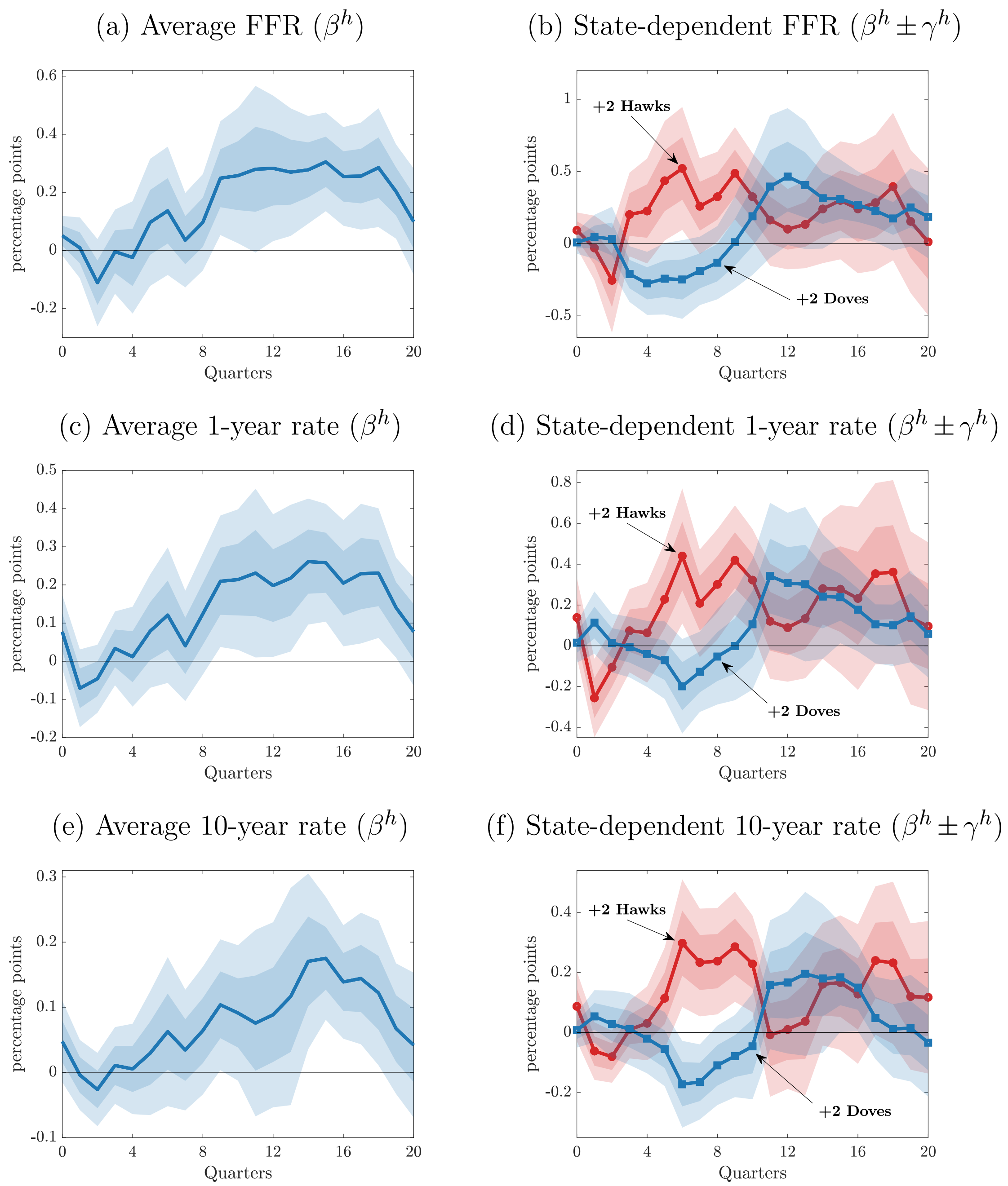
Conventional wisdom says that monetary policy tightens in response to higher government spending in order to mitigate the inflationary pressure. The Federal Reserve can use a range of tools, including the target federal funds rate, the discount rate, balance sheet policies and communication including forward guidance. These tools can affect short- and long-term interest rates, and hence inflation, consumption, and investment.

Nominal interest rates. We study the response of the federal funds rate (FFR) and the annualized yield on 1-year and 10-year Treasury securities to government spending shocks by using our local projection framework (1.10)-(1.11) with interest rates as outcome variable x_{t+h}. We follow the specification in Section 1.4.1 but include four lags of the FFR, 1-year and 10-year Treasury yields, and CPI inflation as additional control variables to control for pre-trends in these outcomes.

Panels (a), (c) and (e) of Figure 1.6 show the IV estimates of β^h, the average response of the three nominal interest rates when $Hawk_t$ equals its sample average. The average FFR response appears muted in the first year, after which it gradually increases and reaches 30 basis points at horizons beyond two years. The average responses of the 1-year and 10-year yields feature similar shapes, albeit at lower magnitudes. Panels (b), (d) and (f) show the IV estimates of $\beta^h \pm \gamma^h$, the state-dependent interest rate responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). All interest rates increase faster and more strongly under a hawkish FOMC. Compared to the average response, the peak in the FFR is reached one year earlier and is almost double in size (about 56 basis points). In contrast, under a dovish FOMC, the FFR falls for almost two years and a reversion to a higher FFR is observed only three years after the shock. Similarly, both 1-year and 10-year Treasury yields increase after two years under a dovish FOMC, suggesting that the monetary regimes also differ in their effects on expected future policy at long horizons.

The delayed FFR response is consistent with the initial uncertainty surrounding the military spending shock and the gradually evolving macroeconomic effects of the shock, see Figure 1.3. Section 1.6 provides narrative evidence from the FOMC historical records suggesting that indeed the FOMC delays action until some uncertainty about the spending

Figure 1.6: Responses of nominal interest rates



Notes: The figure shows responses of the federal funds rate (FFR), as well as the 1-year and 10-year treasury yields to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). All outcomes are annualized interest rates. We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.5.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

plans and their potential effect on the economy and inflation is resolved. Furthermore, a delayed differential policy response that extends for several quarters beyond the term of the FOMC and the associated rotation present at the time of the shock, is consistent with the decision dynamics in the FOMC. For example, Laurence Meyer, member of the Board of Governors from 1996 to 2002, describes these dynamics during his term at the Fed as follows:

> *So was the FOMC meeting merely a ritual dance? No. I came to see policy decisions as often evolving over at least a couple of meetings. The seeds were sown at one meeting and harvested at the next. [...] Similarly, while in my remarks to my colleagues it sounded as if I were addressing today's concerns and today's policy decisions, in reality I was often positioning myself, and my peers, for the next meeting.*
> Laurence Meyer (2004), A Term at the Fed: An Insider's View, Harper Business

Consistent with Meyer's view that it takes time to influence policy strategies in the FOMC, we find that the FOMC rotation ($Hawk_t^{IV}$) is more important for the policy response to the spending shock and its real effects when the shock occurs closer to the beginning of the FOMC rotation, which takes place in the first quarter of the year. When we drop spending shocks in the second half of the year, we obtain similar findings compared to the baseline, see Figures 1.H.1-1.H.2 in Appendix 1.H. Conversely, the dependence on monetary policy becomes weaker and less significant when dropping spending shocks in the first half of the year.

Inflation rates. We further assess the effects of the military spending shocks on inflation expectations, CPI core inflation (excluding food and energy prices), and CPI headline inflation.[43] We estimate the inflation responses using the specification of our local projection framework (1.10)-(1.11) for nominal interest rates and control for four lags of the inflation measure under consideration. The results are shown in Figure 1.7. Overall, the inflation responses are not precisely estimated. The average response of expected inflation tends to be positive, while the evidence is mixed for core and headline inflation. Turning to the dependence on the Hawk-Dove balance, we find that inflation expectations increase sluggishly under a dovish FOMC and peak at about three years. In contrast, inflation expectations tend to fall under a hawkish FOMC, suggesting that the FOMC is successful in

[43] We use one-year inflation expectations based on the CPI forecasts from the Livingston Survey of the Federal Reserve Bank of Philadelphia. It is the oldest continuous survey on the expectations of economists from industry, government, banking, and academia. For details, see Appendix 1.B.

containing inflation expectations. The response of core inflation follows a similar but even more sluggish pattern, suggesting that policy tightening is successful in containing inflationary pressures. Compared to the interest rate responses, the inflation response appear delayed by one to two years, broadly in line with the lags in the transmission of monetary policy. Finally, the results for headline inflation are more mixed, possibly due to larger transitory fluctuations in energy and food prices.

Real interest rates. In a large class of models, the real effects of monetary policy depend on its ability to affect real interest rates. Under a hawkish FOMC, the response of nominal rates is larger, while the response of inflation is smaller. Hence, the implied response of real interest rates is larger. In response to a government spending shock, real interest rates increase by more if the FOMC is hawkish and by less if the FOMC is dovish. We obtain similar results when directly estimating the real interest rate response. We consider real interest rates constructed by subtracting the expected CPI inflation from the three nominal interest rates considered in Figure 1.6. Figure 1.H.3 in Appendix 1.H presents the IV estimates of the average and state-dependent responses.

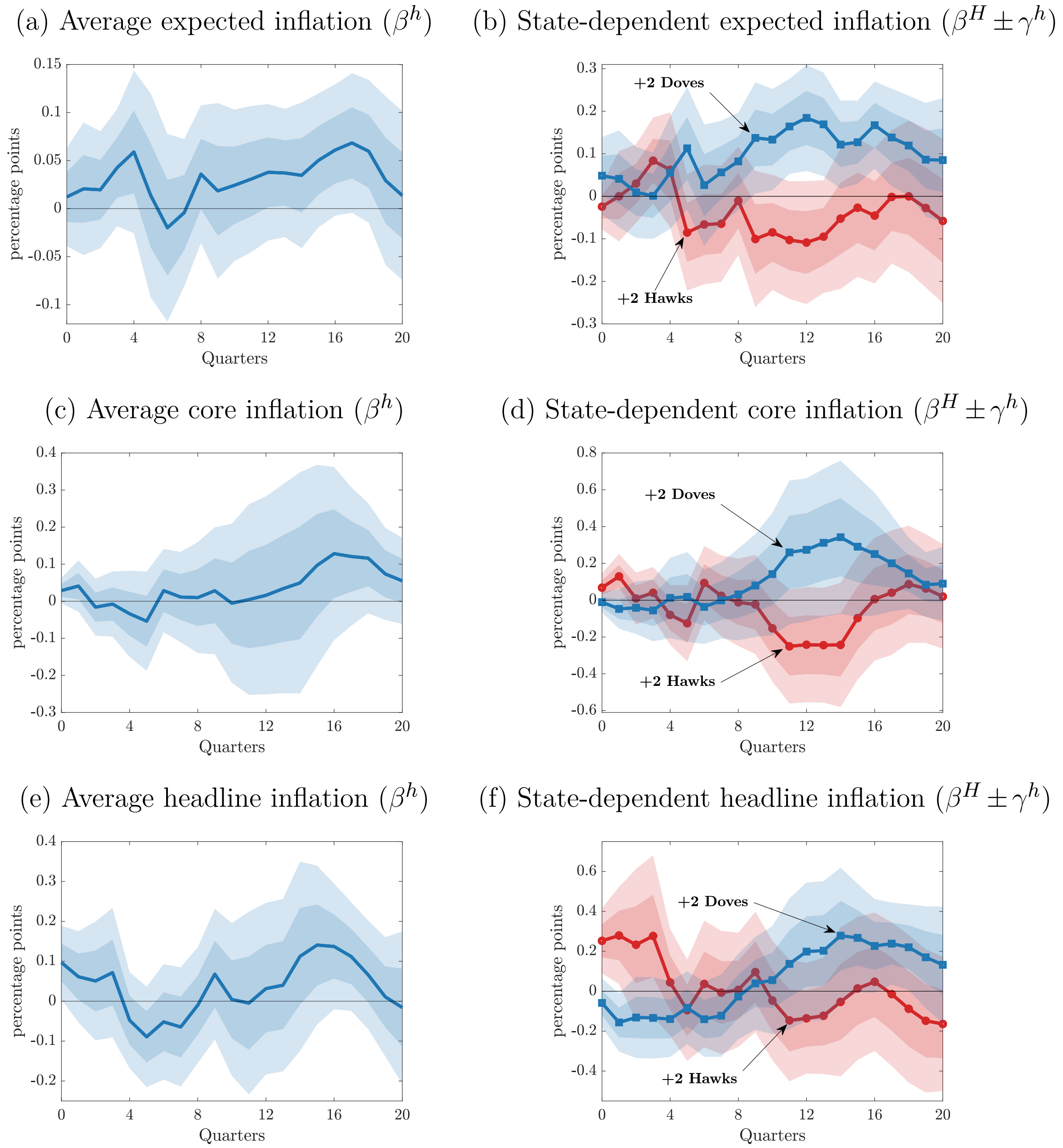
Investment and consumption. We examine the underlying components of the responses of real GDP. The fiscal spending multiplier can be above one when GDP components other than G are crowded in by the spending shock. Conversely, crowding out may lead to multipliers below one. We find that the differential GDP effects are primarily driven by private consumption and somewhat less by private investment, see Figure 1.H.4 in Appendix 1.H.[44] For the average Hawk-Dove balance, we find a mild but insignificant crowding out of private consumption and crowding-in of private investment in the short run. In contrast, the crowding out of consumption is strong and significant under a hawkish FOMC. For investment, we find a similar albeit smaller and less significant pattern. Overall, the strong state-dependence of fiscal multipliers appears to be mainly driven by private consumption.

1.5.2 Relation to the literature

To put our empirical results into perspective, we compare them with prior estimates for the effects of monetary policy shocks and fiscal multipliers.

[44] For details on the definition of consumption and investment, see Appendix 1.B.

Figure 1.7: Responses of inflation rates



Notes: The figure shows responses of expected inflation, CPI core, and CPI headline inflation to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). All outcomes are annualized inflation rates. We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.5.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Relation to monetary policy shocks. Most of the related empirical literature estimates the effects of monetary policy shocks in the economy. Therefore, it may be interesting to compare the effects of such shocks with our estimates.[45] To this aim, we compare the ratio of the peak output to peak interest rate response for various monetary policy shocks with the ratio of the peak differential GDP response to the peak differential interest rate response, formally $(\min_h \gamma_y^h)/(\max_h \gamma_i^h)$, from our estimation. For U.S. monetary policy shocks, recursively identified shocks in Coibion (2012) imply a ratio of -1.56, Romer and Romer (2004) as estimated in Coibion (2012) a ratio of -1.99, and high-frequency identified shocks in Jarociński and Karadi (2020) a ratio of -1.34.[46] Hence, a peak interest rate hike of one percentage point coincides with a peak decline of real GDP between 1 and 2 percentage points. In comparison, our estimates imply a ratio of -1.35, which is within the range implied by evidence on monetary policy shocks.

Relation to fiscal multipliers. The interest rate responses further allow us to relate our fiscal spending multiplier estimates in Table 1.2 with the findings in the related literature. Our spending multiplier is between two and three under the dovish FOMC which is associated with a weak negative response of the nominal (and real) FFR for the first two years. In theory, the multiplier may be far above one (or negative) depending on the response of interest rates (Farhi and Werning, 2016; Woodford, 2011). In an estimated medium-scale DSGE model, Christiano, Eichenbaum, and Rebelo (2011) find multipliers between two and four at the ZLB when the short-run nominal interest rate does not respond.

Our findings also relate to an empirical literature that estimates fiscal spending multipliers. For example, Nakamura and Steinsson (2014) estimate two-year regional multipliers for the U.S. of approximately 1.5. To the extent that regional multipliers correspond to the aggregate multiplier when nominal interest rates do not respond, we can compare their estimates to our two-year multiplier estimates. In particular, we construct a spending multiplier for the case in which the nominal FFR is unresponsive by choosing the Hawk-Dove balance (χ) that minimizes the squared distance of the FFR response from zero in the first two years.[47] This requires a χ slightly below the "+1 Dove" case in Table 1.2. The associated two-year spending multiplier is 1.9, which is similar to the estimates in

[45] As we discuss in the introduction, an advantage of our approach is that it circumvents potential concerns related to the identification of monetary policy shocks and their size.

[46] We compute these numbers based on the (baseline) estimates reported in each article, using the respective replication package.

[47] Formally, we solve $\min_\chi \sum_{h=0}^{8}(\beta_{FFR}^h + \chi \cdot \gamma_{FFR}^h)^2$, where χ indicates a level of the Hawk-Dove balance in deviation from the sample mean $(Hawk_t - \overline{Hawk})$.

Nakamura and Steinsson (2014).

We further compare our results with the estimate of the aggregate spending multiplier when monetary policy is constrained at the ZLB. Ramey and Zubairy (2018) finds a ZLB multiplier of 1.6 after two years (when excluding WWII), while Miyamoto, Nguyen, and Sergeyev (2018) find a ZLB multiplier well above 1.5 for Japan. Notwithstanding the endogeneity of a binding ZLB, our multiplier of 1.9 under a non-responsive FFR is similar to the ZLB multipliers in the literature. Overall, our multiplier estimates and the associated interest rate path are broadly similar to previous quantitative and empirical findings.

1.6 Historical FOMC records

> *Interviewer*: What would have happened, do you think, if the Fed had not raised the discount rate?
> *Chairman Martin*: A golden opportunity to stop inflation in its tracks would have been lost.
> *Interviewer*: It was primarily the projection of Vietnam spending; is that correct?
> *Chairman Martin*: Right. I kept telling him we could not have guns and butter.
> *Interviewer*: When you talked to Lyndon Johnson about this projection, what did he say? Did he disagree with it or did he agree with it?
> *Chairman Martin*: He disagreed. He thought we could have guns and butter.[48]

We complement our quantitative analysis with narrative evidence from the records of discussions and decisions at FOMC meetings. This evidence serves two purposes. First, it confirms that the FOMC members discuss changes in government defense spending, assessing the impact on economic activity and inflation as well as the FOMC's policy response. Second, it shows that the policy response depends on the composition of the FOMC.

To illustrate the FOMC discussion around military spending shocks, the FOMC composition, and the corresponding policy response, we focus on two important events during the 1960s: the acceleration of the U.S. Space Program in 1961 and the Vietnam ground war starting in 1965. The corresponding military shocks are both large while the FOMC

[48] Former Fed Chairman William McChesney Martin: Oral History, Interview I by Michael L. Gillette in 1987, LBJ Library Oral History Collection. The interviewer refers to the decision of the Federal Reserve to raise the discount rate on December 1965. Lyndon B. Johnson was the President of the United States from 1963 to 1969.

composition appears on average hawkish in the first part of the 1960s and dovish in the second part, see Figure 1.2. In this period, the Fed was headed by William McChesney Martin, a consistent hawk whose tenure as chairman from 1951 to 1970 was the longest in history.

For both events, we identify three phases of the FOMC's reaction to military defense spending from the historical FOMC records. First, there is uncertainty about the extent to which the spending plans will be realized and about their impact on the economy. Second, the effects of higher spending on the economy become visible while inflation appears unresponsive, therefore they wait until "all the evidence was in". Third, the effects on inflation become visible but the FOMC delays action. The first two are common for hawkish and dovish committees while the third phase is more pronounced under a dovish one, broadly in line with our empirical findings.

We summarize the key aspects below and discuss the complete case studies in Appendix 1.I. The sources for our narrative evidence are the FOMC Historical Minutes until 1967 and the Memoranda of Discussion thereafter.

1.6.1 The U.S. Space Program

In the first half of 1961, Ramey and Zubairy (2018) identify two expansionary shocks related to President Kennedy's defense spending plans, including the Space Program to "go to the Moon". In the FOMC meeting of August 1, 1961, the staff presents the following assessment:

> *On top of substantial increases in expenditures to finance space exploration and longer-run defense measures [...] the President has found it necessary to recommend an increase of $3-1/2 billion in current defense expenditures [...]. More important, the President accompanied his recommendations with a very firm statement regarding his intentions with respect to the 1963 budget. These factors have certainly tended to minimize the immediate inflationary expectations and the urgency of the need for counter-measures. As of this moment in time, actual developments do not seem to call for any change in monetary policy.* (p.8)

The majority of the FOMC members argued similarly for no change in policy because the effects could not yet be evaluated. Hawkish FOMC members suggested the need for alertness to avoid getting into an inflationary situation while agreeing to no policy change

in this meeting. In this regard, New York Fed first-vice president, William Treiber noted: *If expenditures and related private spending result in an upsurge of activity with inflationary aspects, we may have to modify our policy of basic monetary ease sooner than we would otherwise have done. In the coming period undue ease should be avoided.* (p.22-23)
FOMC members started to acknowledge the expansionary impact on employment and business sentiment in defense-related industries by the end of 1961 and later in 1963 on prices. On May 7, 1963, the FOMC voted to firm policy as a preemptive move against inflation.[49] In this meeting, Chairman Martin said:

> *If the Committee waited too long, however, it might have to deal with an active problem of inflationary pressures. In his opinion, there was already a good bit of pressure in some areas that could build up rapidly. If one waited until after the resulting price movements actually occurred, he might wonder why he had not done something about it before. It would be too late at that juncture.* (p.61)

In this period, the FOMC composition was hawkish on average. This helped the hawkish Chairman Martin to reach a consensus for tighter policy to act preemptively against inflationary pressures.

1.6.2 The Vietnam War

In 1965, the U.S. entered the ground war in Vietnam leading to a series of expansionary military spending shocks lasting until 1967Q1. In the FOMC meeting of August 10, 1965, the staff's presentation explicitly accounted for the intended increase of military spending:

> *Further stimulus to the economy will come from expanded Government procurement for Vietnam hostilities. [...] the increases in spending and in the armed forces now proposed do not appear significant enough to touch off [...] widespread price increases. [...] The market response to Vietnam developments doesn't suggest any widespread fears of shortages, rationing, or inflation. On balance, then, the domestic evidence isn't clear enough to me to justify a significant policy move in either direction at this juncture.* (p. 28-29).

[49]The FOMC shifted the emphasis of monetary policy toward slightly less ease and toward maintaining a moderately firm tone in the money market in June 1962, mentioning balance-of-payments concerns. In this period, FOMC members interested in a tighter, inflation-focused monetary policy often cited the balance-of-payments criterion to bolster their case (Bordo and Humpage, 2014).

Several FOMC members agreed with the staff's assessment and argued for an unchanged policy due to significant uncertainties related to the developments in Vietnam. In contrast, few hawkish FOMC members noted that the Vietnam hostilities were already affecting industrial prices. Two meetings later, on September 28, the dovish members dissented against the "status quo", arguing that, in their judgment, evidence of inflationary pressure was lacking and hence, they preferred an easier policy. In contrast, Alfred Hayes (New York Fed), a hawk, argued in the meeting of October 12, 1965 that: *Looking ahead, I think we have a real basis for concern about potential inflationary pressures* (p.25). Chairman Martin shared similar thinking on inflation while sensing that he did not have a majority to firm policy:

> *While the evidence was not clear, he thought there were many signs of inflation and of inflationary psychology in the economy. [...] But the Committee had a tendency to feel that it was best to wait until all the evidence was in before making a policy change. The difficulty was that when all the evidence was in it was likely to be too late. [...] With a divided Committee and in face of strong Administration opposition he did not believe it would be appropriate for him to lend his support to those who favored a change in policy now.* (p.68-69)

On December 5, 1965, the discount rate was raised with a narrow majority in order to prevent the risk of inflation. However, the tightening signal by the Fed was not enough to contain the buildup of inflationary pressures. While this had become clear for most members, the U.S. President had promised an anti-inflationary fiscal program and the FOMC delayed action in support of promised fiscal restraint. On September 13, 1966, Governor James Robertson summarized the situation as follows: *Inflationary pressures are persisting, as the staff materials have underlined. [...] To counter these inflationary pressures, we now have the promise of help from a somewhat greater degree of fiscal restraint.* (p.72). Hoping on the legislative action to raise taxes in 1967, by the last quarter of 1966 and throughout the first part of 1967, the FOMC eased policy, despite two large expansionary military spending shocks hitting in 1966Q4 and 1967Q1. In the FOMC meeting of September 12, 1967, Chairman Martin acknowledged that tightening had been delayed for too long because of the tendency to *underestimate the strains being put on economic resources by the hostilities in Vietnam. A "guns and butter" economy was not feasible; the country's resources were not sufficient for that.* (p.73). The FOMC decided to tighten the policy on December 12, 1967. Once again, Chairman Martin admitted delayed action as follows:

> *It was his feeling that the Committee had in a sense been caught in a trap [...] From the standpoint of economic considerations alone, it would have been desirable to adopt a firmer monetary policy a number of months ago.* (p.96)

In the period between 1965 and 1967, the FOMC is categorized as dovish on average. Both, the dovish committee and the political pressure against tighter policy made it more difficult for Chairman Martin to reach a consensus for firm policy within the FOMC. Indeed, we observe that even when the expansionary effects of military spending related to the Vietnam War became evident, the FOMC initially hesitated, then tightened modestly but soon erred toward loose policy.

Overall, the narrative evidence from the 1960s supports the important elements that we highlight in this chapter: the Fed's reaction to military spending and the role of the FOMC composition for this reaction. In Appendix 1.I, we provide the complete case studies.

1.7 Conclusion

This chapter proposes an identification design to estimate the effects of systematic monetary policy on the propagation of macroeconomic shocks. Our design combines the narrative classification of FOMC members' policy preferences from Istrefi (2019) with a novel FOMC rotation instrument for systematic monetary policy. The identification design opens up myriad research opportunities, such as revisiting the effects of various fiscal, technology, and oil shocks and their dependence on systematic monetary policy.

We use our identification design to study government spending shocks in the U.S. and find that fiscal spending multipliers depend strongly and significantly on systematic monetary policy. We inspect the mechanism behind our result and find consistent interest rate and inflation responses. In recent years, we have observed large fiscal expansions related to COVID and, more recently, related to Russia's war against Ukraine. In the same period, the FOMC was rather dovish. Applied to these years, our findings suggest that the combination of fiscal and monetary policy contributed to the robust recovery of GDP. However, a potentially misleading conclusion from our results is that the government should increase spending when the FOMC is dovish. This could be misleading because such responses of government spending to systematic monetary policy are not random shocks. This is a case of the Lucas (1976) critique. To avoid misleading conclusions, a promising avenue for future research is to use our results to discipline micro-founded models to study optimal fiscal stabilization policy.

Finally, while our identification design is specific to U.S. monetary policy, a promising avenue for future research is to study other countries or currency areas in which committees decide monetary policy. In fact, since 2015 the European Central Bank's governing council allocates voting rights to its members through a rotation mechanism.

Appendix

1.A New Keynesian model

1.A.1 Equilibrium dynamics

In the following, we derive equation (1.3). Denoting by lower case letters (log) deviations from steady state, we obtain three equilibrium conditions for the model described in Section 1.2:

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \lambda \left(\varphi + \frac{1}{1-\gamma} \right) y_t - \frac{\lambda\gamma}{1-\gamma} x_t^s - \lambda(1+\varphi) x_t^a, \tag{1.13}$$

$$y_t = \mathbb{E}_t[y_{t+1}] - (1-\gamma)(i_t - \mathbb{E}_t[\pi_{t+1}]) + \gamma(1-\rho_s) x_t^s, \tag{1.14}$$

$$i_t = \widetilde{\phi}_t \pi_t, \tag{1.15}$$

where $\lambda = (1-\theta)(1-\beta\theta)/\theta$ and where $\widetilde{\phi}_t = \phi + \phi_t$ follows

$$\phi_t = \rho_\phi \phi_{t-1} + \zeta^s \varepsilon_t^s + \zeta^a \varepsilon_t^a + \eta_t, \quad |\rho_\phi| < 1.$$

We assume the macroeconomic shocks $(\varepsilon_t^a, \varepsilon_t^s)$ and the exogenous shifter η_t are mutually independent and identically distributed over time. We combine the equations to obtain

$$\begin{aligned} y_t = \frac{1-\gamma}{1+\lambda(\varphi(1-\gamma)+1)\phi_t} \Bigg[& \frac{\mathbb{E}_t[y_{t+1}]}{1-\gamma} + (1-\beta\phi_t)\mathbb{E}_t[\pi_{t+1}] \\ & + \frac{\gamma}{1-\gamma}(\phi_t\lambda + (1-\rho_s)) x_t^s + \phi_t \lambda(\varphi+1) x_t^a \Bigg]. \end{aligned} \tag{1.16}$$

Combining (1.13) and (1.16), the model dynamics follow $\mathcal{Y}_t = A(\phi_t)\ \mathbb{E}_t[\mathcal{Y}_{t+1}] + B(\phi_t)\ \mathcal{X}_t$, with $\mathcal{Y}_t = (y_t, \pi_t)'$, $\mathcal{X}_t = (x_t^s, x_t^a)'$ and $A(\phi_t), B(\phi_t)$ depending only on model parameters. A first-order approximation around $\phi_t = 0$ yields

$$\mathcal{Y}_t = A\mathbb{E}_t\left[\mathcal{Y}_{t+1}\right] + B\mathcal{X}_t + \Big(\partial_{\phi_t} A\mathbb{E}_t\left[\mathcal{Y}_{t+1}\right] + A\mathbb{E}_t\left[\partial_{\phi_t}\mathcal{Y}_{t+1}\right] + \partial_\phi B\mathcal{X}_t\Big)\phi_t, \tag{1.17}$$

where $A \equiv A(0), B \equiv B(0)$, $\partial_{\phi_t}(\cdot)$ denotes a derivative with respect to ϕ_t that is evaluated at $\phi_t = 0$. We next guess the solution to (1.17) satisfies $\mathcal{Y}_t = \mathcal{A} + \mathcal{B}\mathcal{X}_t + \mathcal{C}\mathcal{X}_t\phi_t + \mathcal{D}\phi_t$, which is straightforward to verify. The coefficients of the guess depend on the deep structural parameters of the model and can be determined via the method of undetermined coefficients. This fully describes the approximate state-dependent model dynamics with respect to systematic monetary policy ϕ_t and provides equation (1.3) in the main text, where $a = \mathcal{A}_1$, $b_s = \mathcal{B}_{11}$, $b_a = \mathcal{B}_{12}$, and analogously for $\mathcal{C}$ and $\mathcal{D}$. In the special case $\rho_s = \rho_a = \rho_\phi = 0$, the coefficients in (1.3) are given by (1.6).

1.A.2 Identification

We next describe the identification results in Section 1.2 in more detail. Using (1.2), (1.3), and the laws of motion for x_t^s and x_t^a, we obtain

$$v_{t+h}^h = F^h \cdot z_{t+h}^h,$$

where F^h is a coefficient vector and z_{t+h}^h is the following vector of variables:

$$z_{t+h}^h = \Big[x_{t-1}^s, \{\varepsilon_{t+i}^s\}_{i=1}^h, x_{t-1}^s\phi_{t+h}, \varepsilon_t^s\{\eta_{t+i}\}_{i=1}^h, \varepsilon_t^s\{\varepsilon_{t+i}^s\}_{i=1}^h, \varepsilon_t^s\{\varepsilon_{t+i}^a\}_{i=1}^h,$$
$$\{\varepsilon_{t+i}^s\phi_{t+h}\}_{i=1}^h, \{\eta_{t+i}\}_{i=1}^h, \{\varepsilon_{t+i}^a\}_{i=1}^h, x_{t+h}^a, x_{t+h}^a\phi_{t+h}\Big]',$$

where $\{\varepsilon_{t+i}^s\}_{i=1}^h$ denotes the vector of all ε_{t+i}^s for $i = 1$ through $i = h$, and analogously for all terms in braces. Defining the vector of regressors (excluding the intercept) in (1.4) by $X_t = [\varepsilon_t^s,\ \varepsilon_t^s\phi_t,\ \phi_t]'$, consistency of the OLS estimates of $(\beta^h, \gamma^h, \delta^h)$ requires

$$E[X_t(z_{t+h}^h)'] = \mathbf{0},$$

where $\mathbf{0}$ denotes a zero matrix with conforming dimension. This orthogonality condition is satisfied if $\zeta^s = \zeta^a = 0$. We next turn to the IV estimator of $(\beta^h, \gamma^h, \delta^h)$. Consider an instrument ϕ_t^{IV} with the following properties:

$$E[\phi_t^{IV}\varepsilon_{t+i}^s] = E[\phi_t^{IV}\varepsilon_{t+i}^a] = 0 \quad \forall i, \qquad E[\phi_t^{IV}\eta_t] \neq 0, \qquad E[\phi_t^{IV}\eta_{t+i}] = 0 \quad \forall i \neq 0.$$

Defining as instrument vector $Q_t = \left[\varepsilon_t^s, \varepsilon_t^s\phi_t^{IV}, \phi_t^{IV}\right]'$, consistency of the IV estimator requires

$$E[Q_t(z_{t+h}^h)'] = \mathbf{0}.$$

This condition is satisfied given the properties of the instrument.[50] Hence, the IV estimator consistently estimates $(\beta^h, \gamma^h, \delta^h)$ even absent strong exogeneity assumptions for ϕ_t.

1.B Data

1.B.1 Narrative data

We use the narrative classification from Istrefi (2019), which is a panel containing the policy preferences of voting FOMC members at each FOMC meeting for 1960-2023.
The news coverage of FOMC members is relatively sparse during the first six years in our sample, leaving us with relatively more unclassified FOMC members in this period. For example, we observe the preferences for 115 out of 195 member-meeting pairs in 1960. Fortunately, the share of observed preferences increases quickly and from 1966 onward, we reach an average share of 88 percent. Specifically for the first six years, we account for some of the missing data by assuming that the unobserved preferences coincide with the first observed preference of the respective FOMC member.
Occasionally, voting FOMC members do not attend the meetings personally, but are replaced by a substitute. We believe a plausible assumption is that short-term substitutes act in the best interest of the person that is substituted, partly because substitutes are often direct subordinates of the original voting member. More specifically, we assume that short-term substitutes act as if the original member attended the meeting if the following three criteria hold: (i) the substitution period is no longer than six months when the substitute is from the same Federal Reserve bank, (ii) the substitution period is no longer than three months if the substitute is not from the same Federal Reserve bank, (iii) the

[50] Note that $E[Q_t(z_{t+h}^h)'] = 0$ requires not only $E[\phi_t^{IV}\varepsilon_{t+i}] = 0$ but also $E[\phi_t^{IV}(\varepsilon_{t+i})^2] = 0$. However, given the assumption that ε_t and η_t are mutually independently distributed, and given the law of motion for ϕ_t, the second condition is satisfied if the first condition is satisfied.

substitution does not take place at the beginning or the end of a rotation cycle within a rotation group.[51] However, it frequently holds that the preferences of the substitute and the original voter coincide which implies that the procedure above does not change the data. We change less than 1% of preferences when a substitution occurs and our results are insensitive to these changes.

1.B.2 Macroeconomic data

We take the series for potential output (*rgdp_pott6*), real GDP (*rgdp*), nominal government spending (*ngov*), the GDP deflator (*pgdp*) and the military spending news shock (*news*) from the replication package of Ramey and Zubairy (2018). We follow their data preparation steps to create the aggregate series as in their paper.[52]

From FRED, we use headline CPI (*CPIAUCSL*) and CPI core (*CPILFESL*) inflation defined as the year-over-year growth rate of the respective price index, and the effective federal funds rate (*DFF*). The 10-year treasury market yield (*DGS10*) starts only in 1962q1 and is therefore combined with the very same variable from Romer and Romer (2010) to obtain a series that starts in 1960q1. Similarly, we use the 1-year market yield from Liu and Wu (2021) and impute the first four observations (1960q1 to 1960q4) with a similar 1-year treasury market yield from Fred (*DTB1YR*). Personal consumption expenditures (*PCE*), gross private domestic investment (*GPDI*), and federal government defense expenditures (*FDEFX*) is divided by the GDP deflator and by real potential GDP, both taken from Ramey and Zubairy (2018), see above. We compute non-military government spending by subtracting the defense spending from total government spending. Variables are averaged to quarterly frequency, if applicable.

We use inflation expectations from the Livingston survey. Our measure of inflation expectation is the annualized expected growth rate of CPI forecasts from 6 to 12 months ahead. Because the survey is biannual, we assume that inflation expectations remain constant in quarters in which no new data is available. Formally, we let $\pi_t^e = \pi_{t-1}^e$, whenever there is no survey conducted in quarter t. The (ex-ante) real rates are computed as $i_t^r = i_t^n - \pi_t^e$ where i_t^n is a nominal rate of interest.

The validation exercise in Appendix 1.D is based on forecasts from the Fed's Greenbook.

[51] For example, suppose the Chicago president had the voting right until meeting τ and the Cleveland president thereafter. If Chicago exercises the voting right in $\tau+1$ on behalf of Cleveland, we would use the preference of the Chicago president in $\tau+1$.

[52] The fiscal shock is computed as $news_t/(pgdp_{t-1} \times rgdp\_pott6_{t-1}) \times 100$. Detrended real GDP is $rgdp_t/rgdp\_pott6_t \times 100$ and detrended real government spending is $ngov_t/(pgdp_t \times rgdp\_pott6_t) \times 100$.

We use the average of the one- and two-quarter ahead inflation forecast, following Coibion and Gorodnichenko (2011). For the Blanchard and Perotti (2002) shock, we account for anticipation in government spending by including the one-quarter projected growth rate of government spending from Ramey's (2011) data.[53] We further consider as control variable the primary surplus (*svt_q*) from Cochrane (2022), seasonally adjusted via X-13 ARIMA-SEATS procedure from the U.S. Census Bureau.

1.C Hawk-Dove decompositions

We decompose fluctuations in $Hawk_t$ and $Hawk_t^{IV}$ finding that the FOMC rotation is a key source of variation for both time series.

Decomposition of $Hawk_t$. We derive a decomposition of the aggregate Hawk-Dove balance similar to the aggregate productivity decomposition in Baily, Hulten, and Campbell (1992). We first rewrite the aggregate Hawk-Dove balance in equation (1.8) as[54]

$$Hawk_t = \sum_{i \in \mathcal{M}_t} s_t Hawk_{it}, \quad s_t = \frac{1}{|\mathcal{M}_t|}. \tag{1.18}$$

We define a decomposition over p-period changes in the balance:

$$\Delta^p Hawk_t = Hawk_t - Hawk_{t-p} = \sum_{i \in \mathcal{M}_t} s_t Hawk_{it} - \sum_{i \in \mathcal{M}_{t-p}} s_{t-p} Hawk_{it-p} \tag{1.19}$$

We next partition the set $\mathcal{M}_t$ into the set of "surviving" FOMC members S_t present in $t-p$ and t, the set of entering FOMC members E_t present in t but not in $t-p$, and the set of exiting FOMC members X_t present in $t-p$ but not in t to rewrite:

[53] The SPF provides the government spending forecasts only from 1981q3 onward. Ramey (2011) imputes the government spending forecasts with defense spending forecasts to extend the sample until 1968q4.

[54] To be precise, we consider the first FOMC meeting in each quarter t.

$$
\begin{aligned}
\Delta^p Hawk_t &= \sum_{i\in S_t}(s_t Hawk_{it} - s_{t-p}Hawk_{it-p}) + \sum_{i\in E_t} s_t Hawk_{it} - \sum_{i\in X_t} s_{t-p}Hawk_{it-p} \\
&= \sum_{i\in S_t} s_{t-p}(Hawk_{it} - Hawk_{it-p}) + \sum_{i\in S_t}(s_t - s_{t-p})Hawk_{it} \\
&\quad + \sum_{i\in E_t} s_t Hawk_{it} - \sum_{i\in X_t} s_{t-p}Hawk_{it-p} \qquad (1.20)
\end{aligned}
$$

The first term captures changes in preferences of surviving FOMC members, the second term captures changes in the number of FOMC members, the third term captures entry into the FOMC, and the last term captures exit from the FOMC.
Finally, we further distinguish between the rotating and non-rotating FOMC members in the set of entering and exiting FOMC members, denoted E_t^R, E_t^N, X_t^R and X_t^N to obtain our decomposition of interest:

$$
\begin{aligned}
\Delta^p Hawk_t = &\sum_{i\in S_t} s_{t-p}(Hawk_{it} - Hawk_{it-p}) + \sum_{i\in S_t}(s_t - s_{t-p})Hawk_{it} + \sum_{i\in E_t^N} s_t Hawk_{it} \\
&- \sum_{i\in X_t^N} s_{t-p}Hawk_{it-p} + \sum_{i\in E_t^R} s_t Hawk_{it} - \sum_{i\in X_t^R} s_{t-p}Hawk_{it-p} \qquad (1.21)
\end{aligned}
$$

The third and fourth terms capture changes in the aggregate Hawk-Dove balance due to the entry and exit of rotating FOMC members, while the fifth and sixth terms capture the contribution of entry and exit of non-rotating FOMC members. The variance in yearly changes of the aggregate Hawk-Dove balance ($p = 4$) is 0.083. The variance of the first term of (1.21), which captures intensive margin changes of preferences, corresponds to 9% of the total variance. Changes in the weights, the second term, are negligible in size. The variance of the third and fourth term, capturing extensive margin changes of non-rotating FOMC members, corresponds to 22% of the total variance. The variance of the fifth and sixth term, capturing extensive margin changes of rotating FOMC members, corresponds to 53% of the total variance. Finally, the covariances between these terms account for 15% of the total variance. The results differ little for quarterly changes ($p = 1$). Notably, extensive margin changes of rotating FOMC members still account for 52% of the total variance.

Decomposition of $Hawk_t^{IV}$. Analogously, we propose a decomposition for the FOMC rotation instrument

$$\begin{aligned}\Delta^p Hawk_t^{IV} = & \sum_{i\in S_t^R} s_{t-p}^R(Hawk_{it} - Hawk_{it-p}) + \sum_{i\in S_t^R}(s_t^R - s_{t-p}^R)Hawk_{it} \\ & + \left(\sum_{i\in E_t^{RA}} s_t^R Hawk_{it} - \sum_{i\in X_t^{RA}} s_{t-p}^R Hawk_{it-p}\right) \\ & + \left(\sum_{i\in E_t^{RI}} s_t^R Hawk_{it} - \sum_{i\in X_t^{RI}} s_{t-p}^R Hawk_{it-p}\right), \end{aligned} \tag{1.22}$$

with the weights given by $s_t^R = 1/|\mathcal{R}_t|$, S_t^R the set of surviving rotating FOMC members, and distinguishing between the sets of entering rotating FOMC members whose appointments start or end in t (A), and incumbent (I) regional FRB presidents.

For yearly changes in the rotation instrument, we find that 93% of the variance is due to the rotation of incumbent members, while 7% is due to appointments starting or ending. All other variances and covariances are negligible in size. Yearly changes mechanically mute the importance of intensive margin changes, because current rotating FOMC members are typically not FOMC members a year later. Therefore, we also study quarterly changes ($p = 1$). Intensive margin changes now explain 4% of the variance, appointments account for 23%, and rotations of incumbent members account for 71%. Appointments become relatively more important for $p = 1$ because only every fourth quarter of $\Delta^1 Hawk_t^{IV}$ features a rotation. Compared to $\Delta^4 Hawk_t^{IV}$ for which the rotation affects all quarters, we mechanically lower the importance of rotations and the overall variance for $p = 1$.

1.D Validation exercise

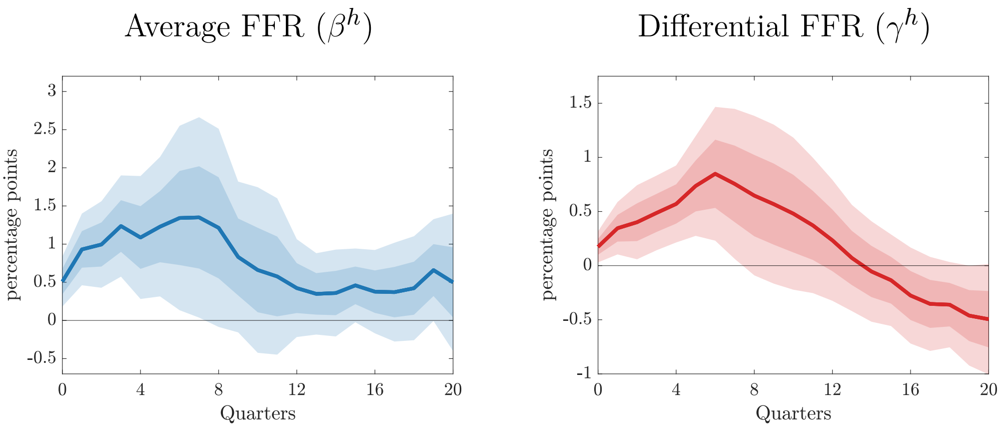
We use the Hawk-Dove balance and the FOMC rotation instrument to estimate the federal funds rate (FFR) response to inflation forecasts as a function of the hawkishness of the FOMC. We find that a hawkish FOMC is associated with a more pronounced hike of the federal funds rate in the face of inflationary pressure.

We estimate a state-dependent local projection specification that is akin to a forward-looking Taylor rule. Formally, we estimate a set of regressions

$$FFR_{t+h} = \alpha^h + \beta^h \hat{\pi}_t + \gamma^h \hat{\pi}_t (Hawk_t - \overline{Hawk}) + \zeta^h Z_{t-1} + v^h_{t+h}, \tag{1.23}$$

for $h = 0, 1, ..., H$, and FFR_{t+h} and $\hat{\pi}_t$ denote the federal funds rate and the average of the one- and two-quarter ahead Greenbook inflation forecast, respectively. The control vector includes four lags of the federal funds rate and the inflation forecast. The data is at a quarterly frequency, and the sample runs from 1969 to 2008 due to the availability of inflation forecasts and the reaching of the zero lower bound in 2008.
Figure 1.D.1 presents IV estimates where we use the FOMC rotation instrument interacted with the inflation forecast as an instrument for the interaction term in the specification above. We show estimates that are normalized to represent the inflation forecast being one percentage point above the sample average. The left panel displays the response under the average FOMC (β^h). The right panel displays the differential response (γ^h) when there are 2 more hawks in the FOMC relative to the average composition.

Figure 1.D.1: FFR response to inflation and the FOMC hawkishness



Notes: The figure shows responses of the federal funds rate to an inflation Greenbook forecast that is one percentage point above its sample average, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on (1.23). The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

On average, the FOMC reacts with a federal funds rate hike. The response is statistically significant at the five percent level for six quarters. The response builds up over time, consistent with interest rate smoothing. Incidentally, it satisfies the Taylor principle for almost two years and peaks at 1.48 percentage points. The response turns stronger when the FOMC is more hawkish, as indicated by the differential effects in Panel (b). The estimates of the interaction coefficient γ^h are hump-shaped and peak after two years at 0.92 percentage points. The response is significant at five percent for almost two years. This result suggests that a more hawkish FOMC is associated with a stronger and more persistent federal funds rate hike. Conversely, a more dovish FOMC implies a substantially weaker response.

Finally, this validation exercise lends itself to assessing the relevance condition of our instrument more formally. We use the weak instruments test from Montiel Olea and Pflueger (2013).[55] We can reject the null of weak instruments. More formally, we compute p-values for the bias exceeding 10% percent of the benchmark, see Montiel Olea and Pflueger (2013) for details. The p-values are bounded from above by 0.055 and are below the 0.05 level at most horizons. Moreover, for a test of whether the bias exceeds 20%, we can reject the null at 1% for all horizons.

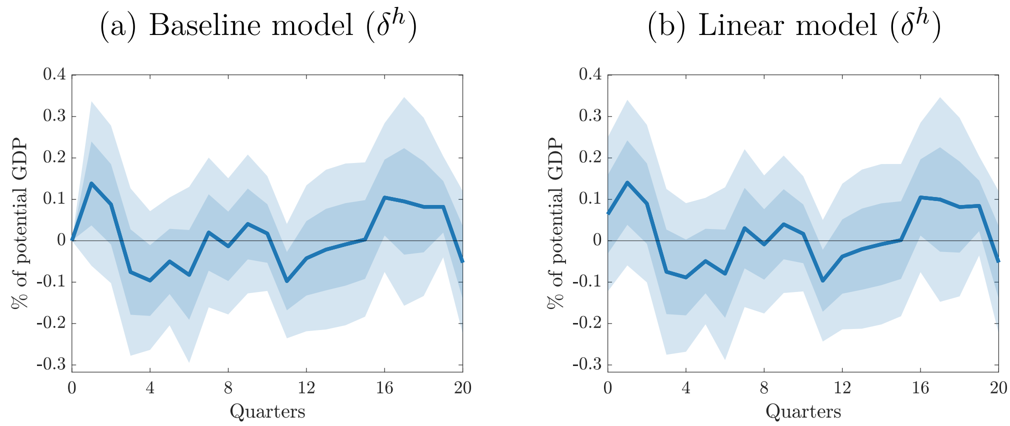
Overall, we show that the federal funds rate response to inflation correlates positively with the hawkishness of the FOMC, $Hawk_t$. The responses are consistent with our measurement of the stance of systematic monetary policy and are further in line with Bordo and Istrefi (2023). We see this result as a validation that our measurement of systematic monetary policy, through $Hawk_t$, captures important aspects of the Federal Reserve's monetary policy-making.

1.E Additional results for Section 1.4

This appendix contains additional results for Section 1.4.1-1.4.3 in the main text.

[55] With a single endogenous regressor, this is equivalent to the Lewis and Mertens (2022) test.

