Subjective Models of the Macroeconomy: Evidence From Experts and a Representative Sample*

Peter Andre† Carlo Pizzinelli‡

Christopher Roth§ Johannes Wohlfart‖

October 22, 2019

Abstract

Using a sample of 2,200 households representative of the US population and a sample of more than 1,000 experts, we measure beliefs about how aggregate unemployment and inflation respond to different macroeconomic shocks. Expert predictions are quantitatively close to standard DSGE models and VAR evidence. While households' beliefs are directionally aligned with those of experts in the case of oil supply shocks and government spending shocks, they predict an opposite reaction of inflation to monetary policy and income tax shocks. A substantial fraction of deviations of household predictions can be explained by the use of a simple affective heuristic.

JEL Classification: D83, D84, E31, E52, E71

Keywords: Expectation Formation, Subjective Models, Heuristics, Macroeconomic Shocks, Monetary Policy, Fiscal Policy.

---

*We thank Rudi Bachmann, Carola Binder, Benjamin Born, Olivier Coibion, Francesco D'Acunto, Stefano DellaVigna, Andreas Fuster, Dimitris Georgarakos, Yuriy Gorodnichenko, Thomas Graeber, Michael Haliassos, Lukas Hensel, Chi Hyun Kim, Michael Kosfeld, Fabian Krüger, Matt Lowe, Michael McMahon, Valerie Ramey, Sonja Settele, Andrei Shleifer, Uwe Sunde, Johannes Stroebel, Giorgio Topa, Michael Weber, Mirko Wiederholt, Basit Zafar, the many experts who participated in our survey as well as seminar and conference participants at Bonn, Leiden, the ifo Insitute (Munich), Frankfurt, the Bank of Finland and CEPR Joint Conference on Monetary Economics and Reality, the International Monetary Fund, the Deutsche Bundesbank and Banque de France Joint Conference on Household Expectations, the Oxford Monetary Economics conference, the Berlin Fiscal Policy Conference, the Theories and Methods in Macroeconomics conference (Nuremberg), the Workshop on Expectations and Decisions of Households (Berlin), the IAAE Conference (Nicosia) and the SEM Conference (Frankfurt) for helpful comments. We thank Dorine Boumans, Johanna Garnitz, Andreas Peichl, and the ifo Institute for including our module in the World Economic Survey. We thank the Joachim Herz Foundation for financial support. The activities of the Center for Economic Behavior and Inequality (CEBI) are funded by the Danish National Research Foundation. Anna Lane and David Zeimentz provided excellent research assistance. We received ethics approval from Goethe University Frankfurt. The experimental instructions can be found at the following link: `https://www.dropbox.com/s/tovmpbqhp0xm9pq/instructions_subjectivemodels.pdf?dl=0`. **Disclaimer**: The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. The AEA RCT identification number is AEARCTR-0004851.

†Peter Andre, University of Bonn, e-mail: p.andre@uni-bonn.de

‡Carlo Pizzinelli, IMF, e-mail: cpizzinelli@imf.org

§Christopher Roth, Department of Economics, University of Warwick, briq, CESifo, CAGE, email: Christopher.Roth@warwick.ac.uk

‖Johannes Wohlfart, Department of Economics, University of Copenhagen, CEBI, CESifo, e-mail: johannes.wohlfart@econ.ku.dk

1 Introduction

A core assumption underlying rational expectations macroeconomic models is