Figure 1.E.1: Responses of military spending shocks to systematic monetary policy



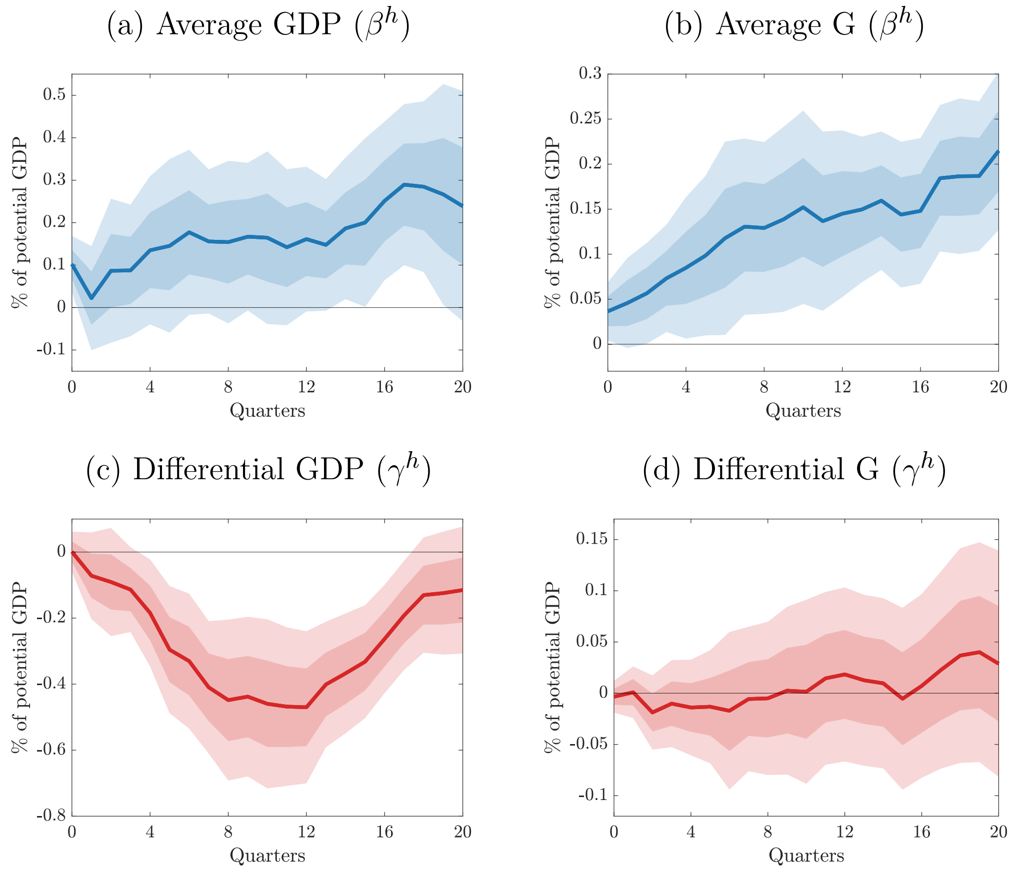
Notes: The figure shows responses of the military spending shock to systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The δ^h captures the response when $Hawk_t$ exceeds the sample average by two hawks. Panel (a) shows the results for our baseline model whereas Panel (b) shows the results when we restrict $\beta^h = \gamma^h = 0$ in the local projection (1.10). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Table 1.E.1: Testing for differences across regimes, p-values

Outcome	+2 Hawk vs. Average	+1 Hawks vs. Average	Average vs. +1 Dove	Average vs. +2 Doves
	Two-year horizon			
Multiplier	0.223	0.119	0.102	0.104
GDP (*cum*)	0.000			
G (*cum*)	0.080			
	Four-year horizon			
Multiplier	0.245	0.122	0.041	0.041
GDP (*cum*)	0.008			
G (*cum*)	0.448			

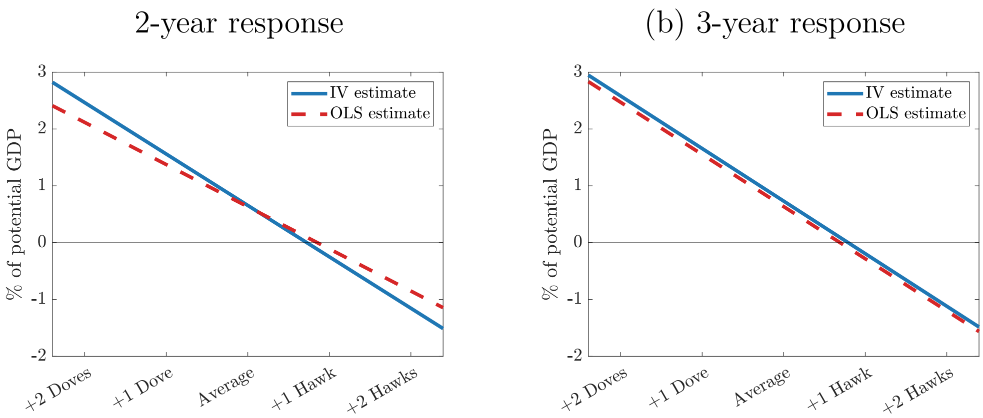
Notes: The table shows p-values corresponding to statistical tests for whether the fiscal multiplier or its components are significantly different across monetary regimes ($Hawk_t$). The tests are based on the multiplier estimates reported in Table 1.2 in Section 1.4.3, using Driscoll-Kraay standard errors, see Appendix 1.F for details.

Figure 1.E.2: Responses of GDP and government spending, OLS



Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show OLS estimates based on the local projection framework (1.10) as specified in Section 1.4.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 1.E.3: GDP responses for OLS and IV



Notes: The figure shows the yearly real GDP response to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV and OLS estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The displayed estimates are computed as $\sum_{h=H-3}^{H}[\beta^h + \gamma^h(Hawk_t - \overline{Hawk_t})]$ for H =8 quarters in Panel (a) and $H = 12$ quarters in Panel (b).

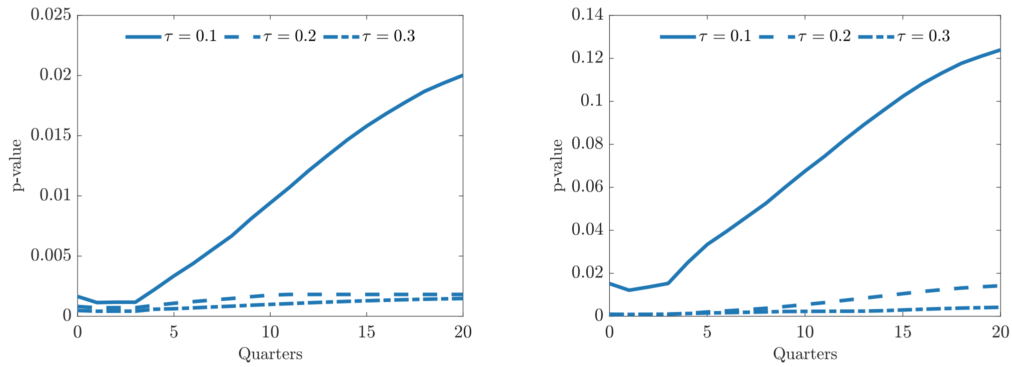
1.F Weak instruments and robust inference

This appendix contains additional results for Section 1.4.4. The first subsection presents diagnostics on instrument strength. The second section presents robust inference regarding weak instruments for impulse responses and fiscal multipliers.

1.F.1 Weak instrument tests

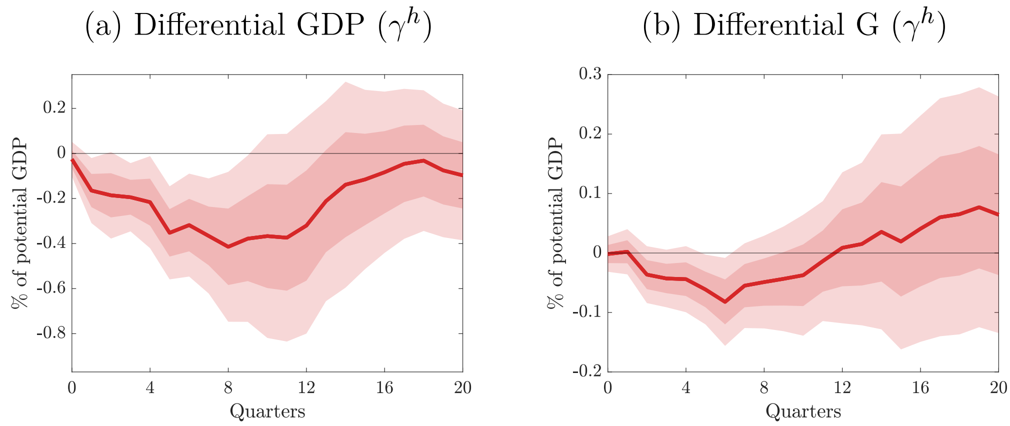
Figure 1.F.1: Weak instrument tests

(a) Montiel Olea and Pflueger (2013) (b) Lewis and Mertens (2022)



Notes: The figure shows p-values for rejecting the null of weak instruments for the responses of real government spending (G), based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The Montiel Olea and Pflueger (2013) test evaluates the null of the bias in γ^h exceeding a threshold τ. Similarly, the Lewis and Mertens (2022) test evaluates the null of the ℓ^2 norm of the bias in γ^h and δ^h exceeding a threshold τ. For the former, the endogenous regressor $Hawk_t$ is not tested but directly replaced by its first stage fitted value. The critical values and associated p-values are based on Newey-West standard errors.

Figure 1.F.2: Differential responses of GDP and government spending, reduced-form



Notes: The figure shows differential responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show reduced-form estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. Moreover, testing whether γ^h is statistically significant from zero is equivalent to testing for zero relevance of the instrument, as explained in the main text. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Table 1.F.1: Responses of GDP and government spending, incl. first-stage

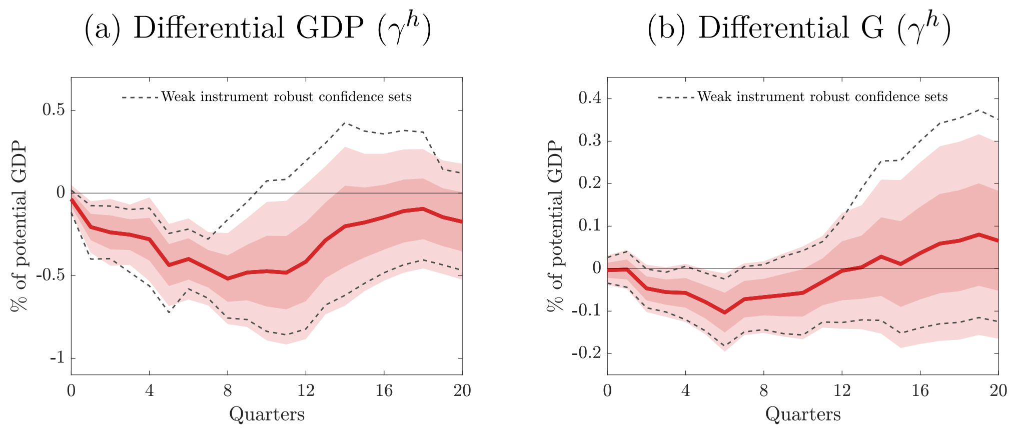
	GDP responses				**G responses**				**First-stage results**	
Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ε_t^s	0.142	0.166	0.185	0.283	0.092	0.140	0.157	0.152	0.050	0.010
	(0.096)	(0.095)	(0.085)	(0.130)	(0.047)	(0.056)	(0.051)	(0.054)	(0.039)	(0.007)
$\varepsilon_t^s(Hawk_t - \overline{Hawk_t})$	-1.672	-3.099	-2.485	-0.873	-0.342	-0.401	-0.030	0.220		
	(0.775)	(0.841)	(1.433)	(1.174)	(0.209)	(0.258)	(0.416)	(0.653)		
$Hawk_t - \overline{Hawk_t}$	-2.770	-3.698	-4.247	-4.562	-0.593	-0.985	-1.389	-0.948		
	(1.220)	(1.728)	(2.216)	(2.217)	(0.322)	(0.650)	(1.020)	(1.135)		
$\varepsilon_t^s(Hawk_t^{IV} - \overline{Hawk_t}^{IV})$									0.290	-0.019
									(0.053)	(0.021)
$Hawk_t^{IV} - \overline{Hawk_t}^{IV}$									-0.008	0.402
									(0.017)	(0.042)
ε_{t-1}^s	0.024	0.057	0.086	0.245	0.044	0.076	0.092	0.124	0.007	0.011
	(0.157)	(0.216)	(0.221)	(0.153)	(0.033)	(0.046)	(0.043)	(0.043)	(0.003)	(0.006)
ε_{t-2}^s	0.110	0.035	0.078	0.150	0.032	0.052	0.063	0.092	-0.012	0.007
	(0.125)	(0.185)	(0.205)	(0.160)	(0.030)	(0.030)	(0.041)	(0.049)	(0.011)	(0.008)
ε_{t-3}^s	0.045	0.036	0.126	0.188	0.038	0.036	0.037	0.073	-0.000	0.008
	(0.149)	(0.163)	(0.153)	(0.144)	(0.018)	(0.028)	(0.045)	(0.052)	(0.006)	(0.008)
ε_{t-4}^s	0.001	0.033	0.152	0.224	0.023	0.037	0.060	0.139	-0.018	0.004
	(0.141)	(0.125)	(0.117)	(0.144)	(0.022)	(0.027)	(0.041)	(0.038)	(0.012)	(0.010)
GDP_{t-1}	1.314	0.777	0.424	0.037	0.033	0.103	0.135	0.124	-0.000	-0.012
	(0.182)	(0.243)	(0.252)	(0.282)	(0.053)	(0.075)	(0.100)	(0.121)	(0.013)	(0.017)
GDP_{t-2}	-0.406	-0.166	-0.110	0.149	0.006	0.060	0.035	0.039	-0.016	0.013
	(0.190)	(0.209)	(0.159)	(0.197)	(0.054)	(0.072)	(0.077)	(0.094)	(0.020)	(0.014)
GDP_{t-3}	-0.240	-0.012	-0.093	0.081	0.062	0.034	0.044	0.084	0.004	-0.005
	(0.203)	(0.180)	(0.223)	(0.171)	(0.055)	(0.068)	(0.068)	(0.062)	(0.016)	(0.010)
GDP_{t-4}	-0.164	-0.440	-0.284	-0.444	-0.103	-0.218	-0.279	-0.355	0.003	-0.026
	(0.183)	(0.267)	(0.313)	(0.333)	(0.051)	(0.095)	(0.138)	(0.167)	(0.007)	(0.011)
G_{t-1}	-0.639	0.012	0.336	0.864	1.340	1.311	1.121	1.138	0.028	-0.073
	(0.714)	(1.012)	(0.909)	(0.940)	(0.195)	(0.241)	(0.308)	(0.387)	(0.022)	(0.055)
G_{t-2}	1.177	0.596	0.194	-0.223	0.008	-0.042	0.078	0.097	0.008	0.062
	(0.617)	(0.734)	(0.602)	(0.479)	(0.195)	(0.220)	(0.277)	(0.278)	(0.039)	(0.047)
G_{t-3}	-0.391	-0.347	-0.233	-0.519	-0.079	-0.106	-0.136	-0.138	0.017	-0.091
	(0.651)	(0.706)	(0.618)	(0.491)	(0.203)	(0.248)	(0.273)	(0.272)	(0.054)	(0.042)
G_{t-4}	0.022	0.020	0.018	0.162	-0.346	-0.308	-0.268	-0.343	-0.039	0.140
	(0.920)	(0.911)	(0.888)	(0.791)	(0.202)	(0.314)	(0.437)	(0.486)	(0.049)	(0.060)
Observations	196	196	196	196	196	196	196	196	196	196
R^2	0.577	0.347	0.201	0.138	0.934	0.843	0.730	0.646	0.452	0.547
R^2 excl. IVs									0.036	0.154
F-statistic	16.398	4.243	3.418	2.630	94.688	22.683	11.316	15.287	43.691	28.077
F-statistic excl. IVs									4.935	5.804

Notes: The table shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. Columns (1) to (4) and (5) to (8) display the one, two, three, and four-year ahead responses, respectively. Regressor ε_t^s captures the responses when $Hawk_t$ equals its sample average and $\varepsilon_t^s(Hawk_t - \overline{Hawk_t})$ captures the differential responses. Columns (9) and (10) display the first-stage results for $\varepsilon_t^s(Hawk_t - \overline{Hawk_t})$ and $(Hawk_t - \overline{Hawk_t})$, respectively. Newey-West standard errors are in parenthesis.

1.F.2 Robust inference

Differential responses. We compute robust inference for the differential GDP and government spending effects based on Andrews (2018).

Figure 1.F.3: Responses of GDP and government spending, robust inference



Notes: The figure shows differential responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. The dashed bands provide 95% confidence sets, robust to weak identification based on Andrews (2018), constructed via the refined projection method from Chaudhuri and Zivot (2011).

Baseline multiplier inference. To obtain multiplier estimates and conduct inference about them, we first estimate the responses of cumulative GDP and cumulative government spending (G). Formally, we estimate

$$\tilde{x}_t = \tilde{\alpha}_x + \tilde{\beta}_x \varepsilon_t^s + \tilde{\gamma}_x \varepsilon_t^s (Hawk_t - \overline{Hawk}) + \tilde{\delta}_x (Hawk_t - \overline{Hawk}) + \tilde{\zeta}_x Z_{t-1} + \tilde{v}_{t+j}, \qquad (1.24)$$

where $\tilde{x}_t$ is either cumulative GDP ($\tilde{x}_t = \sum_{h=0}^{H} GDP_{t+h}$) or cumulative G ($\tilde{x}_t = \sum_{h=0}^{H} G_{t+h}$). This yields estimates $\tilde{\beta}_{\text{GDP}} = \sum_{h=0}^{H} \beta_{\text{GDP}}^h$, $\tilde{\beta}_{\text{G}} = \sum_{h=0}^{H} \beta_{\text{G}}^h$, with β_{GDP}^h and β_{G}^h being the coefficients in (1.10). The coefficients $\tilde{\alpha}_x$, $\tilde{\gamma}_x$, $\tilde{\delta}_x$, $\tilde{\zeta}_x$ are analogously related to (1.10). These estimates allow us to estimate the fiscal multiplier in (1.12).

To obtain a covariance matrix for the IV estimates $\hat{\vartheta} = (\tilde{\beta}_{\text{GDP}}, \tilde{\beta}_{\text{G}}, \tilde{\gamma}_{\text{GDP}}, \tilde{\gamma}_{\text{G}})'$, we estimate the two regressions (i.e., for GDP and G) jointly via seemingly unrelated regressions. For our baseline inference, we use the Driscoll and Kraay (1998) covariance estimator, allowing for serial correlation and cross-correlation between GDP and G. We use the covariance matrix to compute standard errors for the fiscal multiplier by applying the Delta method to the fiscal multiplier in (1.12).

Anderson-Rubin multiplier inference. We construct robust confidence sets for the fiscal multiplier by inverting an Anderson and Rubin (1949) test (AR henceforth) following Andrews, Stock, and Sun (2019). We build the test based on two sets of regressions. First, consider the reduced-form regressions

$$\tilde{x}_t = \tilde{\alpha}_x^{rf} + \tilde{\beta}_x^{rf} \varepsilon_t^s + \tilde{\gamma}_x^{rf} \varepsilon_t^s (Hawk_t^{IV} - \overline{Hawk}^{IV}) + \tilde{\delta}_x^{rf} (Hawk_t^{IV} - \overline{Hawk}^{IV}) + \tilde{\zeta}_x^{rf} Z_{t-1} + \tilde{v}_{t+j}^{rf}, \tag{1.25}$$

and ϱ denotes the OLS estimator of parameters $(\tilde{\beta}_{\text{GDP}}^{rf}, \tilde{\gamma}_{\text{GDP}}^{rf}, \tilde{\beta}_{\text{G}}^{rf}, \tilde{\gamma}_{\text{G}}^{rf})'$. Second, consider the first-stage regressions

$$\begin{aligned} \varepsilon_t^s &= \tilde{\alpha}^{fs1} + \tilde{\beta}^{fs1} \varepsilon_t^s + \tilde{\gamma}^{fs1} \varepsilon_t^s (Hawk_t^{IV} - \overline{Hawk}^{IV}) \\ &\quad + \tilde{\delta}^{fs1} (Hawk_t^{IV} - \overline{Hawk}^{IV}) + \tilde{\zeta}_x^{fs1} Z_{t-1} + \tilde{v}_{t+j}^{fs1}, \qquad (1.26) \\ \varepsilon_t^s (Hawk_t - \overline{Hawk}) &= \tilde{\alpha}^{fs2} + \tilde{\beta}^{fs2} \varepsilon_t^s + \tilde{\gamma}^{fs2} \varepsilon_t^s (Hawk_t^{IV} - \overline{Hawk}^{IV}) \\ &\quad + \tilde{\delta}^{fs2} (Hawk_t^{IV} - \overline{Hawk}^{IV}) + \tilde{\zeta}_x^{fs2} Z_{t-1} + \tilde{v}_{t+j}^{fs2}, \qquad (1.27) \end{aligned}$$

and π denotes the OLS estimator of the 2×2 parameter matrix $((\tilde{\beta}^{fs1}, \tilde{\gamma}^{fs1})', (\tilde{\beta}^{fs2}, \tilde{\gamma}^{fs2})')$. We further define $\Pi = I_2 \otimes \pi$ with I_2 the 2×2 identity matrix, which corresponds to the OLS estimators of the stacked first stage regressions for GDP and G. The AR statistic builds on the identity $\varrho = \Pi \vartheta$ where ϑ is the IV estimator of the coefficients of interest, see Andrews, Stock, and Sun (2019). The test statistic for H_0: $\vartheta = \vartheta_0$ is given by

$$AR(\vartheta_0) = \hat{g}(\vartheta_0)' \; \hat{\Omega}(\vartheta_0)^{-1} \; \hat{g}(\vartheta_0), \tag{1.28}$$

$$\text{with} \qquad \hat{g}(\vartheta_0) = \hat{\varrho} - \hat{\Pi} \vartheta_0, \tag{1.29}$$

$$\text{and} \qquad \hat{\Omega}(\vartheta_0) = \hat{\mathbb{E}}\left[\varrho \; \varrho'\right] - \hat{\mathbb{E}}\left[\varrho \; \vartheta_0' \; \Pi'\right] - \hat{\mathbb{E}}\left[\Pi \; \vartheta_0 \; \varrho'\right] + \hat{\mathbb{E}}\left[\Pi \; \vartheta_0 \; \vartheta_0' \; \Pi'\right], \tag{1.30}$$

where hats denote respective estimates. We estimate all covariance terms in $\hat{\Omega}(\vartheta_0)$ accounting for cross-correlations between estimators as well as for serial correlation using the Driscoll-Kraay covariance estimator. Under weak assumptions, it holds that $AR(\vartheta_0) \xrightarrow{d} \chi^2(4)$, since ϑ is 4×1, see Andrews, Stock, and Sun (2019). This holds regardless of the strength of the instrument and regardless of whether the denominator of the fiscal multiplier is zero. We compute the AR confidence set $CS^{FM}(\chi)$ for the fiscal multiplier $FM^H(\chi)$ from equation (1.12) by inverting the AR test. This requires four steps.

1. Define set Θ that contains the confidence region of $FM^H(\chi)$.

2. Define discrete set $\Theta_N \subset \Theta$ that contains N vectors of ϑ.

3. Construct the set $CS^{\vartheta} = \left\{\vartheta \in \Theta_N \mid AR(\vartheta) \leq c_{1-\alpha,\chi^2(4)}\right\}$.

4. Compute the confidence set for the fiscal multiplier as

$$CS^{FM}(\chi) = \left\{FM \;\middle|\; FM = \frac{\tilde{\beta}_{\text{GDP}} + \chi\ \tilde{\gamma}_{\text{GDP}}}{\tilde{\beta}_{\text{G}} + \chi\ \tilde{\gamma}_{\text{G}}},\ \forall\ (\tilde{\beta}_{\text{GDP}},\ \tilde{\gamma}_{\text{GDP}},\ \tilde{\beta}_{\text{G}},\ \tilde{\gamma}_{\text{G}})' = \vartheta \in CS^{\vartheta}\right\}.$$

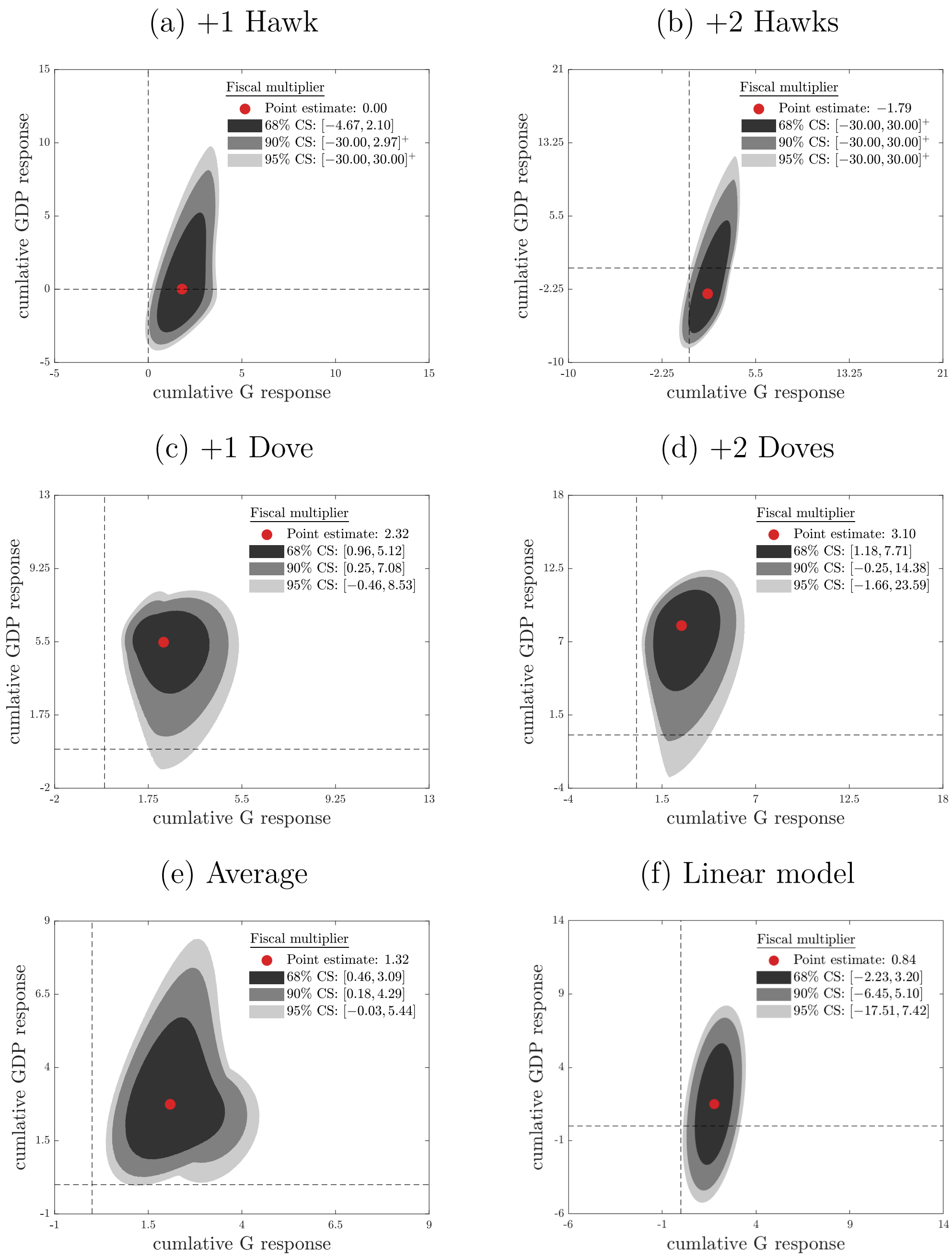
Note that $c_{1-\alpha,\chi^2(4)}$ is the $1-\alpha$ quantile of a χ^2 distribution with four degrees of freedom. We implement step 1 by choosing a closed interval for each entry of the vector ϑ. The set Θ is then defined by the Cartesian product of the four closed intervals. Specifically for entry i of ϑ, which we denote by ϑ_i, we use the interval $[-1.5\ \hat{\vartheta}_i,\ 3.5\ \hat{\vartheta}_i]$, when $\hat{\vartheta}_i > 0$, and $[3.5\ \hat{\vartheta}_i,\ -1.5\ \hat{\vartheta}_i]$ when $\hat{\vartheta}_i < 0$, where $\hat{\vartheta}_i$ denotes the IV estimate, based on (1.24). We verify that the chosen intervals are not binding in the sense that the upper or lower bound of CS^{θ} is not the boundary of Θ.[56] For step 2, we define Θ_N based on a Sobol sequence of length $N = 2,000,000,000$. Finally, we have verified that increasing or decreasing N by 5% does not affect our results.

1.G Sensitivity analysis

This appendix contains the results of our sensitivity analysis in Section 1.4.5.

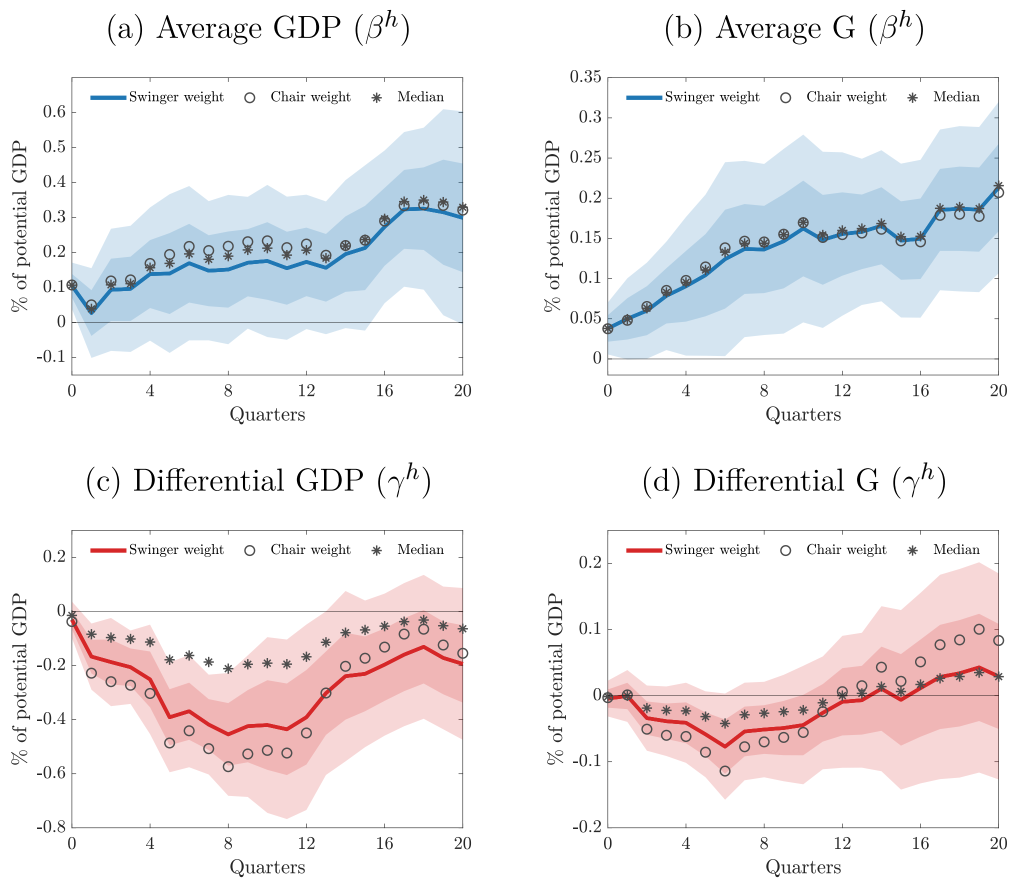
[56] For the multiplier in the linear model, we require a larger set Θ with $[-4\ \hat{\vartheta}_i,\ 10\ \hat{\vartheta}_i]$ if $\hat{\vartheta}_i > 0$ and analogously if $\hat{\vartheta}_i < 0$.

Figure 1.F.4: Anderson-Rubin confidence sets for four-year fiscal multipliers



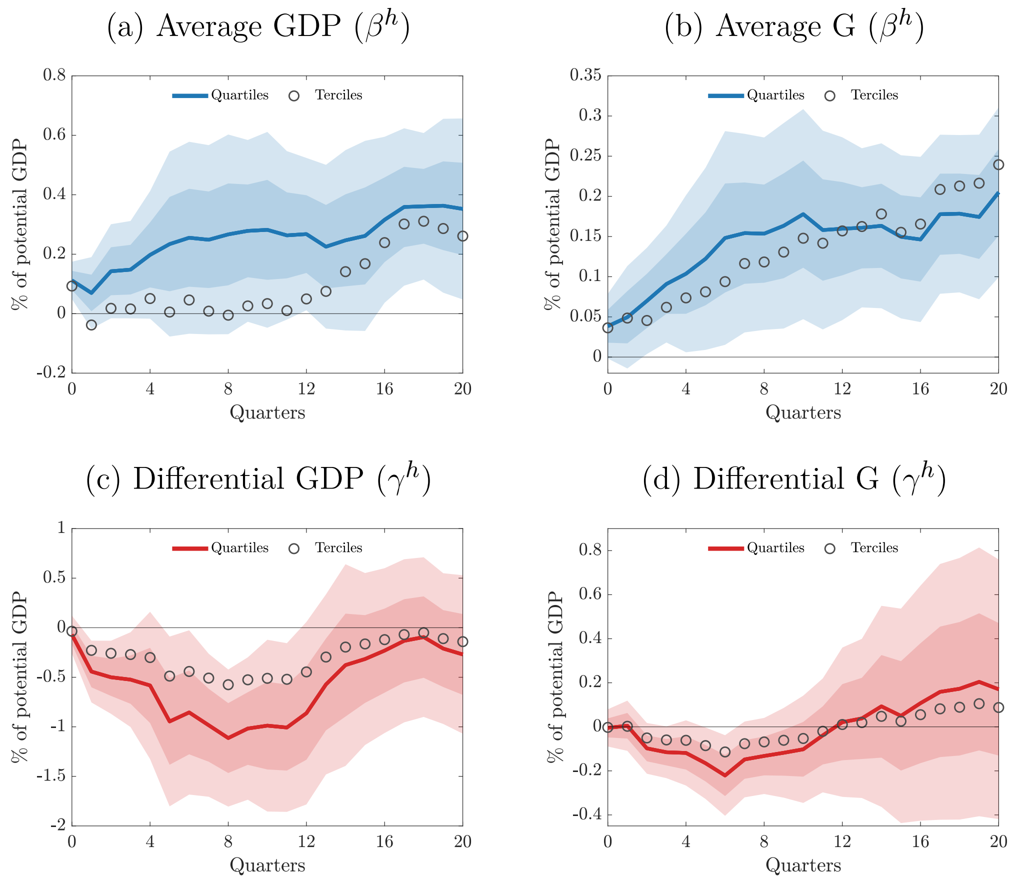
Notes: This figure shows Anderson-Rubin type confidence sets for the cumulative four-year fiscal multiplier. We depict the numerator and denominator of the multiplier on the vertical and horizontal axis, respectively. The shaded areas depict the confidence sets and various levels of significance. The red circle is the baseline point estimate from Table 1.2. The dashed lines indicate the zero values on each axis, respectively. The confidence sets reported in the legend are defined by the minimum and maximum fiscal multiplier that is contained in the respective confidence set, capped at ±30 for readability. Panels (a)-(d) correspond to the fiscal multipliers when $Hawk_t$ exceeds the sample average by either one or two hawks or doves. Panel (e) corresponds to the fiscal multiplier when $Hawk_t$ equals its sample average. Panel (f) corresponds to the fiscal multiplier estimate when we restrict $\tilde{\gamma}_{\mathrm{GDP}} = \tilde{\delta}_{\mathrm{GDP}} = \tilde{\gamma}_{\mathrm{G}} = \tilde{\delta}_{\mathrm{G}} = 0$.

Figure 1.G.1: Responses of GDP and government spending, aggregation schemes



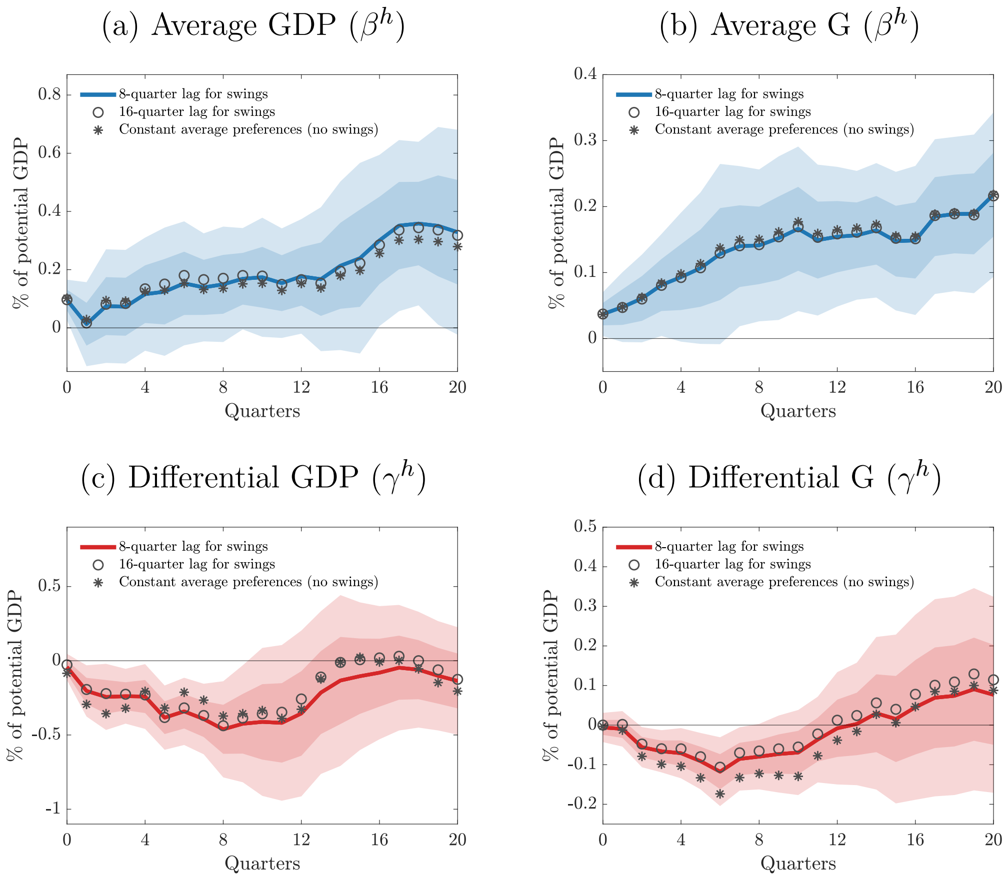
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We use three variants of $Hawk_t$. Swinger weight: We do not discriminate between swingers and consistent members. Chair weight: We assign the preferences of the Fed Chair twice the weight of an ordinary member when aggregating to $Hawk_t$. Median: We aggregate the cross-section of FOMC members by the median, instead of the arithmetic average.

Figure 1.G.2: Responses of GDP and government spending, discrete Hawk-Dove balance



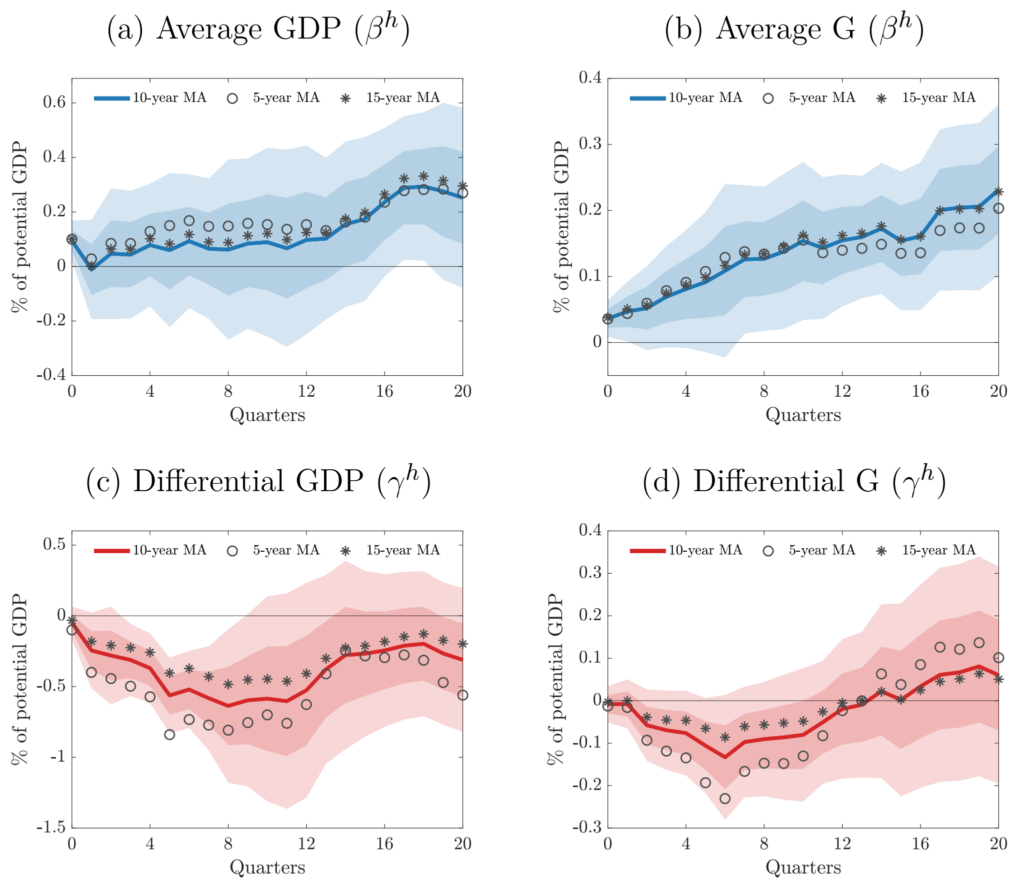
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We use two discrete variants of $Hawk_t$. We define that the discrete $Hawk_t$ equals -1 if $Hawk_t$ falls below the first quartile or tertile of the distribution of $Hawk_t$ over time, +1 if above the highest quartile or tertile, and zero else.

Figure 1.G.3: Responses of GDP and government spending, alternative IVs



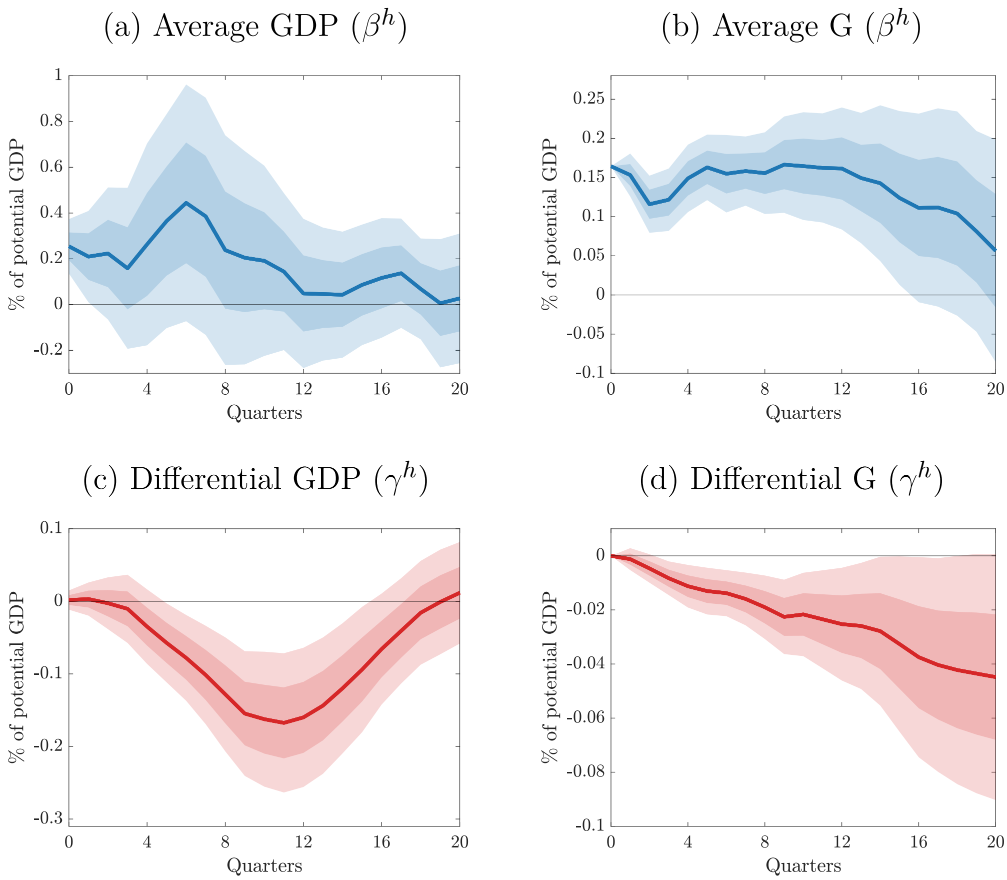
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We use an alternative definition of the instrumental variable $Hawk_t^{IV}$ where swings affect the individual preference only 8 or 16 quarters after the date of the swing, or where no swing occurs because we set the individual preference to the average, rendering them time-invariant.

Figure 1.G.4: Responses of GDP and government spending, accounting for trends



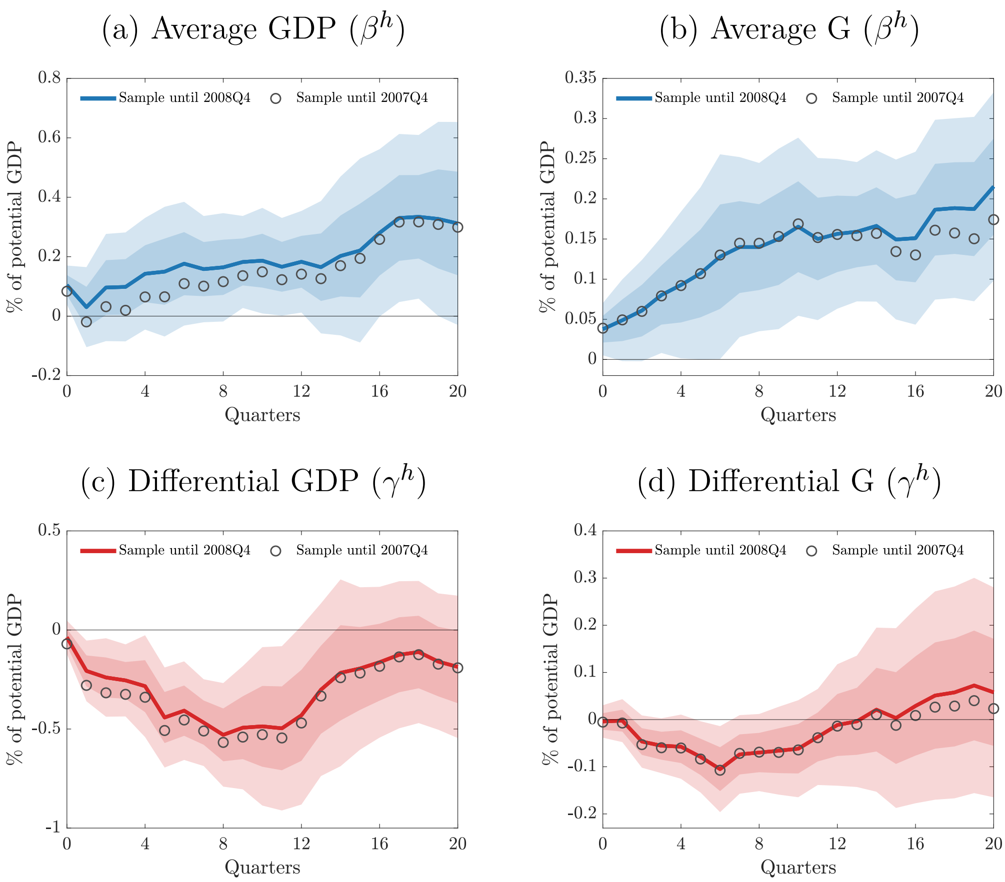
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We use three variants of $Hawk_t$ where we subtract the backward-looking 5, 10, or 15-year moving average from $Hawk_t$ prior estimation.

Figure 1.G.5: Responses of GDP and government spending, Blanchard-Perotti shock



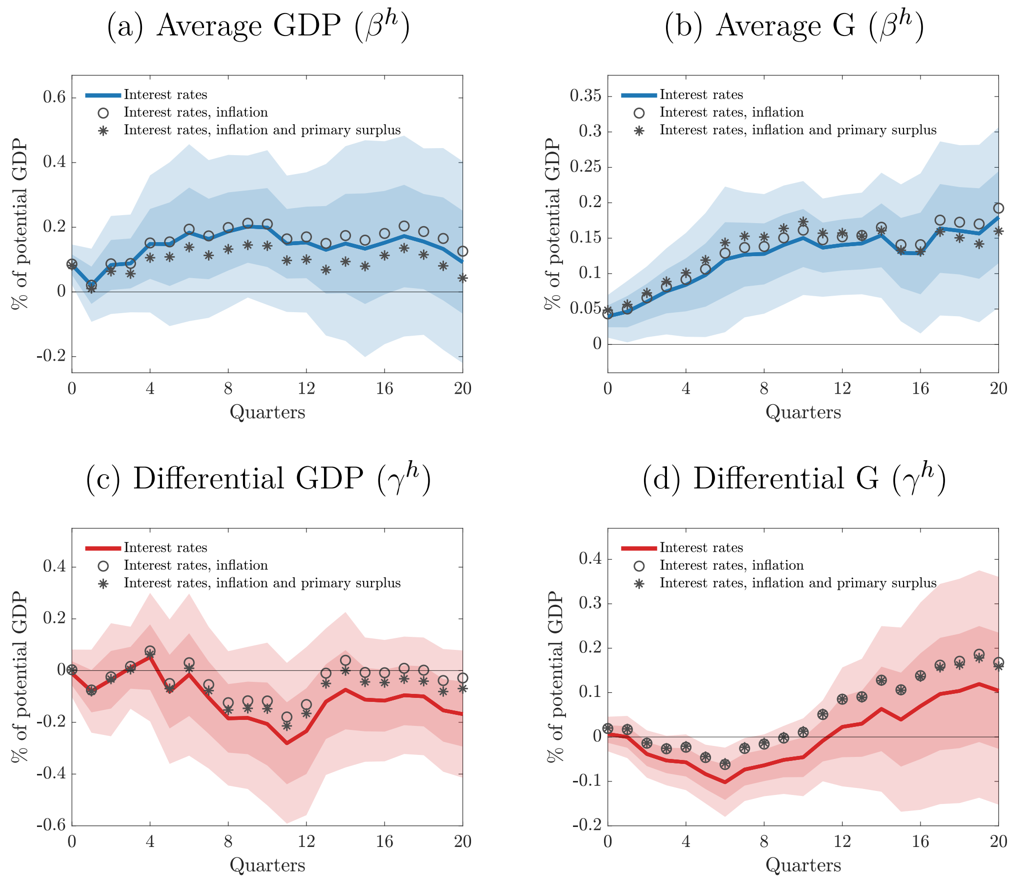
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. The shock is contemporaneous G, conditional on controls that include four lags of real GDP and real government spending, as well as the projected growth rate of real government spending. The projected growth rate is taken from the Survey of Professional Forecasters and is available from 1969 onward, which is the start of our sample, see Appendix 1.B.

Figure 1.G.6: Responses of GDP and government spending, accounting for the ZLB



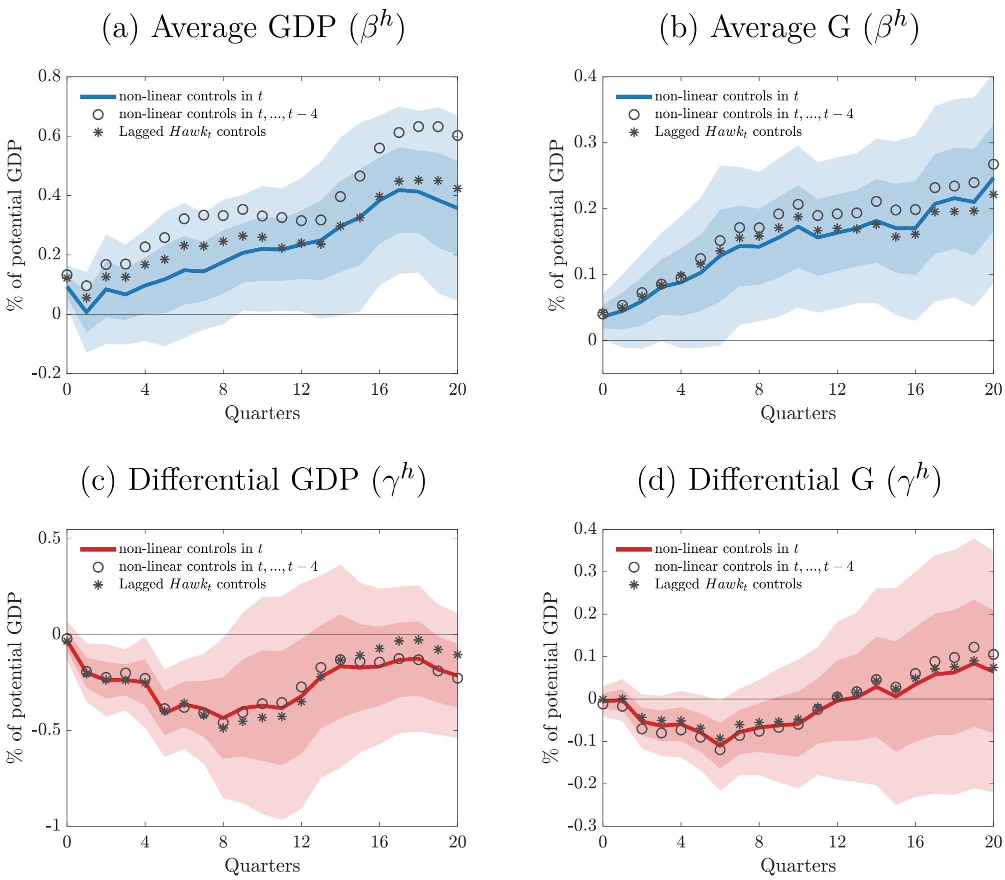
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We use a sub-sample that ends either in 2008Q4 or 2007Q4 to exclude the ZLB, or both the ZLB and the Great Recession.

Figure 1.G.7: Responses of GDP and government spending, additional controls



Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. The different specifications augment the control vector Z_{t-1} gradually by four lags of treasury yields with 1-year and 10-year maturity, the fed funds rate (interest rates), CPI inflation, and the primary surplus from Cochrane (2022).

Figure 1.G.8: Responses of GDP and government spending, non-linear controls



Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The γ^h captures the differential responses when $Hawk_t$ exceeds the sample average by two hawks. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. Non-linear controls in t: Controls Z_{t-1} include four lags of ε_t^s, real GDP and real government spending, both divided by potential GDP in all specifications. All controls are in levels, as well as interacted with $Hawk_t$, and instrumented accordingly. Non-linear controls in $t, ..., t-4$: Augments the control vector by also including and instrumenting lagged interaction terms, i.e. $Hawk_{t-i} \times C_{t-i}$ with $i = 1, ..., 4$ and C_t referring to G, GDP, and ε_t^s. Lagged $Hawk_t$ controls: Baseline controls augmented by four lags of $Hawk_t$ in levels, and instrumented accordingly.

Table 1.G.1: Cumulative 4-year government spending multipliers, Robustness

	Multipliers across regimes			p-values for differences across regimes		
Specification	+2 Hawks	Average	+2 Doves	+2 Hawks vs. +2 Doves	+2 Hawks vs. Average	Average vs. +2 Doves
Baseline	-1.786 (2.636)	1.315 (0.478)	3.105 (1.167)	0.122	0.245	0.041
BP shock	0.849 (1.068)	1.342 (0.839)	1.730 (0.699)	0.077	0.091	0.062
Aggregation schemes						
Median	0.426 (0.569)	1.419 (0.546)	2.232 (0.816)	0.043	0.065	0.036
Chair weight	-1.671 (2.222)	1.538 (0.664)	3.468 (1.518)	0.070	0.175	0.078
Swinger weight	-1.597 (2.168)	1.267 (0.554)	3.046 (1.144)	0.090	0.180	0.043
Accounting for trends						
5-year MA	-11.458 (59.772)	1.290 (1.143)	3.770 (2.106)	0.801	0.829	0.210
10-year MA	-4.716 (10.012)	0.844 (0.868)	3.238 (1.224)	0.439	0.553	0.085
15-year MA	-2.093 (3.051)	0.987 (0.430)	2.977 (0.884)	0.137	0.279	0.035

(Table continues on the next page)

Table 1.G.1 (continued): Cumulative 4-year government spending multipliers, Robustness

	Multipliers across regimes			p-values for differences across regimes		
Specification	+2 Hawks	Average	+2 Doves	+2 Hawks vs. +2 Doves	+2 Hawks vs. Average	Average vs. +2 Doves
Accounting for swings in the IV						
8-quarter lag	-1.534 (2.977)	1.220 (0.541)	2.622 (1.069)	0.233	0.345	0.071
16-quarter lag	-0.728 (2.675)	1.239 (0.599)	2.560 (1.455)	0.338	0.449	0.239
Average preferences	-1.556 (4.946)	1.070 (0.656)	1.872 (1.224)	0.549	0.580	0.494
Accounting for the ZLB						
End sample '08	-1.999 (2.672)	1.306 (0.513)	3.099 (1.138)	0.107	0.225	0.032
End sample '07	-3.378 (4.513)	0.922 (0.531)	3.016 (1.137)	0.204	0.344	0.031
Additional controls						
Interest rates	0.390 (1.269)	1.258 (0.690)	1.861 (0.760)	0.301	0.322	0.351
Interest rates, inflation	0.738 (0.871)	1.260 (0.646)	2.055 (1.046)	0.327	0.306	0.380
Interest rates, inflation, surplus	0.654 (1.107)	1.324 (0.855)	2.210 (1.437)	0.373	0.328	0.463

(Table continues on the next page)

Table 1.G.1 (continued): Cumulative 4-year government spending multipliers, Robustness

	Multipliers across regimes			p-values for differences across regimes		
Specification	+2 Hawks	Average	+2 Doves	+2 Hawks vs. +2 Doves	+2 Hawks vs. Average	Average vs. +2 Doves
Non-linear controls						
in t	-1.019 (2.642)	1.436 (0.553)	2.830 (1.129)	0.206	0.356	0.067
in $t,...,t-4$	0.371 (2.293)	2.026 (0.646)	3.033 (1.157)	0.344	0.486	0.141
Lagged $Hawk_t$	-0.575 (1.734)	1.659 (0.549)	3.216 (1.213)	0.118	0.230	0.070

Notes: The table shows IV estimates of the cumulative fiscal spending multipliers $FM^H(\chi)$ in equation (1.12) for $H=16$ quarters. The last three columns show p-values corresponding to statistical tests for whether the fiscal multiplier is significantly different across monetary regimes ($Hawk_t$). The baseline coefficients are estimated using a cumulative version of the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The columns present different states of the Hawk-Dove balance between "+2 Hawks" ($\chi=+2/12$), "Average" ($\chi=0$), and "+2 Doves" ($\chi=-2/12$). Driscoll-Kraay standard errors are in parenthesis, see Appendix 1.F for details. The various exercises correspond to the impulse responses presented in Figures 1.G.1-1.G.8, see the respective figure notes for details.

Table 1.G.2: Cumulative 4-year government spending multipliers, Discrete Hawk-Dove balance

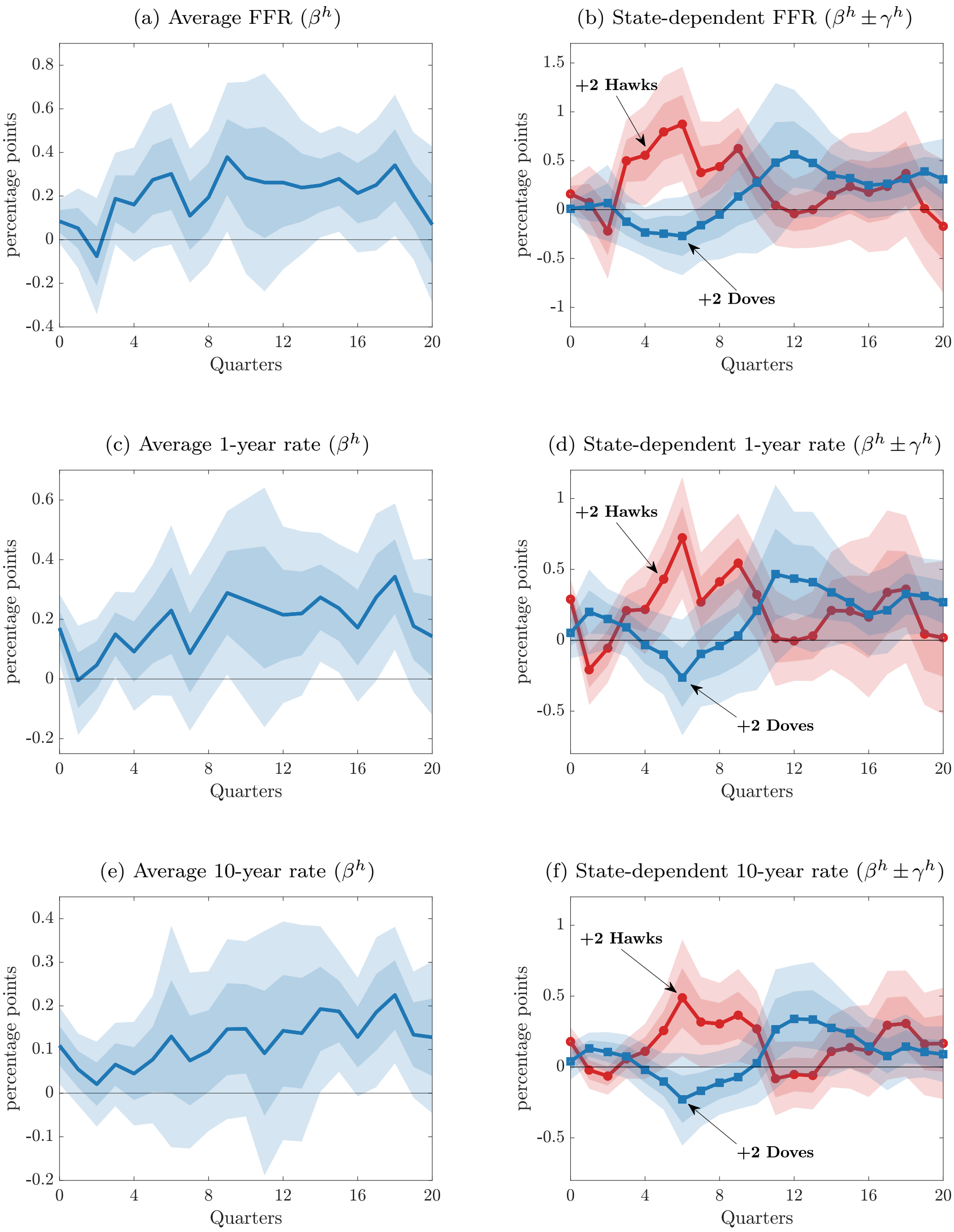
	Multipliers across regimes			p-values for differences across regimes		
Specification	Hawkish	Average	Dovish	Hawkish vs. Dovish	Hawkish vs. Average	Average vs. Dovish
Quartiles	-6.002 (10.343)	1.727 (0.775)	4.814 (2.774)	0.264	0.460	0.201
Tertiles	-3.481 (6.227)	0.490 (0.772)	2.835 (1.083)	0.336	0.488	0.047

Notes: The table shows IV estimates of the cumulative fiscal spending multipliers $FM^H(\chi)$ in equation (1.12) for $H =$ 16 quarters. The last three columns show p-values corresponding to statistical tests for whether the fiscal multiplier is significantly different across monetary regimes ($Hawk_t$). The coefficients are estimated using a cumulative version of the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. We use two discrete variants of $Hawk_t$. We define that the discrete $Hawk_t$ equals -1 if $Hawk_t$ falls below the first quartile or tertile of the distribution of $Hawk_t$ over time, +1 if above the highest quartile or tertile, and zero else. The columns present different states of the Hawk-Dove balance between "Hawkish" (χ within the last quartile or tertile), "Average" (χ between the first and last quartile or tertile) "Dovish" (χ within the first quartile or tertile).

1.H Additional results for Section 1.5

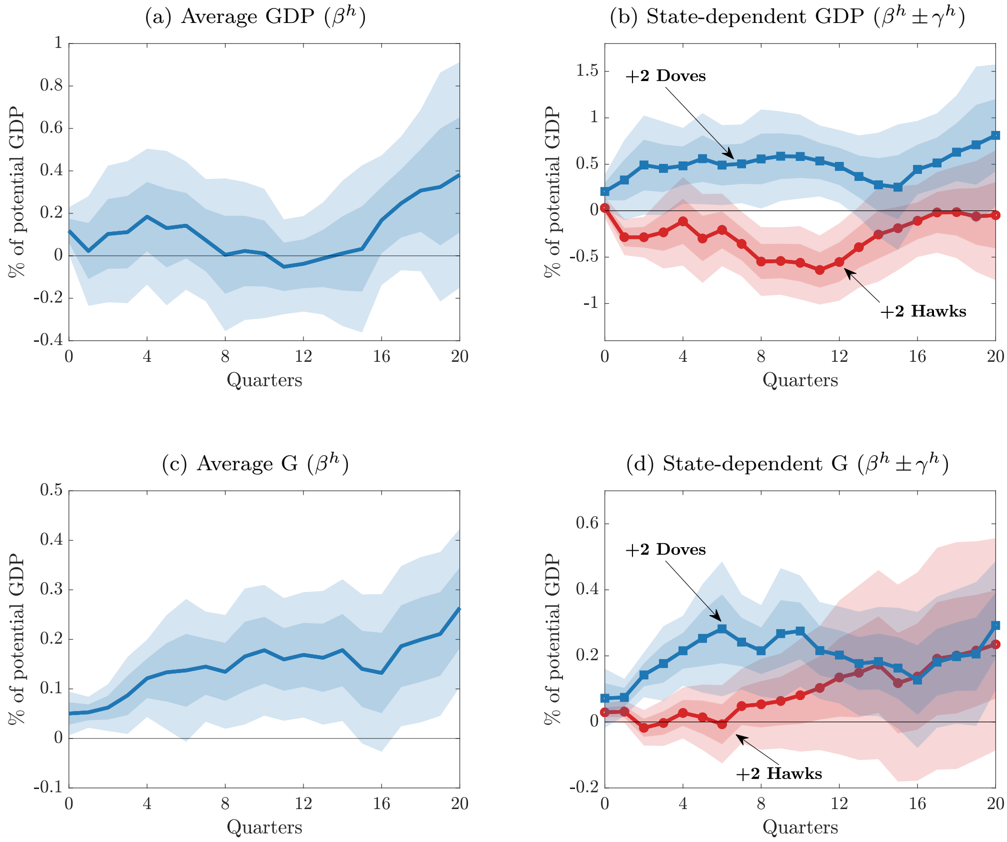
This appendix contains additional findings discussed in the main text.

Figure 1.H.1: Responses of nominal interest rates, omit shocks at end of rotation cycle



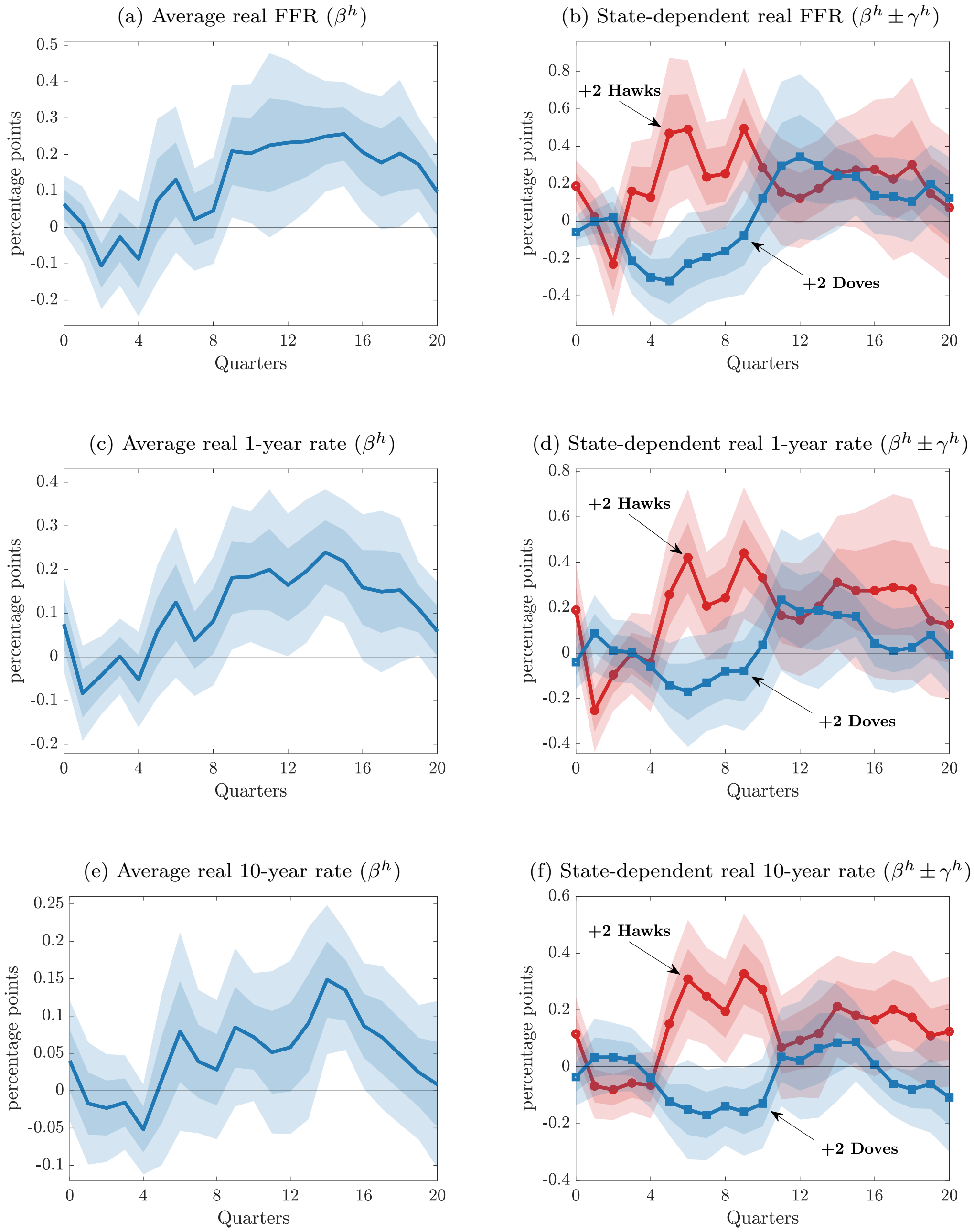
Notes: The figure shows responses of the federal funds rate (FFR), as well as the 1-year and 10-year treasury yields to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). All outcomes are annualized interest rates. We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.5.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We set the military spending shocks occurring in Q3 or Q4 to zero.

Figure 1.H.2: Responses of GDP and government spending, omit shocks at end of rotation cycle



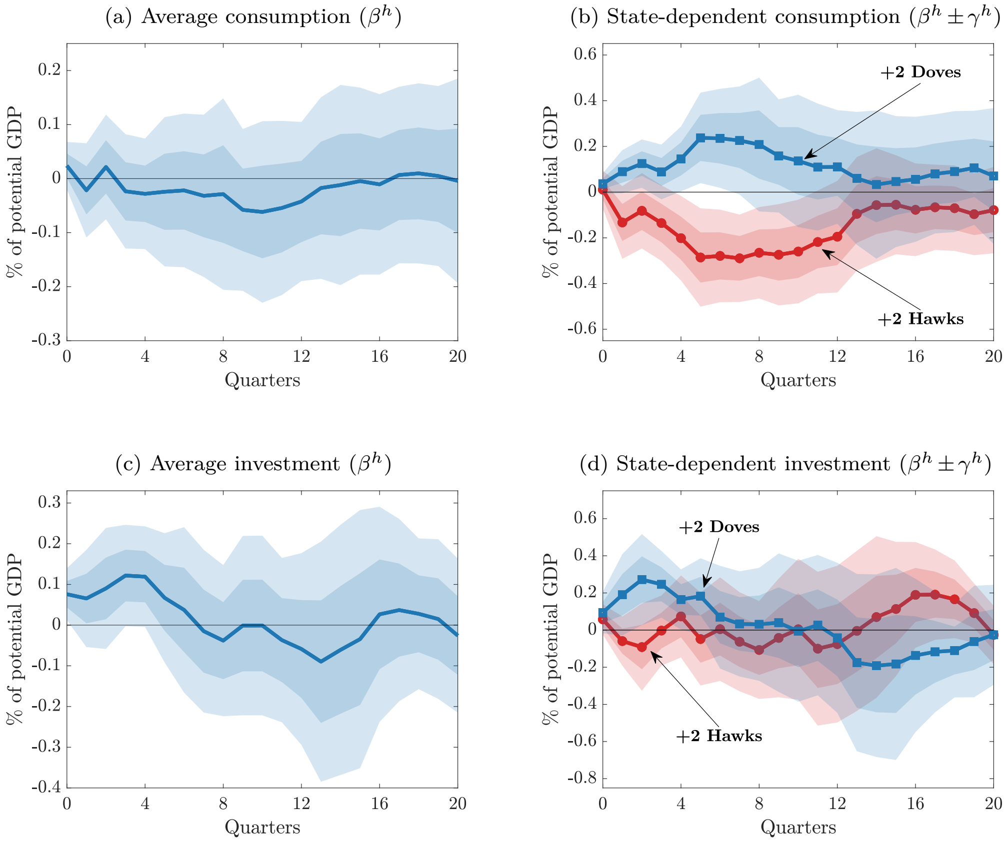
Notes: The figure shows responses of real GDP and real government spending (G) to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1.The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We set the military spending shocks occurring in Q3 or Q4 to zero.

Figure 1.H.3: Responses of real interest rates



Notes: The figure shows responses of the real federal funds rate (FFR), as well as the 1-year and 10-year real treasury yields to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.5.1. All outcomes are annualized ex-ante real interest rates which we compute as nominal rate minus one-year ahead inflation expectations according to the Livingston Survey, see Appendix 1.B for details. The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 1.H.4: Decomposing the GDP response, private spending



Notes: The figure shows responses of real private consumption and real private investment to an expansionary military spending shock, corresponding to one percent of GDP, conditional on systematic monetary policy ($Hawk_t$). We show IV estimates based on the local projection framework (1.10)-(1.11) as specified in Section 1.4.1. The β^h captures the responses when $Hawk_t$ equals its sample average. The $\beta^h \pm \gamma^h$ shows the state-dependent responses when $Hawk_t$ exceeds the sample average either by two hawks (+2 Hawks) or by two doves (+2 Doves). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors. We modify the control vector to include four lags of consumption, investment, and government spending, as well as the shock and a residual component of GDP, which we compute as GDP minus consumption, investment, and government spending.

1.I Two case studies from FOMC records

This section provides the full case studies that are outlined in Section 1.6.

1.I.1 The U.S. Space Program

In the first half of 1961, Ramey and Zubairy (2018) identify two expansionary shocks related to President Kennedy's defense spending plan, including the Space Program to "go to the Moon". Both shocks amount to a total of 6.7% of GDP, see Panel (b) of Figure 1.I.1. In the FOMC meeting of August 1, 1961, the staff presented the following assessment:

> *On top of substantial increases in expenditures to finance space exploration and longer-run defense measures,[...] the President has found it necessary to recom-*

> *mend an increase of $3-1/2 billion in current defense expenditures, [...]. More important, the President accompanied his recommendations with a very firm statement regarding his intentions with respect to the 1963 budget. These factors have certainly tended to minimize the immediate inflationary expectations and the urgency of the need for counter-measures. As of this moment in time, actual developments do not seem to call for any change in monetary policy.*(p.8)

Dovish FOMC members argued similarly for no change in policy because the effects could not yet be evaluated. For example, Eliot J. Swan (San Francisco) noted: *Turning to policy, [...], the available statistics did not yet reflect the impact of recent international developments and the announced plans for increased defense spending on business and consumer expectations* (p.28). Similarly, Edward A. Wayne (Richmond) stated: *As others had pointed out, however, there were two significant uncertainties in the picture. The first was the impact of proposed defense spending, not only directly but on expectations. [...] Nevertheless, until the effects [...] could be better gauged, he felt that maintenance of the present degree of ease was the most appropriate posture for monetary policy* (p.39).
While agreeing to no policy change in the current meeting, hawkish FOMC members such as Frederick L. Deming (Minneapolis) favored *to operate with increased alertness over the forthcoming period* (p.31). Similarly, New York Fed first-vice president, William Treiber noted: *We must be alert, however, to the possibility that stepped-up defense spending and related expansion in private spending may place excessive pressures on the price structure and endanger economic stability [...] If expenditures and related private spending result in an upsurge of activity with inflationary aspects, we may have to modify our policy of basic monetary ease sooner than we would otherwise have done. In the coming period undue ease should be avoided* (p.13;22-23). Malcom H. Bryan (Atlanta), a swinging dove, was "*sympathetic with those who had suggested the need for alertness to avoid getting again into an inflationary situation* (p.47).
By the end of 1961 and the beginning of 1962, some FOMC members started to acknowledge the expansionary impact on employment and business sentiment on defense-related industries. According to the FOMC Minutes of November 14, 1961: *Mr. Swan said that in the Twelfth District defense orders had exerted some impact on employment* (p.48) and *Mr. Ellis* [from Boston] *reported that business sentiment had been tending to become more optimistic, this being traceable partly to the expected increase in defense procurement in the District* (p.54). Further, according to the Minutes of March 6, 1962, Eliot J. Swan said: *Manufacturing employment rose, principally because of further gains in defense-related*

industries (p.50). Chairman Martin remarked that *goods and services were in adequate supply and prices were stable* (p.57), while Alfred Hayes (New York) stated that *Prices continue generally stable, and there are few, if any, signs of inflationary pressures* (p.34). By the end of 1961, real GDP growth accelerated to 6.4%, while inflation hovered around 1%, see Figure 1.I.1.

In June 1962, the FOMC shifted the emphasis of monetary policy toward slightly less ease and toward maintaining a moderately firm tone in the money market, mentioning balance-of-payments concerns. In this period, FOMC members interested in a tighter, inflation-focused monetary policy often cited the balance-of-payments criterion to bolster their case (Bordo and Humpage, 2014). Dovish FOMC members such as Governors James L. Robertson and George W. Mitchell frequently dissented in favor of easier policy throughout the second part of 1962.

By early 1963, several hawkish FOMC members such as Governors Charles N. Shepardson and Canby Balderston began arguing that some inflationary pressures were visible. Alfred Hayes (and William Treiber) dissented frequently in favor of tighter policy. In the FOMC meeting of May 7, 1963, these pressures were acknowledged by more members and Chairman Martin proposed firming of policy, noting the following:

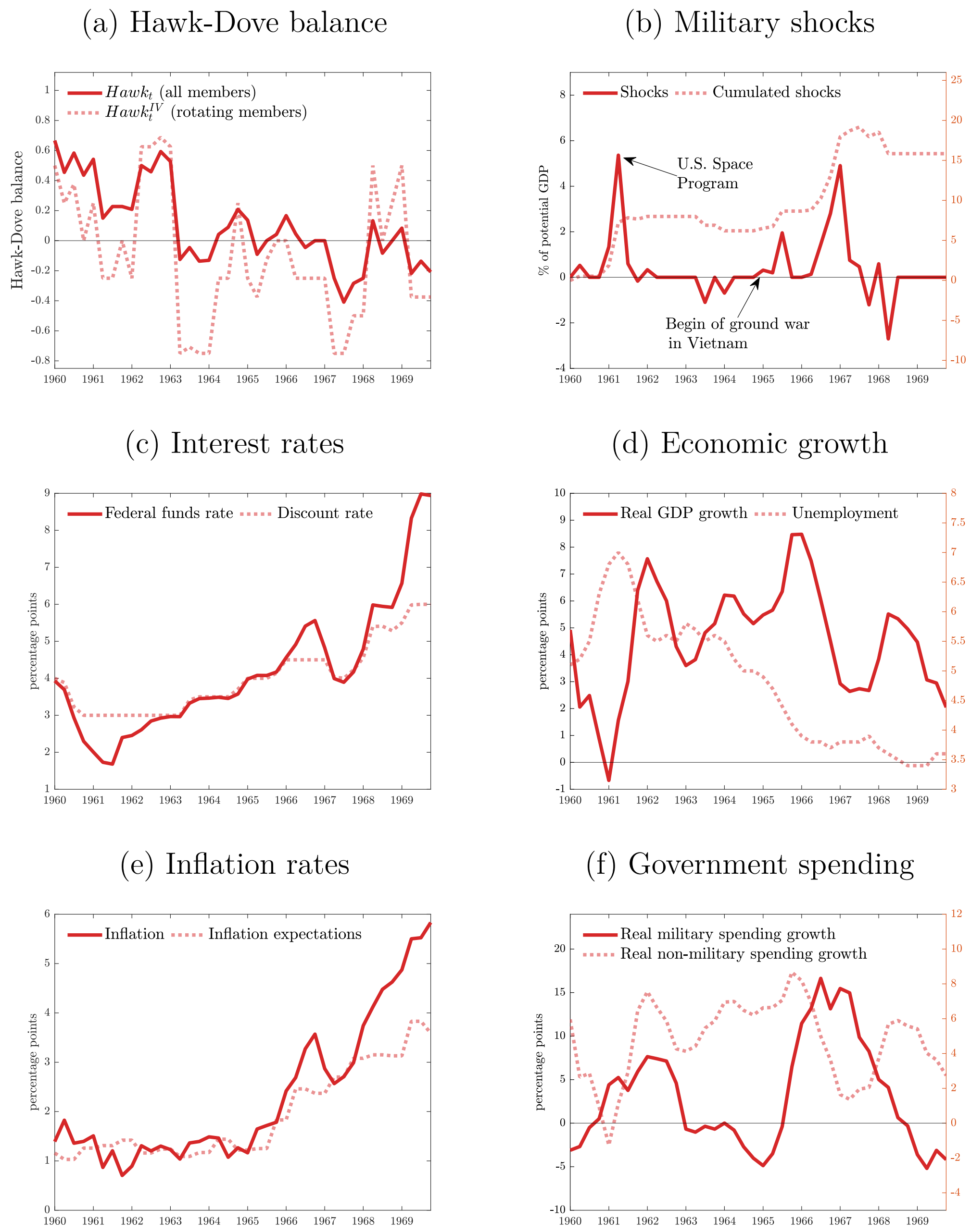
> *If the Committee waited too long, however, it might have to deal with an active problem of inflationary pressures. In his opinion, there was already a good bit of pressure in some areas that could build up rapidly. If one waited until after the resulting price movements actually occurred, he might wonder why he had not done something about it before. It would be too late at that juncture.* (p.61)

The FOMC voted to firm policy, however, with five dissenters favoring an unchanged policy to give the economy a chance to improve. Overall, we observe increased interest rates after the shock, albeit with a time lag, see Figure 1.I.1. Realized and expected inflation rates remained stable, broadly consistent with the hawkish leaning of the FOMC.

1.I.2 The Vietnam War

In 1965, the U.S. entered the ground war in Vietnam, leading to a series of expansionary military spending shocks lasting until 1967Q1. The largest spending shocks occured in 1966Q4 and 1967Q7, with a total of 7.7% of GDP. The Great Society initiatives of U.S. President Johnson which fall in the same period, are often used as an example of the government's effort to produce both "guns and butter". We focus on the FOMC discussion

Figure 1.I.1: Macroeconomic developments in the 1960s



Notes: The figure shows selected quarterly time series from 1960Q1 until 1969Q4. In Panel (b), shocks refer to the military spending shocks. In Panels (c) and (e), rates are annualized, and inflation and expectations refer to the CPI. In Panels (d) and (f), real growth rates are computed year-over-year. In Panels (b), (d) and (f), the scale corresponding to the dotted line is on the right.

with regard to the military expenses related to the Vietnam War. In the FOMC meeting of August 10, 1965, the staff's presentation explicitly accounted for the intended scaling up of military spending:

> *Further stimulus to the economy will come from expanded Government procurement for Vietnam hostilities. [...] the increases in spending and in the armed forces now proposed do not appear significant enough to touch off widespread commodity or labor shortages or widespread price increases. [...] The market response to Vietnam developments doesn't suggest any widespread fears of shortages, rationing, or inflation. On balance, then, the domestic evidence isn't clear enough to me to justify a significant policy move in either direction at this juncture."* (p. 28-29).

Several FOMC members agreed with the staff's assessment and argued for an unchanged policy in light of significant uncertainties related to the developments in Vietnam. Many shared the idea explained by George H. Clay (Kansas City) that *the Committee did not really know what the military program would be in the months ahead, however, and accordingly the effect on the economy would have to be reevaluated constantly* (p.71). In contrast, hawkish members, such as W. Braddock Hickman (Cleveland) noted that *Vietnam had already had an impact on industrial prices* (p.57). Similarly, William Treiber noted that *the amount of the spending and its timing are uncertain. It is quite clear, however, that the changed situation is already having an effect on the thinking of businessmen and of the public in general.* (p.44). In this meeting the staff presented three alternatives for policy directive, one of which considered firming of policy in relation to the war in Vietnam, arguing that *it is the Federal Open Market Committee's current policy to help defend the international position of the Dollar, and to avoid the emergence of inflationary pressures, by moderating growth in the reserve base, bank credit, and the money supply.* (p.1-2, Attachment A). However, the FOMC voted for no change in policy.

Two meetings later, on September 28, the dovish Governors Sherman J. Maisel, George W. Mitchell and James L. Robertson dissented against the "status quo", arguing that evidence of inflationary pressure was lacking and hence, they would have preferred an easier policy. For example, Sherman L. Mitchell said *he thought he fell within the group that was somewhat skeptical about the prevailing optimism on the business outlook. It seemed to him that the optimism was based primarily on a somewhat exaggerated notion of the economic stimulus that would be provided by the Vietnam hostilities.* (p.55-56).

In contrast, Alfred Hayes (New York Fed), a hawk, argued in the meeting of October 12, 1965 that: *Looking ahead, I think we have a real basis for concern about potential inflationary pressures*" (p.25). Similarly, hawkish Governor Balderston noted: *Wage pressures* *combined with Government spending for war and welfare activities both suggest to businessmen that things will cost more later on than now* (p.64). Chairman Martin shared similar thinking while sensing that he did not have (a hawkish) majority on his side to firm policy:

> *While the evidence was not clear, he thought there were many signs of inflation and of inflationary psychology in the economy. [...] But the Committee had a tendency to feel that it was best to "wait until all the evidence was in" before making a policy change. The difficulty was that when all the evidence was in it was likely to be too late. [...] With a divided Committee and in face of strong Administration opposition he did not believe it would be appropriate for him to lend his support to those who favored a change in policy now.* (p.68-69)

On December 5, 1965, the discount rate was raised with a narrow majority in order to prevent the risk of inflation. The dovish dissenters were arguing that inflationary pressures could be dealt with eventually but higher rates might push the economy toward a recession (cf. Meltzer, 2005). However, the tightening signal by the Fed was not enough to contain the buildup of inflationary pressures. While this had become clear for most members, the U.S. President had promised an anti-inflationary fiscal program and the FOMC delayed action in support of promised fiscal restraint. In the FOMC meeting of September 13, 1966, the staff economist presented:

> *Other fragmentary bases for these increased projections are the rapid increase in actual defense spending in August, continued increases in defense orders, and further rises in draft calls. All of this means that even if the President's new fiscal program is adopted in its entirety as outlined last week, the overall Federal contribution to economic activity will likely continue to shift to a more stimulative position over the rest of this year, and probably into next year unless further tax measures are adopted.* (p.20)

James L. Robertson summarized the situation: *Inflationary pressures are persisting [...] To counter these inflationary pressures, we now have the promise of help from a somewhat greater degree of fiscal restraint* (p.72).

Hoping on the legislative action to raise taxes in 1967, the FOMC eased policy in the last quarter of 1966 and throughout the first part of 1967, despite two large expansionary military spending shocks hitting in 1966Q4 and 1967Q1, see Panel (b) of Figure 1.I.1. In the FOMC meeting of September 12, 1967, Chairman Martin acknowledged the delay in action:

> *With fiscal policy strongly stimulative pending action on the President's tax program, the simple logic of the economic situation implied the desirability of changing monetary policy, as it probably had as much as two months ago. But the overriding need at this point was to get some restraint from fiscal policy through a tax increase, and in his judgment that would be less likely if Congress came to believe that adequate restraint was being exercised by monetary policy. The country was engaged in a major war, yet there had been an unfortunate tendency to underestimate the strains being put on economic resources by the hostilities in Vietnam. A "guns and butter" economy was not feasible; the country's resources were not sufficient for that.* (p.73).

The FOMC decided to tighten the policy on December 12, 1967. Once again, Chairman Martin admitted delayed action:

> *It was his feeling that the Committee had in a sense been caught in a trap [...] From the standpoint of economic considerations alone, it would have been desirable to adopt a firmer monetary policy a number of months ago. It had been clear then, however, that the overriding need was for a tax increase, and that a firming of monetary policy would make Congressional action on taxes less likely.* (p.96).

Overall, we observe that interest rates temporarily fall after the largest spending shocks were realized, see Figure 1.I.1. Unemployment remained below 4% but inflation increased substantially, broadly consistent with the dovish leaning of the FOMC. Indeed, the dovish FOMC responded very differently from the hawkish FOMC in the first half of the 1960s, despite the same Fed Chair being in power.

Finally, the Great Society initiatives are an important confounder. The program entailed an increase in spending for non-military purposes to achieve a variety of domestic goals. In Panel (f) of Figure 1.I.1, we display the growth rates of real military and real non-military spending, showing, however, that the growth rate of military spending largely dominates the growth rate of non-military spending of the federal government when the largest military spending shocks hit in 1966Q4 and 1967Q1.

Chapter 2

The Systematic Origins of Monetary Policy Shocks

Joint with Klodiana Istrefi and Matthias Meier.

2.1 Introduction

Empirical monetary policy shock series form the backbone of a large literature in monetary economics. The estimated responses to these shocks are used to assess the effectiveness of monetary policy, construct policy counterfactuals, study the optimality of monetary policy, estimate structural macroeconomic equations, or estimate DSGE models.[1] These applications require empirical monetary policy shocks that are well identified, meaning they capture exogenous changes in a policy instrument orthogonal to other macroeconomic shocks.

The central point of Chapter 2 is that fluctuations in systematic monetary policy pose a challenge to conventional strategies to identify monetary policy shocks. The fundamental problem of conventional identification strategies is the implicit assumption that systematic monetary policy is constant across time. Under the assumption, any time-variation in systematic monetary policy will be contained in the empirical monetary policy shock, consistent with common views about these shocks.[2]

[1]See, e.g., Barnichon and Mesters (2020), Barnichon and Mesters (2023), Bernanke and Blinder (1992), Bernanke, Gertler, and Watson (1997), Christiano, Eichenbaum, and Evans (1999, 2005), Gertler and Karadi (2015), McKay and Wolf (2023), Romer and Romer (1989), and Romer and Romer (2004).

[2]Similarly, in the first Handbook of Macroeconomics, Christiano, Eichenbaum, and Evans (1999, p.71-72) argue that an empirical monetary policy shock *[..] reflects exogenous shocks to the preferences of the*

> *We do not have many good economic theories for what a structural monetary policy shock should be. Other than "random coin flipping," the most frequently discussed source of monetary policy shocks is shifts in central bank preferences, caused by changing weights on inflation vs unemployment in the loss function or by a change in the political power of individuals on the FOMC.* Ramey (2016, Handbook of Macroeconomics, Vol. 2A, p.89)

This chapter makes a theoretical and empirical contribution to the identification of monetary policy shocks. Our theoretical contribution is to formally show that conventional empirical identification strategies do not isolate monetary policy shocks in an environment with time-varying systematic monetary policy. Instead, they are contaminated by systematic monetary policy and other macroeconomic shocks, leading to contamination bias in estimated impulse response functions. The empirical contribution is threefold. First, we show that monetary policy shocks as estimated in the seminal Romer and Romer (2004) are predictable by a time-varying measure of systematic monetary policy. Second, we propose a new monetary policy shock series that is orthogonal to measured fluctuations in systematic monetary policy. Third, we find that inflation and output respond more quickly and strongly relative to the Romer and Romer (2004) shock.

Our theoretical analysis starts from the assumption that monetary policy follows a general type of Taylor rule. The rule determines a policy instrument as a function of inputs to the rule, e.g., inflation and output, time-varying slope coefficients describing the policy response to macroeconomic conditions, i.e., systematic monetary policy, and a stochastic intercept, i.e., a monetary policy shock.[3] In contrast, many conventional empirical identification strategies implicitly assume a Taylor rule with time-invariant slope coefficients. Empirical monetary policy shocks are estimated as deviations from such rule. Identification strategies following this approach include Taylor rule-type regressions (e.g., Romer and Romer, 2004) and monetary VAR models using exclusion restrictions (e.g., Christiano, Eichenbaum, and Evans, 1999), sign restrictions (e.g., Uhlig, 2005), narrative restrictions (e.g., Antolín-Díaz and Rubio-Ramírez, 2018), or external instruments (e.g., Gertler and Karadi, 2015).

Against the backdrop of a time-varying Taylor rule, we show that the empirical monetary policy shock contains the (true) monetary policy shock but also time variation in systematic monetary policy interacted with the inputs to the policy rule. To the extent that other

monetary authority, perhaps due to stochastic shifts in the relative weight given to unemployment and inflation. These shifts could reflect shocks to the preferences of the members of the Federal Open Market Committee (FOMC), or to the weights by which their views are aggregated.

[3] For evidence on fluctuations in the coefficients of the policy rule, see, e.g., Clarida, Gali, and Gertler (2000), Orphanides (2004), Bordo and Istrefi (2023), and Chapter 1 of this dissertation.

macroeconomic shocks affect the inputs to the policy rule, the empirical monetary policy shock is contaminated by these other macroeconomic shocks.

The contaminated shocks lead to biased impulse response estimates. We formally show that they do not identify the causal effects of (true) monetary policy shocks. We analytically characterize three sources of bias reflecting endogeneity and attenuation. The estimated impulse response function remains biased even if time variation in systematic monetary policy is exogenous, i.e., if time variation in the slope coefficients of the Taylor rule is independent of other macroeconomic shocks.

Our theoretical insights similarly apply to empirical monetary shocks identified using high-frequency data (e.g., Gertler and Karadi, 2015). Identification rests on the implicit assumption that systematic monetary policy, as perceived by financial market participants, is constant in a time window around monetary announcements. Otherwise, the shocks are contaminated and lead to biased estimates. Bauer and Swanson (2023b) provide evidence consistent with such high-frequency belief changes. We contribute to this debate by showing analytically that regressing high-frequency monetary surprise on publicly available macroeconomic forecasts (Bauer and Swanson, 2023a) or Greenbook forecasts (Miranda-Agrippino and Ricco, 2021) does not resolve the contamination problem.

While previous work has noted that time-varying systematic monetary policy may complicate the identification of monetary policy shocks (e.g., Bauer and Swanson, 2023b; Coibion, 2012; McMahon and Munday, 2023), our work is the first to formally characterize (i) how time-varying systematic monetary policy leads to contamination in the monetary policy shocks obtained from a wide set of conventional empirical identification strategies, and (ii) how contamination leads to biased impulse response estimates. We further go beyond previous work by providing new empirical evidence on shock contamination and a new identification strategy that tackles this problem.

Our empirical analysis starts from the testable prediction of the theory that conventional monetary policy shocks are predictable by time variation in systematic monetary policy interacted with the inputs to the policy rule. We measure time variation in systematic U.S. monetary policy through the historical composition of hawks and doves in the Federal Reserve's Federal Open Market Committee (FOMC), as introduced in Chapter 1. We briefly repeat the data explanation for the convenience of the reader. Our measure builds on the narrative classification of FOMC members by Istrefi (2019). Hawks are more concerned about inflation. Doves are more concerned about supporting employment and growth.[4] We

[4] Istrefi (2019) shows that these preferences match with narratives on monetary policy, preferred interest

consider two measures of systematic monetary policy, the Hawk-Dove balance across all voting FOMC members and the balance across the four FOMC members currently with voting rights through the annual rotation. The former is a more comprehensive measure, whereas the latter reflects exogenous variation through the rotation.

We test our prediction using empirical monetary policy shocks as estimated in Romer and Romer (2004), RR in the following.[5] We regress the RR shock on the Taylor rule inputs considered by RR, notably Greenbook forecasts for various macroeconomic variables and horizons, interacted with measured fluctuations in systematic monetary policy. We consider the original RR sample 1969-1996, the extended Wieland and Yang (2020) sample 1969-2007, and the post-Volcker disinflation sample 1983-2007. The regression explains between 10 and 54% of the variance of RR shocks depending on sample and regressors (contemporaneous or lagged). Using the regressors lagged by one FOMC meeting yields the highest R^2, ranging between 0.33 and 0.54. Overall, our evidence strongly suggests that RR shocks are contaminated by fluctuations in systematic monetary policy.[6]

The empirical evidence motivates us to construct a new series of empirical monetary policy shocks that are not predictable by fluctuations in measured systematic monetary policy. We estimate an extension of the Taylor rule regression in RR that includes the interaction of the Hawk-Dove balance with the Taylor rule inputs. The correlation between the original RR shock and our new shock is 0.67. The sign-correlation between the two series is lower, meaning many shocks flip signs. The distribution of new shocks is less dispersed, with a standard deviation of 0.23, compared to 0.34 for the RR shock.

Finally, we compare impulse responses between our new monetary policy shock and the RR shock. We focus on the post-Volcker disinflation sample 1983-2007 because the estimated responses to many conventional monetary policy shock series appear puzzling in this sample (e.g., Ramey, 2016).[7] For comparability, we normalize the size of both shocks to the same impact increase of the FFR. The dynamic FFR response to our new shock is less persistent and smaller at peak. In contrast, the decline in GDP and inflation is substantially larger

rates, dissents, and forecasts of FOMC members. Bordo and Istrefi (2023) study the origins of these preferences, linking them to early-life experiences and education. Chapter 1 use the Hawk-Dove classification to study the effects of systematic monetary policy on the propagation of macroeconomic shocks.

[5] The RR identification strategy has been applied to the U.K. (Cloyne and Hürtgen, 2016), Germany (Cloyne, Hürtgen, and Taylor, 2022), Norway (Holm, Paul, and Tischbirek, 2021), and many other countries (Choi, Willems, and Yoo, 2024).

[6] We show that measured systematic monetary policy also has predictive power for the refined RR shocks in Aruoba and Drechsel (2024), who use textual analysis to create sentiment indicators about the Fed staff's assessment of the economy to better capture the Fed's information set about the state of the economy.

[7] Relatedly, Barakchian and Crowe (2013) show that a variety of conventional monetary policy shock series raise GDP when raising the federal funds rate in a post-1988 sample.

for the new shock. The trough GDP response is about twice as large for the new shock compared to the RR shock. The differences between the responses to the two shocks are statistically significant at the five percent level for many horizons. Importantly, the RR shock seems to operate with a long lag, not affecting inflation up until two years after the shock. The GDP response is broadly insignificant. In contrast, inflation and GDP respond to our new shock with a lag of one year. Beyond the first year, the responses of inflation and GDP are significantly different from zero at the five percent level.[8] Our findings suggest that the puzzling effects of RR shocks in the 1983-2007 sample may reflect contamination from time-varying systematic monetary policy.

This chapter highlights the importance of accounting for the time-varying nature of systematic monetary policy when identifying monetary policy shocks. An alternative approach addresses time-varying systematic monetary policy by modeling it as latent variable or time-varying coefficients, see, for example, regime-switching models (e.g., Owyang and Ramey, 2004; Sims and Zha, 2006), time-varying coefficient monetary VAR models (e.g., Primiceri, 2005), and Taylor rules with time-varying coefficients (e.g., Bauer, Pflueger, and Sunderam, 2022; Boivin, 2006; Coibion, 2012). Particularly related is Coibion (2012) who uses the latter approach to estimate a monetary policy shock series. The estimated shock is highly correlated with the RR shock and yields similar impulse responses as the RR shock. The difference between this finding and ours might reflect the challenge of time-varying coefficient models to identify genuine time variation in the parameters of interest while avoiding overfitting.

2.2 Identification challenge in theory

In this section, we study the identification of monetary policy shocks in an environment with time-varying systematic monetary policy. We formally show that a wide spectrum of identification strategies to estimate monetary policy shocks yield shocks that are contaminated by other macroeconomic shocks. Using these shocks to estimate impulse response functions generally leads to biased estimates.

[8] In the 1969-2007 sample, we also find that output and inflation respond more strongly to the new shock, albeit with a sluggish inflation response. We further find that orthogonalizing the Aruoba and Drechsel (2024) shock with respect to systematic monetary policy leads to similar differences in the estimated responses.

2.2.1 Time-varying systematic monetary policy

Departing from the common assumption that systematic monetary policy is constant across time, we assume monetary policy follows the time-varying Taylor rule

$$i_t = \alpha + (\phi + \tilde{\phi}_t)' x_t + w_t^m, \quad \mathbb{E}[\tilde{\phi}_t] = \mathbb{E}[x_t] = \mathbb{E}[w_t^m] = \mathbb{E}[\phi_t w_t^m] = 0, \tag{2.1}$$

where $i_t \in \mathbb{R}$ is a policy instrument, $x_t \in \mathbb{R}^{n \times 1}$ are the n inputs of the policy rule, e.g., present and lagged (forecasts of) GDP and inflation, $\tilde{\phi}_t \in \mathbb{R}^{n \times 1}$ is a vector of time-varying coefficients describing fluctuations in systematic monetary policy, with $\phi \in \mathbb{R}^{n \times 1}$ the average coefficient vector, and w_t^m denotes a random monetary policy shock. We assume the inputs in x_t are mean zero and set $\alpha = -\mathbb{E}[\tilde{\phi}_t x_t]$, which simplifies some subsequent derivations but is not critical for our results.[9]

Time variations in the coefficients of the rule $\tilde{\phi}_t$ may be driven by changes in the preferences of central bankers that may occur for exogenous reasons, e.g., the FOMC rotation of voting rights, or for endogenous reasons, e.g., monetary policy may become more responsive to inflation when inflation is high (Davig and Leeper, 2008). Our main results hold irrespective of whether $\tilde{\phi}_t$ fluctuates for exogenous or endogenous reasons. Finally, we assume that ϕ_t does not co-move with monetary policy shocks, $\mathbb{E}[\phi_t w_t^m] = 0$, which allows for a sharp conceptual distinction between systematic monetary policy and monetary policy shocks, but is otherwise not critical for our results.

2.2.2 Conventional identification of monetary policy shocks

In this section, we show that time-varying systematic monetary policy implies that conventional identification strategies yield contaminated monetary policy shocks under general conditions. Our result derives from the above monetary policy rule, no further structural assumptions about the macroeconomy are needed.

Many conventional identification strategies estimate monetary policy shocks as residual from a time-invariant Taylor rule-type regression

$$i_t = b' x_t + e_t^m, \tag{2.2}$$

[9] A richer formulation of (2.1) may contain time-varying target variables, e.g., $i_t = \alpha + (\phi + \tilde{\phi}_t)'(x_t - x_t^*) + w_t^m$, where $x_t^* \in \mathbb{R}^{n \times 1}$ is the target, e.g., the inflation target. Shocks to the target generate a third type of monetary policy shock, the effect of which is correlated with fluctuations in systematic monetary policy. In the scope of this chapter, we abstract from fluctuations in the target.

where the estimated regression residual, $\hat{e}_t^m$, is an empirical monetary policy shock. This is a broad description of a wide variety of identification strategies which differ mainly in how the coefficients in equation (2.2), and thus the residual, are estimated. Romer and Romer (2004) propose to directly estimate (2.2) via OLS. A common alternative approach is to use monetary VAR models. Irrespective of identifying assumptions and estimation method, monetary VAR models contain an equation consistent with equation (2.2).[10] This equation is typically identified via internal instruments (Shapiro and Watson, 1988), external instruments, or estimation methods for set-identified models. The estimation method depends on the identifying assumptions, which may be exclusion restrictions (e.g., Christiano, Eichenbaum, and Evans, 1999), sign restrictions (e.g., Uhlig, 2005), narrative restrictions (e.g., Antolín-Díaz and Rubio-Ramírez, 2018), or external instruments (e.g., Gertler and Karadi, 2015).

Against the backdrop of the time-varying monetary policy rule in (2.1), the time-invariant regression in (2.2) is misspecified. In general, this misspecification leads to contamination in the estimated monetary policy shock. The following proposition formally characterizes the estimated empirical shocks for a given estimate $\hat{b}$.

Proposition 1 (Monetary policy shock). *Let monetary policy follow* (2.1)*. Given an estimate $\hat{b}$, let $\hat{e}_t^m$ be the estimated residual from* (2.2)*. The residual satisfies*

$$\hat{e}_t^m = w_t^m + \omega_t^{\hat{b}} + \omega_t^{\tilde{\phi}}$$

where the two wedges are defined by

$$\omega_t^{\hat{b}} = (\phi - \hat{b})' x_t, \quad \text{and} \quad \omega_t^{\tilde{\phi}} = \tilde{\phi}_t' x_t - \mathbb{E}[\tilde{\phi}_t' x_t].$$

The proof is straightforward when combining (2.1) and (2.2). The proposition characterizes two wedges between the actual monetary policy shock w_t^m and the estimated shock $\hat{e}_t^m$. The first wedge, $\omega_t^{\hat{b}}$, arises whenever the estimate $\hat{b}$ does not equal the average policy coefficient ϕ. This wedge may be present even in the absence of time-variation in systematic monetary policy $\tilde{\phi}_t = 0$. For example, if b is estimated via OLS, a well-known endogeneity bias arises if the monetary policy shock correlates with x_t (Carvalho, Nechio, and Tristão,

[10] A (structural) monetary VAR model is defined by $B(L)Y_t = W_t$, where Y_t is a vector of variables, $B(L)$ a lag polynomial, and W_t a vector of structural shocks. Y_t includes the policy instrument i_t and W_t includes a monetary policy shock, *wlog* the first element of W_t. Then, the first equation of the VAR is a monetary policy rule that is identical with equation (2.2) given a corresponding specification of Y_t.

2021; Cochrane, 2011). In addition, the presence of time-varying systematic monetary policy generates a second type of endogeneity bias. Formally, the OLS estimate $\hat{b}$ of the regression model (2.2) satisfies $\hat{b} \xrightarrow{p} \phi + \mathbb{E}[x_t x_t']^{-1} \mathbb{E}[x_t w_t^m] + \mathbb{E}[x_t x_t']^{-1} \mathbb{E}\left[x_t x_t' \tilde{\phi}_t\right]$.[11] Whatever the method by which (2.2) is estimated, if $\hat{b} \neq \phi$ then the estimated monetary policy shock $\hat{e}_t^m$ correlates with x_t.
The second wedge is novel. It arises because (2.2) is misspecified in the sense that fluctuations in systematic monetary policy are not modeled. Fluctuations in $\tilde{\phi}_t$ interacted with x_t must therefore be captured by the regression residual. The wedge disappears if we assume away fluctuations in systematic monetary policy $\tilde{\phi}_t = 0$. Note that the wedge is present for any estimate $\hat{b}$. Even if $\hat{b} = \phi$, i.e., even if the first wedge is nil, the estimated monetary policy shock is still contaminated by other macroeconomic shocks through $\tilde{\phi}_t' x_t$. In general, x_t reflects all present and past macroeconomic shocks. Thus $\hat{e}_t^m$ is contaminated by other macroeconomic shocks through the two wedges.

The discussion has so far omitted a popular type of conventional identification strategy. High-frequency identification uses interest rate futures (or swaps) to approximate expectations about future interest rates in a narrow time window around a monetary announcement in t. A high-frequency identified monetary policy shock is constructed as $\hat{e}_t^m = \mathbb{E}_{t+\Delta}[i_{t+\tau}] - \mathbb{E}_{t-\Delta}[i_{t+\tau}]$, where $\mathbb{E}_{t+\Delta}[i_{t+\tau}]$ denotes the period $t+\Delta$ expectation of period $t+\tau$ interest rates as measured by the price of an interest rate future contract. If monetary policy follows (2.1), we can rewrite high-frequency monetary policy surprises as

$$\hat{e}_t^m = w_t^m + \mathbb{E}_{t+\Delta}[(\phi + \tilde{\phi}_t)' x_t] - \mathbb{E}_{t-\Delta}[(\phi + \tilde{\phi}_t)' x_t]. \tag{2.3}$$

If systematic monetary policy as perceived by financial market participants varies between $t-\Delta$ and $t+\Delta$, then the monetary policy surprise is contaminated by variation in $\tilde{\phi}_t$ in interaction with x_t. This result has previously been noted by Bauer and Swanson (2023b), who also provide evidence consistent with such contamination. Different from Bauer and Swanson (2023b), we show that the contamination arising from time-varying systematic monetary policy may afflict a wide range of identification strategies.
The contamination result for the high-frequency identified monetary policy shock is closely related to Proposition 1. Different from Proposition 1, however, it is not sufficient for systematic monetary policy to vary at some point(s) in time. Instead, the contamination of high-frequency identified monetary policy shocks requires that perceived systematic

[11] In the New Keynesian model with time-varying systematic monetary policy we study in Section 2.2.4, it generally holds that $\mathbb{E}\left[x_t x_t' \tilde{\phi}_t\right]$ is non-zero.

monetary policy differs at the end of a narrow window around monetary announcements relative to expectations at the beginning of the time window.[12] Whether (perceived) systematic monetary policy varies outside those narrow windows is irrelevant for high-frequency identification, but not for the other conventional identification strategies.

Importantly, regressing high-frequency identified monetary policy shocks on x_t, whether that is publicly available macroeconomic forecasts (Bauer and Swanson, 2023a) or Greenbook forecasts (Miranda-Agrippino and Ricco, 2021), does not resolve the contamination problem. A simple way to see that is to regress $\tilde{\phi}_t' x_t$ on x_t. The residual will be $\tilde{\phi}_t' x_t - \hat{\gamma} x_t$, with $\hat{\gamma}$ the estimated coefficient. In general, for any $\hat{\gamma}$, the residual still contains variation in $\tilde{\phi}_t' x_t$. Hence, if high-frequency monetary policy shocks are contaminated by time-varying systematic monetary policy, regressing the estimated shock on x_t does not fundamentally heal the problem.

2.2.3 Impulse response estimate

Empirical monetary policy shocks are often not the object of interest *per se*, but rather the impulse response function (IRF) that is estimated based on these shocks. We analytically show that the contamination of monetary policy shocks generally leads to biased IRF estimates, including relative IRF estimates.

Suppose we are interested in the causal effects of the monetary policy shock w_t^m on some scalar outcome z_{t+h}, h periods after the shock, and where z_t may or may not be contained in the vector x_t. Let z_t follow the Moving Average (MA) process

$$z_t = \gamma_z + \sum_{h=0}^{\infty} \left(\delta_z^h w_{t-h}^m + v_{z,t-h}^h \right), \qquad \mathbb{E}[v_{z,t-h}^h] = \mathbb{E}[w_{t-h}^m v_{z,t-j}^h] = 0 \quad \forall h, j, \tag{2.4}$$

where δ_z^h denotes the causal effect of w_t^m on z_{t+h} and γ_z is a constant. The second term, $v_{z,t+h}^h$, may contain, for example, the linear effects of macroeconomic shocks other than the monetary policy shock, and the effects of all macroeconomic shocks interacted with time-varying systematic monetary policy $\tilde{\phi}_t$ (see Section 2.2.4 for an example). When the true causal effect of w_t^m on z_{t+h} depends on systematic monetary policy, then δ_z^h can be defined as the best linear prediction and $v_{z,t+h}^h$ contains the residual non-linear effects of w_t^m. The MA process is general in the sense that we put no restriction on what is contained

[12] The high-frequency shocks can also be contaminated if perceived systematic monetary policy remains unchanged around the monetary announcement but expectations over x_t update, the information effect of monetary policy (e.g., Jarociński and Karadi, 2020; Nakamura and Steinsson, 2018).

in $v^h_{z,t+h}$ other than assuming $\mathbb{E}[w^m_{t-h}v^h_{z,t-j}] = 0 \;\; \forall h, j$. It will be convenient to rewrite (2.4) as

$$z_{t+h} = \gamma_z + \delta^h_z w^m_t + \tilde{v}^h_{z,t+h}, \quad \tilde{v}^h_{z,t+h} = \sum_{j=0}^{\infty} v^h_{z,t+h-j} + \sum_{j=0,\; j\neq h}^{\infty} \delta^h_z w^m_{t+h-j} \tag{2.5}$$

and it follows from (2.4) that $\mathbb{E}[w^m_t \tilde{v}^h_{z,t+h}] = 0 \;\; \forall h \geq 0$.[13]
Next, suppose an econometrician aims to estimate the effects of monetary policy via the local projection

$$z_{t+h} = c^h_z + d^h_z \hat{e}^m_t + u^h_{z,t+h}, \tag{2.6}$$

where $\hat{e}^m_t$ denotes the estimated monetary policy shock as described in Proposition 1.[14] If $\hat{e}^m_t = w^m_t$, the econometrician could easily uncover the causal effect via the OLS estimate $\hat{d}^h_z \xrightarrow{p} \delta^h_z$. In general, however, the estimate $\hat{d}^h$ will be biased, as the following proposition shows.

Proposition 2 (IRF bias). *Let monetary policy follow* (2.1) *and z_t follow the MA process in* (2.4)*. Consider the local projection in* (2.6) *with $\hat{e}^m_t$ as described in Proposition 1. As $T \to \infty$, the OLS estimate $\hat{d}^h_z$ of the local projection satisfies*

$$\hat{d}^h_z \xrightarrow{p} \delta^h_z + \vartheta^{\hat{b}}_z + \vartheta^{\tilde{\phi}}_z + \vartheta^a_z$$

where the three bias terms are defined by

$$\begin{aligned}
\vartheta^{\hat{b}}_z &= \mathbb{E}\left[(\hat{e}^m_t)^2\right]^{-1} \left(\phi - \hat{b}\right)' \left(\delta^h_z \mathbb{E}\left[x_t w^m_t\right] + \mathbb{E}\left[x_t \tilde{v}^h_{z,t+h}\right]\right), \\
\vartheta^{\tilde{\phi}}_z &= \mathbb{E}\left[(\hat{e}^m_t)^2\right]^{-1} \left(\delta^h_z \mathbb{E}\left[\tilde{\phi}'_t x_t w^m_t\right] + \mathbb{E}\left[\tilde{\phi}'_t x_t \tilde{v}^h_{z,t+h}\right]\right), \\
\vartheta^a_z &= \mathbb{E}\left[(\hat{e}^m_t)^2\right]^{-1} \delta^h_z \left(\mathbb{E}\left[(w^m_t)^2\right] - \mathbb{E}\left[(\hat{e}^m_t)^2\right]\right).
\end{aligned}$$

The proof is straightforward and follows from inserting $\hat{e}^m_t$ into the population counterpart of $\hat{d}^h_z$ rewriting.[15] We first discuss the three bias terms in general and then discuss an

[13] We assume $\tilde{\phi}_t$, x_t, z_t, $\tilde{v}^h_{z,t+h}$ jointly follow a stable and ergodic process with finite fourth moments.

[14] We further consider an extension of the local projection in (2.6) that includes lagged control variables. This leads to broadly similar results, as we discuss further below.

[15] In addition, one needs to employ the appropriate law of large numbers (the Ergodic Theorem) and the Continous Mapping Theorem to show that the numerator and denominator of $\hat{d}^h_z$ converge to their population counterparts.

extension of the local projection.

The first bias $\vartheta_z^{\hat{b}}$ arises from the wedge $\omega_t^{\hat{b}}$ and is zero if $\hat{b} = \phi$ or if $\mathbb{E}\left[x_t w_t^m\right] = \mathbb{E}\left[x_t \tilde{v}_{z,t+h}^h\right] = 0$. For example, a sufficient condition is that the monetary policy shock w_t^m and all other macroeconomic shocks that affect z_{t+h} are uncorrelated with x_t. We argue that strong assumptions are required to satisfy both conditions. For example, in monetary VAR models, x_t commonly includes endogenous variables such as period t inflation and GDP, where a period is commonly a month or a quarter. Assuming a zero response of these variables to a monetary policy shock in the same period is a strong assumption. In contrast, under the Romer and Romer (2004) strategy, x_t includes forecasts shortly before a monetary policy decision in which w_t^m realizes. In that case, $\mathbb{E}\left[x_t w_t^m\right] = 0$ seems plausible. However, the second condition still requires a strong assumption. The outcome variable z_{t+h} is commonly a macroeconomic variable observed at a monthly or quarterly frequency. Assuming persistence, the residual $\tilde{v}_{z,t+h}^h$ then contains all period t macroeconomic shocks other than w_t^m. For example, an oil supply shock at the beginning of the period will affect the outcome z_{t+h} but also the forecasts in x_t if the monetary policy meeting was later in the period. Overall, we require strong assumptions to eliminate the first bias $\vartheta_z^{\hat{b}} = 0$.

The second bias $\vartheta_z^{\tilde{\phi}}$ arises from the wedge $\omega_t^{\tilde{\phi}}$, which captures the misspecification of the linear Taylor rule regression. The bias depends on two expectations, $\mathbb{E}\left[\tilde{\phi}_t' x_t w_t^m\right]$ and $\mathbb{E}\left[\tilde{\phi}_t' x_t \tilde{v}_{z,t+h}^h\right]$. Similarly to the first wedge, the first expectation is plausibly zero in the case of a Romer and Romer (2004) identification strategy. The second expectation, however, is generally non-zero. Consider again an oil supply shock at the beginning of the period, before the monetary policy meeting and associated forecast x_t. The shock correlates with x_t, but its effect on z_{t+h} also generally depends on $\tilde{\phi}_t$. Therefore, the second expectation is non-zero. In monetary VAR models, both expectations are generally non-zero, for similar reasons as discussed above. Hence, the second bias is present unless we impose strong assumptions.

The third term, ϑ^a, can be interpreted as a type of attenuation bias. If the estimated monetary policy shock satisfies $\mathbb{E}\left[(\hat{e}_t^m)^2\right] > \mathbb{E}\left[(w_t^m)^2\right]$, the estimate $\hat{d}_z^h$ will be biased toward zero relative to δ_z^h. However, given that w_t^m may correlate with the wedges $\omega_t^{\hat{b}}$ and $\omega_t^{\tilde{\phi}}$, the estimated monetary policy shock is not classical measurement error. If $\mathbb{E}\left[(\hat{e}_t^m)^2\right] < \mathbb{E}\left[(w_t^m)^2\right]$, the estimate $\hat{d}_z^h$ will be biased away from zero. Even if the first two biases are zero, the third bias remains non-zero as long as $\tilde{\phi}_t \neq 0 \ \forall t$.

The local projection, as specified in (2.6), is highly parsimonious, not including any endogenous control variables. Consider instead the extended local projection $z_{t+h} = c_z^h + d_z^h \hat{e}_t^m +$

$\Gamma(L)Y_t + u^h_{z,t+h}$, where $\Gamma(L) = \sum_{i=0}^{\infty}\Gamma_i L^i$ is a lag polynomial, and Y_t a vector of control variables. The additional control vector means we need to replace $\hat{e}^m_t$ and $\tilde{v}^h_{z,t+h}$ by projections of these variables on $\{Y_{t-1}, Y_{t-2}, ...\}$ in the bias terms in Proposition 2. While the controls may quantitatively change the bias, they do not eliminate the bias. In fact, in the above discussion of the bias terms, we have discussed bias arising from contemporaneous macroeconomic shocks. This bias cannot disappear by regressing on lagged variables.

In many empirical applications, the econometrician aims to identify the relative effect of monetary policy shocks rather than its absolute effect. If δ_{z_1} is the absolute causal effect of w^m_t on z_{1t}, the relative causal effect is $\delta_{z_1}/\delta_{z_2}$, where z_2 denotes another outcome. For example, it is common to study the effects of monetary policy shocks that raise the nominal interest rate by 25 or 100 basis points. This requires dividing the response of some outcome variable of interest by the interest rate response. For many empirical questions, a bias in the estimated absolute effect may be acceptable as long as the bias cancels out in the estimated relative effect. The following proposition provides a condition for the relative estimate to be unbiased.

Proposition 3 (Relative IRF bias). *Let monetary policy follow* (2.1) *and* z_{1t} *and* z_{2t} *follow MA processes as in* (2.4). *Consider two local projections, as in* (2.6)*, to estimate the effects of* $\hat{e}^m_t$ *on* z_{1t+h} *and* z_{2t+h}. *The two OLS estimates* $\hat{d}^h_{z_1}$ *and* $\hat{d}^h_{z_2}$ *satisfy*

$$\frac{\hat{d}^h_{z_1}}{\hat{d}^h_{z_2}} \xrightarrow{p} \frac{\delta^h_{z_1}}{\delta^h_{z_2}}$$

if and only if

$$\frac{\left(\phi - \hat{b}\right)'\mathbb{E}\left[x_t \tilde{v}^h_{z_1,t+h}\right] + \mathbb{E}\left[\tilde{\phi}'_t x_t \tilde{v}^h_{z_1,t+h}\right]}{\delta^h_{z_1}} = \frac{\left(\phi - \hat{b}\right)'\mathbb{E}\left[x_t \tilde{v}^h_{z_2,t+h}\right] + \mathbb{E}\left[\tilde{\phi}'_t x_t \tilde{v}^h_{z_2,t+h}\right]}{\delta^h_{z_2}}.$$

The result follows from rearranging the bias terms from Proposition 2 for both outcomes. The condition under which the relative IRF is not biased is a knife-edge condition, which is generally not satisfied. A sufficient condition is $\hat{b} = \phi$ and $\tilde{\phi}_t = 0$, which also yields unbiased absolute IRF estimates.

2.2.4 A non-linear New Keynesian model

We revisit the results in Proposition 1-3 through the lens of a stylized model. Whereas the propositions provide general conditions for contamination and bias, the model illustrates the bite of the general result and allows us to discuss the bias in special cases of the model.

We consider a dynamic New Keynesian model in which systematic monetary policy may fluctuate for exogenous and endogenous reasons. Monetary policy follows the Taylor rule

$$i_t = \alpha + (\phi + \tilde{\phi}_t)\pi_t + x_t^m, \tag{2.7}$$

where x_t^m follows a stable AR(1) process $x_t^m = \rho_m \log x_{t-1}^m + w_t^m$, where $w_t^m \overset{iid}{\sim} (0, \sigma_m^2)$ denotes a monetary policy shock. As above, we conveniently set $\alpha = -\mathbb{E}[\tilde{\phi}_t \pi_t]$. A second exogenous driving variable is technology, denoted by x_t^a, which also follows a stable AR(1) process $x_t^a = \rho_a \log x_{t-1}^a + w_t^a$, where $w_t^a \overset{iid}{\sim} (0, \sigma_a^2)$ denotes a technology shock. Systematic monetary policy fluctuates according to a stable AR(1) process

$$\tilde{\phi}_t = \rho_\phi \tilde{\phi}_{t-1} + \psi_a w_t^a + q_t, \tag{2.8}$$

which features endogenous movements in response to w_t^a and an exogenous policy shifter $q_t \overset{iid}{\sim} (0, \sigma_q^2)$.

The exogenous drivers of the model are $\{w_t^m, w_t^a, q_t\}$, which we assume to be mutually independent. Absent time-varying systematic monetary policy, the model is a textbook New Keynesian model (Gali, 2015). Based on the derivation in Chapter 1, the approximate equilibrium dynamics of GDP y_t and inflation π_t follow

$$y_t = a^y + b_m^y x_t^m + b_a^y x_t^a + c_m^y x_t^m \tilde{\phi}_t + c_a^y x_t^a \tilde{\phi}_t + d^y \tilde{\phi}_t, \tag{2.9}$$
$$\pi_t = a^\pi + b_m^\pi x_t^m + b_a^\pi x_t^a + c_m^\pi x_t^m \tilde{\phi}_t + c_a^\pi x_t^a \tilde{\phi}_t + d^\pi \tilde{\phi}_t. \tag{2.10}$$

Given this process, the residual in equation (2.5) for $z = y$ is

$$\begin{aligned}\tilde{v}_{y,t+h}^h =& b_m^y \left(\rho_m^{h+1} x_{t-1}^m + \sum_{i=1}^{h} \rho_m^{h-i} w_{t+i}^m \right) + b_a^y x_{t+h}^a + c_m^y \left(x_{t+h}^m \tilde{\phi}_{t+h} - \mathbb{E}[x_{t+h}^m \tilde{\phi}_{t+h}] \right) \\ &+ c_a^y \left(x_{t+h}^a \tilde{\phi}_{t+h} - \mathbb{E}[x_{t+h}^a \tilde{\phi}_{t+h}] \right) + d^y \tilde{\phi}_{t+h}. \end{aligned} \tag{2.11}$$

We next use the model to revisit Proposition 2. In general, it is straightforward to verify that the estimate $\hat{d}_y^h$ is biased and that all three bias terms are non-zero.

The first bias term, $\vartheta^{\hat{b}}$, and thus both expectations $\mathbb{E}[\pi_t w_t^m]$ and $\mathbb{E}\left[\pi_t \tilde{v}_{y,t+h}^h\right]$ are zero only under strong assumptions, such as assuming inflation does not respond to monetary policy shocks, or technology shocks, or systematic monetary policy. The second bias term,

$\vartheta^{\tilde{\phi}}$, requires similarly extreme assumptions to be zero. For example, if the inflation rate indeed depends on $x_t^m \tilde{\phi}_t$ and $x_t^a \tilde{\phi}_t$, i.e., if $c_m^\pi, c_a^\pi \neq 0$, both expectations $\mathbb{E}\left[\tilde{\phi}_t' \pi_t w_t^m\right]$ and $\mathbb{E}\left[\tilde{\phi}_t' \pi_t \tilde{v}_{y,t+h}^h\right]$ are non-zero. The third bias term, ϑ^a, is even more robust. It is generally non-zero whenever $\tilde{\phi}_t$ varies over time. Note that whether or not fluctuations in $\tilde{\phi}_t$ are fully exogenous, $\psi_a = 0$, or allow for endogenous movements is irrelevant for the question of whether $\hat{d}_y^h$ is biased.

We next consider a special case, an environment without persistence, $\rho_m = \rho_a = \rho_\phi = 0$. For $h = 0$, both expectations in $\vartheta^{\hat{b}}$ and $\vartheta^{\tilde{\phi}}$ remain non-zero, respectively. For $h > 0$, the second expectations equal zero, respectively. However, bias remains through the first expectations in $\vartheta^{\hat{b}}$ and $\vartheta^{\tilde{\phi}}$, respectively, and through the third bias, ϑ^a. For $h > 0$, the causal effect is $\delta_y^h = 0$ and bias means we estimate a non-zero effect.

2.3 Empirical evidence on the systematic origins of monetary policy shocks

In this section, we provide empirical evidence suggesting that U.S. monetary policy shocks as identified by the Romer and Romer (2004) (henceforth RR) approach are partly explained by fluctuations in the Federal Reserve's systematic monetary policy.

2.3.1 RR monetary policy shocks

The RR shock is estimated as the residual $\hat{e}_\tau^{rr}$ when estimating a Taylor rule-type regression

$$i_\tau = a + b' x_\tau + e_\tau^{rr}, \tag{2.12}$$

via OLS and where τ denotes FOMC meetings. We have reproduced (2.2) here for the convenience of the reader. RR specify i_τ as the change in the intended federal funds rate between two FOMC meetings. The right-hand side x_τ includes 18 variables: the Greenbook forecast of output growth and inflation, prepared in advance of FOMC meeting τ, respectively for the quarter preceding the FOMC meeting, the current and the two subsequent quarters; the revision of all 8 Greenbook forecasts relative to the same forecasts prepared for the preceding FOMC meeting; the Greenbook forecast of the unemployment rate in the current quarter; and the intended federal funds rate before FOMC meeting τ. We use the estimated monetary policy shocks $\hat{e}_\tau^{rr}$ and associated regressors x_τ from

Wieland and Yang (2020) who extend the RR sample 1969-1996 to 1969-2007.[16]

2.3.2 Measuring time-varying systematic monetary policy

We describe two time series of systematic monetary policy, the Hawk-Dove balance among all Federal Open Market Committee (FOMC) members and the Hawk-Dove balance among the subset of rotating FOMC members. The subsequent description will be relatively brief, with further details and discussion in Chapter 1.

The FOMC has authority over U.S. monetary policy and consists of 12 members, among which four members serve one-year terms on a rotating basis. We use the narrative classification of FOMC members as hawks and doves in Istrefi (2019). The hawk-dove classification is a panel that tracks FOMC members over time at FOMC meeting frequency. Hawks are perceived to be more concerned with inflation, while doves are more concerned with employment and growth.[17] Istrefi (2019) shows that the perceived policy preferences match well with policy tendencies that are unknown in real-time to the public, as expressed by preferred interest rates, with forecasting patterns of individual FOMC members, and with dissents. In addition, Bordo and Istrefi (2023) show that the FOMC members' educational background and early life experience have predictive power for individual policy preferences.

To measure variation in systematic monetary policy over time, we aggregate the individual FOMC member preferences into a Hawk-Dove balance for each meeting (cf. Istrefi, 2019). We do so because the nature of monetary policy-making involves the aggregation of diverse individual policy preferences in a collective decision. We first map the qualitative hawk-dove classification on a numerical scale for FOMC member i at meeting τ ranging from $Hawk_{i\tau}$ =+1 for consistent hawks, +1/2 for hawks who have been doves before, 0 for unclassified member, and -1/2 (-1) for swinging (consistent) doves.[18] We then construct

[16] We end the sample just before the Great Financial Crisis, thus avoiding periods for which policy rules may have changed fundamentally. Likewise, we avoid the period when interest rates reached the effective lower bound. The FFR was kept constant from December 2008 to December 2015 at the zero lower bound.

[17] Among the 147 FOMC members between 1960 and 2023, 129 are classified as hawk or dove. The news coverage for the remaining 18 members is insufficient for classification. 95 classified members are consistently hawks or doves, while the others switch camps at least once. The 34 swinging members switch camps at 1.8% of member-meeting pairs.

[18] In Chapter 1, we show that alternative aggregation schemes lead to similar empirical findings.

the aggregate Hawk-Dove balance in the FOMC by

$$Hawk_{\tau}^{\mathcal{F}} = \frac{1}{|\mathcal{F}_{\tau}|} \sum_{i \in \mathcal{F}_{\tau}} Hawk_{i\tau} \tag{2.13}$$

where $\mathcal{F}_{\tau}$ denotes the (full) set of FOMC members i at meeting τ.[19]

The Hawk-Dove balance may respond to the state of the economy. For example, the Federal Reserve may become more dovish in response to high unemployment or more hawkish in response to high inflation (cf. Davig and Leeper, 2008). Systematic monetary policy may also change in response to political pressure (e.g., Abrams, 2006; Bianchi, Gómez-Cram, Kind, and Kung, 2023). To address the endogeneity of the Hawk-Dove balance, we construct the Hawk-Dove balance among the set of FOMC members who currently have voting rights through the annual rotation.[20] The mechanical nature of the rotation renders it orthogonal to the state of the economy and political cycles. Formally, the Rotation Hawk-Dove balance is defined by

$$Hawk_{\tau}^{\mathcal{R}} = \frac{1}{|\mathcal{R}_{\tau}|} \sum_{i \in \mathcal{R}_{\tau}} Hawk_{i\tau}, \tag{2.14}$$

where $\mathcal{R}_{\tau}$ denotes the set of rotating FOMC members at FOMC meeting τ.[21] While $Hawk_{\tau}^{\mathcal{F}}$ is a more comprehensive measure of systematic monetary policy, $Hawk_{\tau}^{\mathcal{R}}$ has the advantage of reflecting exogenous variation through the rotation.

We present the evolution of $Hawk_{\tau}^{\mathcal{F}}$ and $Hawk_{\tau}^{\mathcal{R}}$ from 1960 through 2023 in Figure 2.3.1. Both balances vary considerably, featuring hawkish and dovish majorities. The variation reflects the turnover of rotating FOMC members, the turnover of non-rotating FOMC members, and changes in policy preferences of incumbent FOMC members. The correlation between $Hawk_{\tau}^{\mathcal{F}}$ and $Hawk_{\tau}^{\mathcal{R}}$ is 0.60, see Table 2.A.1 for further descriptive statistics. Fluctuations in $Hawk_{\tau}^{\mathcal{R}}$ are more short-lived, reflecting the annual rotation of voting rights.

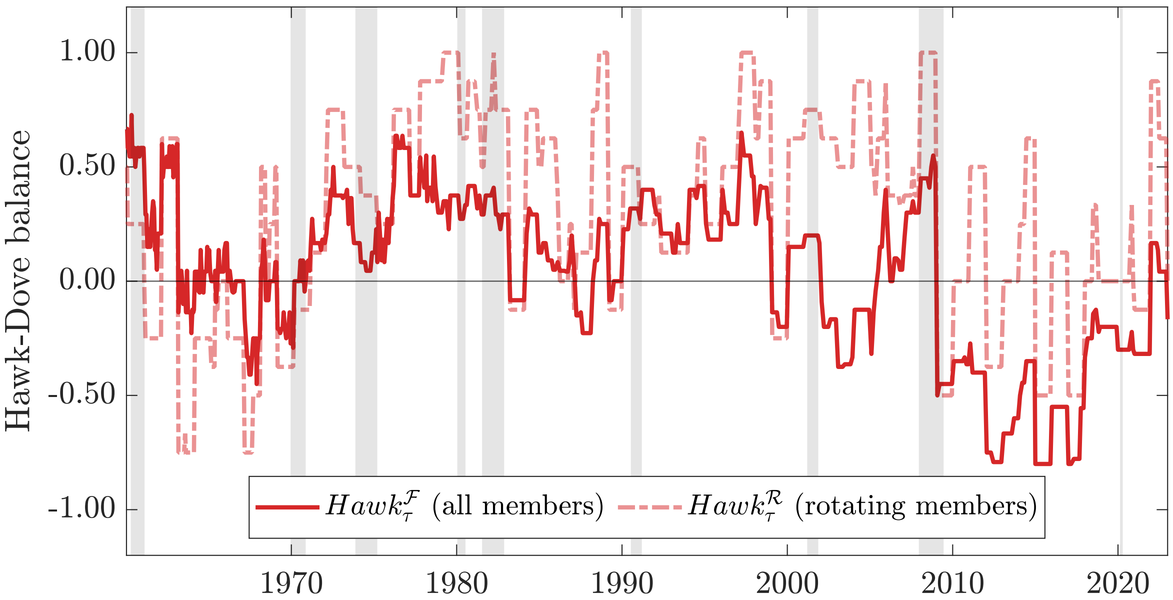
The Hawk-Dove balance is informative about systematic monetary policy. First, the classification matches well with narratives of monetary policy in the U.S. (Istrefi, 2019). Second, a hawkish FOMC responds to higher inflation by raising the policy rate more aggressively (Bordo and Istrefi, 2023, see also Appendix 1.D of Chaper 1). Finally, a hawkish FOMC tightens monetary policy more aggressively in response to expansionary government spend-

[19] Occasionally, $|\mathcal{F}_{\tau}| \leq 12$ because of absent members and vacant positions. Note that the set $\mathcal{F}_{\tau}$ is identical to $\mathcal{A}_{\tau}$ in Chapter 1.

[20] This is the same variable as the FOMC rotation instrument from Chapter 1.

[21] In our sample, $|\mathcal{R}_{\tau}| = 4$ for 625 out of 634 FOMC meetings and $|\mathcal{R}_{\tau}| = 3$ for the remaining meetings.

Figure 2.3.1: Hawk-Dove balance in the FOMC



Notes: The solid red line shows the aggregate Hawk-Dove balance of the full FOMC $Hawk^{\mathcal{F}}_{\tau}$ at FOMC meeting frequency from 1960 through 2023. The dashed red line shows the aggregate Hawk-Dove balance of the rotation panel $Hawk^{\mathcal{R}}_{\tau}$. Grey bars indicate NBER dated recessions.

ing shocks, leading to a significantly dampened GDP expansion, see Chapter 1.

2.3.3 Predictability of RR shocks

In this section, we show that RR monetary policy shocks are predictable by fluctuations in measured systematic monetary policy in a way that supports our theoretical results. If systematic monetary policy is time-varying as in (2.1), the estimated residuals $\hat{e}^{rr}_{\tau}$ based on the regression in (2.2) will include fluctuations in systematic monetary policy multiplied with the inputs of the Taylor rule, the term $\tilde{\phi}_{\mathcal{T}} x_{\mathcal{T}}$ in Proposition 1. Hence, a testable prediction of a time-varying Taylor rule is that fluctuations in measured systematic monetary policy multiplied with x_{τ} partly explain conventional RR shocks $\hat{e}^{rr}_{\tau}$. To test this prediction we estimate the following regression

$$\begin{aligned} \hat{e}^{rr}_{\tau} = & \beta_0 + \beta_1' \, x_{\tau-p} Hawk_{\tau-p} + \beta_2' \, x_{\tau-p} \Delta Hawk_{\tau-p} \\ & + \beta_3' \, Hawk_{\tau-p} + \beta_4' \, \Delta Hawk_{\tau-p} + \beta_5' \, x_{\tau-p} + u_{\tau}, \end{aligned} \tag{2.15}$$

where τ denotes an FOMC meeting, $\hat{e}^{rr}_{\tau}$ is the RR shock and x_{τ} the RR regressors, both as defined in Section 2.3.1, $Hawk_\tau$ is either $Hawk_\tau^{\mathcal{F}}$ or $Hawk_\tau^{\mathcal{R}}$, and $\Delta Hawk_\tau$ is the first difference of $Hawk_\tau$.[22] We consider contemporaneous regressors ($p=0$) or lags up to two meetings ($p=1,2$). Our motivation to consider lags is to capture that it may take time for FOMC members to affect policy decisions, as argued in Chapter 1.[23]

Table 2.3.1 presents the R^2 for various specifications of (2.15), as well as the p-values for the null hypothesis that all coefficient estimates are jointly zero, estimated for different sample periods. Across regression specifications, we obtain an R^2 between 0.1 and 0.54. Using a one-meeting lag ($p=1$) yields the largest R^2 ranging from 0.33 to 0.54. In other words, a sizable fraction of the variation in RR shocks can be explained by past variables, irrespective of the type of Hawk-Dove balance and the three sample specifications. For $p=1$, we can reject the null hypothesis that all coefficient estimates are zero at the 1% significance level. Except the post-Volcker sample, the R^2 is lower for $p=0$, and the R^2 is also lower for $p=2$. Our finding that lagged regressors raise predictability is consistent with the nature of decision-making in the FOMC.

We next investigate the contribution of the individual regressors for explaining variation in RR shocks. We focus on the regression specification of (2.15) that yields the largest (total) R^2 in Table 2.3.1, i.e., $Hawk_{\tau-p}^{\mathcal{R}}$ and $p=1$, but we obtain similar results for the other specifications. Table 2.3.2 reports the R^2 and p-value when regressing the RR shock $\hat{e}_\tau^{rr}$ separately on subsets of the regressors included in equation (2.15). The first key takeaway is that the interactions between $x_{\tau-1}$ and, respectively, $Hawk_{\tau-1}^{\mathcal{R}}$ and $\Delta Hawk_{\tau-1}^{\mathcal{R}}$ account for the bulk of the total R^2. Both $Hawk_{\tau-1}^{\mathcal{R}}$ and $\Delta Hawk_{\tau-1}^{\mathcal{R}}$ are informative (in interaction with $x_{\tau-1}$), with the latter having more predictive power. The second key takeaway is that the (non-interacted) level of $Hawk_{\tau-1}^{\mathcal{R}}$ and $\Delta Hawk_{\tau-1}^{\mathcal{R}}$ has practically no predictive power for the RR shock. Importantly, this finding further supports the interpretation of the Hawk-Dove balance as capturing variation in systematic monetary policy. If, in contrast, the Hawk-Dove balance also captured significant information about the intercept of the monetary policy rule, we could expect the level of the Hawk-Dove balance to have predictive power for the RR shock. Finally, the regressor $x_{\tau-1}$ has some predictive power in explaining

[22] In Chapter 1, the rotation Hawk-Dove balance is proposed as an instrument to provide causal evidence on the state-dependent effects of macroeconomic shocks with variation in respect to systematic monetary policy. In this chapter, we do not use the rotation Hawk-Dove balance explicitly as an instrument because the high number of regressors in our empirical application would render an IV approach unreliable.

[23] For example, former Governor Laurence Meyer remarks: *I came to see policy decisions as often evolving over at least a couple of meetings. The seeds were sown at one meeting and harvested at the next. [...] Similarly, while in my remarks to my colleagues it sounded as if I were addressing today's concerns and today's policy decisions, in reality I was often positioning myself, and my peers, for the next meeting.* Laurence Meyer (2004), A Term at the Fed: An Insiders' View, Harper Business.

Table 2.3.1: Explaining RR shocks by systematic monetary policy

	$Hawk_\tau^{\mathcal{F}}$			$Hawk_\tau^{\mathcal{R}}$		
Sample	69-07	69-96	83-07	69-07	69-96	83-07
(a) Contemporaneous FOMC meeting (p=0)						
R^2	0.098	0.133	0.432	0.167	0.216	0.464
p-value	0.248	0.239	0.000	0.003	0.000	0.000
T	354	266	200	354	266	200
(b) One FOMC meeting lag (p=1)						
R^2	0.333	0.430	0.451	0.429	0.541	0.443
p-value	0.002	0.002	0.000	0.000	0.000	0.000
T	350	262	200	350	262	200
(c) Two FOMC meetings lag (p=2)						
R^2	0.236	0.313	0.373	0.279	0.360	0.422
p-value	0.000	0.000	0.000	0.000	0.000	0.000
T	348	260	200	348	260	200

Notes: The table shows results from regressions based on (2.15). The rows of the three subtables show R^2, the p-values for the null hypothesis that all coefficient estimates are jointly zero, and the number of observations T. The three left columns show results for the Hawk-Dove balance across all FOMC members, the three right columns for the Hawk-Dove balance across all rotating FOMC members with voting rights. Columns one to three show differ by the sample period between 1969-2007, 1969-1996, and 1983-2007, and analogously for columns four to six. The three subtables differ by the specification of FOMC meeting lag p.

RR shocks, in particular for the post-Volcker sample. The results in Table 2.3.2 differ by little across the three samples. Overall, our results suggest that a substantial fraction of the conventional RR shocks can be explained by variation in systematic monetary policy. Our evidence thus supports the notion that empirical monetary policy shocks have systematic monetary policy origins.

A potential concern with our results is that the large set of regressors we include might lead to overfitting. We may mechanically absorb variation, although there is no systematic relationship in the data. To address this concern, we use a Lasso estimation, which minimizes the sum of squared residuals (as in OLS) but additionally penalizes the number of estimated parameters to keep the set of included regressors small. We choose the penal-

Table 2.3.2: Explaining RR shocks by subsets of regressors

	Interactions			Levels		
Sample	69-07	69-96	83-07	69-07	69-96	83-07
	(a) $Hawk^{\mathcal{R}}_{\tau-1} \times x_{\tau-1}$			(b) $Hawk^{\mathcal{R}}_{\tau-1}$ & $\Delta Hawk^{\mathcal{R}}_{\tau-1}$		
R^2	0.112	0.138	0.117	0.006	0.010	0.002
p-value	0.087	0.058	0.034	0.370	0.330	0.826
	(c) $\Delta Hawk^{\mathcal{R}}_{\tau-1} \times x_{\tau-1}$			(d) $x_{\tau-1}$		
R^2	0.248	0.289	0.065	0.090	0.133	0.255
p-value	0.000	0.000	0.000	0.031	0.005	0.000
	(e) All interactions			(f) All level terms		
R^2	0.341	0.399	0.193	0.096	0.151	0.255
p-value	0.000	0.000	0.000	0.039	0.001	0.000
T	350	262	200	350	262	200

Notes: The table shows results from regressions based on (**2.15**), considering different subsets of the regressors. The rows of the three subtables show R^2 and the p-values for the null hypothesis that all coefficient estimates are jointly zero, and the number of observations T. The three left columns show results for the interactions between the Hawk-Dove balance and $x_{\tau-1}$, the three right columns show the results for the non-interacted (level) regressors. Columns one to three show differ by the sample period between 1969-2007, 1969-1996, and 1983-2007, and analogously for columns four to six.

ization parameter to gradually increment the number of regressors from one to five. We present the results for the sample 1969 - 2007 in Table 2.B.1, in Appendix 2.B. We find that five (scalar) regressors are sufficient to yield an R^2 of 0.15. All five regressors selected by the Lasso estimation involve interactions of elements in $x_{\tau-1}$ with $\Delta Hawk^{\mathcal{R}}_{\tau-1}$. The elements included are one Greenbook forecasts of output growth, two inflation forecasts, and two inflation forecast revisions.

In related work, Aruoba and Drechsel (2024) refine the RR shock by using a large vector x_t with the goal of better capturing the Fed's information set about the state of the economy. They use textual analysis to create sentiment indicators about the Fed staff's assessment of the economy before FOMC meetings. The sentiment indicators are used as additional regressors in Taylor rule-type Ridge regression. We regress their shock on our right-hand side variables in (2.15). We find an R^2 between 0.26 and 0.35 depending on lag order

$(p = 0,1,2)$ and the type of Hawk-Dove balance. We obtain the largest R^2 for $p = 2$ and $Hawk_\tau^{\mathcal{R}}$. Thus, even their refined shock is predictable and, hence, may be contaminated by time variation in systematic monetary policy.

2.4 A new monetary policy shock

The results in Section 2.3 motivate us to construct a new monetary policy shock that is no longer predictable by measured systematic monetary policy. The shock is the estimated residual when regressing policy rate changes on Greenbook forecasts as well as the interaction between Greenbook forecasts and measured time variation in systematic monetary policy. We find that our new monetary policy shocks affect output and inflation more strongly with a substantially shorter delay, and at higher statistical significance compared to the RR shock, in particular for a post-Volcker sample.

2.4.1 Shock identification

We estimate a new monetary policy shock series via the augmented Taylor rule regression

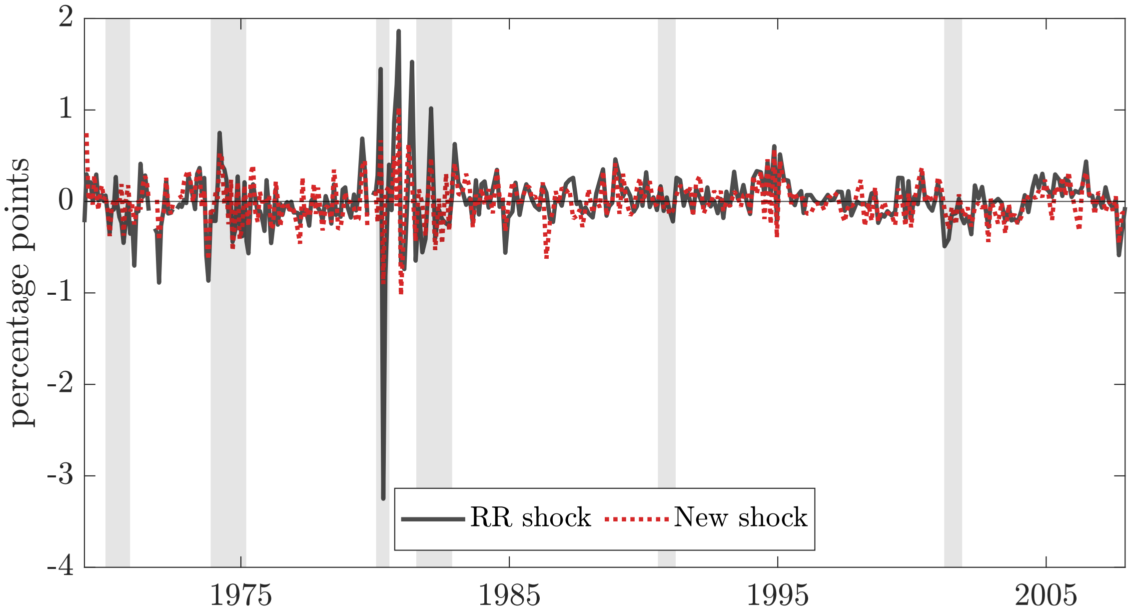
$$\begin{aligned} i_\tau = &\beta_0 + \beta_1' x_\tau + \beta_2' x_{\tau-1} + \beta_3' \, x_{\tau-1} Hawk_{\tau-1} + \beta_4' \, x_{\tau-1} \Delta Hawk_{\tau-1} \\ &+ \beta_5' \, Hawk_{\tau-1} + \beta_6' \, \Delta Hawk_{\tau-1} + e_\tau^{new}, \end{aligned} \tag{2.16}$$

where the policy instrument i_τ and Greenbook forecast x_τ are specified as in Section 2.3 and $Hawk_\tau$ is the Rotation Hawk-Dove balance. Our new monetary policy shock is the estimated residual $\hat{e}_\tau^{new}$ when estimating (2.16) via OLS.[24] The specification nests the original Romer and Romer (2004) regression if we restrict $\beta_j = 0 \; \forall j > 1$, in which case we denote the estimated residual by $\hat{e}_\tau^{rr}$. Our baseline sample to identify the shock is the full sample from 1969 through 2007. We discuss the robustness of our results for alternative samples in Section 2.4.3.

Figure 2.4.1 shows the new shock series (red dotted line) in comparison with the original RR shock series (black solid line). The new series is substantially less dispersed than the RR shock with the standard deviation falling from 0.34 to 0.23 (see Table 2.4.1). The overall correlation between new and RR shock is 0.67. The correlation between the sign of

[24] While we follow RR in using OLS, this leads to endogeneity bias as discussed in Section 2.2. However, Carvalho, Nechio, and Tristão (2021) argue that the endogeneity bias is quantitatively negligible. Finally note that estimating (2.16) via IV is practically challenging because it would require a large number of instruments.

Figure 2.4.1: Time series of monetary policy shocks



Notes: The solid black line shows the RR shock $\hat{e}^{rr}_{\tau}$ based on the regression in (2.16) when restricting $\beta_j = 0 \; \forall j > 1$. The dotted red line shows the new shock $\hat{e}^{new}_{\tau}$ based on the regression in (2.16). The sample period is 1969 through 2007. Grey bars indicate NBER recession.

both shocks is 0.42, meaning new and RR shock frequently have the opposite sign. Both the new shock and the RR shock exhibit practically no serial correlation.[25]

Table 2.4.1: Descriptive statistics of monetary policy shocks

	Mean	Median	SD	Autocorr	Corr	Sign-corr	Min	Max	T
RR shock	0.00	-0.01	0.34	0.12	-	-	-3.25	1.86	354
New shock	-0.00	-0.01	0.23	-0.09	0.67	0.42	-1.03	1.03	350

Notes: The table shows descriptive statistics for the new shock ($\hat{e}^{new}_{\tau}$) and the RR shock ($\hat{e}^{rr}_{\tau}$) at FOMC meeting frequency from 1969 through 2007. "Autocorr" refers to the meeting-over-meeting autocorrelation. "Corr" refers to the correlation between new and RR shock. "Sign-corr" refers to the correlation of the sign of both shock series.

The two shock series most visibly differ during 1979-1982 with our new shock being smaller in magnitude. RR argue that their shocks in this period reflect changes in the Federal Reserve's operating procedures and an increased distaste for inflation. In fact, we do observe a relatively hawkish FOMC, in particular among rotating FOMC members (see

[25] For four FOMC meetings, x_{τ} is missing because not all Greenbook forecasts are available. The regression for the new shock (2.16) includes $x_{\tau-1}$ creating four additional missing observations.

Figure 2.3.1). Hence, a plausible reason for our new shocks being smaller is that accounting for variation in systematic monetary policy better explains variation in monetary policy during this episode.

2.4.2 Impulse responses

We next compare impulse response estimates for alternative monetary policy shock series.

Econometric framework. We estimate impulse responses using the local projections

$$z_{t+h} - z_{t-1} = \alpha_z^h + \beta_z^h \ \hat{e}_t + \Gamma \ Y_t + v_{t+h}^h, \quad h = 0, \ldots, H, \tag{2.17}$$

where z_t is an outcome variable of interest. The main outcomes of our analysis are the federal funds rate, the inflation rate, and log real GDP. The monetary policy shock $\hat{e}_t$ is either the new shock $\hat{e}_t^{new}$ or the RR shock $\hat{e}_t^{rr}$. The control vector Y_t includes twelve lags of the federal funds rate, the inflation rate, the log of real GDP, and a linear time trend. A period t is a month. This is a common choice in the related literature and limits the need to aggregate the monetary policy shocks.[26] Monthly log real GDP and the monthly GDP deflator inflation rate are obtained by interpolation using the procedure of Chow and Lin (1971).[27] Section 2.4.3 explores the sensitivity of our results on monthly interpolation and the specification of the control vector. The baseline sample of our analysis is 1983 through 2007, so post-Volcker disinflation and pre-Great Recession. We consider this sample particularly interesting because the estimated responses to many conventional monetary policy shock series appear implausible in such sample (e.g., Ramey, 2016). This sample further avoids potential structural breaks around the Great Inflation episode. Section 2.4.3 explores the sensitivity of our results regarding the sample.

Responses of main outcomes. Figure 2.4.2 presents the estimated responses of our main outcome variables, the federal funds rate, the inflation rate, and log real GDP, to the new shock and the RR monetary policy shock. The key takeaway is that the two shocks

[26] Only 4 months (all between 1969 through 1971) contain more than one FOMC meeting with a monetary policy shock $\hat{e}_\tau$, while a large fraction of quarters across the entire sample contain multiple $\hat{e}_\tau$. In months in which we observe at least one $\hat{e}_\tau$, we construct $\hat{e}_t$ as the sum of $\hat{e}_\tau$ contained in t. Otherwise, we set $\hat{e}_t = 0$.

[27] The related monthly series we use for interpolating GDP and the GDP deflator are CPI, industrial production, one-year treasury yield, and excess bond premium.

differ substantially in the estimated lag of monetary policy transmission, the magnitude of the responses, and statistical significance.

The two panels in the first row of Figure 2.4.2 show the estimated response of the federal funds rate (FFR) to the two shocks. While the left-hand side panel shows the 68% and 95% confidence bands for the new shock, the right-hand side panel shows the corresponding confidence bands for the RR shock. The confidence bands are based on standard errors that are robust to heteroskedasticity and serial correlation. Both shocks are normalized to increase the FFR by 100 basis points on impact. This facilitates comparability of the inflation and GDP responses. The dynamic FFR responses differ markedly in magnitude and persistence. The new shock leads to a peak FFR response of 2.4 percentage points after 6 months and quickly reverts to zero. The RR shock leads to a peak FFR response of 3.7 percentage points after 8 months and remains significantly above zero for 18 months. Figure 2.4.3(a) shows that the difference between the FFR responses is statistically significant at the 5% level for some horizons after 15 months, precisely where the RR shock has more persistent effects.[28] Overall, the new monetary policy shock leads to a less strong and more transitory dynamic federal funds rate response than the RR shock. If both shocks are well identified, we might expect a larger demand contraction from the RR shock.

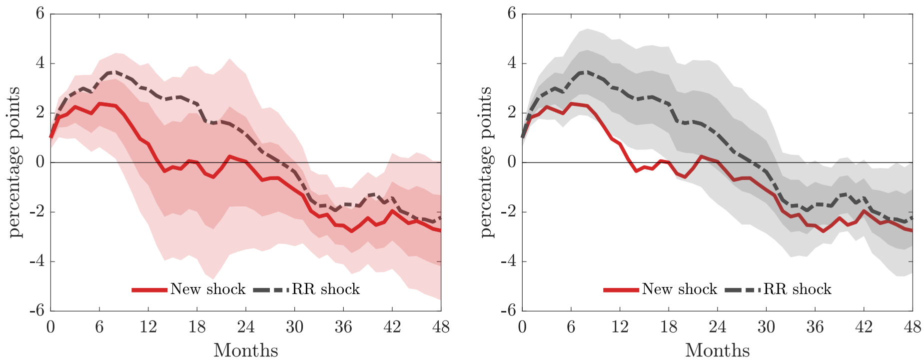
The second row of Figure 2.4.2 shows the estimated responses of real GDP. The new shock leads to a contraction of GDP that is mostly significant at the 5% level between 14 months and 37 months after the shock. In contrast, the GDP response to the RR shock is statistically insignificant at the 5% level for all horizons we consider. If we use the (much) lower 32% significance standard, then the new shock leads to a significant GDP contraction starting 8 months after the shock, while it takes 20 months for the RR shock. In addition, the RR shock generates a short-lived significant expansion around 6 months after the shock (an output puzzle). The difference between the two GDP responses is statistically significant for most horizons between 6 and 20 months, see Figure 2.4.3(b). The shocks further differ strongly in the magnitude of the GDP response. Despite the larger and more persistent FFR increase for the RR shock, the GDP contraction is substantially larger for the new shock. The trough response is -3.2 percent for the new shock and -1.4 percent for the RR shock.

Finally, the third row shows the estimated responses of inflation. Arguably, the most striking finding of Figure 2.4.2 is the difference in the lag of monetary policy affecting

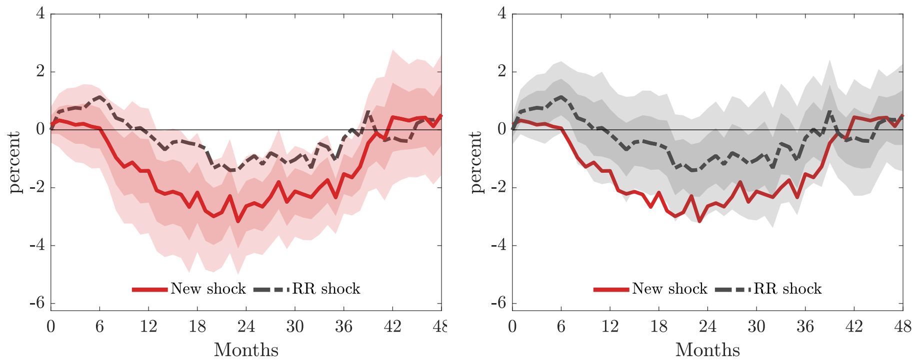
[28] The standard errors for the difference across impulse responses are constructed by estimating both local projections as seemingly unrelated regressions and estimating the joint covariance matrix via Driscoll-Kraay.

Figure 2.4.2: Responses of main outcomes to monetary policy shocks

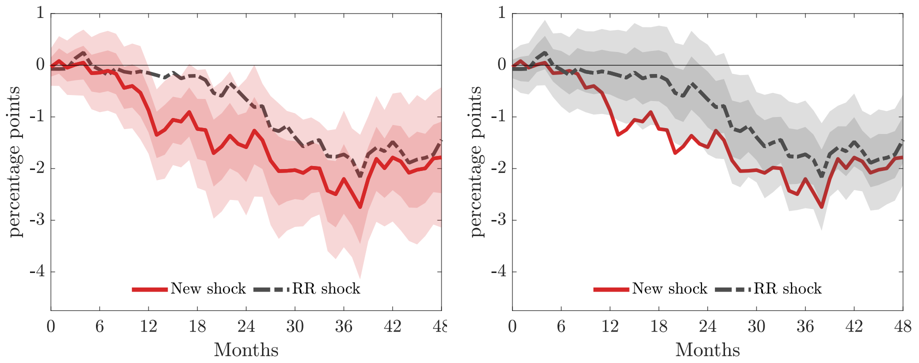
(a) Federal funds rate



(b) Real GDP



(c) Inflation

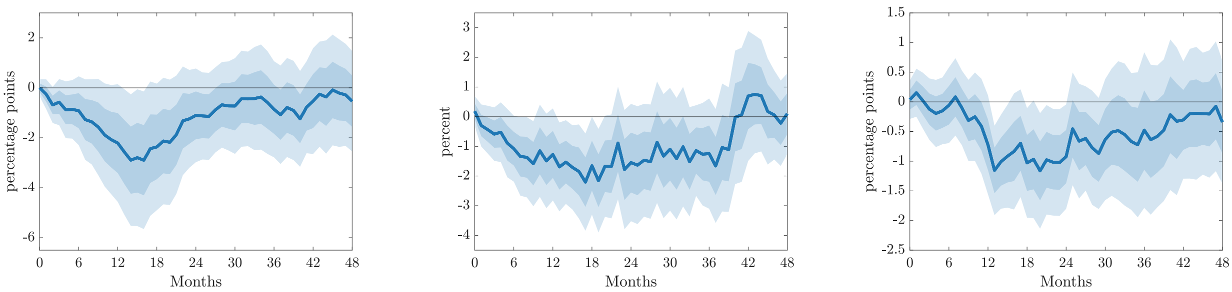


Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The shaded areas indicate 68% and 95% confidence bands using standard errors robust to serial correlation and heteroskedasticity.

inflation. The inflation response becomes 5% significant only after 27 months for the RR shock but after 13 months for the new shock. Our new shock shows that monetary policy

Figure 2.4.3: Response to new shock "minus" response to RR shock

(a) FFR (b) Real GDP (c) Inflation



Notes: The figure shows the differences across impulse responses for the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The difference is computed as the response to the new shock minus the response to the old shock for each outcome, respectively. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The shaded areas indicate 68% and 95% confidence bands using standard errors robust to serial correlation and heteroskedasticity.

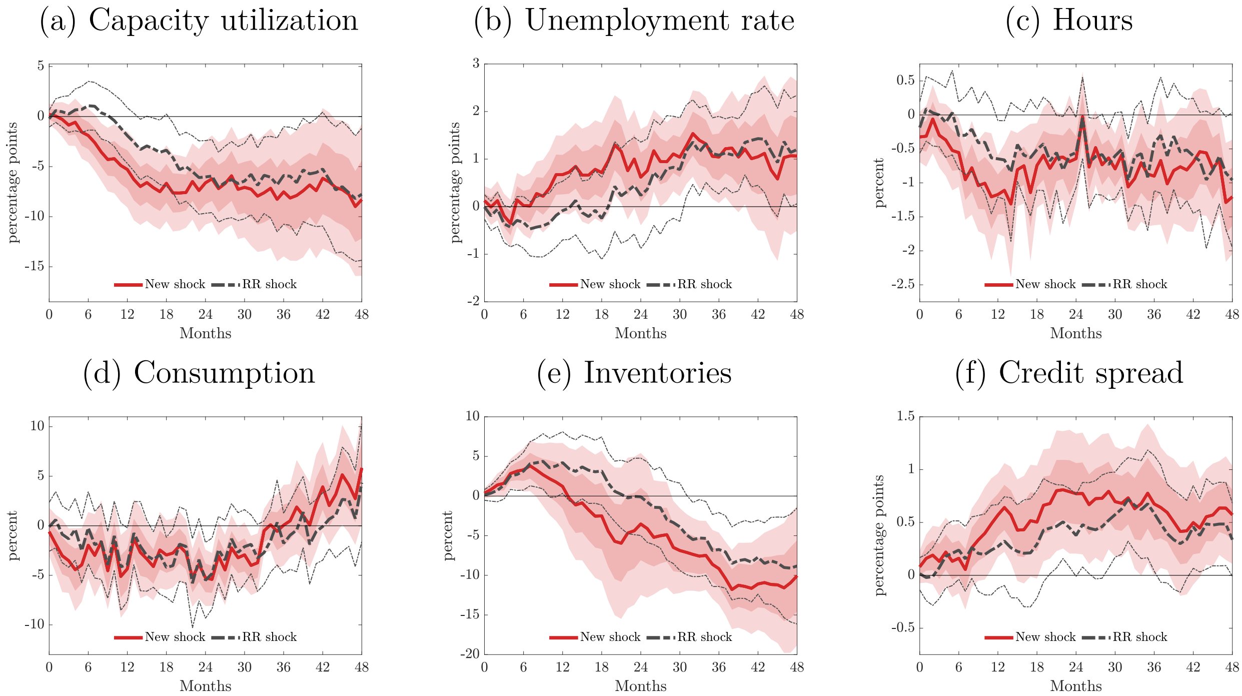
shocks affect inflation at substantially shorter lags than what the RR shock suggests. The difference between the inflation responses is particularly significant between 12 and 28 months, see Figure 2.4.3(c). We further uncover some differences in magnitudes. The trough response is -2.7% for the RR shock and -2.2% for the new shock.

Overall, our results suggest that accounting for time variation in systematic monetary policy is critically important when identifying monetary policy shocks. Disregarding variation in systematic monetary policy may lead to strongly biased impulse response estimates and an inaccurate assessment of the effectiveness of monetary policy. It may further bias analyses using impulse responses estimates to estimate DSGE models, construct policy counterfactuals, or investigate the optimality of monetary policy.

Further outcome variables. In Figure 2.4.4, we extend the analysis to further outcome variables, notably capacity utilization, unemployment, hours worked, consumption, inventories, and a credit spread. The variables are informative about the transmission mechanism of monetary policy. In addition, they further show that accounting for systematic monetary policy matters.

In response to the new shock, we find a significant decrease in capacity utilization, increase in the unemployment rate, and decrease in hours worked. All measures suggest an increase of slack in the economy. The responses to the RR shock are broadly similar. However, they suggest (again) a substantially longer lag of monetary policy and the responses are less precisely estimated. We report the differences of responses and confidence bands in

Figure 2.4.4: Response of further outcomes to monetary policy shocks



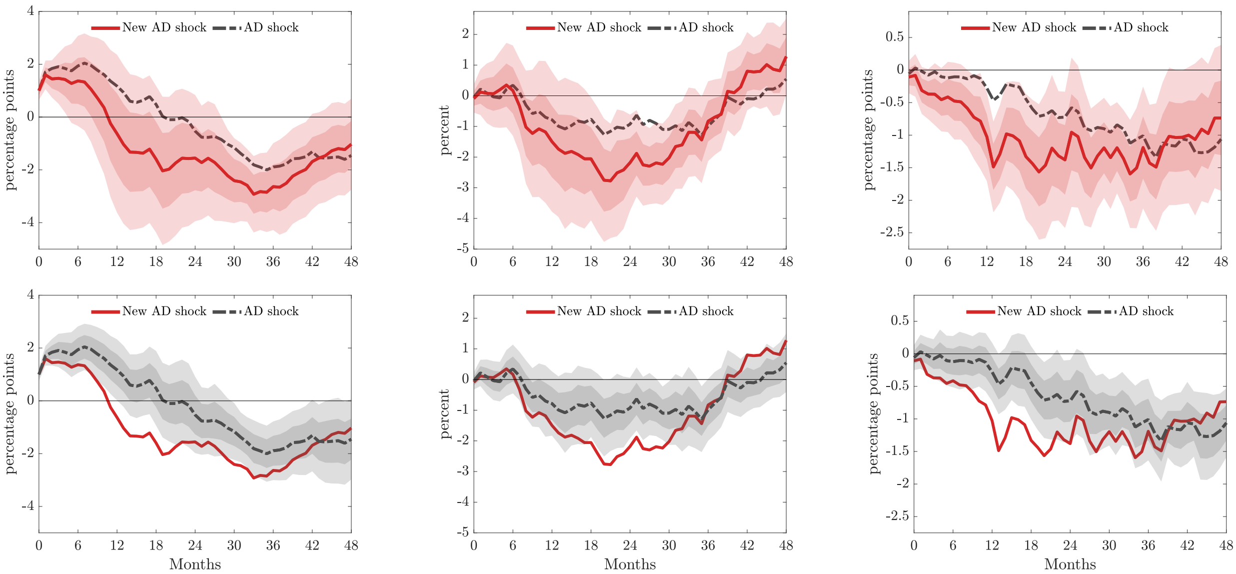
Notes: The figure shows responses of capacity utilization, the unemployment rate, log consumption expenditures, log business inventories, log hours (in manufacturing), and credit spreads (BAA- minus AAA-rated corporate bond yield) to a monetary policy shock based on the local projection as specified along with (2.17). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The shaded areas indicate 68% and 95% confidence bands for the new shock, and the dotted lines indicate the 95% confidence band for the conventional shock using standard errors robust to serial correlation and heteroskedasticity for all bands.

Figure 2.C.1 in Appendix 2.C. The response of consumption expenditures to the new shock is much quicker and occurs within the first six months. Beyond the short-run, however, the response of consumption is highly similar across the new shock and the RR shock, suggesting that investment, government spending, or net exports respond quite differently to the two shocks. Business inventories initially increase, consistent with a surprise reduction in demand, and then fall. The reduction in inventories is significantly more pronounced for the new shock consistent with the more rapid decline in capacity utilization. Finally, the yield spread between BAA- and AAA-rated corporate bonds responds more strongly and significantly to the new shock. Overall, we find substantial differences in the response of these outcomes, and the effects of the new shock tend to be stronger and more significant.

Aruoba and Drechsel (2022) shock. Section 2.3 provides evidence suggesting that the shock constructed by Aruoba and Drechsel (2024) (AD shock henceforth) may be

Figure 2.4.5: Comparison of main responses with Aruoba and Drechsel (2024) shock

(a) FFR (b) Real GDP (c) Inflation



Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) where we put the Aruoba and Drechsel (2024) shock on the left-hand-side, whereas the conventional monetary policy shock is taken directly from Aruoba and Drechsel (2024). The shaded areas indicate 68% and 95% confidence bands using standard errors robust to serial correlation and heteroskedasticity.

contaminated by systematic monetary policy. We next compare the impulse response estimates between the AD shock and a new AD shock, which is the residual when regressing the AD shock on the right-hand side variables of (2.16). Figure 2.4.5 shows that the new AD shock leads to a more short-lived response of the federal funds rate, a stronger decline of real GDP declines, and a substantially shorter lag in the inflation response, when compared to the original AD shock. The differences are sizable and statistically significant at some horizons, see Figure 2.C.2 in Appendix 2.C.

2.4.3 Sensitivity analysis

In this section, we provide a sensitivity analysis to assess the robustness of our baseline results. We investigate how our results depend on varying sample periods, the inclusion of additional control variables, and alternative measures of economic activity and prices. We summarize our findings in the following, but delegate all figures to Appendix 2.C.

Alternative sample periods. Our baseline shock measure is estimated on the full sample of Greenbook forecasts from 1969 through 2007, but the impulse responses presented above are estimated on the post-1983 sub-sample. We analyze whether our estimated responses differ if the shock identification regressions (2.16) for both the RR shock and our new shock are estimated on the same post-1983 sub-sample. Figure 2.C.3 shows that the inflation response to the RR shock features a similarly long lag as in the baseline. The GDP response to the RR shock is insignificant but rather expansionary. In contrast, the response of inflation to our new shock remains similar to the baseline, while the GDP response remains contractionary but is less significant than in the baseline. We further estimate impulse responses on the full sample (1969-2007) instead of the post-1983 sample and report the results in Figure 2.C.4. Similar to the baseline, we find that the new shock delivers a significantly stronger contraction in real GDP. Interestingly, the inflation response is similar across both shocks for around two years and features a price puzzle.[29] At longer horizons, however, the new shock leads to a stronger inflation decline. The large difference between estimating the impulse responses on the full-sample vis-à-vis the post-1983 may potentially arise because of the structural breaks around the Great Inflation and subsequent disinflation, which our linear local projection does not model.

Additional control variables. Romer and Romer (2004) and Coibion (2012) impose a recursiveness assumption by including contemporaneous real GDP and inflation as control variables. In effect, these variables cannot contemporaneously respond to the monetary policy shock. Figure 2.C.5 shows that our results are highly similar to the baseline imposing the recursiveness assumption. Parts of the related literature control for lags of the log S&P 500 and the Gilchrist and Zakrajšek (2012) excess bond premium (e.g., Jarociński and Karadi, 2020). Figure 2.C.6 shows that our estimated responses are similar to the baseline when adding twelve lags of the two control variables. Finally, some of the related literature controls for lags of the RR shocks (see, e.g., Ramey, 2016). Figure 2.C.7 shows that our results hardly change when adding twelve lags of the shock under consideration to the baseline set of control variables.

Alternative outcome variables. Our baseline results use interpolated real GDP and the GDP deflator to measure economic activity and prices at monthly frequency as similarly done in Aruoba and Drechsel (2024) and Jarociński and Karadi (2020). An alternative is

[29]Including twelve lags of the log commodity price index (or its growth rate) resolves the price puzzle.

to use industrial production (IP) and CPI inflation, which are readily available at monthly frequency (e.g. Bauer and Swanson, 2023a; Gertler and Karadi, 2015). Figure 2.C.8 shows the responses of IP and CPI. The differences between the new and the RR shock remain similar to the baseline. However, the IP response is not very precisely estimated for the new shock. If we further control for twelve lags of the Gilchrist and Zakrajšek (2012) excess bond premium and the log S&P 500, the IP response is more precisely estimated, see Figure 2.C.9.

2.5 Conclusion

This chapter revisits conventional empirical strategies to estimate monetary policy shock series. We show theoretically that fluctuations in systematic monetary policy lead to misidentified shocks and bias in the estimated impulse responses. We provide empirical evidence to support the theory. We find that Romer and Romer (2004) monetary policy shocks are predictable by fluctuations in measured systematic monetary policy. We construct a new shock series that is orthogonal to systematic monetary policy and assess its effects on the U.S. economy. Our shock suggests monetary policy has shorter lags and stronger effects on inflation and output relative to comparable evidence for the Romer and Romer (2004) shock.

Appendix

2.A Data

Table 2.A.1: Descriptive statistics of the Hawk-Dove balances

	Mean	Median	SD	Autocorr	Corr	Min	Max	T
$Hawk_{\tau}^{\mathcal{F}}$	0.06	0.10	0.34	0.95	-	-0.80	0.73	630
$Hawk_{\tau}^{\mathcal{R}}$	0.24	0.25	0.47	0.91	0.60	-0.75	1.00	630

Notes: This table shows descriptive statistics for the time series at FOMC meeting frequency from 1960 through 2023. $Hawk_{\tau}^{\mathcal{F}}$ is the average Hawk-Dove balance of the FOMC. $Hawk_{\tau}^{\mathcal{R}}$ is the FOMC rotation instrument. "Autocorr"refers to the meeting-over-meeting autocorrelation. "Corr" refers to the correlation between both series.

2.B Additional results for Section 2.3

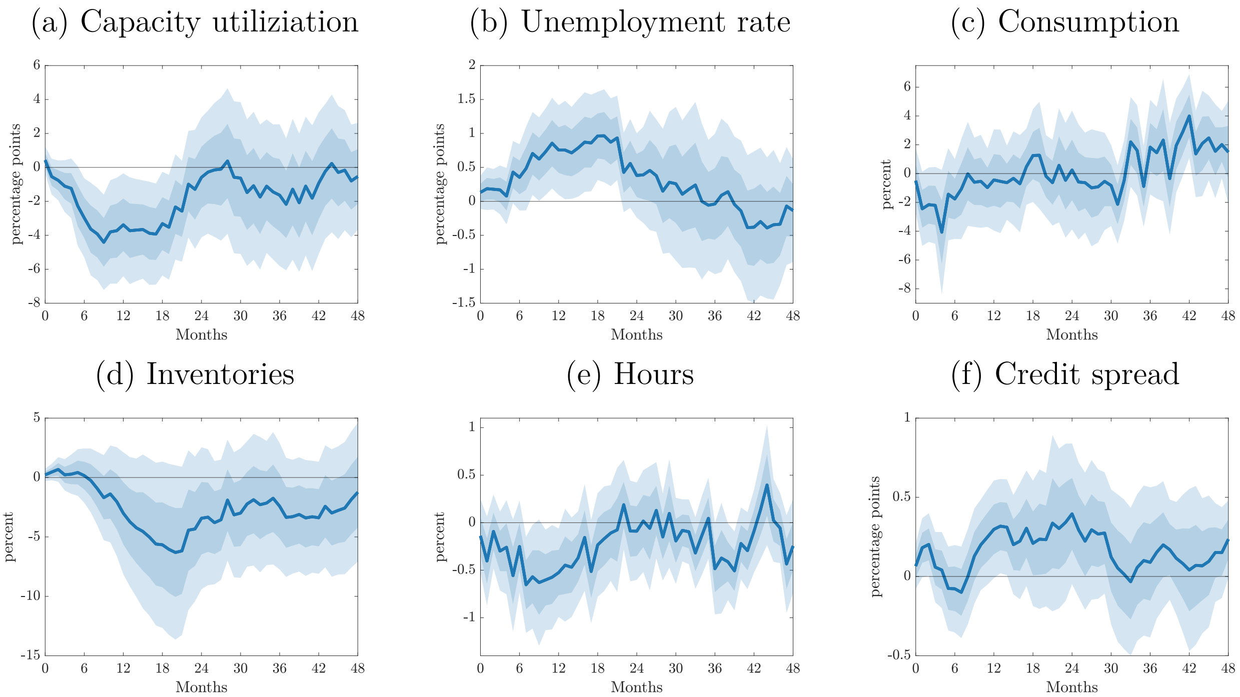
Table 2.B.1: Lasso estimation to explain RR shocks

	(1)	(2)	(3)	(4)	(5)
$\Delta Hawk^{\mathcal{R}}_{\tau-1} \times y_{\tau-1,2}$	-0.195	-0.148	-0.138	-0.107	-0.082
	(0.300)	(0.368)	(0.346)	(0.301)	(0.367)
$\Delta Hawk^{\mathcal{R}}_{\tau-1} \times \Delta\pi_{\tau-1,-1}$		0.149	0.111	0.233	0.224
		(0.137)	(0.244)	(0.047)	(0.054)
$\Delta Hawk^{\mathcal{R}}_{\tau-1} \times \pi_{\tau-1,1}$			0.133	0.076	-0.226
			(0.262)	(0.338)	(0.400)
$\Delta Hawk^{\mathcal{R}}_{\tau-1} \times \Delta\pi_{\tau-1,1}$				0.222	0.273
				(0.032)	(0.026)
$\Delta Hawk^{\mathcal{R}}_{\tau-1} \times \pi_{\tau-1,2}$					0.325
					(0.267)
Constant	0.007	0.002	0.003	0.007	0.006
	(0.713)	(0.917)	(0.864)	(0.678)	(0.715)
T	350	350	350	350	350
R^2	0.046	0.067	0.086	0.145	0.154

Notes: The table shows Lasso regression results based on (2.15). The Lasso shrinkage parameter is chosen to increment the number of regressors from one to five, and the associated results are presented in columns one to five, respectively. The time sample runs from 1969 through 2007, and standard errors robust to serial correlation and heteroskedasticity are in parentheses.

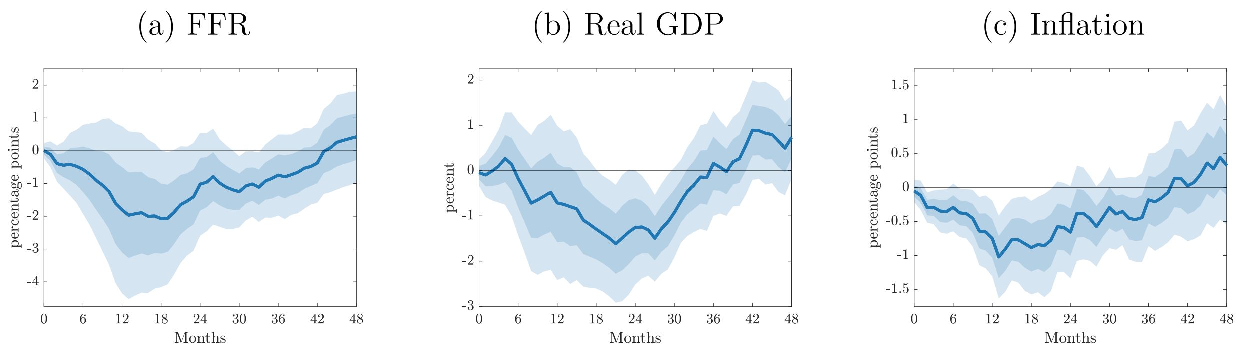
2.C Additional results for Section 2.4

Figure 2.C.1: Response to new shock "minus" response to RR shock



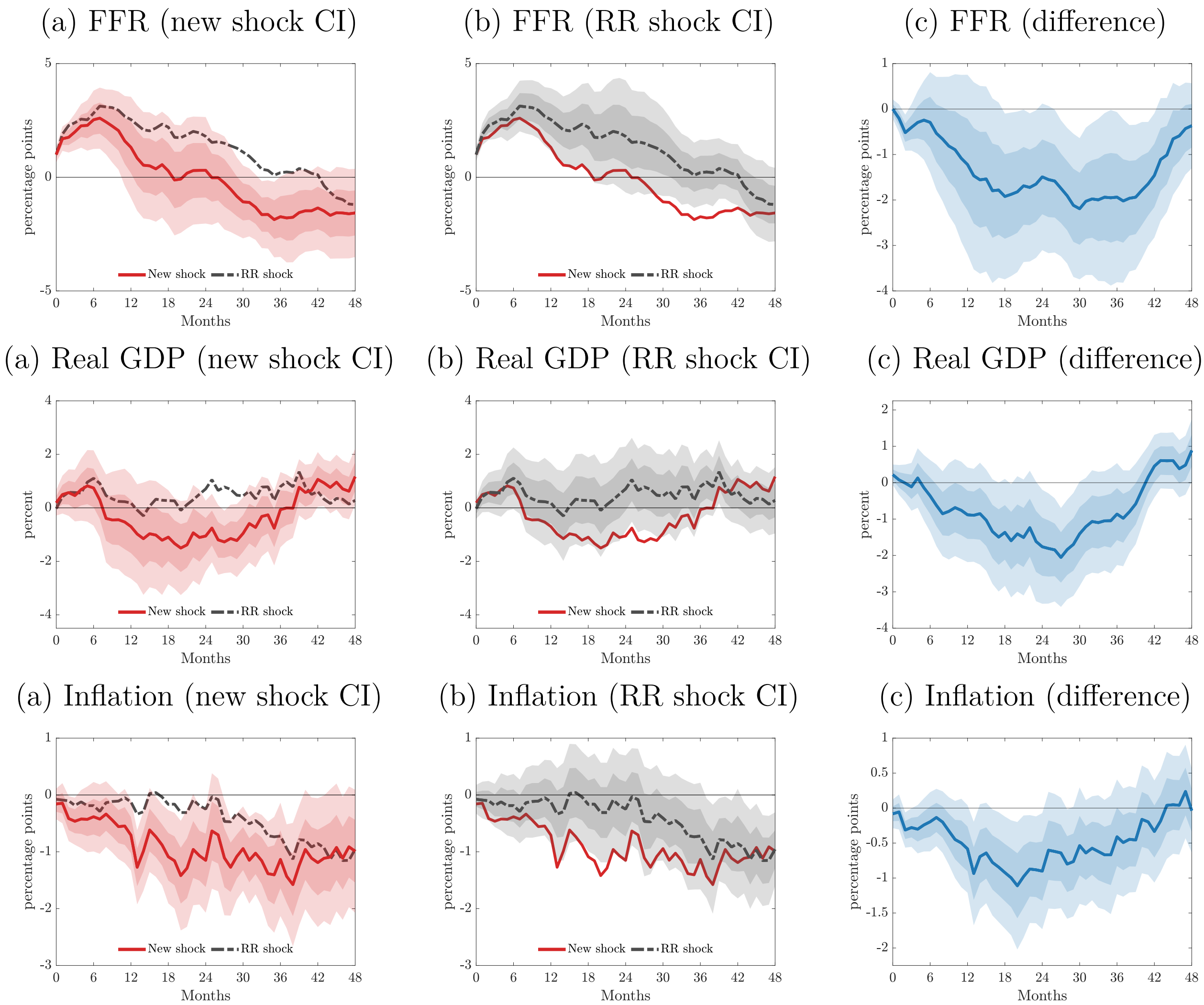
Notes: The figure shows the differences across impulse responses for capacity utilization, the unemployment rate, log consumption expenditures, log business inventories, log hours (in manufacturing), and credit spreads (BAA- minus AAA-rated corporate bond yield) to a monetary policy shock based on the local projection as specified along with (2.17). The difference is computed as the response to the new shock minus the response to the old shock for each outcome, respectively. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.2: Response to new shock “minus”response to Aruoba and Drechsel (2024) shock



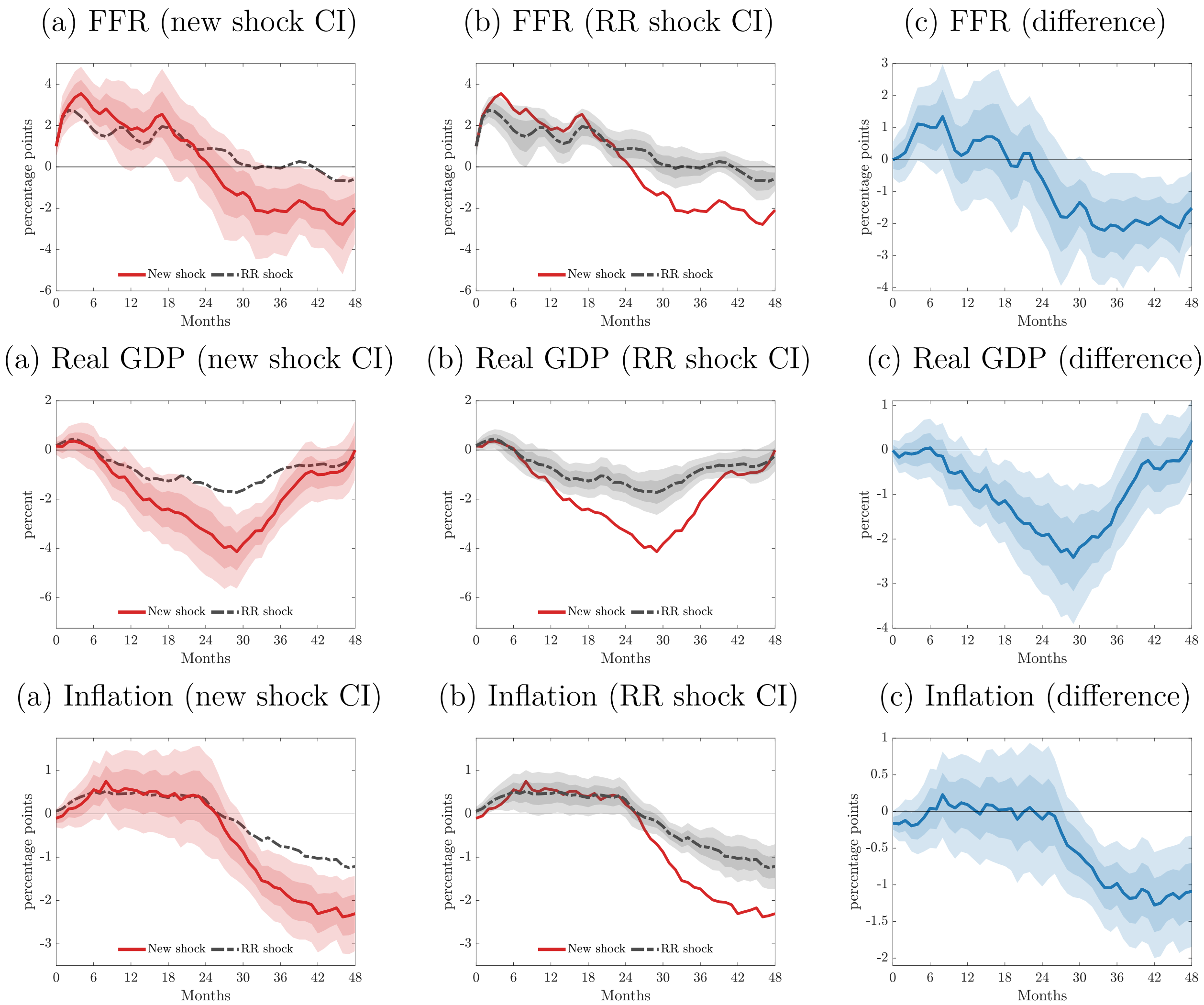
Notes: The figure shows the differences across impulse responses for the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The difference is computed as the response to the new shock minus the response to the old shock for each outcome, respectively. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) where we put the Aruoba and Drechsel (2024) shock on the left-hand-side, whereas the conventional monetary policy shock is taken directly from Aruoba and Drechsel (2024). The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.3: Responses for identification sample 1983-2007



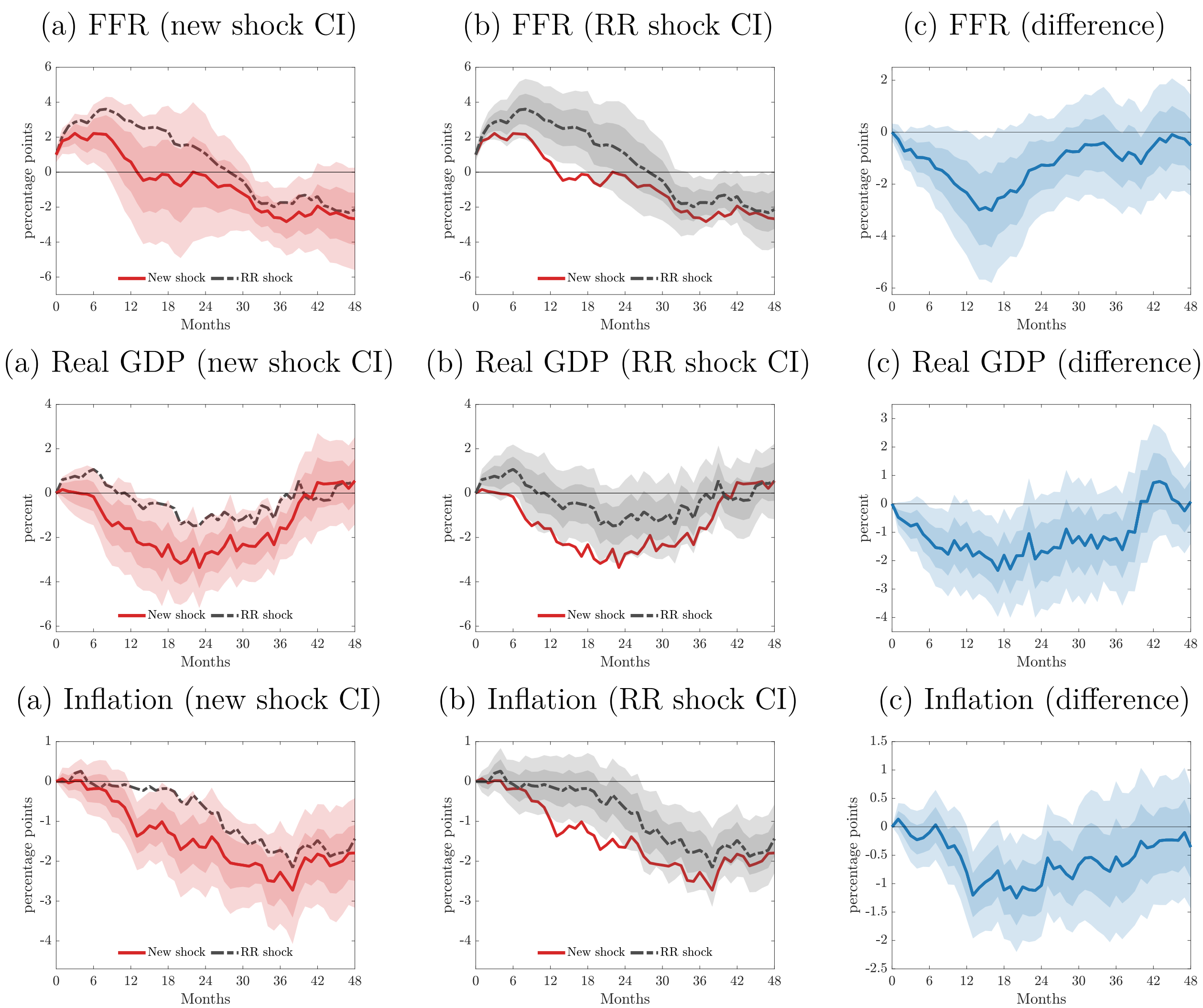
Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The estimation sample for shock identification coincides with the impulse response estimation sample, running from 1983 until 2007. Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.4: Responses for estimation sample 1969-2007



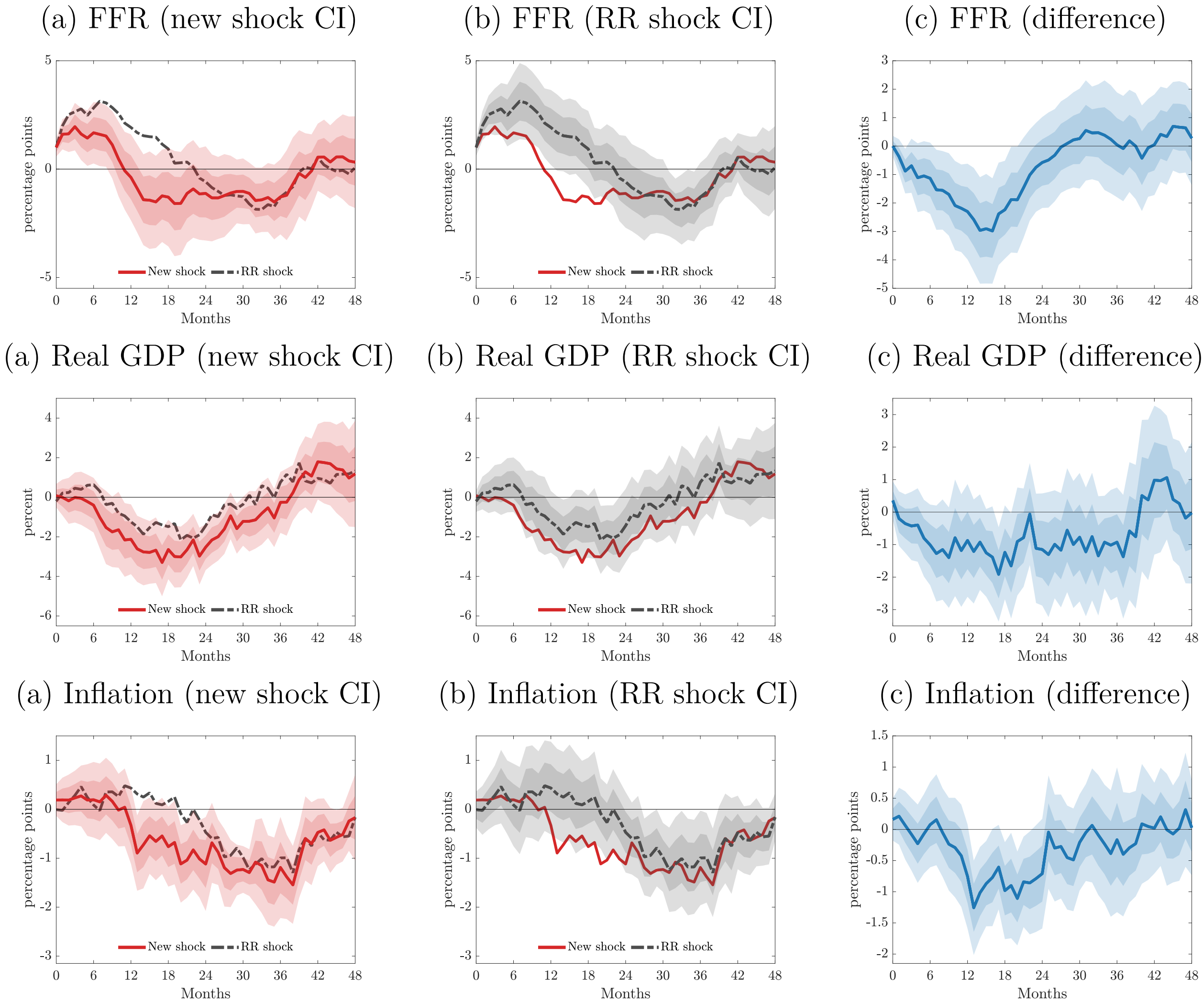
Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). The results correspond to the full sample, running from 1969 until 2007. Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.5: Responses when imposing recursiveness assumption



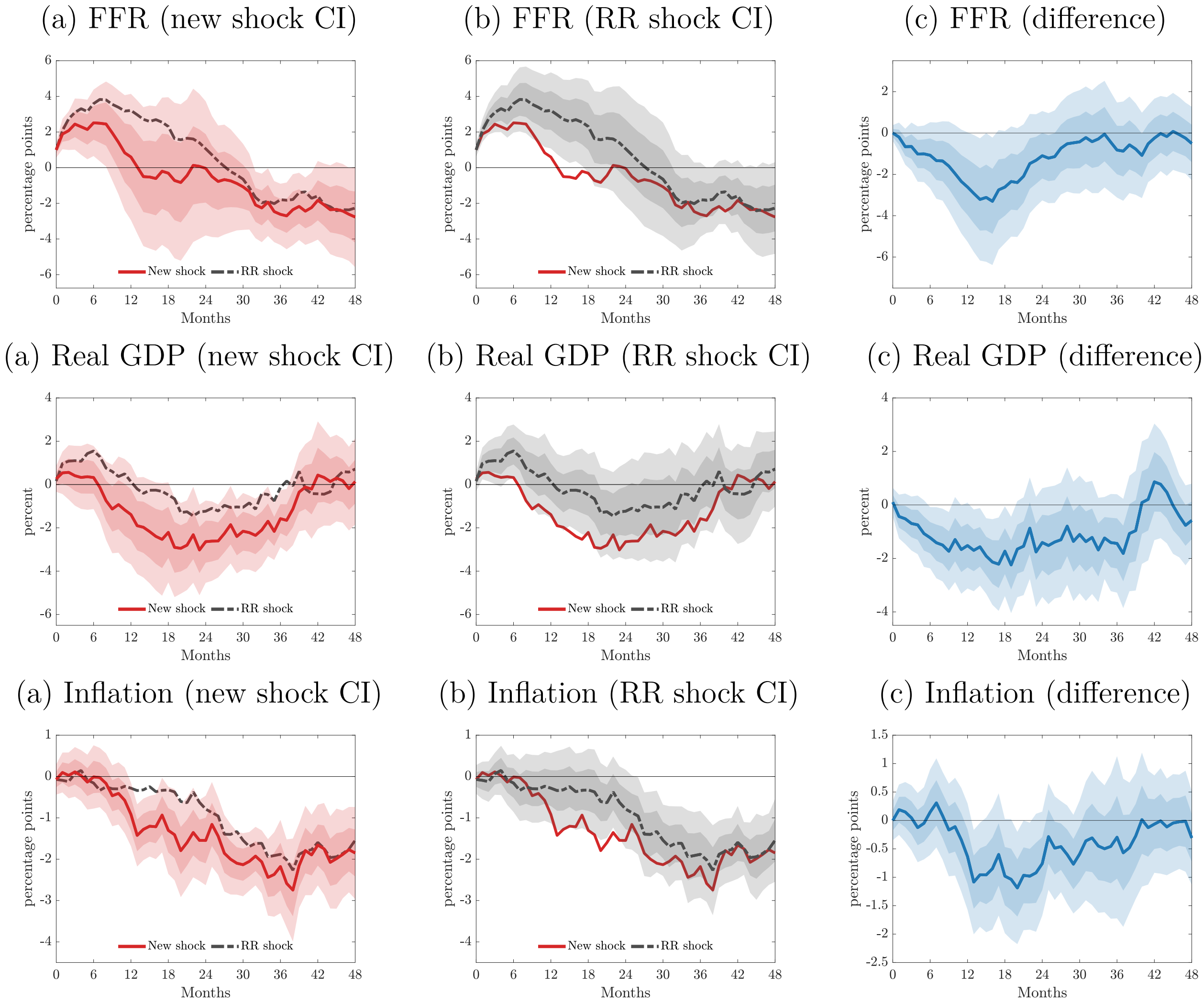
Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). Additionally, we control for contemporaneous log real GDP and inflation imposing the recursiveness assumption that monetary policy shocks affect these variables only with a one-month lag. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.6: Responses when controlling for S&P 500 and EBP



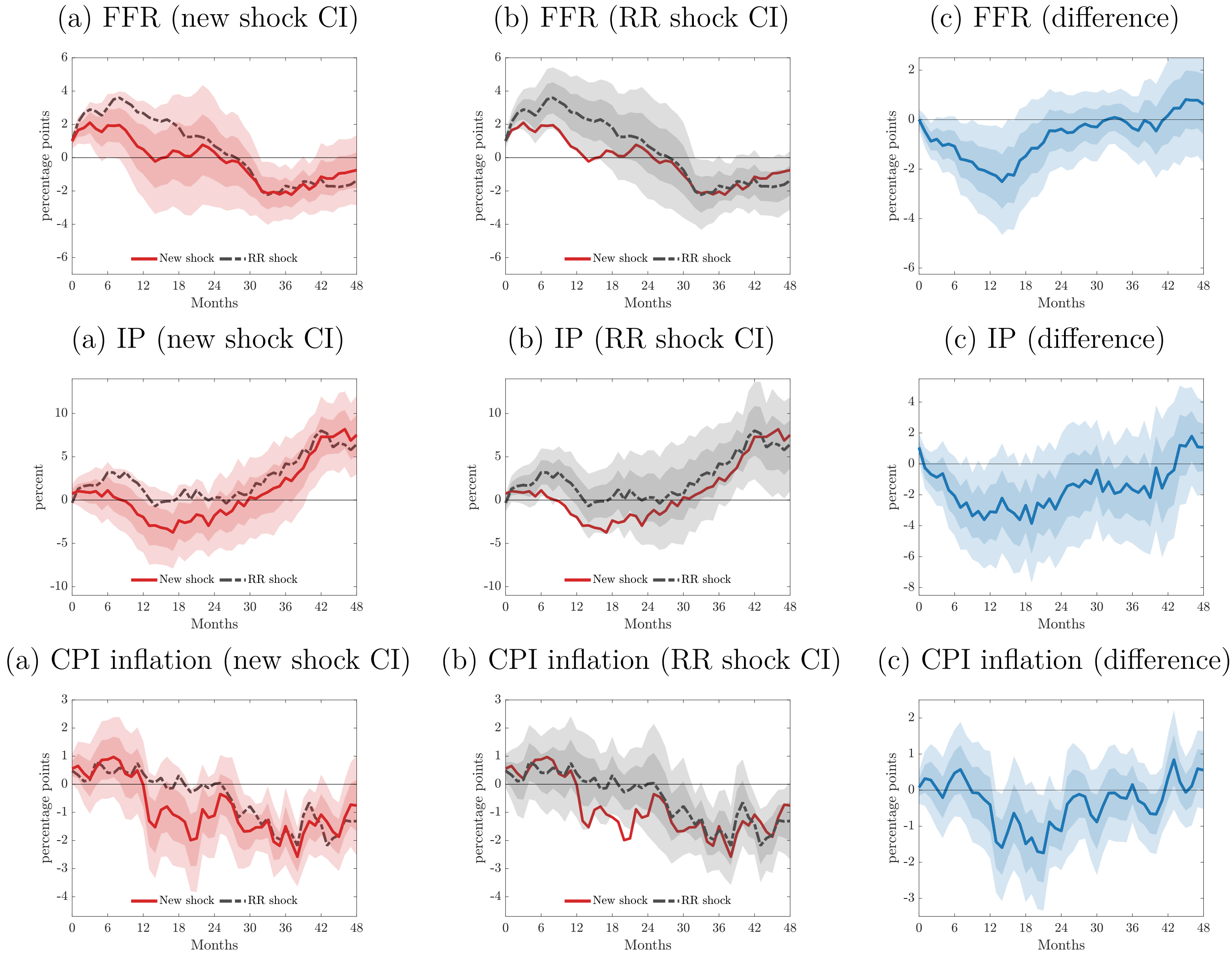
Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). Additionally, we control for 12 lags of both the S&P 500 and the excess bond premium from Gilchrist and Zakrajšek (2012). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.7: Responses when controlling for lagged shocks



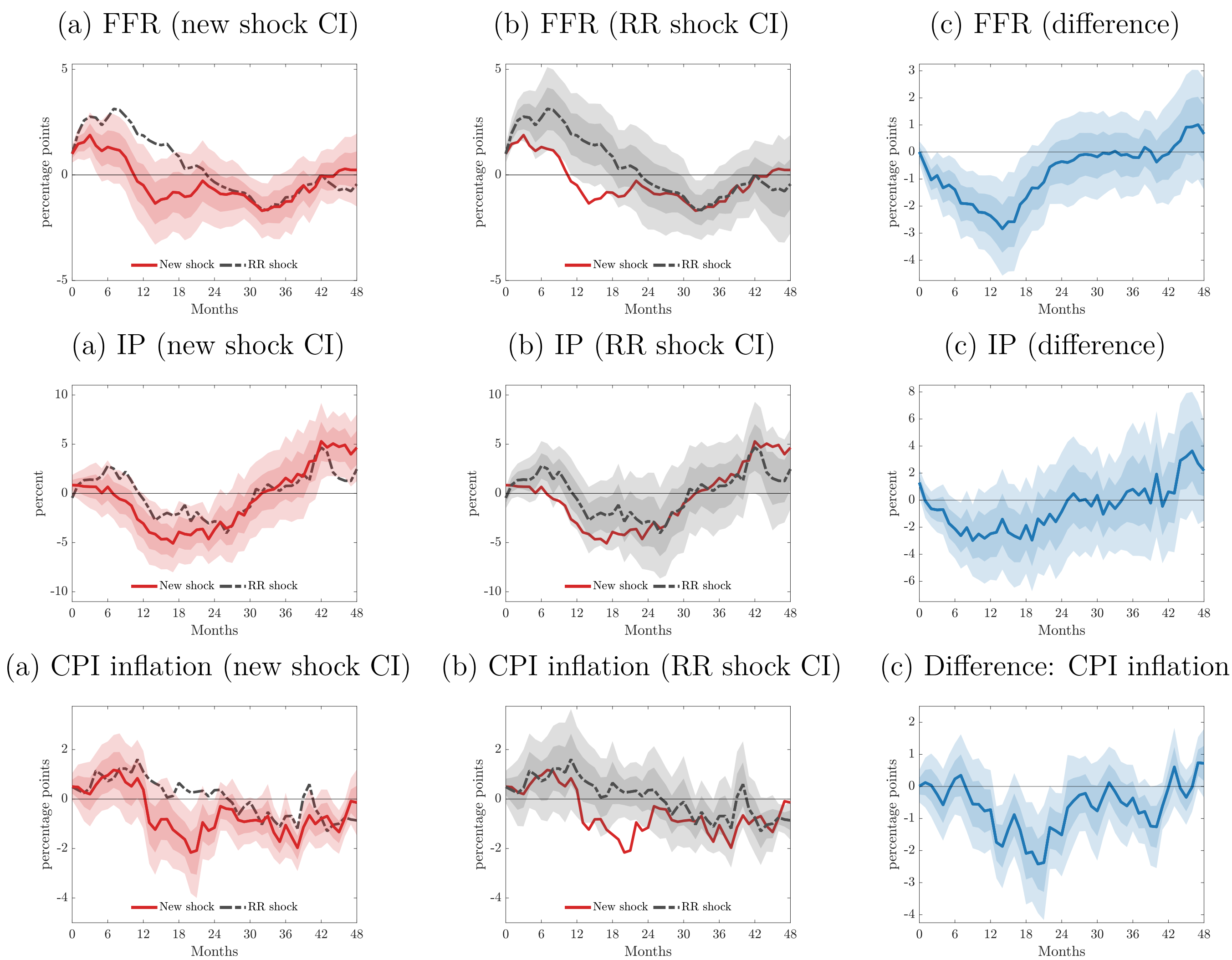
Notes: The figure shows responses of the federal funds rate, log real GDP and the inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). Additionally, we control for 12 lags of monetary policy shock under consideration. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.8: Responses of IP and CPI



Notes: The figure shows responses of the federal funds rate, log industrial product and the CPI inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). We control for 12 lags of both, the log of industrial production and CPI inflation instead of real GDP and inflation based on the GDP deflator. The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Figure 2.C.9: Responses of IP and CPI when controlling for S&P 500 and EBP



Notes: The figure shows responses of the federal funds rate, log industrial product and the CPI inflation rate to a monetary policy shock based on the local projection as specified along with (2.17). We control for 12 lags of both, the log of industrial production and CPI inflation instead of real GDP and inflation based on the GDP deflator. Additionally, we control for 12 lags of both the S&P 500 and the excess bond premium from Gilchrist and Zakrajšek (2012). The new monetary policy shock is identified as the residual from the Taylor rule regression in (2.16) whereas the conventional monetary policy shock is based on the same regression when $b_j = 0$ for $j > 1$, as in Romer and Romer (2004). Columns 1 and 2 display the response to the new shock and conventional shock, respectively. Column 3 display the response to the new shock minus the response to the conventional shock. The shaded areas indicate 68% and 95% confidence bands using Newey-West standard errors.

Chapter 3

Progressive Income Taxation and Inflation: The Macroeconomic Effects of Bracket Creep

3.1 Introduction

Most theoretical models of progressive income taxation consider a tax function that is a mapping from real taxable income to average and marginal tax rates (e.g. Heathcote, Storesletten, and Violante, 2017). In practice, however, tax rates are a function of nominal taxable income. The distinction matters when the tax code is not indexed to inflation, that is, when the tax parameters are not adjusted to account for changes in prices.[1] In this chapter, I refer to bracket creep as any wedge in tax rates between real and nominal taxation. Such wedges may exist when tax rates are not merely a function of real taxable income.[2]

Bracket creep implies that tax rates are differently affected by macroeconomic shocks. Demand shocks move real income and prices in the same direction, implying that nominal income responds stronger than real income. The stronger response in the nominal tax base translates into a greater tax rate change. Conversely, tax rates become less respon-

[1] Many developed countries still lack annual indexation schemes that automatically adjust the tax code. For example, ten out of twenty Euro Area member states that account for 63% of Euro Area GDP had no automatic annual indexation implemented by the end of 2022 (see, e.g., https://taxfoundation.org/data/all/global/income-tax-inflation-adjustments-europe/).

[2] Bracket creep in the literal sense refers to taxpayers who get pushed into the next tax bracket with a higher tax rate, even when only nominal income but not real income grows. The notion of bracket creep entertained in this chapter encompasses this effect.

sive to supply shocks as real income and prices move in opposite directions, implying a weaker nominal income response.[3] This may matter for the macroeconomy since income tax changes have large aggregate effects (e.g. Mertens and Ravn, 2013). Moreover, in the context of monetary policy, it implies a new dimension of fiscal-monetary interaction because monetary policy shocks partly propagate through fiscal instruments, i.e., tax rates via inflation.

This chapter investigates the quantitative consequences of bracket creep on the macroeconomy. Empirically, I isolate bracket creep from other sources of tax rate changes based on a non-parametric decomposition of changes in tax rates. Applying the decomposition to German administrative tax records yields sizable bracket creep episodes. While the overall importance of bracket creep has decreased over time due to institutional changes, the post-Covid inflation surge led to a resurgence. Motivated by the empirical evidence, I analytically characterize how bracket creep affects labor supply decisions in a partial equilibrium framework. Further, I estimate a theory-consistent measure of bracket creep, the indexation gap, which is used to discipline a New Keynesian model with incomplete markets. The New Keynesian model predicts that a given reduction in inflation via a contractionary monetary policy shock leads to substantially smaller short-run output costs in an economy with bracket creep. Put differently, the output costs are aggravated when the tax code is indexed to inflation, revealing a potential caveat of such indexation schemes.

To obtain a tax rate decomposition that separates bracket creep from other sources of tax rate changes, I propose to measure the adjustments that compensate for inflation as the actual change in tax rates that a taxpayer with constant real income faces relative to the change in tax rates she would face if the nominal tax code was not adjusted at all – the latter being a benchmark of "full" bracket creep. When this ratio is zero, then the taxpayer is fully compensated since tax rates remain unchanged, implying full indexation, i.e., there is no bracket creep. Conversely, when this ratio is one, then there is full bracket creep, implying no indexation because the taxpayer is not compensated at all. I use this measure to decompose year-over-year changes in average and marginal tax rates into three distinct components: (i) bracket creep, (ii) real income growth, and (iii) discretionary tax changes. Importantly, the decomposition imposes no restrictions on the tax schedule beyond progressivity.

[3]For example, consider a contractionary supply shock such that real income falls, but prices rise. When real income was the tax base, tax rates would decline. However, the rise in prices implies that nominal income declines by less, leading to less reduction in tax rates. In the knife-edge case of constant nominal income, there would be no tax rate decline at all because both forces perfectly cancel each other.

Empirically, I implement the decomposition based on German administrative tax records from 2002 until 2018. The administrative tax data is desirable because I need to know the entire distribution of gross incomes and claimed deductions to compute tax rates accurately, which is a prerequisite for reliable decomposition results. Further, the German setting is suitable because there are multiple years in which the tax system was not adjusted, inevitably leading to bracket creep. Moreover, I can evaluate a 2012 tax reform that aimed at reducing bracket creep. The reform requires the government to publish a mandatory bracket creep report, along with suggestions to undo bracket creep. While not mandated by law, since then, the government has aimed to adjust the tax code based on inflation forecasts, which may only address bracket creep due to anticipated inflation. Imputing the tax data until 2023 allows me to evaluate this policy regime in the presence of a large inflation surprise.

Focusing on the long-run average effects in my sample, I find that bracket creep accounts for an annual increase in average and marginal tax rates between 0.10 and 0.12 percentage points. For comparison, the average annual inflation rate in this sample was 1.43, which implies that a percentage point increase in inflation corresponds to a 7-9 basis point increase in tax rates per year. These bracket creep effects are relatively uniform across the income distribution. In contrast, tax changes due to real income growth are negligible, accounting for less than 0.05 percentage points. Finally, discretionary tax changes account for an annual decrease in tax rates between 0.14 and 0.17 percentage points. Importantly, discretionary tax changes do not compensate for bracket creep because they occur infrequently.

The above results characterize the trend in tax rate changes. For macroeconomic stabilization, however, the fluctuations around the trend matter. The most pronounced fluctuations arise when the government does not adjust the tax schedule for multiple consecutive periods. There are two such bracket creep episodes before 2012. During these episodes, bracket creep accounts for a total increase in average and marginal tax rates between 0.64 and 0.86 percentage points, cumulated over each three-year bracket creep episode. In contrast, from 2013 until 2018, I find very little bracket creep, suggesting that the 2012 tax reform successfully eliminated bracket creep during a period of low and stable inflation. However, the post-Covid inflation surge, a large inflation surprise, led to a resurgence of bracket creep with sizable effects on average and marginal tax rates, which increased by 0.51 and 0.66 percentage points, respectively. Overall, this shows that bracket creep accounts for sizable changes in tax rates paid by households.

To understand how household choices respond to bracket creep, I propose an analytical model with a tax schedule that nests the one from Heathcote, Storesletten, and Violante (2017) but allows for bracket creep (HSV-type tax schedule, henceforth). I study the labor-leisure choice of a household facing this tax schedule and a government that may return a fraction of the tax revenues to the household via a transfer. I show that the labor supply response to bracket creep is theoretically ambiguous and crucially depends on how the government uses tax revenues. Intuitively, when all tax revenues (from bracket creep) are returned to the households, then income effects are eliminated, and only a substitution effect prevails, reducing labor supply. Conversely, labor supply may increase when the government is not giving back tax revenues via transfers. Incidentally, the previous literature on bracket creep neglects the important role of transfers and assumes that all tax revenues are given back without (convincing) justification.[4]

An appealing feature of the proposed HSV-type tax schedule is that bracket creep can be conveniently summarized by a scalar statistic, the indexation gap. The time series of indexation gaps can be estimated based on restrictions derived from the HSV-type tax schedule, delivering a theory-consistent measurement of bracket creep that captures the government's adjustment or indexation choices as well as the prevailing inflation rate. Finally, I use the indexation gap series to provide reduced-form evidence that supports my tax schedule formulation and that can be used to discipline the quantitative analysis.

The quantitative analysis of bracket creep is based on a standard New Keynesian model with incomplete markets (e.g. Auclert, Bardóczy, Rognlie, and Straub, 2021), which also nests my analytical model. Households consume, supply labor, and may save in a liquid asset. The production side features nominal price rigidities. A fiscal authority uses the tax revenues to finance spending, transfers and interest payments, and a monetary policy authority controls the nominal interest rate. I calibrate the model to the German economy before the 2012 tax reform, using the empirically observed indexation gaps to discipline how the fiscal authority adjusts the tax schedule.

In this setup, I study the responses to a monetary policy shock and compare it with the counterfactual responses under full indexation. The results suggest that indexation amplifies the effects of monetary policy on output, whereas the impact on inflation dynamics is negligible. Quantitatively, the impact response under full indexation is roughly thirty percent larger than with bracket creep, but the difference vanishes roughly within a year. The intuition for the differential output effects is that tax rates are more responsive to demand

[4] It applies to the few papers that use New Keynesian models with complete markets (Edge and Rudd, 2007; Keinsley, 2016), or a money growth model with incomplete markets (Heer and Süssmuth, 2013).

shocks in an economy with bracket creep. The resulting substitution effects of taxation dominate income effects under the calibration to the German economy. This discourages labor supply, which depresses the production response. Finally, one interpretation of the result is that the output costs of reducing inflation via monetary policy are aggravated when the tax code is indexed to inflation, revealing a potential caveat of such indexation schemes.

Related literature. This research relates to the surprisingly scant literature on bracket creep. Empirically, most papers study bracket creep based on micro-simulations focusing on particular historical episodes in the European (e.g., Immervoll, 2005; Paulus, Sutherland, and Tasseva, 2020), or specifically in the German context e.g., Blömer, Dorn, and Fuest, 2023; Zhu, 2014.[5] My contribution lies in a comprehensive and transparent documentation of bracket creep effects over twenty years for Germany and a comparison with other sources of tax rate changes through my decomposition approach. Theoretically, bracket creep has been studied in New Keynesian models with complete markets (Edge and Rudd, 2007; Keinsley, 2016) and in a money growth model with incomplete markets (Heer and Süssmuth, 2013). Relative to these papers, I analytically show that transfers crucially shape the labor supply response to bracket creep, and I offer a quantitative analysis of bracket creep using a workhorse New Keynesian model that accounts for household heterogeneity along the income distribution.

More broadly, I relate to studies that focus on inflation and its interaction with taxation in general, (e.g., Altig, Auerbach, Eidschun, Kotlikoff, and Ye, 2024; Cloyne, Martinez, Mumtaz, and Surico, 2023; Süssmuth and Wieschemeyer, 2022), with capital taxation specifically (e.g., Feldstein, 1983; Gavin, Keen, and Kydland, 2015; Gavin, Kydland, and Pakko, 2007), or on inflation and its impact on households (e.g., Adam and Zhu, 2016; Doepke and Schneider, 2006; Erosa and Ventura, 2002; Pallotti, 2022; Pallotti, Paz-Pardo, Slacalek, Tristani, and Violante, 2023). Further, I relate to the broad literature on progressive taxation (e.g., Benabou, 2002; Conesa and Krueger, 2006; Heathcote, Storesletten, and Violante, 2017, 2020; Mattesini and Rossi, 2012; McKay and Reis, 2021), as well as to the literature on New Keynesian models with incomplete markets (e.g., Auclert, Bardóczy, Rognlie, and Straub, 2021; Auclert, Rognlie, and Straub, 2023; Kaplan, Moll, and Violante, 2018).

[5] For a current discussion of bracket creep during the recent inflation surge, see Bundesbank (2022).

3.2 Empirical analysis

In this section, I propose a new approach to measure bracket creep based on tax data. My approach rests on measuring the degree of indexation of the tax schedule in reduced-form to compute a decomposition of the changes in tax rates into three distinct components: (i) real income growth, (ii) discretionary tax changes, and (iii) bracket creep. I apply this approach to German administrative tax records and show that bracket creep effects are of similar quantitative importance as discretionary tax changes. I further identify two sizable bracket creep episodes before 2012 and a decline in the quantitative importance of bracket creep thereafter. However, the 2022 inflation surge led to a sizable resurgence of bracket creep because of imperfect inflation adjustments.

3.2.1 Measuring bracket creep

I derive a decomposition of the year-over-year changes in average and marginal tax rates of a single taxpayer. Let $Y_t > 0$ be nominal pre-tax income in year t. Taxable income is $Z_t = Y_t - D_t$ with $D_t \geq 0$ being the amount of deductions. The average or marginal tax rate can be represented as a mapping $\tau_t : \mathbb{R}^+ \to \mathbb{R}^+$ from nominal taxable income to the respective tax rate. This mapping to tax rates incorporates tax exemptions.[6] I assume that the tax schedule is progressive, which implies that $\tau_t(Z_t)$ is strictly increasing in income when it refers to the average tax rate. When $\tau_t(Z_t)$ refers to the marginal tax rate, then progressivity only demands that it exceeds the average tax rate for any Z_t, but it need not be strictly increasing for any Z_t. I further assume that taxable income Z_t is sufficiently large to ensure $\tau_t(Z_t) > 0$, focusing on individuals who actually pay taxes. Put differently, this rules out incomes below the tax exemption threshold. Let $Z_t^\Pi = Y_{t-1}\,\Pi_t - D_t^\Pi$ be taxable income of a taxpayer who has the same real pre-tax income as in the previous year, i.e., $Y_t = Y_{t-1}\Pi_t$, and $\Pi_t = P_t/P_{t-1}$ is the gross inflation rate. Deductions $D_t^\Pi \in [D_{t-1}, D_{t-1}\Pi_t]$ may or may not be adjusted to inflation, as I explain below. I define a tax function that gives the tax rate in year t as a function of nominal taxable income in years t and $t-1$ for the taxpayer with constant real pre-tax income.

[6] The tax exemption implies that $\tau_t(Z_t) = 0$ for all $Z_t \leq \underline{Z}$, where $\underline{Z}$ is the exemption amount. This exemption amount is a parameter of the tax function that the government may adjust. Indeed, anticipating the empirical application, the exemption amount is adjusted on an annual basis in Germany.

Definition 1.

$$\tau_t^{\mathcal{I}}\left(Z_t^{\Pi},\ Z_{t-1}\right) = \alpha_t\,\tau_{t-1}\left(Z_{t-1}\right) + \left(1-\alpha_t\right)\tau_{t-1}\left(Z_t^{\Pi}\right). \tag{3.1}$$

The tax function is a (point-wise) convex combination of two cases where $\alpha_t \in [0,1]$ measures the degree of indexation of the tax function. With $\alpha_t = 1$, the tax function is perfectly indexed to inflation since the tax rate in period t coincides with the tax rate in the previous period. In other words, tax rates are unaffected by nominal income growth that compensates for inflation. Conversely, when $\alpha_t = 0$, the tax function is not indexed at all because nominal income in t is evaluated at the schedule from the previous year. Thus, there is no inflation adjustment in the tax schedule, and the tax rate may change whenever $Z_t^{\Pi} \neq Z_{t-1}$.[7]

Decomposition. Given the tax function from Definition 1, I decompose the year-over-year changes in tax rates paid by the taxpayer as follows.

$$\tau_t\left(Z_t\right) - \tau_{t-1}\left(Z_{t-1}\right) = \Psi_t^{bc} + \Psi_t^{rg} + \Psi_t^{tc}, \tag{3.2}$$

and

$$\Psi_t^{bc} = \underbrace{\tau_t^{\mathcal{I}}\left(Z_t^{\Pi},\ Z_{t-1}\right) - \tau_{t-1}\left(Z_{t-1}\right)}_{\textbf{b}\text{racket }\textbf{c}\text{reep}},$$

$$\Psi_t^{rg} = \underbrace{\tau_t\left(Z_t\right) - \tau_t\left(Z_t^{\Pi}\right)}_{\textbf{r}\text{eal income }\textbf{g}\text{rowth}},$$

$$\Psi_t^{tc} = \underbrace{\tau_t\left(Z_t^{\Pi}\right) - \tau_t^{\mathcal{I}}\left(Z_t^{\Pi},\ Z_{t-1}\right)}_{\text{discretionary }\textbf{t}\text{ax }\textbf{c}\text{hange}},$$

The first term measures bracket creep, that is, changes in tax rates due to a lack of indexation of the tax schedule. The second term captures the changes in tax rates due to real income growth. Taken together, both terms show increases in tax rates due to nominal income growth. The third term captures discretionary changes to the tax schedule that are not captured by indexation through $\tau_t^{\mathcal{I}}(\cdot)$. Next, I can characterize the bracket creep term.

[7] Marginal tax rates may also stay constant when both Z_t^{Π} and Z_{t-1} fall in a tax bracket with the same constant marginal tax rate. The average tax rate necessarily adjusts under a progressive schedule when $Z_t^{\Pi} \neq Z_{t-1}$ because the average tax rate is strictly increasing in nominal income.

Proposition 1. *Given Definition 1, the bracket creep term from* (3.2) *is given by*

$$\Psi_t^{bc} = (1-\alpha_t)\left[\tau_{t-1}\left(Z_t^{\Pi}\right) - \tau_{t-1}\left(Z_{t-1}\right)\right].$$

If $\Pi_t > 1$*, and* $\alpha_t < 1$ *and* $\frac{\partial \tau_{t-1}(Z)}{\partial Z} > 0$ *for all* $Z \in [Z_{t-1},\ Z_t^{\Pi}]$*, then it holds that* $\Psi_t^{bc} > 0$*.*

The proof is in Appendix 3.A. Note that under a progressive tax system, the third condition, $\frac{\partial \tau_{t-1}(Z)}{\partial Z} > 0$, is always satisfied for the average tax rate but not necessarily for the marginal tax rate.[8] Suppose $\tau_t(\cdot)$ is indeed the average tax rate to illustrate the proposition. Then, the proposition states that the bracket creep term in the decomposition is strictly positive when three conditions apply: (i) there is positive inflation, (ii) the tax code is not perfectly indexed to the actual rate of inflation, and (iii) the tax code is progressive. In contrast, the bracket creep term is zero under full indexation or absent inflation or under a linear tax schedule. This suggests that the bracket creep term captures only bracket creep effects when expected.

Degree of indexation. To operationalize the decomposition, I need to measure α_t. I propose a measurement of α_t that leverages observed changes in the tax schedule, irrespective of whether these adjustments are discretionary or implemented via an automatic indexation scheme. This corresponds to a notion of *effective* indexation measured as

$$\alpha_t = \begin{cases} 1 - \max\left\{\min\left\{\frac{\tau_t\left(Z_t^{\Pi}\right) - \tau_{t-1}(Z_{t-1})}{\tau_{t-1}\left(Z_t^{\Pi}\right) - \tau_{t-1}(Z_{t-1})}, 1\right\}, 0\right\} & \text{if} \quad \tau_{t-1}\left(Z_t^{\Pi}\right) > \tau_{t-1}\left(Z_{t-1}\right), \\ 1 & \text{otherwise.} \end{cases} \tag{3.3}$$

Focusing on the first case in (3.3), the denominator captures the amount of bracket creep under constant real income that prevails when the tax code is not adjusted at all, i.e., $\tau_t(Z) = \tau_{t-1}(Z)$, $\forall Z$. This is the benchmark of "full" bracket creep, or equivalently, no indexation. The numerator measures the change in the tax rate, accounting for actual adjustments in the tax function. The ratio can be interpreted as a measure of the "distance" between the actual change in the tax rate and the full bracket creep benchmark. Therefore, I refer to this ratio as the degree of bracket creep and, conversely, to α_t as the degree of indexation. Consider the empirically relevant case of positive inflation, where $Z_t^{\Pi} >$

[8] When marginal tax rates are constant within a given tax bracket and, under constant real income, inflation does not push a taxpayer into the next bracket, then there is no bracket creep in terms of the marginal tax rate

Z_{t-1}. In this case, when the numerator is larger than the denominator, then the degree of indexation is zero. Any increase in the tax rate larger than the full bracket creep benchmark must be a discretionary tax hike. Conversely, when the numerator is negative, the degree of indexation is unity because the taxpayer is fully compensated for bracket creep. Any (additional) reduction in the tax rate must be a discretionary tax cut. Further, the degree of indexation is also unity when the denominator is zero. It may only happen when considering marginal tax rates that are constant within a tax bracket, and both Z_t^{Π} and Z_t fall in the same bracket with a constant marginal tax rate. Finally, note that this approach works equally well in the case of deflation, although this never occurs in my sample.[9]

This approach to quantifying the degree of indexation is appealing because it imposes no parametric restriction on the tax schedule. It can be implemented with relatively mild information requirements. One only needs to measure taxable income, deductions, inflation, and the exact tax schedule as specified in the tax law, including tax exemptions.

Aggregation. The presented decomposition applies to a single taxpayer. It may be a single person who files taxes on her behalf or married spouses who file their taxes jointly. The decomposition does not require to distinguish between these two cases. Aggregation to sample averages is straightforward since the decomposition is additive. Thus, I can readily compute arithmetic averages $\bar{\Psi}_A^k = \sum_{(i,t)\in A} \Psi_{i,t}^k$ for any decomposition term k, where A denotes the set of individuals and time periods over which the average is computed.

Alternative mechanical decomposition. An alternative decomposition may measure bracket creep as changes in tax rates under constant real income, keeping the tax schedule constant. From Proposition 1, it becomes clear that this naive mechanical decomposition is nested when imposing $\alpha_t = 0 \; \forall t$. Based on this, one could still compute the bracket creep term and subtract the discretionary tax change term after aggregation to check whether there is bracket creep that is not compensated with tax function changes. However, even when both terms net out, one cannot conclude that all taxpayers got compensated for bracket creep every year because it does not take into account how the compensation via discretionary tax changes is distributed across taxpayers and time. For example, it could be that a fraction of taxpayers is benefiting from large tax cuts (that over-compensate bracket

[9] In this case, the "full" bracket creep benchmark is negative. There is full indexation when actual tax rates do not change (or even increase). When tax rates fall, there is incomplete indexation or, equivalently, bracket creep. Naturally, in this case, bracket creep lowers the tax rates relative to full indexation.

creep), whereas others receive no compensation and, therefore, see tax rates changing due to bracket creep. Whether these composition effects matter is an empirical question. Thus, I also report the results of the mechanical decomposition.

Deductions. It may be important to account for deductions because many fixed-amount deductions are specified in the tax law and only infrequently adjusted. For example, this applies to work-related deductions that are lump-sum or calculated based on commuting distance in Germany.[10] Accounting for deductions is particularly important for the constant real income scenario, where I aim to measure how taxable income would have evolved when the taxpayer's behavior is kept constant. In this case, deduction amounts may only increase when the deductions reflect actual nominal payments that increase with inflation (itemized deductions) or when the government raises the deduction amounts specified in the tax law. Unfortunately, discriminating these two cases is infeasible in the data.[11] Thus, I will present two versions of the decomposition. As a conservative baseline, I assume that all deductions grow with inflation, i.e., $D_t^{\Pi} = D_{t-1}\Pi_t$. Alternatively, I present results where the deductions are kept constant, i.e., $D_t^{\Pi} = D_{t-1}$. The former may be a lower bound on the quantitative importance of bracket creep, while the latter delivers an upper bound. In practice, the appropriate value of $\mathcal{D}_t^{\Pi}$ should be in between these two extreme cases. Reporting both reveals to what extent my results depend on deductions. Finally, note that tax exemptions are part of the tax rate function $\tau_t(\cdot)$. Thus, I account for the empirically observed changes in exemption amounts, irrespective of the treatment of deductions.

3.2.2 Administrative tax records

Institutional setting. I analyze administrative tax records from Germany, where income from most sources is subject to the progressive income tax schedule.[12] A fixed

[10] In practice, the available deduction possibilities may affect economic choices, e.g., the work location and commuting distance. While the decomposition does not take a stand on these incentive effects, I abstract from this in the theoretical models presented in Sections 3.3-3.4. This is a common assumption when one studies the (macroeconomic) consequences of taxation (see, e.g., Heathcote, Storesletten, and Violante, 2017).

[11] This is because the actual computation of deductions in the tax data is extremely complex since the tax declarations involve more than 2000 variables, of which most matter for this computation. While it is ex-ante unclear, it turns out that deductions are not crucial for the empirical results.

[12] A noteworthy exception is that capital income has been taxed at a flat rate of 25% since 2009. A further special case is that taxpayers may opt to pay regular income taxes on their capital income (as opposed to the flat rate) when the progressive tax schedule implies a lower tax rate. In practice, these cases are likely negligible. It only applies to taxpayers with sufficiently low taxable income (including capital income) so that the regular income tax rate (based on the progressive tax schedule) does not exceed 25%, but capital

amount of around 10,000 euros (varying over time) is exempt. Any taxable income beyond the exemption is taxed. Figure 3.B.1 in Appendix 3.B illustrates the schedule for different years in my sample. An important feature of the schedule is that marginal tax rates increase linearly within each tax bracket, except for taxable income above around 60,000 euros. It implies that bracket creep can increase marginal tax rates for any taxpayer below the top brackets, even if she stays within her tax bracket. Moreover, all taxpayers may experience bracket creep effects in terms of the average tax rate as this rate is always increasing in income, regardless of whether the marginal tax rate is constant. This matters because an increase in the average tax rate reduces the real disposable income of taxpayers.

Germany also offers a preferential tax scheme for married spouses. Under this joint taxation scheme, the tax function is evaluated only at the average taxable income of both spouses.[13] This implies that married spouses are typically in a lower tax bracket than the higher income earner under individual taxation. Thus, under joint taxation, the tax function is evaluated at lower taxable income where average and marginal tax rates respond more strongly to income changes; see also Figure 3.B.1 for an illustration. Thus, bracket creep may also be important for middle-class households. This is especially relevant when one spouse is the only breadwinner.

Turning to indexation, Germany has no automatic inflation adjustments to the tax code. However, a tax reform in 2012 mandated the government to prepare a bracket creep report every other year (e.g. Bundesbank, 2022). Along with this obligation, the Federal Ministry of Finance regularly adjusts the tax schedule for inflation based on inflation forecasts for the subsequent two years. The tax parameters are adjusted for both years separately, given the inflation forecast. The adjustment procedure applies only to the statutory tax code but not to deductions.[14] Below, I study to which extent this reform eliminated bracket creep.

Taxpayer panel. The data is an annual panel of income taxpayers in Germany from 2002 until 2018. It contains administrative tax records that are provided by the German Federal Statistical Office.[15] An individual taxpayer may be an individual or married spouses who

income still exceeds the exemption amount on capital income, the so-called “Sparerfreibetrag”.

[13] The final nominal tax payment is given by two times the tax payment that a single taxpayer with this average income would have to pay.

[14] Note that I account for changes in tax exemptions, which are part of the statutory tax code. This is important because the exemption amount is frequently adjusted to not tax a subsistence level of income.

[15] To be precise, the data source is the Research Data Center of the Federal Statistical Office and Statistical Offices of the Laender, Taxpayer Panel, 2002-2018. All presented results are based on my own calculations.

file their taxes jointly. Specifically, I define a taxpayer conditional on filing status. For example, a taxpayer filing taxes individually is considered a different cross-sectional unit than the same taxpayer when filing jointly later in life. The data is a five percent random sample taken from the universe of taxpayers. For my analysis, I focus on taxpayers who file their tax declarations, have positive tax liabilities, and do not apply specific widow tax schemes.[16] The resulting sample contains around 14 million tax records. Restricting the attention to taxpayers with observations available for at least two consecutive years leaves me with around 10 million tax records as the baseline sample for the tax rate decomposition.

Variables. The Taxpayer panel contains all variables that can be filed in German income tax declarations. For my analysis, I mainly use gross income, taxable income, and the final tax liability. Gross income refers to all reported income before deductions are subtracted, whereas taxable income is gross income minus deductions. I use these two variables to compute the total amount of deductions. Finally, I use the tax liability to verify that my implementation of the tax schedule is accurate for all years.[17] The descriptive statistics are presented in Table 3.B.1 in Appendix 3.B.

Additional data. I use the inflation rate of the German CPI to obtain nominal income that maintains constant real value. To study the post-Covid inflation surge, I use average household income growth from the German Federal Statistical Office to impute the income distribution for years beyond 2018. Specifically, I take the 2018 cross-sectional distribution of taxpayers as given and assume that all Euro-value variables grow at the rate at which average household income was growing; see Table 3.B.2 for the data used for imputation.

3.2.3 Results

I report empirical results based on the decomposition developed in Section 3.2.1. First, I present averages over the entire sample, which reflect long-run trends in tax rate changes. Then, I show how the effects vary across years to understand the cyclical properties. I find two sizable bracket creep episodes before 2012 and a resurgence of bracket creep during

[16] Not all taxpayers in Germany need to file their taxes. This applies when taxpayers have only one source of (labor) income such that the monthly withholding tax can be expected to be close to the tax liability when filing (these are the so-called "Lohnsteuerfaelle"). I exclude these taxpayers to maintain a consistent sample because they are not in the tax data before 2012.

[17] The difference between the actual tax liability and my own calculations is less than 1 euro per average monthly income of each taxpayer for more than 99% of the tax declarations.

the post-Covid inflation period. All reported results are significant at the 5% level, and standard errors are reported in Tables 3.B.3-3.B.4 in Appendix 3.B.

Average effects. Panels (a) and (b) of Figure 3.2.1 provide the decomposition averaged across all taxpayers between 2002 and 2018 for average and marginal tax rates, respectively. The baseline decomposition results are depicted as blue bars. The (smaller) red bars indicate how the decomposition would change when deductions are assumed to be constant in nominal terms.[18] Finally, the cross-markers give the total effect under constant deductions. I find that bracket creep accounts for 0.10 percentage points when measured in average tax rates and 0.11 percentage points when measured in marginal tax rates. Keeping deductions constant raises these numbers by one basis point only, suggesting that the treatment of deductions is quantitatively not relevant. The average annual inflation rate in this sample was 1.43, which implies that a percentage point inflation corresponds to a 7-9 basis point increase in tax rates per year. In comparison, real income growth leads to modest increases in tax rates of up to 0.05 percentage points and even less when keeping deductions fixed. Tax rate increases due to both bracket creep and real income growth fundamentally reflect growth in nominal incomes. The estimates suggest that the former is more important for the average changes in tax rates.[19] Finally, discretionary tax changes contribute negatively with -0.14 and -0.17 percentage points, respectively. Below, I show that discretionary tax rate changes do not cancel out with tax rate changes due to bracket creep because they occur in different years.

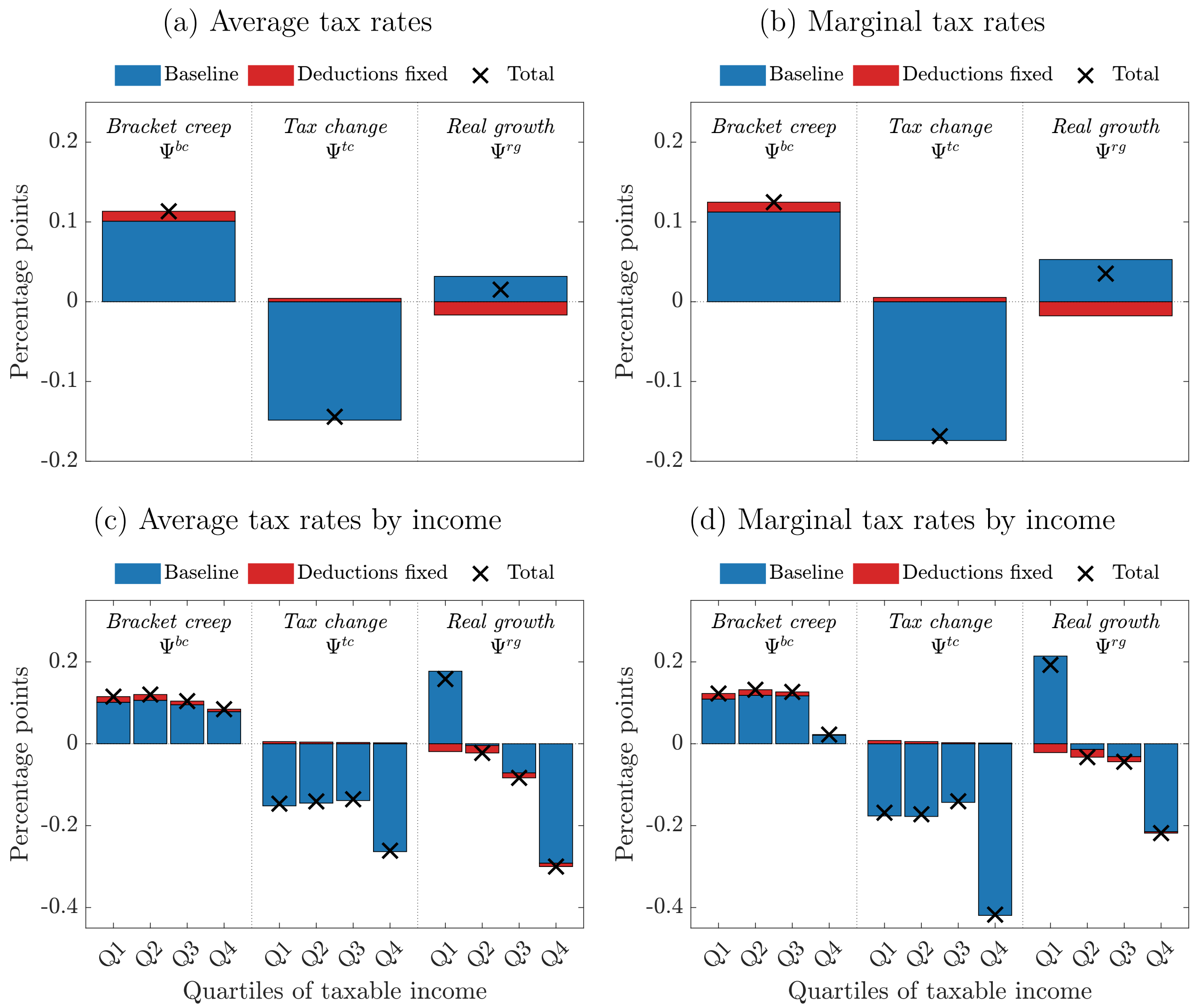
Distributional effects. Panels (c) and (d) of Figure 3.2.1 unpack how the average decomposition results varies along the taxable income distribution. Specifically, I group households in quartiles of their taxable income from the previous year and divide the taxable income of jointly filing spouses by two to obtain a per-person measure of taxable income. Irrespective of the tax rate for which the decomposition is implemented, bracket creep appears to be relatively uniform across the income distribution. Only the top quartile is slightly less exposed to bracket creep.[20] While absent tax adjustments, one should expect

[18] Recall the baseline decomposition assumes that deductions grow at the rate of inflation. The exemption is directly included in $\tau_t(\cdot)$ and not kept constant, even when nominal deductions are fixed.

[19] Note that the results reflect aggregate effects that net out idiosyncratic shocks. Thus, tax rate changes due to real income growth may be more important when idiosyncratic shocks are considered.

[20] Fixed exemption amounts for capital income are unlikely to affect the results for two reasons. First, capital income is typically not included in taxable income since 2009. Second, before 2009, capital income makes up a small fraction of total gross income for most households. For those with large capital income, in turn, the exemption amount may be negligible.

Figure 3.2.1: Decomposition of tax rates



Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018; all estimates are significant at the 5% level, for standard errors, see Table 3.B.3. The top row presents the results for the full sample and the bottom row distinguishes taxpayers by quartiles of per-person taxable income in the previous calendar year. For reference, the average annual inflation rate over this sample period was 1.43 percentage points.

that bracket creep continuously decreases along the income distribution; it is consistent with households in the second and third quartiles being less compensated by tax adjustments. A noteworthy heterogeneity is that bracket creep in terms of the marginal tax rate is very small at the top quartile. This is in line with the German tax schedule because the marginal tax rate is constant in the two top tax brackets. In comparison, tax rate changes due to real income growth are more pronounced in the tails of the income distribution, consistent with mean-reversion of idiosyncratic shocks. Finally, the tax rate changes from discretionary tax function changes suggest that high-income earners benefited most from discretionary tax reforms.

Time variation. The results presented so far are averages over all years, which mask potentially important time variation. In Figure 3.2.2, I present the baseline tax rate decomposition for the average and marginal tax rate for each year separately. The shaded areas indicate the time sample before and after the reform that mandated the government to publish a bracket creep report, as well as the time sample that is imputed, as explained in Section 3.2.2.

In 2004 and 2005, there were large tax cuts implemented, leading to a sizable decline in tax rates that fully compensated bracket creep and also over-compensated tax rate increases due to real income growth. In contrast, between 2006 and 2008, there were virtually no changes to the tax function, so nominal income growth led to higher tax rates. Decomposing these increases into real income growth and bracket creep suggests that bracket creep accounts for the larger fraction of tax rate increases during these years. Since there were no tax adjustments, the increases in bracket creep cumulate over multiple years: on average, in 2008, a taxpayer with the same real income as in 2005 faces an average tax rate that is 0.86 percentage points higher and a marginal tax rate that is 0.77 percentage points higher.[21] To see the cross-sectional impact of this episode, I display how the 2005 distribution of tax rates would have shifted until 2008 when all taxpayers had the same real income as in 2005 in Panels (a) and (b) of Figure 3.B.3 in Appendix 3.B.

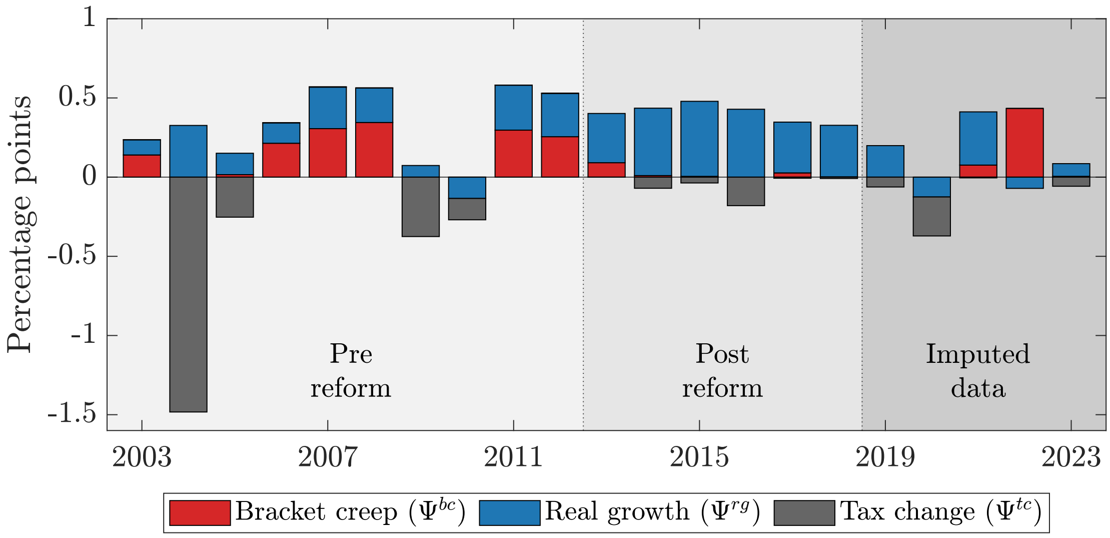
While bracket creep played no role in 2009 and 2010, there is a second bracket creep episode from 2011 until 2013. During these years, bracket creep accumulated to 0.64 and 0.67 percentage points in terms of the average and marginal tax rates, respectively. The distribution of tax rates, assuming constant real income, shifts similarly as for the first bracket creep episode; see Panels (c) and (d) of Figure 3.B.3. Comparing tax rate changes due to real income growth and bracket creep, I find that real income growth turns out to be more important. Yet, bracket creep still implies a sizable amplification relative to tax rate changes under real income growth only.

From 2013 until 2018, I find only smaller bracket creep effects, suggesting that the 2012 tax reform successfully induced regular adjustments to the tax code. In Figure 3.B.4 in

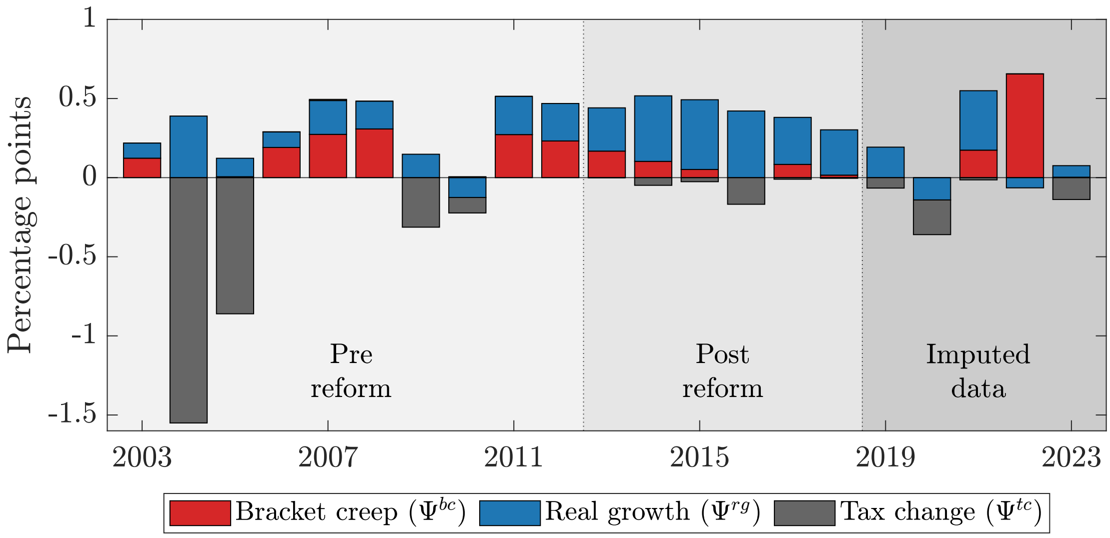
[21] These numbers can be computed by summing up the red bars for 2006 until 2008 from Figure 3.2.2.

Figure 3.2.2: Time variation in the tax rate decomposition

(a) Average tax rate



(b) Marginal tax rate



Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018; all non-zero estimates are significant at the 5% level, for standard errors, see Table 3.B.4. The imputed data is based on average household income growth as explained in Section 3.2.3.. Panel (a) and (b) show the results for the average and marginal tax rate, respectively.

the appendix, I present the average results before and after the reform, which support this conclusion.

The tax code adjustments that regularly occur since 2012 are based on inflation forecasts, as discussed in Section 3.2.2. It is not surprising that adjustments based on inflation forecasts successfully compensate for bracket creep during a period of low and stable inflation. Thus, I use the imputed data to evaluate to what extent there is bracket creep during the post-Covid inflation surprise episode. I find a sizable resurgence of bracket creep because inflation was underestimated for 2021 and especially for 2022. Cumulated over both years, this amounts to an increase of 0.51 and 0.83 percentage points in terms of average and marginal tax rates, respectively.

Finally, note that the results from Figure 3.2.2 use the baseline treatment of deductions, which are assumed to be perfectly indexed to inflation. The results when keeping deductions fixed are similar and shown in Figure 3.B.2 in Appendix 3.B.

Mechanical decomposition. The presented decomposition estimates α_t based on equation (3.3). Thus, the bracket creep term already accounts for inflation adjustments of the tax function, if they occur. An alternative is the naive mechanical decomposition, where I set $\alpha_t = 0$ directly. This decomposition computes the hypothetical bracket creep that may occur when the nominal tax code is kept constant, ignoring inflation adjustments. The resulting decomposition is presented in Figure 3.B.6 in Appendix 3.B. Two observations are noteworthy. First, during the above-highlighted bracket creep episodes, there is virtually no change in the tax function, which implies that the mechanical decomposition coincides with my baseline. Second, I can use the mechanical results and subtract the tax change term, Ψ^{tc}, from the mechanical bracket creep term, Ψ^{bc}, and set the difference to zero if negative because bracket creep is fully compensated in this case (circular markers in Figure 3.B.6). Then, this difference mostly coincides with my baseline bracket creep term (plus markers in Figure 3.B.6), suggesting that the mechanical decomposition leads to similar conclusions as the baseline version. Finally, note that this is not true for the (long-run) effects where I also average across time. There, the naive decomposition would mask important time variation because discretionary tax changes exceed bracket creep when averaged across all years.

Overall, my findings show that bracket creep accounts for a non-negligible fraction of the changes in tax rates, irrespective of whether it is measured via average or marginal tax rates. It suggests that analysis based on a tax function that only depends on real income misses important aspects of taxation that may matter empirically.

3.3 A tractable model with bracket creep

I develop an analytical partial equilibrium model that encompasses (i) a progressive income tax schedule allowing for bracket creep, (ii) a household choosing consumption and labor, and (iii) a government that uses tax revenues for government spending or transfers to the household. The model is useful to understand what we miss when abstracting from bracket creep by modeling the tax system only in real terms. I characterize the labor supply response to bracket creep and use the administrative data to estimate a theory-consistent measure of bracket creep, the indexation gap.

3.3.1 Model

Progressive tax schedule. I consider the following generalized version of the progressive income tax schedule in Heathcote, Storesletten, and Violante (2017) where

$$T(Y) = Y - \lambda \, \frac{Y^{1-\tilde{\tau}}}{1-\tilde{\tau}} \, (\mathcal{P}^g)^{\tilde{\tau}} \tag{3.4}$$

maps nominal income $Y \geq \underline{Y}$ into a nominal tax liability $T(Y)$. The parameter $\tilde{\tau} \in [0,1)$ measures the degree of tax progressivity, and $\lambda \geq 0$ captures the average level of taxation. In the special case of no progressivity, i.e., $\tilde{\tau} = 0$, the system reduces to a linear tax on Y at rate $1-\lambda$. The degree of indexation to inflation is captured by $\mathcal{P}^g > 0$, which denotes the price level to which the tax code is anchored.[22] Full indexation requires that $\mathcal{P}^g$ coincides with the (market) price level $P > 0$, whereas there will be bracket creep effects when $\mathcal{P}^g \neq P$.[23] Throughout, I assume that $\underline{Y} > \left(\frac{\lambda}{1-\tilde{\tau}}\right)^{1/\tilde{\tau}} \mathcal{P}^g$ to ensure that income is sufficiently high to have only positive tax payments because I focus on progressive income taxation and bracket creep, and not on the entire tax and transfer system. The tax schedule implies that real net-of-tax income y^{net} is given by

$$y^{net} = \lambda \, \frac{y^{1-\tilde{\tau}}}{1-\tilde{\tau}} \, x^{-\tilde{\tau}}, \tag{3.5}$$

[22] The parameter $\mathcal{P}^g$ could be subsumed in λ. However, I aim to distinguish between taxation under full indexation, as captured by $\lambda, \tilde{\tau}$, and additional bracket creep effects that will crucially depend on $\mathcal{P}^g$. In Section 3.2.3, I also show that German administrative tax data support my formulation of the tax schedule.

[23] Bracket creep in the sense of higher tax rates because of higher nominal income despite no real income gains is captured by $\mathcal{P}^g < P$, i.e., the tax code is not adjusted to increases in the price level.

where $y = Y/P$ is real pre-tax income, and $x \equiv P \,/\, \mathcal{P}^g$ is the *indexation gap* that measures the "distance" between the market price level and the level to which the tax code is anchored. The tax schedule in real terms coincides with the version from Heathcote, Storesletten, and Violante (2017) when the indexation gap is closed, i.e., $x = 1$. The tax schedule further implies

$$ATR = 1 - \lambda \frac{(yx)^{-\tilde{\tau}}}{1-\tilde{\tau}} \qquad \text{and} \qquad MTR = 1 - \lambda \, (yx)^{-\tilde{\tau}}\,, \tag{3.6}$$

where $ATR \equiv T(Y)\,/\,Y$ and $MTR \equiv T'(Y)\,/\,P$ denote the average and (real) marginal tax rate, respectively. Both tax rates increase in the indexation gap, i.e., $\partial_x ATR > 0$ and $\partial_x MTR > 0$, where ∂_x denotes the partial derivative regarding the indexation gap x. The magnitude of the increase declines in income as $\partial_{x,y} ATR < 0$ and $\partial_{x,y} MTR < 0$, where $\partial_{x,y}$ denotes the second partial derivative regarding the indexation gap x and income y.[24] It implies that bracket creep effects are stronger at the bottom of the income distribution because tax rates are more sensitive to changes in nominal income. Finally, it is worthwhile reiterating that I refer to bracket creep as the difference between real and nominal taxation. Such a difference exists whenever the indexation gap is not unity. For example, it encompasses the circumstances where the price level increases but nominal income stays constant, leading to a decline in $y = Y/P$. If the tax schedule is perfectly indexed ($x = 1$), then average and marginal tax rates would fall, partly compensating for the decline in real income. However, when the tax code is not adjusted, then the indexation gap exceeds unity providing a counteracting force leading to higher tax rates. In the knife-edge case of constant nominal incomes, both forces cancel, leaving tax rates unchanged despite the real income decline.

Household decision problem. I consider a single household that decides on consumption and labor supply, subject to the progressive income tax schedule from (3.4),

$$\max_{c,\ell} \frac{c^{1-\sigma}}{1-\sigma} - \varphi \frac{\ell^{1+\gamma}}{1+\gamma} + \log(G) \qquad \text{s.t.} \qquad c = \lambda \, \frac{(w\ell)^{1-\tilde{\tau}}}{1-\tilde{\tau}} \, (x)^{-\tilde{\tau}} + \mathcal{T}, \tag{3.7}$$

[24] Both properties are consistent with the German tax schedule where the ATR and the MTR increase with decreasing magnitude in nominal taxable income within and across tax brackets. The MTR only jumps from 42 to 45 percent at an income level of around 250,000 euros since 2009, affecting only very few taxpayers. Further, Heathcote, Storesletten, and Violante (2017) argues that their schedule provides a good approximation for the U.S. despite marginal tax rates being constant within tax brackets.

where c and ℓ are consumption and labor supply. The real wage is given by $w = W/P$, with W being the nominal wage. Parameters $\sigma \geq 0$ and $1/\gamma \geq 0$ denote relative risk aversion and the Frisch elasticity of labor supply, respectively. The parameter φ shifts disutility from labor and may be used to normalize equilibrium labor supply to unity. Finally, $\mathcal{T}$ is a transfer from the government, and G is the level of public good provision by the government. Both are taken as given by the household.

Government. The government returns a fraction $\theta \in [0, 1]$ of the tax revenues to the household which yields the real transfer

$$\mathcal{T} = \theta \left(w\ell - \lambda \, \frac{(w\ell)^{1-\tilde{\tau}}}{1-\tilde{\tau}} \, (x)^{-\tilde{\tau}} \right). \tag{3.8}$$

The remaining tax revenues are used for government spending to finance the public good.[25] The parameter θ matters for the effects of bracket creep because it governs the strength of the income effects of taxation on labor supply.[26] For the comparative statics below, I assume the indexation gap x is exogenous. In Section 3.4, I embed the tax schedule into a quantitative general equilibrium framework with endogenous indexation gaps where the government chooses a time path for $\mathcal{P}^g$, and changes in the price level P are determined in general equilibrium.

3.3.2 Theoretical results

I study how the household's labor supply responds to bracket creep. Throughout, I let (c^*, ℓ^*) denote the optimal consumption and labor supply choice of the household. I further set φ such that labor supply equals unity in the stationary equilibrium.[27] The first proposition characterizes how the labor supply response to bracket creep depends on the share of tax revenues that are given back. The proof of this and all other propositions is in Appendix 3.A.

[25] The results below focus exclusively on labor supply, which is not affected by public good provision since preferences are additively separable.

[26] All previous theoretical papers that study bracket creep assume lump-sum redistribution of tax revenues due to bracket creep, i.e., $\theta = 1$ (Edge and Rudd, 2007; Heer and Süssmuth, 2013; Keinsley, 2016).

[27] Labor supply may still respond to exogenous changes in variables, e.g., to exogenous changes in x or w.

Proposition 2. *Let $\sigma \geq 1$. Then, there exists a threshold value $\bar{\theta} \in [0,1]$ such that*

$$\frac{d\ell^*}{dx} \geq 0 \Longleftrightarrow \theta < \bar{\theta} \equiv \frac{\chi}{1+\chi}, \tag{3.9}$$

with $\chi = (\sigma - 1)\frac{\lambda}{1-\tilde{\tau}}(x\,w)^{-\tilde{\tau}} \geq 0$.

The proposition states that the labor supply response to bracket creep is ambiguous.[28] Intuitively, an increase in the indexation gap x raises average and marginal tax rates, giving rise to income and substitution effects on labor supply. It also generates additional tax revenues for the government. Returning these tax revenues to the household diminishes the income effect. Hence, when θ is too high, the substitution effect dominates, and the household works less due to higher marginal tax rates. Moreover, the threshold $\bar{\theta}$ increases in the relative risk aversion σ since the income effect increases in this parameter. This means that more tax revenues must be given back to obtain a reduction in labor supply in response to bracket creep when risk aversion is high.

Next, I focus on bracket creep and real wage fluctuations jointly. It is useful to consider the first-order dynamics of labor supply around a stationary equilibrium without bracket creep, i.e., $x = 1$. The approximate labor supply response is

$$\hat{\ell} = \Gamma_w \hat{w} + \Gamma_x \hat{x}, \tag{3.10}$$

where $\hat{\ell} = \ell^*/\ell_0^* - 1$, and ℓ_0^* denotes optimal labor supply at the point of approximation, the stationary equilibrium, and similarly for $\hat{w}$ and $\hat{x}$. The following proposition characterizes how the coefficients Γ_w and Γ_x depend on relative risk aversion σ and the usage of tax revenues θ.

Proposition 3. *Conditional on σ, there exist threshold values $\bar{\theta}_x(\sigma)$ and $\bar{\theta}_w(\sigma)$ such that*

$$\theta \geq \bar{\theta}_x(\sigma) \Longrightarrow \Gamma_x \leq 0 \quad \textit{and} \quad \theta \geq \bar{\theta}_w(\sigma) \Longrightarrow \Gamma_w \leq 0.$$

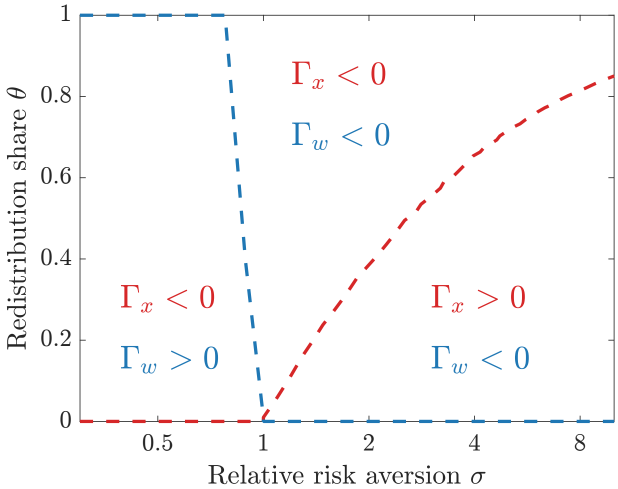
Moreover, if $\sigma > 1$, then $\bar{\theta}_x(\sigma) \in (0,1)$ and $\bar{\theta}_w(\sigma) = 0$.

The proposition states that there are threshold values for the share of tax revenues returned to the household, the redistribution share θ, that determine the labor supply response to

[28] When the derivative $d\ell^*/dx$ is evaluated at the stationary equilibrium, then w and x in χ refer to the values of w and x in the stationary equilibrium that prevails absent the marginal increase dx. Similarly, one could also study the labor supply response where $x = w = 1$ in the stationary equilibrium.

changes in the real wage and to changes in the indexation gap. In the empirically plausible case that relative risk aversion exceeds unity, we have that the threshold value for the indexation gap strictly exceeds the threshold for real wages.

Figure 3.3.1: Labor supply response regions



Notes: The figure shows parameter regions for which an increase in the real wage leads to higher labor supply ($\Gamma_w > 0$), and for which an increase in the indexation gap leads to higher labor supply ($\Gamma_x > 0$), based on the first-order dynamics from (3.10). The remaining parameters are $\tilde{\tau} = 0.2$, $\lambda = 0.6$, $\gamma = 2$ and $w = 1$.

Examples. To illustrate the implications of the proposition, I display a numerical example in Figure 3.3.1. There are distinct parameter regions that govern whether or not increases in the real wage and in the indexation gap have opposing effects on labor supply. For concreteness, suppose that the price level P and the nominal wage W increase with inflation at rate $\Pi > 1$, but the government keeps $\mathcal{P}^g$ constant. Thus, the real wage stays constant, but the indexation gap increases, i.e., $\hat{x} = \Pi - 1 > 0$. This exemplifies what is commonly understood as bracket creep, i.e., an increase in the nominal wage that only compensates for inflation leads to higher average and marginal tax rates, which affect labor supply. The effect on labor supply is only negative when relative risk aversion is sufficiently small or the redistribution share is high enough, as illustrated in Figure 3.3.1.

Alternatively, consider that nominal wages remain unchanged, but the price level still grows at the rate Π, implying that the indexation gap increases and the real wage falls.[29] The overall effect on labor supply now depends on the response to both variables. When risk

[29] This gives rise to bracket creep effects because tax rates do not fall as they would under full indexation to partly compensate for the real income loss; see the discussion along with (3.6).

aversion is sufficiently low, then the decline in the real wage leads to less labor supply. This is amplified by bracket creep, which further reduces labor supply.[30] When risk aversion and the redistribution share are sufficiently high, then the real wage loss leads to a higher labor supply, but bracket creep effects lead to a lower labor supply, dampening the overall effect. Finally, when risk aversion is sufficiently high and redistribution is not too high, then both the real wage loss and bracket creep lead to a higher labor supply. Thus, bracket creep amplifies the effects of the price level increase in the latter case.

Overall, this shows that bracket creep can either amplify or dampen the effects of progressive taxation that would prevail when the tax system is perfectly indexed to inflation.

3.3.3 Empirical indexation gaps

Next, I develop an approach to measure the indexation gap in the data. Based on this, I present two pieces of evidence that support the parametric tax schedule, which I introduced above. First, the resulting time series of the indexation gap is highly correlated with the bracket creep term based on the non-parametric decomposition. Second, I present regression results that directly deliver a test of a restriction derived from the tax schedule. Finally, the indexation gap series is useful to discipline the quantitative model in Section 3.4.

Estimation strategy. I use the empirical measure of the degree of indexation as defined in equation 3.3 to obtain an empirical measurement of the indexation gap. This only requires assuming the tax schedule specified in equation 3.4, but not the remainder of the analytical model. Consider the change in the average tax rate

$$ATR_t(y,x_t) - ATR_{t-1}(y,x_{t-1}) = \frac{\lambda}{1-\tilde{\tau}}\,(yx_{t-1})^{-\tilde{\tau}}\left(1-\left(\frac{\Pi_t}{\Pi_t^g}\right)^{-\tilde{\tau}}\right), \tag{3.11}$$

where $\Pi_t^g \equiv \mathcal{P}_t^g/\mathcal{P}_{t-1}^g$ determines the strength of bracket creep. When $\Pi_t \geq \Pi_t^g = 1$, then this corresponds to the change in average tax rates in the model under a full bracket creep benchmark, analogously to the measurement of indexation in Section 3.2.1. Next, I define

[30]Note that the effects of $\hat{w}$ and $\hat{x}$ on labor supply have the same sign when Γ_w and Γ_x have opposite signs as x increases but w decreases in this example.

a model counterpart to the degree of indexation measured in the data based on

$$\tilde{\alpha}_t = 1 - \frac{ATR_t(y, x_t) - ATR_{t-1}(y, x_{t-1})}{ATR_t(y, x_{t-1}) - ATR_{t-1}(y, x_{t-1})} = 1 - \frac{1 - \left(\Pi_t / \Pi_t^g\right)^{-\tilde{\tau}}}{1 - (\Pi_t)^{-\tilde{\tau}}} \tag{3.12}$$

Imposing the empirical degree of indexation equals the model counterpart, i.e., $\alpha_t = \tilde{\alpha}_t$, yields $\Pi_t^g = \left(\alpha_t \Pi_t^{\tilde{\tau}} + (1 - \alpha_t)\right)^{1/\tilde{\tau}}$. Hence, given the empirical degree of indexation α_t and an estimate of the progressivity parameter $\tilde{\tau}$, I can measure the growth rate of the indexation parameter $\mathcal{P}_t^g$. Under the assumption of no bracket creep at date zero, i.e., $\mathcal{P}_0^g = P_0$, I compute the indexation parameter as

$$\mathcal{P}_t^g = \begin{cases} \mathcal{P}_{t-1}^g \Pi_t^g & \text{if} \quad \alpha_t > 0 \\ P_t & \text{otherwise,} \end{cases} \tag{3.13}$$

where full make-up for past bracket creep is assumed when $\alpha_t = 0$. I follow this approach because the empirical measure does not capture make-up for accumulated bracket creep from previous periods. Therefore, the implied indexation gap $x_t = P_t / \mathcal{P}_t^g$ is a lower bound of the true unobserved indexation gap, understating bracket creep effects. This requires the identifying assumption that the government only compensates past bracket creep but does not compensate for future bracket creep through over-compensating contemporaneous bracket creep. Maintaining this assumption is needed to disentangle bracket creep and indexation from discretionary tax changes.[31]

Implementation. To obtain the tax progressivity $\tilde{\tau}$, I follow Heathcote, Storesletten, and Violante (2017) and run an OLS regression of log real net-of-tax income on log real pre-tax income, using the full sample of taxpayers from 2002 until 2018. The coefficient on log pre-tax income is an estimate of $1 - \tilde{\tau}$. The results are given in Column (1) of Table 3.3.1, with standard errors clustered at the taxpayer level in parenthesis. The estimate is significant at any conventional level and implies $\tilde{\tau} = 0.14$. It is further reassuring that my estimate is close to the estimate of 0.16 presented in Heathcote, Storesletten, and Violante (2020) for Germany in the year 2005. Given this estimate, I can readily compute indexation gaps at the taxpayer level.[32]

[31] Without such an assumption, it is not possible to discriminate between compensation for bracket creep and discretionary tax code changes that are unrelated to inflation. This is because any tax cut can be interpreted as only compensating for future inflation.

[32] I use the baseline results for α_t where deductions are assumed to grow at the inflation rate.

Table 3.3.1: Estimated tax parameters

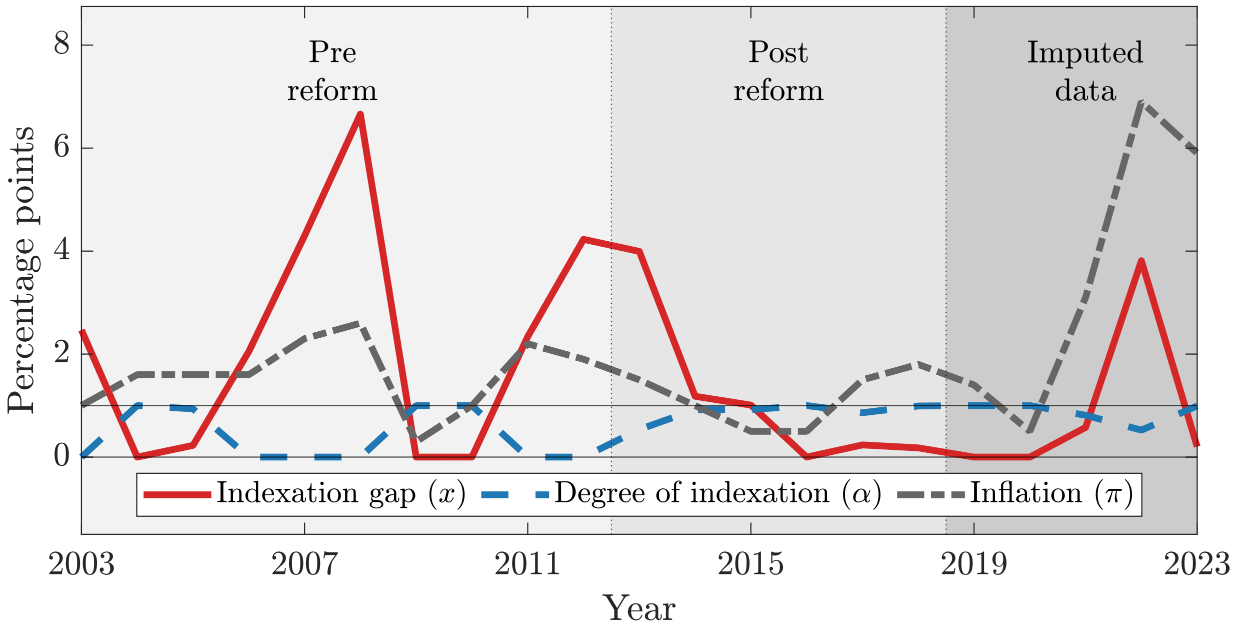
	(1)	(2)
Real pre-tax income: $\log(y)$	0.86	0.86
	(0.0001)	(0.0001)
Indexation gap: $\log(x)$		-0.12
		(0.0004)
Constant	0.70	0.70
	(0.0005)	(0.0007)
R^2	0.999	0.999
Taxpayer FE	✓	✓
Observations (Mio.)	14.371	9.203

Notes: The table shows OLS regression results based on equation (3.14) using the administrative tax records from 2002 until 2018, as presented in Section 3.2.2. Standard errors are clustered at the taxpayer level and provided in parentheses.

I compute indexation gaps at the taxpayer level (since α_t is measured at this level) and then aggregate in the cross-section of taxpayers for each year. The resulting time series is presented in Figure 3.3.2. I further display the degree of indexation and the inflation rate to illustrate how both relate to indexation gaps. There is sizable time variation in the indexation gap, which peaks at 6.7 percentage points in 2008. Despite the tax reform in 2012, it took until 2016 to close the measured indexation gap. During the 2022 inflation surge, a sizable indexation gap is visible, albeit less pronounced than the peak gap in 2008. This reflects the fact that the lack of indexation was very transitory during the recent inflation surge. In this period, the increase in the indexation gap is primarily driven by the inflation surge and less by the degree of indexation, which remains relatively high. Overall, the indexation gap is strongly correlated with the bracket creep term from the decomposition in Section 3.2.3. It suggests that the parametric tax schedule from (3.4) and the resulting indexation gaps are consistent with the non-parametric decomposition results.

Testing the tax schedule. My approach of estimating indexation gaps imposes the tax function from equation 3.4. To check whether this tax function is supported by the data,

Figure 3.3.2: Time variation in the indexation gap



Notes: The figure shows the time series of the indexation gap, based on equations (3.11)-(3.13); the empirical degree of indexation based on equation (3.3); and the inflation rate. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018. The imputed data is based on average household income growth as explained in Section 3.2.3.

I consider the following regression

$$\log y_{i,t}^{net} = a_i + b\log(y_{i,t}) + c\log(x_{i,t}) + v_{i,t} \tag{3.14}$$

where i and t index individual taxpayers and years, and a_i are taxpayer fixed effects. A testable prediction of the model is that $1-b = -c = \tilde{\tau}$. It follows from taking the log of real net-of-tax income from equation (3.5). The OLS estimates of this specification are presented in Column (2) of Table 3.3.1. The results are remarkably close to the theoretical prediction as the OLS estimates of b and c imply a degree of progressivity of 0.14 and 0.12, respectively.[33] I interpret this result as supporting the formulation of the proposed tax schedule.

[33] Ex-ante, a concern could be that the regression to obtain $\tilde{\tau}$ in the first place (without indexation gaps) was misspecified. Ex-post, however, it turns out that including indexation gaps in the specification does not change the implied $\tilde{\tau}$ estimate.

3.4 A New Keynesian model with bracket creep

I study how nominal progressive taxation affects the propagation of macroeconomic shocks through the bracket creep channel. The analysis is based on a New Keynesian model with incomplete markets. Calibrating the model to the German economy before the indexation reform in 2012, I show that indexation of the tax schedule amplifies the short-run output effects of monetary policy.

3.4.1 Model

The model is a closed economy populated by a continuum of households with unit mass and time is discrete.

Households. I consider a generalized version of the household setup in the analytical model from Section 3.3. Households consume a final good, supply labor, and may save in a liquid bond, $b_{i,t}$. All households are ex-ante identical and solve the following dynamic problem

$$V_t(e_{i,t}, b_{i,t-1}) = \max_{c_{i,t}, \ell_{i,t}, b_{i,t}} \left(\frac{c_{i,t}^{1-\sigma}}{1-\sigma} - \varphi \frac{\ell_{i,t}^{1-\gamma}}{1-\gamma} + \log(G_t) + \beta \mathbb{E}_t \Big[V_{t+1}(e_{i,t+1}, b_{i,t}) \Big] \right) \tag{3.15}$$

$$\text{s.t. } c_{i,t} + b_{i,t} = \lambda \, \frac{(w_t e_{i,t} \ell_{i,t})^{1-\tilde{\tau}}}{1-\tilde{\tau}} \, (x_t)^{-\tilde{\tau}} + b_{i,t-1}(1+r_t) + \mathcal{T}_{i,t} + d_{i,t},$$

$$-\underline{b} \leq b_{i,t}, \quad \text{and} \quad \ln(e_{i,t+1}) = \rho_e \ln(e_{i,t}) + \upsilon_{i,t+1},$$

where $e_{i,t}$ is the idiosyncratic endowment of labor efficiency units and $\upsilon_{i,t} \overset{iid}{\sim} N(0, \sigma_e)$, and $|\rho_e| < 1$. Borrowing must not exceed $-\underline{b}$. The government transfer $\mathcal{T}_{i,t}$ and dividends from the firms $d_{i,t}$ are distributed proportionally to labor efficiency units as in (Auclert, Bardóczy, Rognlie, and Straub, 2021). Both are taken as given by households. The level of public good provision is given by G_t. Aggregate variables w_t, r_t, and x_t denote the real wage, the real interest rate, and the indexation gap, respectively. Flow utility is additively separable and depends on constant relative risk aversion σ, inverse Frisch elasticity of labor supply γ, and a labor disutility shifter φ. The beginning-of-period bond holdings $b_{i,t-1}$ are given, and households choose labor supply $\ell_{i,t}$ and allocate the disposable income to consumption $c_{i,t}$ and liquid bond holdings $b_{i,t}$.

Production. The final consumption good is produced from a continuum of varieties by a representative final good firm based on a technology with constant elasticity of substitution given by $\mu/(\mu-1)$. The varieties are produced by intermediate good firms $j \in [0,1]$ that use a constant returns-to-scale technology, and labor is the only production input. The intermediate good producers are monopolistically competitive and take the demand schedule of the final good firm as given when setting retail prices $p_{j,t}$, subject to quadratic adjustment costs $\mathcal{C}_t = \mathcal{K}\ \log(p_{j,t}/p_{j,t-1})^2\ Y_t$, with Y_t being aggregate output and the firm discount rate is given by the real interest rate.[34] The adjustment cost parameter $\mathcal{K} = \frac{\mu}{2(\mu-1)\kappa}$ is defined such that κ will represent the slope of the Phillips curve. Solving the firm problem and imposing a symmetric equilibrium across intermediate good producers gives rise to a standard New Keynesian Phillips curve

$$\log(\Pi_t) = \kappa\left(\frac{w_t}{Z_t} - \frac{1}{\mu}\right) + \frac{1}{1+r_{t+1}}\frac{Y_{t+1}}{Y_t}\log(\Pi_{t+1}), \tag{3.16}$$

where $\Pi_t = P_t/P_{t-1}$ is the gross inflation rate, and Z_t is total factor productivity. The slope of the Phillips curve is given by κ, which measures the price adjustment costs. Finally, profits or losses of the firm are distributed to households through dividends $d_t = Y_t - w_tN_t - \mathcal{C}_t$, where N_t is total labor demand.[35]

Government. The government consists of a fiscal and a monetary authority. The fiscal authority issues the liquid bond, collects real tax revenues from progressive income taxation R_t, and decides on the amount of aggregate transfers $\mathcal{T}_t$ and government spending for the public good G_t. It faces the following period-by-period real budget constraint

$$R_t = r_tB^g + G_t + \mathcal{T}_t. \tag{3.17}$$

Let variables without time subscript denote the steady-state values, then the government behavior (for deviations from the steady-state) is described by

$$G_t - G = \phi_g(R_t - r_tB^g - G - \mathcal{T}) \quad \text{and} \quad \mathcal{T}_t - \mathcal{T} = (1-\phi_g)(R_t - r_tB^g - G - \mathcal{T}).$$

[34] I assume a price and not a wage rigidity because endogenous labor supply is key for my mechanism. In contrast, rigid wages are typically implemented via a labor union that sets wages and hours uniformly for all households, implying that agents are not on their individual labor supply curves (e.g., Auclert, Bardóczy, Rognlie, and Straub, 2021).

[35] Note that firms must always serve demand and there is no firm exit. Negative profits imply a negative dividend, which can be interpreted that the firms are raising additional equity from the household sector.

The equations state that a fraction ϕ_g from the tax revenues after paying interest, as well as steady-state transfers and government spending, is used for additional government spending, and the remainder constitutes an additional transfer to households. I follow this approach because it allows me to calibrate ϕ_g such that the composition of government expenses for public good provision and transfers is kept constant in response to shocks. This is important because it ensures that the results are not driven by the government using bracket creep tax revenues to alter the composition of government expenses.[36] Finally, the fiscal authority decides the indexation parameter $\mathcal{P}_t^g$ that determines the indexation gap $x_t = P_t/\mathcal{P}_t^g$ such that

$$(x_t - x) = \phi_x(x_{t-1} - x) + (1-\alpha)(\Pi_t - 1), \tag{3.18}$$

where α determines how much of inflation is instantaneously compensated and ϕ_x governs how fast $\mathcal{P}^g$ is adjusted given the already accumulated indexation gap x_{t-1}. Importantly, the tax system is fully indexed when $\alpha = 1$ such that the indexation gap is always at its steady-state value, i.e., $x_t = x$, $\forall t$. Finally, the model is closed with a standard Taylor rule that governs the behavior of the monetary authority

$$i_t = \phi_\Pi(\Pi_t - \Pi) + m_t, \tag{3.19}$$

where i_t is the nominal interest rate that maps into the real rate via the Fisher equation $r_t = (1+i_t)/\Pi_t - 1$, and m_t is an autocorrelated monetary policy shock that follows a stable auto-regressive process with $m_t = \rho_m m_{t-1} + \varepsilon_t^{mp}$ and standard normal innovations ε_t^{mp}.

Model solution. An equilibrium consists of sequences for all household, firm, and governmental variables such that all private agents behave optimally (given prices and the transfer) and such that the goods-, labor- and asset markets clear at any date t. The model is solved based on a first-order perturbation in sequence space (Auclert, Bardóczy, Rognlie, and Straub, 2021) around a steady state with zero inflation and zero indexation gaps, i.e., $\Pi_t - 1 = x_t - 1 = 0$. By studying an economy with zero trend inflation, I follow most of the New Keynesian literature (e.g., Auclert, Rognlie, and Straub, 2023; Gali, 2015).

[36] It would be desirable to measure how tax revenues due to bracket creep are used in the data. Unfortunately, this is infeasible because it would require exogenous variation in bracket creep that does not affect the government budget constraint through other channels.

Calibration. The model is calibrated to the German economy before the 2012 tax reform that reduced bracket creep and a period is a quarter. All parameters are summarized in Table 3.4.1. Relative risk aversion and the inverse Frisch elasticity of labor supply are set to conventional values. The discount factor and the labor disutility shifter are set to match an annualized real interest rate of two percent and an effective steady-state labor supply of unity under the normalization that steady-state TFP is $Z = 1$. The parameters that govern the endowment with idiosyncratic labor efficiency units are set to match the annual moments according to the GRID database, which is constructed from German administrative data. The remaining supply-side parameters are set to standard values. The borrowing limit corresponds to the average monthly income in the steady-state. The supply of government bonds from the government matches an annual debt-to-GDP ratio of 60%. The tax progressivity parameter is set to $\tilde{\tau} = 0.14$, in line with my estimation results from above, and the tax level is set to match the ratio of aggregate tax payments to income in the same data.[37] The government spending to GDP ratio is set to 12.5%, which implies that 40% of the tax revenues net of interest payments are used for government spending to provide the public good G_t, and the remainder is redistributed to households. The parameter ϕ_g is set to keep the ratio of government spending to transfers constant in response to shocks. Finally, I set $\alpha = 0$ as there were multiple periods with no compensation for bracket creep before 2012 and estimate the degree of mean reversion ϕ_x based on the computed indexation gaps.[38] The Taylor rule coefficients and the autocorrelation of the monetary shock are set to conventional values.

3.4.2 Results

The baseline calibration corresponds to the German economy before the 2012 tax reform that reduced bracket creep. In this setup, I study how the propagation of macroeconomic shocks is altered through the presence of bracket creep. The underlying idea is that any shock that affects inflation impacts tax rates through the bracket creep channel. In the context of monetary policy, it implies a new dimension of fiscal-monetary interaction because monetary shocks partly propagate through fiscal instruments via inflation.

In Figure 3.4.1, I display the responses to an expansionary monetary policy shock that equals a 25 basis point rate cut on impact. The blue solid line shows the baseline economy

[37] Note that the tax data does not include social security contributions. Thus, I only consider progressive income taxes but not the entire tax and transfer system.

[38] Since the indexation gap series is annual, I interpolate to quarterly frequency by assuming that the indexation gap remains constant within a given year and estimate ϕ_x using OLS.

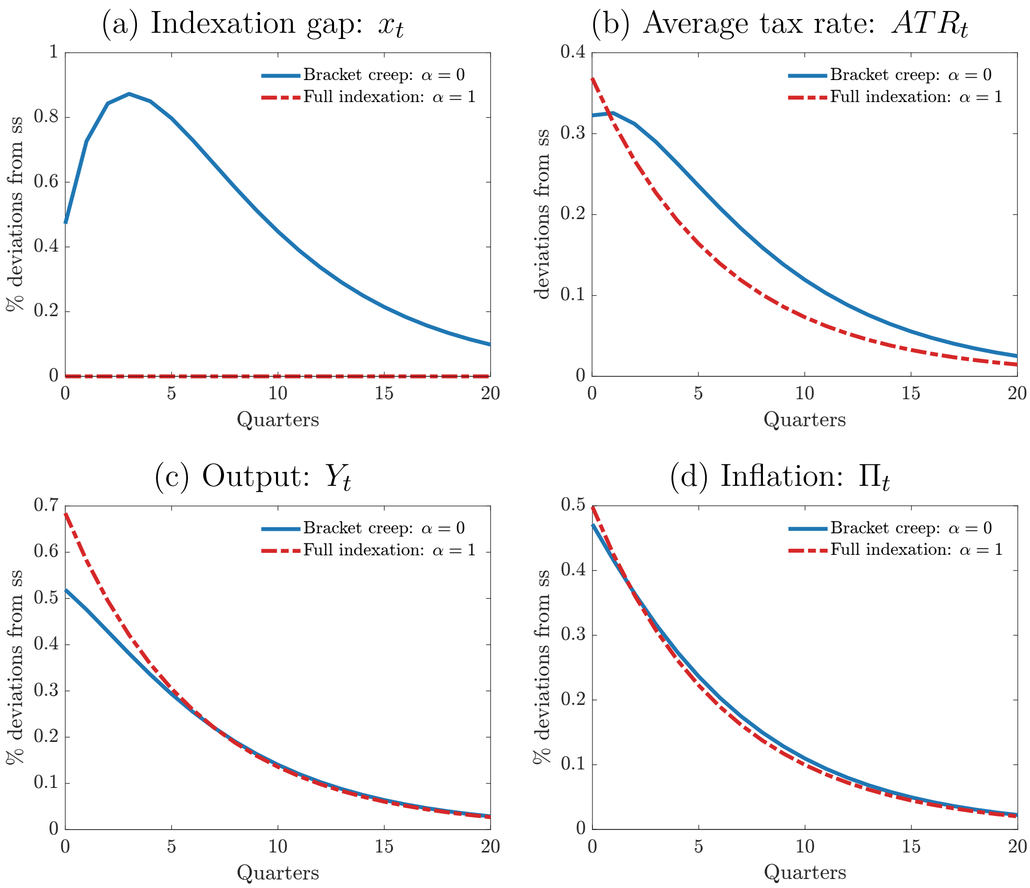
Table 3.4.1: Calibration

Variable		Value	Target/Source
Relative risk aversion	σ	2.00	standard value
Inverse Frisch elasticity	γ	4.00	standard value
Discount factor	β	0.99	Annual real rate of 2%
Labor disutility shifter	φ	0.66	Steady-state labor $\int e_i \ell_i di = 1$
Borrowing limit	$\underline{b}$	0.33	Steady-state monthly average income
SD of idiosy. endowment	σ_e	0.25	Annual SD of income
Autocorr. of idiosy. endowment	ρ_e	0.90	Annual autocorr. of income
Steady-state markup	μ	1.10	10% markup
Slope of the Phillips curve	κ	0.025	standard value
Steady-state TFP	Z	1.00	Normalization
Tax progressivity	$\tilde{\tau}$	0.14	Estimated based on tax data
Tax level	λ	0.65	Tax-income ratio in tax data
Gov't bond supply	B^g	2.40	Debt-to-GDP of 60%
Steady-state share of G	G/Y	0.125	Spending-transfer ratio of 40%
G spending response	ϕ_g	0.40	Steady-state spending-transfer ratio
Degree of indexation	α	0.00	Full bracket creep
Autocorr. of indexation gap	ϕ_x	0.66	Estimated based on tax data
Taylor rule coefficient	ϕ_Π	1.50	standard value
Autocorr. of MP shock	ρ_m	0.85	standard value

Notes: Calibration for the baseline economy, corresponding to Germany for the period 2003 until 2012 before the bracket creep tax reform. The annual income data moments are taken from the GRID database.

without indexation, and the dashed red line shows the same economy but with a perfectly indexed tax code, i.e., $\alpha = 1.0$. Panel (a) displays the response of the indexation gap. In the bracket creep economy, the indexation gap equals the rate of inflation on impact but then further builds up because the fiscal authority adjusts the tax code only slowly to the new price level. In contrast, under full indexation, the indexation gap is always closed

Figure 3.4.1: Responses to an expansionary monetary policy shock



Notes: The figure shows impulse responses based on the one asset New Keynesian model with incomplete markets as specified in Section 3.4.1. The expansionary monetary policy shock is a nominal interest rate cut of 25 basis points.

and remains at the steady-state level. In Panel (b), I show the implied effects on average tax rates. Already under full indexation, the average tax rate increases because of larger real income. In the bracket creep economy, this is further amplified through bracket creep effects. Quantitatively, the increase in the average tax rate is 36% larger after one year. Note that the (first-order) response of the marginal tax rate coincides with the average tax rate response under the tax schedule that I assume. Moving to the output response in Panel (c), one can see that bracket creep dampens the expansionary effects of monetary policy. Intuitively, the increase in the tax rate induces a substitution effect that discourages labor supply. On the other hand, the income effect is not sufficiently strong under the calibration assumption that the composition of government spending and transfers is kept constant. Finally, the inflation responses are given in Panel (d). While ex-ante, it is unclear whether the presence of bracket creep meaningfully alters inflation dynamics, it turns out that it is quantitatively irrelevant for a monetary policy shock.

The results for a contractionary monetary policy shock are symmetric. A given reduction in inflation via a contractionary monetary policy shock leads to substantially smaller short-run output costs in an economy with bracket creep. Put differently, the output costs are aggravated when the tax code is indexed to inflation, revealing a potential caveat of such indexation schemes.

The magnitude of the presented effects depends on various parameters. In particular, the slope of the Phillips curve determines the inflation response to the shock and, hence, the strength of bracket creep. While many recent estimates suggest fairly low parameters (e.g., Hazell, Herreno, Nakamura, and Steinsson, 2022), the recent inflation surge may be hard to reconcile with a very flat Phillips curve.

Overall, the model results suggest that bracket creep (or indexation) may alter the transmission of monetary policy shocks in a meaningful way.

3.5 Conclusion

Bracket creep effects occur when inflation changes tax rates because the progressive income tax schedule is not adjusted. I document the quantitative importance of bracket creep over time using German administrative tax records. I find that bracket creep played an important role in changes in tax rates until around 2012. In 2012, a tax reform led to a substantial decline in bracket creep because tax code adjustments based on inflation forecasts performed well when inflation was relatively low and stable. However, the post-Covid inflation surge led to a resurgence with sizable bracket creep effects. Moving to the theoretical results, I characterize how bracket creep affects labor supply decisions in a partial equilibrium framework. Further, I estimate a theory-consistent measure of bracket creep, the indexation gap, which is used to discipline a New Keynesian model with incomplete markets. The model predicts that a given reduction in inflation via a monetary contraction leads to less output costs in an economy with bracket creep.

Going forward, there are several avenues for expanding this research. First, not only the income tax schedule but also many other government policies are specified in nominal terms, including unemployment insurance, childcare subsidies, and more. Inflation adjustments are often infrequent and incomplete. Thus, quantifying the effects of imperfect inflation adjustments would be valuable to investigate whether they impact shock transmission, and to understand where large gains from indexation are available. Second, potentially inefficient fluctuations in taxes due to bracket creep and delayed compensation may amplify the welfare costs of inflation. This provides a motive for a lower inflation target by the central bank when imperfectly indexed taxes are taken as given. Future work may quantify the importance of this channel. Finally, extending the empirical analysis to more countries within and beyond the Euro Area would be important to quantify bracket creep effects more broadly.[39]

[39] For example, as of 2022, Italy and Spain have not adjusted their tax parameters since 2007 and 2015, respectively (see OECD: Tax Database Table I.1 Central Government Personal Income Tax Rates and Thresholds).

Appendix

3.A Derivations

Proof of Proposition 1. Inserting Definition 1 in Ψ_t^{bc} yields

$$\begin{aligned}
\Psi_t^{bc} &= t_t^{\mathcal{I}}\left(Z_t^{\Pi},\ Z_{t-1}\right) - t_{t-1}\left(Z_{t-1}\right) \\
&= \left[(1-\alpha_t)\, t_{t-1}\left(Z_t^{\Pi}\right) + \alpha_t\, t_{t-1}\left(Z_{t-1}\right)\right] - t_{t-1}\left(Z_{t-1}\right) \\
&= (1-\alpha_t)\left[t_{t-1}\left(Z_t^{\Pi}\right) - t_{t-1}\left(Z_{t-1}\right)\right].
\end{aligned}$$

When $\Pi_t > 1$ and $Z_t > 0$ (the latter being assumed throughout in the main text), then $Z_t^{\Pi} > Z_t$. As $\frac{\partial t_{t-1}(Z)}{\partial Z} > 0,\ \forall Z \in [Z_{t-1}, Z_t^{\Pi}]$, it follows that $t_{t-1}(Z_t^{\Pi}) > t_{t-1}(Z_{t-1})$, which completes the proof. $\square$

Proof of Proposition 2. Substituting the budget constraint for c in the household problem yields the first-order condition

$$\begin{aligned}
0 &= \left(c^*\right)^{-\sigma} \lambda w^{1-\tau} x^{-\tau} - \varphi\left(\ell^*\right)^{\gamma+\tau} \\
&= \left(\theta \ell^* w + (1-\theta)\frac{\lambda}{1-\tau}\left(w\ell^*\right)^{1-\tau} x^{-\tau}\right)^{-\sigma} \lambda w^{1-\tau} x^{-\tau} - \varphi\left(\ell^*\right)^{\gamma+\tau},
\end{aligned} \tag{3.20}$$

where the second equality uses the budget constraint and the definition of $\mathcal{T}$. Totally differentiating with respect to x and ℓ^* gives

$$\begin{aligned}
0 = &\left(-\tau (c^*)^{-\sigma} \lambda w^{1-\tau} x^{-\tau-1} \left[1 - \sigma (c^*)^{-1} (1-\theta) \frac{\lambda}{1-\tau} (w\ell^*)^{1-\tau} x^{-\tau}\right]\right) dx \\
&+ \left(-\sigma (c^*)^{-\sigma-1} \left[\theta \lambda w^{2-\tau} x^{-\tau} + (1-\theta) \left(\lambda w^{1-\tau} x^{-\tau}\right)^2 (\ell^*)^{-\tau}\right] - (\gamma+\tau) \varphi (\ell^*)^{\gamma+\tau-1}\right) d\ell^* \\
\iff \frac{d\ell^*}{dx} = &\frac{-\tau (c^*)^{-\sigma} \lambda w^{1-\tau} x^{-\tau-1} \left[1 - \sigma (c^*)^{-1} (1-\theta) \frac{\lambda}{1-\tau} (w\ell^*)^{1-\tau} x^{-\tau}\right]}{\sigma (c^*)^{-\sigma-1} \left[\theta \lambda w^{2-\tau} x^{-\tau} + (1-\theta) (\lambda w^{1-\tau} x^{-\tau})^2 (\ell^*)^{-\tau}\right] + (\gamma+\tau) \varphi (\ell^*)^{\gamma+\tau-1}}.
\end{aligned}$$

The denominator is strictly positive as all parameters, as well as the real wage, the indexation gap and equilibrium choices c^*, ℓ^* are strictly positive. For the same reason, it follows that the sign of $d\ell^*/dx$ is pinned down by the term in square brackets as the multiplicative term in front of it is strictly negative. Hence, the cutoff $\bar{\theta}$ is determined via

$$\begin{aligned}
0 &= 1 - \sigma (c^*)^{-1} (1-\bar{\theta}) \frac{\lambda}{1-\tau} (w\ell^*)^{1-\tau} x^{-\tau} \\
&= \bar{\theta} \ell^* w + (1-\bar{\theta}) \frac{\lambda}{1-\tau} (w\ell^*)^{1-\tau} x^{-\tau} - \sigma (1-\bar{\theta}) \frac{\lambda}{1-\tau} (w\ell^*)^{1-\tau} x^{-\tau} \\
&= \bar{\theta} - (\sigma-1)(1-\bar{\theta}) \frac{\lambda}{1-\tau} (xw\ell^*)^{-\tau} \\
&= \bar{\theta} - (1-\bar{\theta})\chi
\end{aligned}$$

where the second equality follows from multiplying with c^*. Since $\chi \geq 0 \iff \sigma \geq 1$, it follows that $0 < \theta - (1-\theta)\chi \iff \theta > \bar{\theta}$ which implies $d\ell^*/dx < 0$, and analogously for $\theta < \hat{\theta}$. Note that the last equality also uses that $\ell^* = 1$ as stated in the main text.

□

Proof of Proposition 3. I first establish a Lemma that will be useful for this proof.

Lemma 1. *The coefficients from equation* (3.10) *in the main text are*

$$\Gamma_x = \frac{\tau \vartheta_0}{\gamma + \tau + \sigma \vartheta_1} \qquad \textit{and} \qquad \Gamma_w = \frac{1 - \tau - \sigma \vartheta_1}{\gamma + \tau + \sigma \vartheta_1},$$

with

$$\vartheta_0 = \frac{(\sigma-1)(1-\theta)\lambda(1-\tau)^{-1}(w\ell^*)^{-\tau} - \theta}{\theta + (1-\theta)\lambda(1-\tau)^{-1}(w\ell^*)^{-\tau}} \quad \textit{and} \quad \vartheta_1 = \frac{\theta + (1-\theta)\lambda(w\ell^*)^{-\tau}}{\theta + (1-\theta)\lambda(1-\tau)^{-1}(w\ell^*)^{-\tau}}.$$

Proof. A first order approximation of (3.20) around (ℓ^*, w, x) with $x = 1$ yields

$$\begin{aligned}
0 = &\left[-\sigma\left(c^*\right)^{-\sigma-1}\left(\theta w\ell^* + (1-\theta)\lambda\left(w\ell^*\right)^{1-\tau}\right) - (\gamma+\tau)\varphi\left(\ell^*\right)^{\gamma+\tau}\right]\hat{\ell} \\
&+ \left[-\sigma\left(c^*\right)^{-\sigma-1}\left(\theta w\ell^* + (1-\theta)\lambda\left(w\ell^*\right)^{1-\tau}\right)\lambda w^{1-\tau} + \left(c^*\right)^{-\sigma}\lambda(1-\tau)w^{1-\tau}\right]\hat{w} \\
&+ \left[\sigma\left(c^*\right)^{-\sigma-1}(1-\theta)\lambda^2\left(w\ell^*\right)^{1-\tau}(1-\tau)^{-1}\tau w^{1-\tau} - \left(c^*\right)^{-\sigma}\lambda w^{1-\tau}\tau\right]\hat{x}
\end{aligned}$$

Inserting the household budget constraint and equilibrium transfers $\mathcal{T}$ for c^*, and inserting (3.20) for $\varphi\left(\ell^*\right)^{\gamma+\tau}$, and rearranging gives the result. □

Now I turn to the proof of Proposition 3 from the main text.

Existence of $\bar{\theta}_x(\sigma)$. Lemma 1 implies $\vartheta_1 \geq 0$ and that the denominator of Γ_x is strictly positive under the parameter restrictions stated in the main text. It also implies that the denominator of ϑ_0 is strictly positive. It follows that, conditional on σ, the sign of Γ_x is pinned down by

$$f(\theta;\sigma) \equiv \tau\left[(\sigma-1)(1-\theta)\lambda(1-\tau)^{-1}\left(w\ell^*(\theta;\sigma)\right)^{-\tau} - \theta\right],$$

where I make explicit that $\ell^* > 0$ depends on σ and θ. First, consider $\sigma \in [0,1]$. In this case, I have $f(\theta;\sigma) \leq 0$, and hence, $\Gamma_x \leq 0$, regardless of θ. This implies that $\bar{\theta}_x(\sigma) = 0$ for $\sigma \in [0,1]$. Second, consider $\sigma > 1$. Now, I have $f(1;\sigma) < 0$ and $f(0,\sigma) > 0$. The existence of $\bar{\theta}_x(\sigma) \in (0,1)$ such that $f(\bar{\theta}_x(\sigma);\sigma) = 0$ follows from the intermediate value theorem. Taken together, a threshold $\bar{\theta}_x(\sigma)$ exists for all $\sigma \geq 0$ such that $\Gamma_x \leq 0$ if $\theta \geq \bar{\theta}_x(\sigma)$.

Existence of $\bar{\theta}_w(\sigma)$. From Lemma 1 (using the same arguments as for $\bar{\theta}_x(\sigma)$), we can see that the sign of Γ_w is determined by

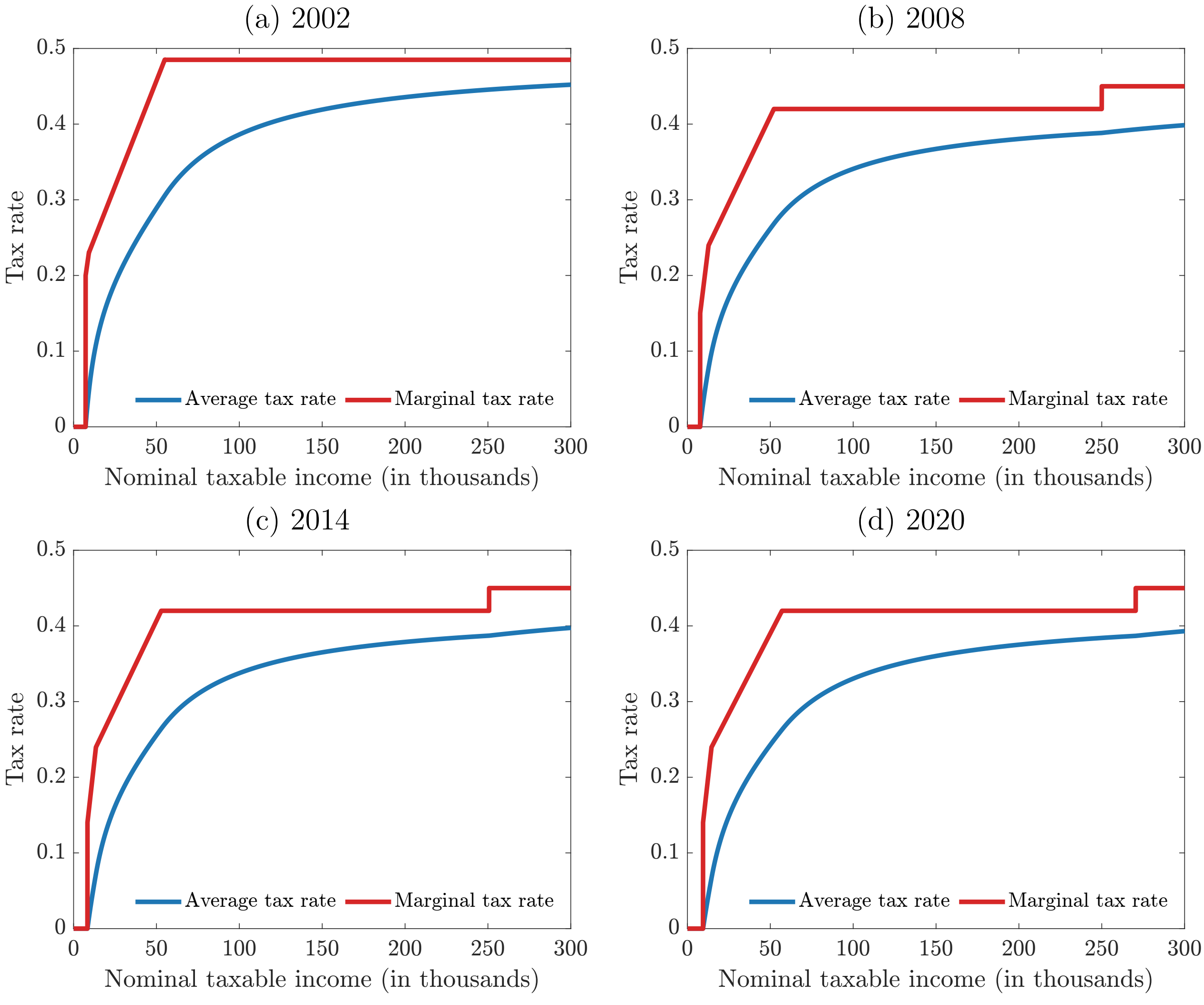
$$g(\theta;\sigma) \equiv 1 - \tau - \sigma\vartheta_1(\theta;\sigma).$$

Note that $\vartheta_1 \equiv \vartheta_1(\theta;\sigma) \in [1-\tau, 1]$ and $\partial\vartheta_1/\partial\theta > 0$, $\forall\theta$. Consider $\sigma \in [0, 1-\tau]$. Then $g(1;\sigma) = 1 - \tau - \sigma \geq 0$. As $\partial\vartheta_1/\partial\theta > 0$, we have $g(\theta;\sigma) \geq 0$, $\forall\theta$ which implies $\bar{\theta}_w(\sigma) = 1$ in this parameter region. Consider $\sigma \in (1-\tau, 1]$ where $g(1;\sigma) < 0$ but $g(0;\sigma) = (1-\tau)(1-\sigma) \geq 0$. The intermediate value theorem implies existence of $\bar{\theta}_w(\sigma)$ such that $g(\bar{\theta}_w(\sigma);\sigma) = 0$ and it is easy to see that $\bar{\theta}_w(1) = 0$. Finally, for $\sigma > 1$, we have $g(0;\sigma) < 0$ and hence, $g(\theta;\sigma) < 0\ \forall\ \theta$ which implies that $\bar{\theta}_w(\sigma) = 0$. Taken together, this establishes the existence of $\bar{\theta}_w(\sigma)$.

□

3.B Empirical analysis

Figure 3.B.1: German tax schedules



Notes: The figure illustrates the personal income tax schedule in Germany for selected years.

Table 3.B.1: Descriptive statistics

	Mean	SD	N
(a) All years (2002-2018)			
Market income: Y	49168.67	90315.72	14394702
Deductions: D	7683.13	12972.97	14394702
Taxable income: Z	41485.54	84983.04	14394702
Tax payment: $T(Z)$	8753.28	36689.79	14394702
(b) Pre reform (2002-2012)			
Market income: Y	45854.74	88134.82	8828636
Deductions: D	6616.06	12698.82	8828636
Taxable income: Z	39238.68	83306.25	8828636
Tax payment: $T(Z)$	8110.09	36043.61	8828636
(c) Post reform (2013-2018)			
Market income: Y	54555.22	93502.15	5566066
Deductions: D	9417.57	13224.18	5566066
Taxable income: Z	45137.65	87517.12	5566066
Tax payment: $T(Z)$	9798.75	37693.08	5566066

Notes: The table shows descriptive statistics of selected variables, computed from German administrative tax records.

Table 3.B.2: CPI inflation and average household income growth

Year	Household income growth	CPI inflation
2019	2.9	1.4
2020	0.0	0.5
2021	5.6	3.1
2022	6.4	6.9
2023	6.5	5.9

Notes: The table shows the average household income growth rate between 2019 and 2023 that is used to extrapolate the administrative tax records. For comparison, the table also includes the CPI inflation rate.

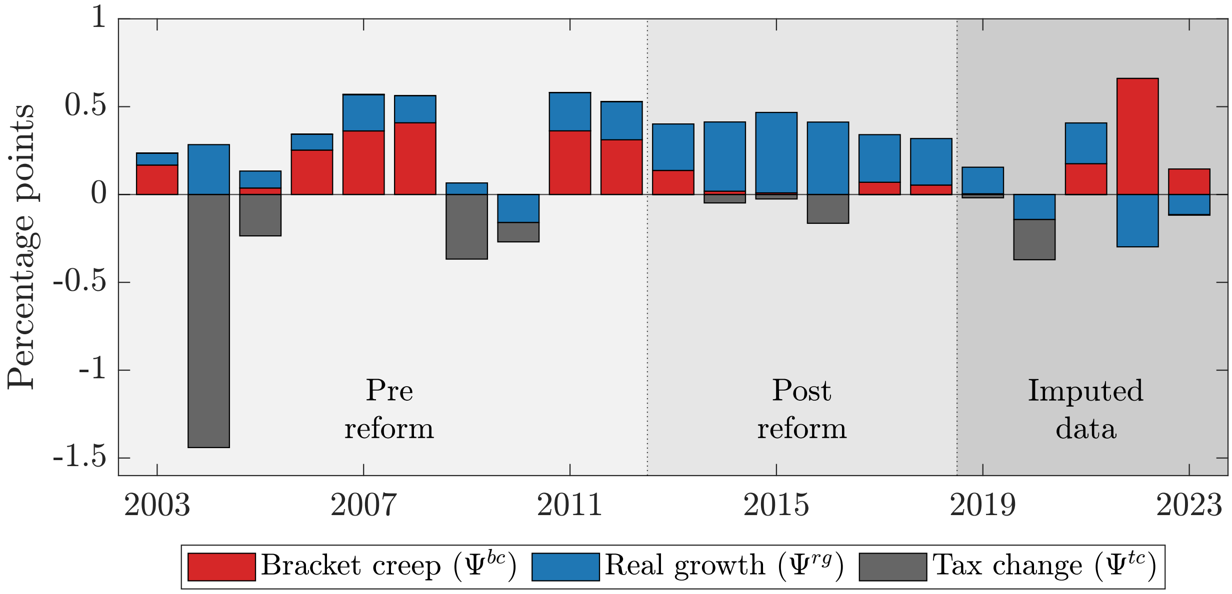
Table 3.B.3: Decomposition of average and marginal tax rates

	ATR			MTR			N
	Bracket creep Ψ^{bc}	Tax change Ψ^{tc}	Real growth Ψ^{rg}	Bracket creep Ψ^{bc}	Tax change Ψ^{tc}	Real growth Ψ^{rg}	
(a) All years (2002-2018)							
All taxpayers	0.11	-0.14	0.02	0.12	-0.17	0.04	9724120
	(0.0000)	(0.0001)	(0.0006)	(0.0000)	(0.0002)	(0.0006)	
Q1 of tax. inc.	0.12	-0.15	0.16	0.12	-0.17	0.19	2431131
	(0.0001)	(0.0002)	(0.0014)	(0.0001)	(0.0005)	(0.0013)	
Q2 of tax. inc.	0.12	-0.14	-0.02	0.13	-0.17	-0.03	2431069
	(0.0001)	(0.0002)	(0.0012)	(0.0001)	(0.0004)	(0.0012)	
Q3 of tax. inc.	0.10	-0.14	-0.08	0.13	-0.14	-0.04	2431055
	(0.0001)	(0.0002)	(0.0010)	(0.0001)	(0.0002)	(0.0009)	
Q4 of tax. inc.	0.08	-0.26	-0.30	0.02	-0.42	-0.22	2430865
	(0.0001)	(0.0004)	(0.0011)	(0.0001)	(0.0007)	(0.0009)	
(b) Pre reform (2002-2012)							
All taxpayers	0.17	-0.21	-0.08	0.16	-0.26	-0.05	5520740
	(0.0001)	(0.0002)	(0.0008)	(0.0001)	(0.0003)	(0.0008)	
Q1 of tax. inc.	0.18	-0.21	0.05	0.15	-0.25	0.10	1380266
	(0.0001)	(0.0004)	(0.0018)	(0.0001)	(0.0007)	(0.0017)	
Q2 of tax. inc.	0.18	-0.21	-0.12	0.17	-0.27	-0.14	1380237
	(0.0001)	(0.0004)	(0.0016)	(0.0002)	(0.0006)	(0.0016)	
Q3 of tax. inc.	0.15	-0.21	-0.16	0.15	-0.23	-0.11	1380212
	(0.0001)	(0.0004)	(0.0013)	(0.0001)	(0.0004)	(0.0013)	
Q4 of tax. inc.	0.11	-0.41	-0.35	0.03	-0.68	-0.32	1380025
	(0.0001)	(0.0006)	(0.0017)	(0.0001)	(0.0011)	(0.0014)	
(c) Post reform (2013-2018)							
All taxpayers	0.03	-0.05	0.14	0.08	-0.04	0.15	4203380
	(0.0000)	(0.0000)	(0.0009)	(0.0000)	(0.0002)	(0.0009)	
Q1 of tax. inc.	0.02	-0.06	0.31	0.08	-0.05	0.32	1050865
	(0.0000)	(0.0001)	(0.0020)	(0.0001)	(0.0006)	(0.0020)	
Q2 of tax. inc.	0.03	-0.05	0.11	0.09	-0.04	0.10	1050832
	(0.0000)	(0.0001)	(0.0017)	(0.0001)	(0.0005)	(0.0019)	
Q3 of tax. inc.	0.04	-0.03	0.02	0.09	-0.02	0.04	1050843
	(0.0000)	(0.0001)	(0.0014)	(0.0001)	(0.0000)	(0.0014)	
Q4 of tax. inc.	0.04	-0.01	-0.21	0.00	-0.00	-0.06	1050840
	(0.0000)	(0.0000)	(0.0015)	(0.0001)	(0.0000)	(0.0006)	

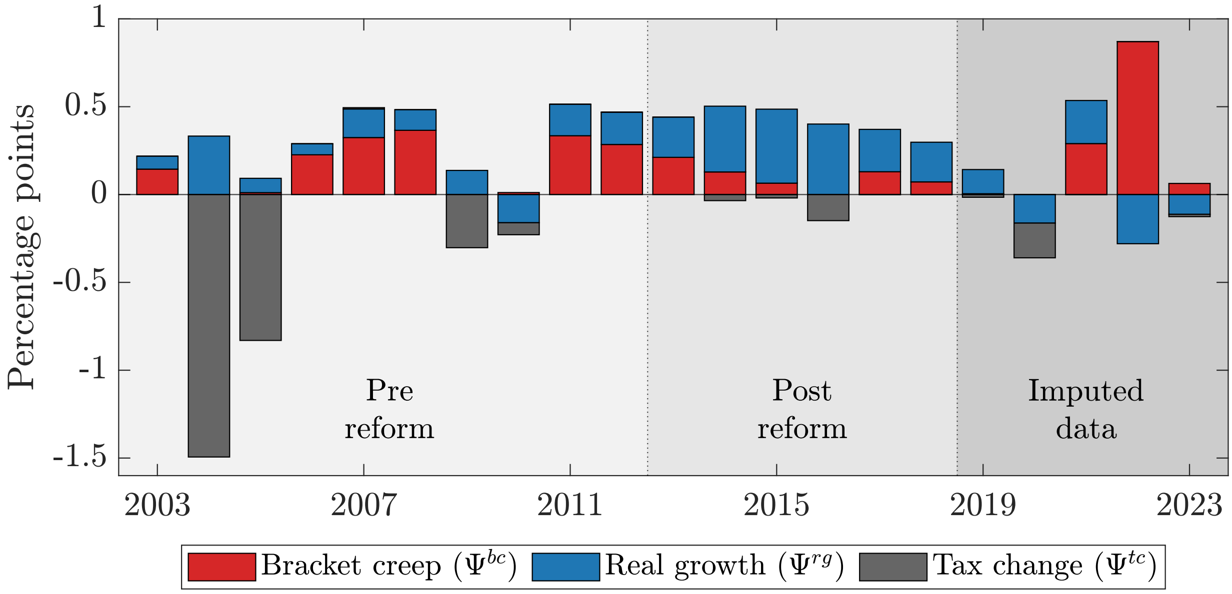
Notes: The table shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule.The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018. Standard errors are in parentheses.

Figure 3.B.2: Time variation in the tax rate decomposition

(a) Average tax rate

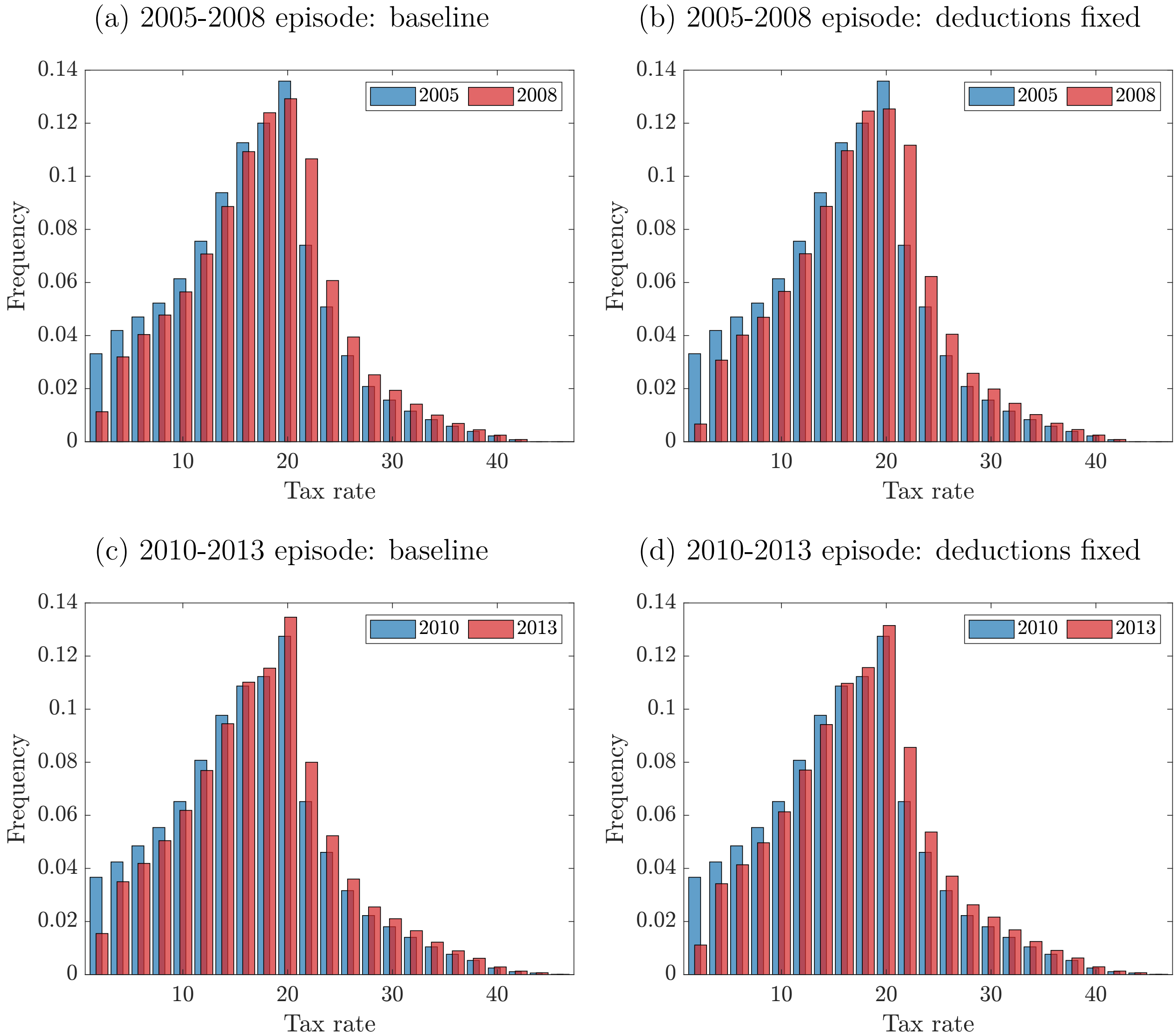


(b) Marginal tax rate



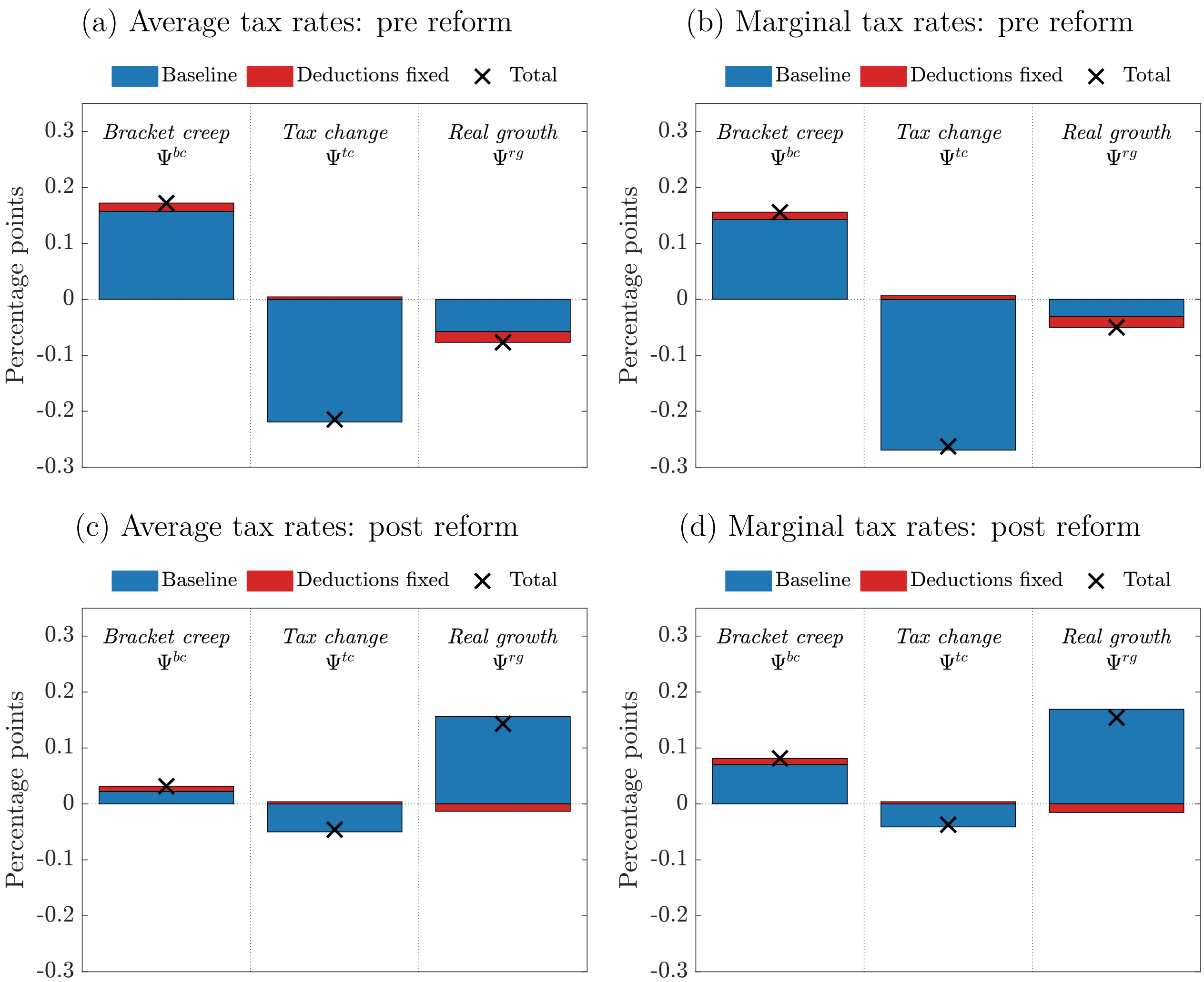
Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018; all non-zero estimates are significant at the 5% level, for standard errors, see Table 3.B.4. The imputed data is based on average household income growth as explained in Section 3.2.3.. Panel (a) and (b) show the results for the average and marginal tax rate, respectively.

Figure 3.B.3: Distributional changes of average tax rates during bracket creep episodes



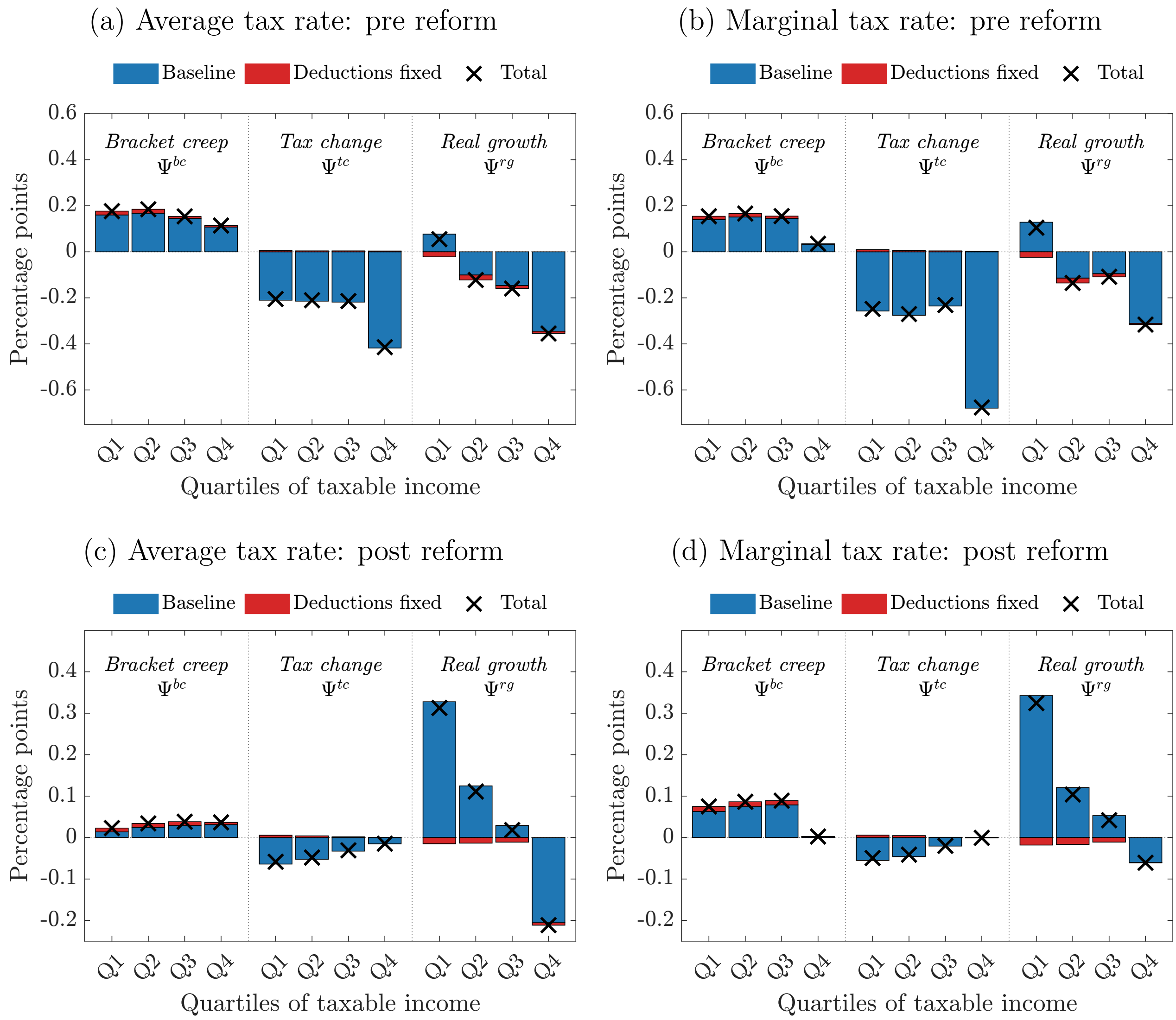
Notes: The figure shows the how the distribution of average tax rates shifts over time under constant real income during the bracket creep episodes from 2005-2008 and 2010-2013, respectively.

Figure 3.B.4: Decomposition of tax rates before and after the 2012 indexation reform



Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018; all estimates are significant at the 5% level, for standard errors, see Table 3.B.3. The top row presents the results for the years 2002 until 2012 and the bottom row for the years 2013 until 2018. For reference, the average annual inflation rate over the two sample periods was 1.61 and 1.13 percentage points, respectively.

Figure 3.B.5: Decomposition of tax rates before and after the 2012 indexation reform by quartiles of taxable income

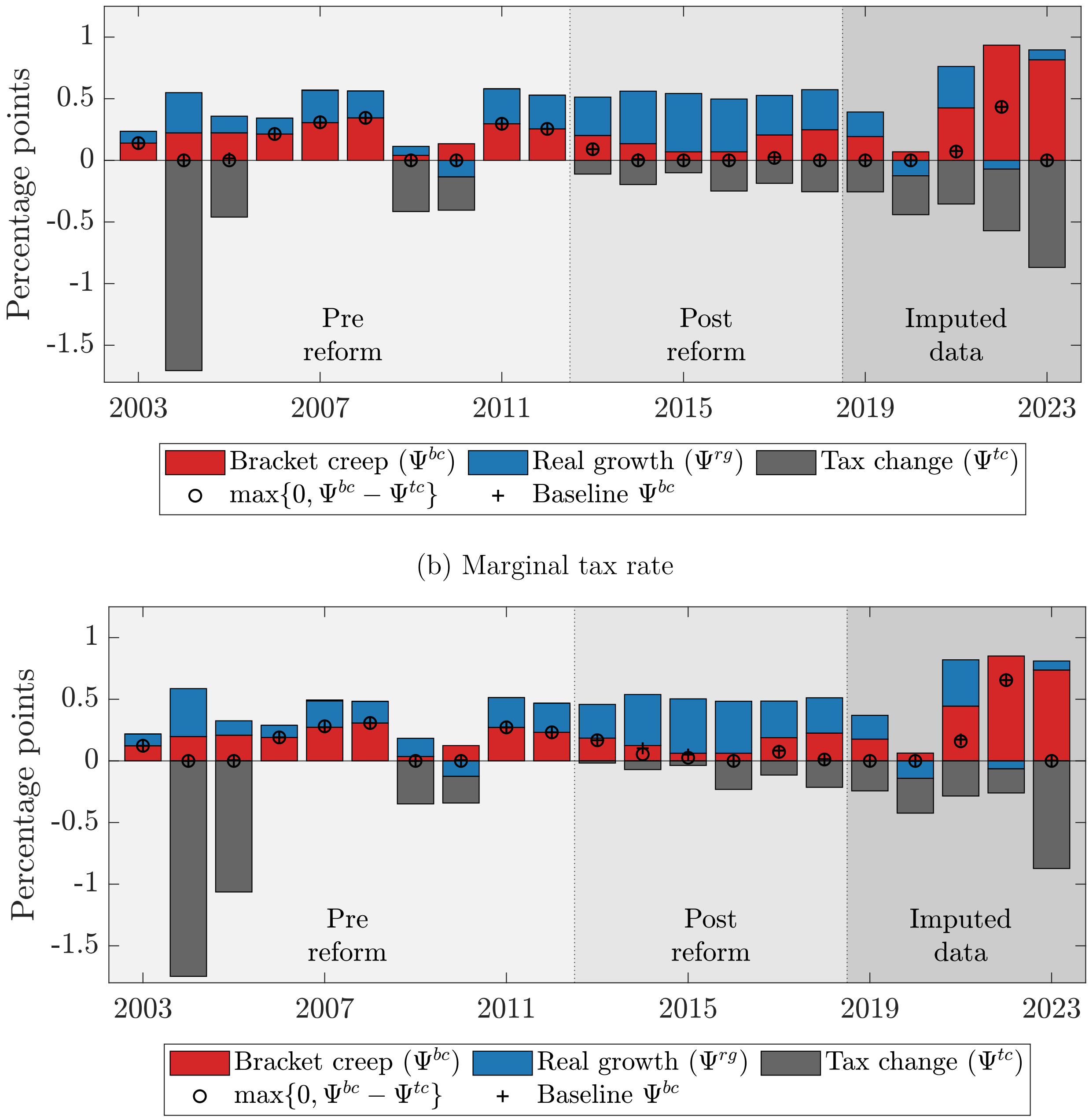


Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018; all estimates are significant at the 5% level, for standard errors, see Table 3.B.3. The top row presents the results for the years 2002 until 2012 and the bottom row for the years 2013 until 2018. For reference, the average annual inflation rate over the two sample periods was 1.61 and 1.13 percentage points, respectively.

Figure 3.B.6: Mechanical decomposition with $\alpha_t = 0$

(a) Average tax rate

(b) Marginal tax rate



Notes: The figure shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule. The results refer to the mechanical decomposition where $\alpha_t = 0$ is imposed. The circle markers indicate the differences between the bracket creep term and the tax change term, which is set to zero when the difference is negative. The plus markers indicate the value of the bracket creep term from the baseline decomposition. The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018. The imputed data is based on average household income growth as explained in Section 3.2.3.. Panel (a) and (b) show the results for the average and marginal tax rate, respectively.

Table 3.B.4: Decomposition of average and marginal tax rates by year

	ATR			MTR			
	Bracket creep	Tax change	Real growth	Bracket creep	Tax change	Real growth	N
Year	Ψ^{bc}	Ψ^{tc}	Ψ^{rg}	Ψ^{bc}	Ψ^{tc}	Ψ^{rg}	
2003	0.17	0.00	0.07	0.14	0.00	0.07	745290
	(0.0000)	(0.0000)	(0.0038)	(0.0001)	(0.0000)	(0.0033)	
2004	0.00	-1.44	0.28	0.00	-1.49	0.33	757885
	(0.0000)	(0.0003)	(0.0038)	(0.0000)	(0.0018)	(0.0038)	
2005	0.04	-0.24	0.10	0.01	-0.83	0.08	797060
	(0.0001)	(0.0004)	(0.0035)	(0.0001)	(0.0009)	(0.0031)	
2006	0.25	0.00	0.09	0.23	0.00	0.06	827996
	(0.0001)	(0.0000)	(0.0034)	(0.0001)	(0.0000)	(0.0030)	
2007	0.36	0.00	0.21	0.32	0.01	0.16	815173
	(0.0001)	(0.0001)	(0.0035)	(0.0002)	(0.0002)	(0.0031)	
2008	0.41	0.00	0.15	0.37	0.00	0.12	807104
	(0.0001)	(0.0000)	(0.0036)	(0.0002)	(0.0000)	(0.0031)	
2009	0.00	-0.37	0.07	0.00	-0.30	0.14	782609
	(0.0000)	(0.0001)	(0.0039)	(0.0000)	(0.0010)	(0.0037)	
2010	0.00	-0.11	-0.16	0.01	-0.07	-0.16	800770
	(0.0000)	(0.0000)	(0.0036)	(0.0000)	(0.0005)	(0.0034)	
2011	0.36	0.00	0.22	0.33	0.00	0.18	839959
	(0.0001)	(0.0000)	(0.0034)	(0.0002)	(0.0000)	(0.0031)	
2012	0.31	0.00	0.22	0.29	0.00	0.18	708472
	(0.0001)	(0.0000)	(0.0036)	(0.0002)	(0.0000)	(0.0033)	
2013	0.14	-0.00	0.26	0.21	0.00	0.23	911690
	(0.0000)	(0.0000)	(0.0032)	(0.0001)	(0.0000)	(0.0029)	
2014	0.02	-0.05	0.39	0.13	-0.03	0.37	932753
	(0.0000)	(0.0001)	(0.0032)	(0.0001)	(0.0007)	(0.0030)	
2015	0.01	-0.03	0.46	0.06	-0.02	0.42	950161
	(0.0000)	(0.0000)	(0.0032)	(0.0000)	(0.0005)	(0.0030)	
2016	0.00	-0.16	0.41	0.00	-0.15	0.40	974531
	(0.0000)	(0.0000)	(0.0031)	(0.0000)	(0.0007)	(0.0030)	
2017	0.07	-0.00	0.27	0.13	-0.00	0.24	981555
	(0.0000)	(0.0000)	(0.0031)	(0.0001)	(0.0001)	(0.0028)	
2018	0.05	-0.00	0.26	0.07	-0.00	0.23	976609
	(0.0000)	(0.0000)	(0.0031)	(0.0001)	(0.0000)	(0.0028)	
2019	0.00	-0.02	0.15	0.00	-0.02	0.14	1051350
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	
2020	0.00	-0.23	-0.14	0.00	-0.20	-0.16	1051350
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0006)	
2021	0.18	-0.00	0.23	0.29	-0.00	0.25	1051350
	(0.0001)	(0.0000)	(0.0001)	(0.0005)	(0.0000)	(0.0007)	
2022	0.66	0.00	-0.30	0.87	0.00	-0.28	1051350
	(0.0002)	(0.0000)	(0.0001)	(0.0005)	(0.0000)	(0.0002)	
2023	0.15	-0.00	-0.11	0.06	-0.01	-0.11	1051350
	(0.0001)	(0.0000)	(0.0001)	(0.0002)	(0.0000)	(0.0001)	

Notes: The table shows the decomposition of year-over-year changes in average and marginal tax rates, see equation (3.2). *Bracket creep* refers to the change in the tax rate that a taxpayer with constant real income experiences, absent discretionary tax reforms, whereas *Tax change* refers to changes due to discretionary tax reforms. *Real growth* refers to changes in tax rates due to real income growth under the contemporaneous tax schedule.The results are arithmetic averages based on 10 Mio. German administrative tax records between 2002 and 2018. Standard errors are in parentheses.

Bibliography

Abrams, Burton (2006). “How Richard Nixon Pressured Arthur Burns: Evidence from the Nixon Tapes.” In: *Journal of Economic Perspectives* 20.4, pp. 177–188.

Abrams, Burton A. and James L. Butkiewicz (2012). “The Political Business Cycle: New Evidence from the Nixon Tapes.” In: *Journal of Money, Credit and Banking* 44, pp. 385–399.

Adam, Klaus and Junyi Zhu (2016). “Price-level changes and the redistribution of nominal wealth across the euro area.” In: *Journal of the European Economic Association* 14.4, pp. 871–906.

Altig, David, Alan J Auerbach, Erin F Eidschun, Laurence Kotlikoff, and Victor Yifan Ye (2024). *Inflation’s Fiscal Impact on American Households.* NBER Working Paper 32482.

Anderson, Theodore W and Herman Rubin (1949). “Estimation of the parameters of a single equation in a complete system of stochastic equations.” In: *The Annals of Mathematical Statistics* 20.1, pp. 46–63.

Andrews, Isaiah (2018). “Valid Two-Step Identification-Robust Confidence Sets for GMM.” In: *Review of Economics and Statistics* 100.2, pp. 337–348.

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019). “Weak Instruments in Instrumental Variables Regression: Theory and Practice.” In: *Annual Review of Economics* 11.1, pp. 727–753.

Antolín-Díaz, Juan and Juan F. Rubio-Ramírez (2018). “Narrative Sign Restrictions for SVARs.” In: *American Economic Review* 108.10, pp. 2802–2829.

Arias, Jonas E, Dario Caldara, and Juan F Rubio-Ramirez (2019). “The systematic component of monetary policy in SVARs: An agnostic identification procedure.” In: *Journal of Monetary Economics* 101, pp. 1–13.

Aruoba, S Borağan and Thomas Drechsel (2024). *Identifying monetary policy shocks: A natural language approach*. NBER Working Paper 32417.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub (2021). “Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models.” In: *Econometrica* 89.5, pp. 2375–2408.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub (2023). “The intertemporal keynesian cross.” In: *Journal of Political Economy* (accepted).

Auerbach, Alan and Yuriy Gorodnichenko (2012). “Measuring the Output Responses to Fiscal Policy.” In: *American Economic Journal: Economic Policy* 4.2, pp. 1–27.

Baily, Martin Neil, Charles Hulten, and David Campbell (1992). “Productivity Dynamics in Manufacturing Plants.” In: *Brookings Papers on Economic Activity. Microeconomics 1992*, pp. 187–267.

Barakchian, S. Mahdi and Christopher Crowe (2013). “Monetary policy matters: Evidence from new shocks data.” In: *Journal of Monetary Economics* 60.8, pp. 950–966.

Barnichon, Regis and Geert Mesters (2020). “Identifying modern macro equations with old shocks.” In: *Quarterly Journal of Economics* 135.4, pp. 2255–2298.

Barnichon, Régis, Davide Debortoli, and Christian Matthes (2022). “Understanding the Size of the Government Spending Multiplier: It’s in the Sign.” In: *Review of Economic Studies* 89, pp. 87–117.

Barnichon, Régis and Geert Mesters (2023). “A Sufficient Statistics Approach for Macro Policy.” In: *American Economic Review* 113.11, pp. 2809–2845.

Barro, Robert J. and Charles J. Redlick (2011). “Macroeconomic Effects from Government Purchases and Taxes.” In: *Quarterly Journal of Economics* 126.1, pp. 51–102.

Barsky, Robert and Eric Sims (2011). “News shocks and business cycles.” In: *Journal of Monetary Economics* 58.3, pp. 273–289.

Barthélemy, Jean and Magali Marx (2017). “Solving endogenous regime switching models.” In: *Journal of Economic Dynamics and Control* 77, pp. 1–25.

Bauer, Michael D and Eric T Swanson (2023a). “A reassessment of monetary policy surprises and high-frequency identification.” In: *NBER Macroeconomics Annual*. Vol. 37. National Bureau of Economic Research, pp. 87–155.

Bauer, Michael D., Carolin Pflueger, and Adi Sunderam (2022). *Perceptions about Monetary Policy*. NBER Working Paper 30480.

Bauer, Michael D. and Eric T. Swanson (2023b). “An Alternative Explanation for the “Fed Information Effect”.” In: *American Economic Review* 113.3, pp. 664–700.

Belden, Susan (1989). “Policy Preferences of FOMC Members as Revealed by Dissenting Votes.” In: *Journal of Money, Credit and Banking* 21.4, pp. 432–441.

Ben Zeev, Nadav, Valerie A Ramey, and Sarah Zubairy (2023). “Do Government Spending Multipliers Depend on the Sign of the Shock?” In: *AEA Papers and Proceedings*. Vol. 113, pp. 382–387.

Benabou, Roland (2002). “Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency?” In: *Econometrica* 70.2, pp. 481–517.

Bernanke, Ben S and Alan S Blinder (1992). “The Federal Funds Rate and the Channels of Monetary Transmission.” In: *American Economic Review* 82.4, pp. 901–921.

Bernanke, Ben S, Mark Gertler, and Mark Watson (1997). “Systematic Monetary Policy and the Effects of Oil Price Shocks.” In: *Brookings Papers on Economic Activity* 1997.1, pp. 91–157.

Bianchi, Francesco (2012). “Evolving monetary/fiscal policy mix in the United States.” In: *American Economic Review* 102.3, pp. 167–172.

Bianchi, Francesco (2013). "Regime Switches, Agents' Beliefs, and post-World War II U.S. Macroeconomic Dynamics." In: *Review of Economic Studies* 80.2, pp. 463–490.

Bianchi, Francesco, Roberto Gómez-Cram, Thilo Kind, and Howard Kung (2023). "Threats to central bank independence: High-frequency identification with Twitter." In: *Journal of Monetary Economics* 135, pp. 37–54.

Bianchi, Francesco and Cosmin Ilut (2017). "Monetary/Fiscal policy mix and agents' beliefs." In: *Review of Economic Dynamics* 26, pp. 113–139.

Blanchard, Olivier and Roberto Perotti (2002). "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." In: *Quarterly Journal of Economics* 117.4, pp. 1329–1368.

Blinder, Alan S (2022). *A Monetary and Fiscal History of the United States, 1961-2021*. Princeton University Press.

— (1999). *Central Banking in Theory and Practice*. MIT Press.

— (2007). "Monetary Policy by Committee: Why and How?" In: *European Journal of Political Economy* 23.1, pp. 106–123.

Blömer, Maximilian, Florian Dorn, and Clemens Fuest (2023). "Kalte Progression in Zeiten hoher Inflation: Wer trägt die Lasten?" In: *ifo Schnelldienst* 2023.2, pp. 3–13.

Boivin, Jean (2006). "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data." In: *Journal of Money, Credit and Banking* 38.5, pp. 1149–1173.

Bordo, Michael D (2016). "Some historical reflections on the governance of The Federal Reserve." In: *Central Bank governance and oversight reform*, pp. 221–238.

Bordo, Michael D and Klodiana Istrefi (2023). "Perceived FOMC: The making of hawks, doves and swingers." In: *Journal of Monetary Economics* 136, pp. 125–143.

Bordo, Michael D. and Owen F. Humpage (2014). *Federal Reserve policy and Bretton Woods*. NBER Working Paper 20656.

Bound, John, David A. Jaeger, and Regina M. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." In: *Journal of the American Statistical Association* 90.430, pp. 443–450.

Broner, Fernando, Daragh Clancy, Aitor Erce, and Alberto Martin (2022). "Fiscal Multipliers and Foreign Holdings of Public Debt." In: *Review of Economic Studies* 89.3, pp. 1155–1204.

Bundesbank (2022). "Inflation-induced bracket creep in the income tax scale." In: *Monthly Report* June 2022.63.

Camous, Antoine and Dmitry Matveev (2021). "Furor over the Fed: A President's Tweets and Central Bank Independence." In: *CESifo Economic Studies* 67.1, pp. 106–127.

Carvalho, Carlos, Fernanda Nechio, and Tiago Tristão (2021). "Taylor rule estimation by OLS." In: *Journal of Monetary Economics* 124, pp. 140–154.

Chang, Yoosoon, Junior Maih, and Fei Tan (2021). "Origins of monetary policy shifts: a new approach to regime switching in DSGE models." In: *Journal of Economic Dynamics and Control* 133, p. 104235.

Chappell, Henry, Rob Roy McGregor, and Todd A. Vermilyea (2005). *Committee Decisions on Monetary Policy: Evidence from Historical Records of the Federal Open Market Committee*. MIT Press.

Chaudhuri, Saraswata and Eric Zivot (2011). "A new method of projection-based inference in GMM with weakly identified nuisance parameters." In: *Journal of Econometrics* 164.2, pp. 239–251.

Chernozhukov, Victor and Christian Hansen (2008). "The reduced form: A simple approach to inference with weak instruments." In: *Economics Letters* 100.1, pp. 68–71.

Choi, Sangyup, Tim Willems, and Seung Yong Yoo (2024). "Revisiting the monetary transmission mechanism through an industry-level differential approach." In: *Journal of Monetary Economics*, p. 103556.

Chow, Gregory C. and An-loh Lin (1971). "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series." In: *Review of Economics and Statistics* 53.4, pp. 372–375.

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo (2011). "When is the Government Spending Multiplier Large?" In: *Journal of Political Economy* 119.1, pp. 78–121.

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

— (2005). "Nominal rigidities and the dynamic effects of a shock to monetary policy." In: *Journal of Political Economy* 113.1, pp. 1–45.

Cieslak, Anna, Stephen Hansen, Michael McMahon, and Song Xiao (2023). *Policymakers' Uncertainty*. NBER Working Paper 31849.

Clarida, R., Jordi Gali, and Mark Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." In: *Quarterly Journal of Economics* 115.1, pp. 147–180.

Cloyne, James and Patrick Hürtgen (2016). "The macroeconomic effects of monetary policy: a new measure for the United Kingdom." In: *American Economic Journal: Macroeconomics* 8.4, pp. 75–102.

Cloyne, James, Patrick Hürtgen, and Alan M Taylor (2022). *Global monetary and financial spillovers: Evidence from a new measure of Bundesbank policy shocks*. NBER Working Paper 30485.

Cloyne, James, Joseba Martinez, Haroon Mumtaz, and Paolo Surico (2023). "Do Tax Increases Tame Inflation?" In: *AEA Papers and Proceedings*. Vol. 113, pp. 377–381.

Cloyne, James S, Òscar Jordà, and Alan M Taylor (2021). *Decomposing the Fiscal Multiplier*. NBER Working Paper 26939.

Cochrane, John H (2011). "Determinacy and Identification with Taylor Rules." In: *Journal of Political Economy* 119.3, pp. 565–615.

— (2022). “The fiscal roots of inflation.” In: *Review of Economic Dynamics* 45, pp. 22–40.

Cochrane, John H. and Monika Piazzesi (2002). “The Fed and Interest Rates - A High-Frequency Identification.” In: *American Economic Review* 92.2, pp. 90–95.

Coibion, Olivier (2012). “Are the effects of monetary policy shocks big or small?” In: *American Economic Journal: Macroeconomics* 4.2, pp. 1–32.

Coibion, Olivier and Yuriy Gorodnichenko (2011). “Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation.” In: *American Economic Review* 101.1, pp. 341–70.

Conesa, Juan Carlos and Dirk Krueger (2006). “On the optimal progressivity of the income tax code.” In: *Journal of Monetary Economics* 53.7, pp. 1425–1450.

Davig, Troy and Eric M. Leeper (2007). “Generalizing the Taylor Principle.” In: *American Economic Review* 97.3, pp. 607–635.

— (2008). “Endogenous Monetary Policy Regime Change.” In: *NBER International Seminar on Macroeconomics 2006*. NBER Chapters. National Bureau of Economic Research, pp. 345–391.

DeLong, B. J. (1997). “America’s Peacetime Inflation: The 1970s.” In: *Reducing Inflation: Motivation and Strategy*. Ed. by Christina Romer and David Romer. University of Chicago Press.

Doepke, Matthias and Martin Schneider (2006). “Inflation and the redistribution of nominal wealth.” In: *Journal of Political Economy* 114.6, pp. 1069–1097.

Drechsel, Thomas (2024). *Estimating the Effects of Political Pressure on the Fed: A Narrative Approach with New Data*. NBER Working Paper 32461.

Driscoll, John C and Aart C Kraay (1998). “Consistent covariance matrix estimation with spatially dependent panel data.” In: *Review of Economics and Statistics* 80.4, pp. 549–560.

Edge, Rochelle M and Jeremy B Rudd (2007). “Taxation and the Taylor principle.” In: *Journal of Monetary Economics* 54.8, pp. 2554–2567.

Ehrmann, Michael, Robin Tietz, and Bauke Visser (2022). *Voting Right Rotation, Behavior of Committee Members and Financial Market Reactions: Evidence from the U.S. Federal Open Market Committee*. IMF Working Paper 2022/105.

Erosa, Andrés and Gustavo Ventura (2002). "On inflation as a regressive consumption tax." In: *Journal of Monetary Economics* 49.4, pp. 761–795.

Farhi, Emmanuel and Ivan Werning (2016). "Fiscal Multipliers: Liquidity Traps and Currency Unions." In: *Handbook of Macroeconomics* 2, pp. 2417–2492.

Feldstein, Martin (1983). "The fiscal framework of monetary policy." In: *Economic Inquiry* 21.1, pp. 11–23.

Ferguson, Niall, Martin Kornejew, Paul Schmelzing, and Moritz Schularick (2023). *The Safety Net: Central Bank Balance Sheets and Financial Crisis, 1987-2020*. Hoover Institution Economics Working Paper 23102.

Ferriere, Axelle and Gaston Navarro (2024). "The heterogeneous effects of government spending: It's all about taxes." In: *Review of Economic Studies*, rdae032.

Gali, Jordi (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton University Press.

Galí, Jordi, J.David López-Salido, and Javier Vallés (2003). "Technology shocks and monetary policy: assessing the Fed's performance." In: *Journal of Monetary Economics* 50.4, pp. 723–743.

Gavin, William T, Benjamin D Keen, and Finn E Kydland (2015). "Monetary policy, the tax code, and the real effects of energy shocks." In: *Review of Economic Dynamics* 18.3, pp. 694–707.

Gavin, William T, Finn E Kydland, and Michael R Pakko (2007). "Monetary policy, taxes, and the business cycle." In: *Journal of Monetary Economics* 54.6, pp. 1587–1611.

Gertler, Mark and Peter Karadi (2015). "Monetary policy surprises, credit costs, and economic activity." In: *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.

Ghassibe, Mishel and Francesco Zanetti (2022). "State dependence of fiscal multipliers: the source of fluctuations matters." In: *Journal of Monetary Economics* 132, pp. 1–23.

Gilchrist, Simon and Egon Zakrajšek (2012). "Credit spreads and business cycle fluctuations." In: *American Economic Review* 102.4, pp. 1692–1720.

Goodfriend, Marvin and Robert G. King (2005). "The incredible Volcker disinflation." In: *Journal of Monetary Economics* 52.5, pp. 981–1015.

Gordon, Robert J. and Robert Krenn (2010). *The End of the Great Depression: VAR Insight on the Roles of Monetary and Fiscal Policy.* NBER Working Paper 16380.

Hazell, Jonathon, Juan Herreno, Emi Nakamura, and Jón Steinsson (2022). "The slope of the Phillips Curve: evidence from US states." In: *Quarterly Journal of Economics* 137.3, pp. 1299–1344.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante (2017). "Optimal Tax Progressivity: An Analytical Framework." In: *Quarterly Journal of Economics* 132.4, pp. 1693–1754.

— (2020). "Presidential address 2019: How should tax progressivity respond to rising income inequality?" In: *Journal of the European Economic Association* 18.6, pp. 2715–2754.

Heer, Burkhard and Bernd Süssmuth (2013). "Tax bracket creep and its effects on income distribution." In: *Journal of Macroeconomics* 38, pp. 393–408.

Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek (2021). "The Transmission of Monetary Policy under the Microscope." In: *Journal of Political Economy* 129.10, pp. 2861–2904.

Ilzetzki, Ethan, Enrique G Mendoza, and Carlos A Végh (2013). "How big (small?) are fiscal multipliers?" In: *Journal of Monetary Economics* 60.2, pp. 239–254.

Immervoll, Herwig (2005). "Falling up the stairs: the effects of "bracket creep" on household incomes." In: *Review of Income and Wealth* 51.1, pp. 37–62.

Istrefi, Klodiana (2019). *In Fed Watchers' Eyes: Hawks, Doves and Monetary Policy.* Banque de France Working Paper 725.

Jarociński, Marek and Peter Karadi (2020). "Deconstructing monetary policy surprises—the role of information shocks." In: *American Economic Journal: Macroeconomics* 12.2, pp. 1–43.

Jordà, Òscar and Alan M. Taylor (2016). "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy." In: *Economic Journal* 126.590, pp. 219–255.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante (2018). "Monetary Policy According to HANK." In: *American Economic Review* 108.3, pp. 697–743.

Keinsley, Andrew (2016). "Indexing the income tax code, monetary/fiscal interaction, and the great moderation." In: *European Economic Review* 89, pp. 1–20.

Kilian, Lutz and Logan T Lewis (2011). "Does the Fed Respond to Oil Price Shocks?" In: *Economic Journal* 121.555, pp. 1047–1072.

Lazarus, Eben, Daniel J Lewis, James H Stock, and Mark W Watson (2018). "HAR Inference: Recommendations for Practice." In: *Journal of Business & Economic Statistics* 36.4, pp. 541–559.

Leeper, Eric M (1991). "Equilibria under 'active'and 'passive'monetary and fiscal policies." In: *Journal of Monetary Economics* 27.1, pp. 129–147.

Leeper, Eric M, Nora Traum, and Todd B Walker (2017). "Clearing up the Fiscal Multiplier Morass." In: *American Economic Review* 107.8, pp. 2409–54.

Leeper, Eric M and Tao Zha (2003). "Modest policy interventions." In: *Journal of Monetary Economics* 50.8, pp. 1673–1700.

Lewis, Daniel J and Karel Mertens (2022). *A Robust Test for Weak Instruments with Multiple Endogenous Regressors*. Federal Reserve Bank of New York Staff Reports 1020.

Liu, Yan and Jing Cynthia Wu (2021). "Reconstructing the yield curve." In: *Journal of Financial Economics* 142.3, pp. 1395–1425.

Lucas, Robert E (1976). "Econometric policy evaluation: A critique." In: *Carnegie-Rochester Conference Series on Public Policy* 1, pp. 19–46.

Mattesini, Fabrizio and Lorenza Rossi (2012). "Monetary policy and automatic stabilizers: the role of progressive taxation." In: *Journal of Money, Credit and Banking* 44.5, pp. 825–862.

McKay, Alisdair and Ricardo Reis (2021). "Optimal automatic stabilizers." In: *Review of Economic Studies* 88.5, pp. 2375–2406.

McKay, Alisdair and Christian K Wolf (2023). "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" In: *Econometrica* 91.5, pp. 1695–1725.

McMahon, Michael and Tim Munday (2023). "Element(s) of Surprise." In: *Essays on Monetary Policy*, pp. 248–340.

Meltzer, Allan H (2005). "Origins of the great inflation." In: *Federal Reserve Bank of St. Louis Review* 87.2, pp. 145–75.

Mertens, Karel and Morten O Ravn (2013). "The dynamic effects of personal and corporate income tax changes in the United States." In: *American Economic Review* 103.4, pp. 1212–47.

Miranda-Agrippino, Silvia and Giovanni Ricco (2021). "The transmission of monetary policy shocks." In: *American Economic Journal: Macroeconomics* 13.3, pp. 74–107.

Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitriy Sergeyev (2018). "Government Spending Multipliers Under the Zero Lower Bound: Evidence from Japan." In: *American Economic Journal: Macroeconomics* 10.3, pp. 247–77.

Montiel Olea, José Luis and Carolin Pflueger (2013). "A Robust Test for Weak Instruments." In: *Journal of Business & Economic Statistics* 31.3, pp. 358–369.

Mountford, Andrew and Harald Uhlig (2009). "What are the effects of fiscal policy shocks?" In: *Journal of Applied Econometrics* 24.6, pp. 960–992.

Nakamura, Emi and Jón Steinsson (2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect." In: *Quarterly Journal of Economics* 133.3, pp. 1283–1330.

Nakamura, Emi and Jon Steinsson (2014). "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." In: *American Economic Review* 104.3, pp. 753–92.

Orphanides, A. (2003). "The quest for prosperity without inflation." In: *Journal of Monetary Economics* 50, pp. 633–663.

Orphanides, Athanasios (2004). "Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches." In: *Journal of Money, Credit and Banking* 36.2, pp. 151–175.

Owyang, Michael and Garey Ramey (2004). "Regime switching and monetary policy measurement." In: *Journal of Monetary Economics* 51.8, pp. 1577–1597.

Pallotti, Filippo (2022). *Winners and losers from unexpected inflation.* SSRN 4124693.

Pallotti, Filippo, Gonzalo Paz-Pardo, Jiri Slacalek, Oreste Tristani, and Giovanni L Violante (2023). *Who Bears the Costs of Inflation? Euro Area Households and the 2021–2022 Shock.* NBER Working Paper 31896.

Paulus, Alari, Holly Sutherland, and Iva Tasseva (2020). "Indexing Out of Poverty? Fiscal Drag and Benefit Erosion in Cross-National Perspective." In: *Review of Income and Wealth* 66.2, pp. 311–333.

Primiceri, Giorgio (2006). "Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy." In: *Quarterly Journal of Economics* 121, pp. 867–901.

Primiceri, Giorgio E (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy." In: *Review of Economic Studies* 72.3, pp. 821–852.

Ramey, Valerie A (2011). "Identifying Government Spending Shocks: It's all in the Timing." In: *Quarterly Journal of Economics* 126.1, pp. 1–50.

— (2016). "Macroeconomic Shocks and Their Propagation." In: *Handbook of Macroeconomics* 2, pp. 71–162.

Ramey, Valerie A and Sarah Zubairy (2018). "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." In: *Journal of Political Economy* 126.2, pp. 850–901.

Ramey, Valerie A. and Matthew D. Shapiro (1998). "Costly capital reallocation and the effects of government spending." In: *Carnegie-Rochester Conference Series on Public Policy* 48, pp. 145–194.

Riboni, Alessandro and Francisco Ruge-Murcia (2023). "The power of the federal reserve chair." In: *International Economic Review* 64.2, pp. 727–756.

Riboni, Alessandro and Francisco J. Ruge-Murcia (2010). "Monetary policy by Committee: Consensus, Chairman dominance, or simple majority?" In: *Quarterly Journal of Economics* 125.1, pp. 363–416.

Romer, Christina and David Romer (1989). "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." In: *NBER Macroeconomics Annual*. Vol. 4. National Bureau of Economic Research, pp. 121–184.

Romer, Christina D and David H Romer (2010). "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." In: *American Economic Review* 100.3, pp. 763–801.

— (2002). *The Evolution of Economic Understanding and Postwar Stabilization Policy.* NBER Working Paper 9274.

— (2004). "A New Measure of Monetary Shocks: Derivation and Implications." In: *American Economic Review* 94.4, pp. 1055–1084.

Sargent, Thomas J. (1979). "Estimating Vector Autoregressions Using Methods not Based on Explicit Economic Theories." In: *Quarterly Review* 3.Sum.

Shapiro, Matthew and Mark Watson (1988). "Sources of Business Cycle Fluctuations." In: *NBER Macroeconomics Annual*. Vol. 3. National Bureau of Economic Research, pp. 111–156.

Sims, Christopher A and Tao Zha (2006). "Were There Regime Switches in US Monetary Policy?" In: *American Economic Review* 96.1, pp. 54–81.

Süssmuth, Bernd and Matthias Wieschemeyer (2022). "Taxation and the distributional impact of inflation: The US post-war experience." In: *Economic Modelling* 111, p. 105813.

Uhlig, Harald (2005). “What are the effects of monetary policy on output? Results from an agnostic identification procedure.” In: *Journal of Monetary Economics* 52.2, pp. 381–419.

Wieland, Johannes F and Mu-Jeung Yang (2020). “Financial dampening.” In: *Journal of Money, Credit and Banking* 52.1, pp. 79–113.

Wolf, Christian K (2023). “Fiscal stimulus and the systematic response of monetary policy.” In: *AEA Papers and Proceedings*. Vol. 113, pp. 388–393.

Woodford, Michael (2011). “Simple Analytics of the Government Expenditure Multiplier.” In: *American Economic Journal: Macroeconomics* 3.1, pp. 1–35.

Zhu, Junyi (2014). “Bracket Creep Revisited-with and without r> g: Evidence from Germany.” In: *Journal of Income Distribution* 23.3, pp. 106–158.

Curriculum Vitae

2019–2024	University of Mannheim (Germany)
	Ph.D. in Economics
2018–2019	University of Vienna (Austria)
	M.Sc. in Economics
2016–2018	University of Vienna (Austria)
	B.Sc. in Economics
2013–2016	Baden-Wuerttemberg Cooperative State University (Germany)
	B.A. in Business Administration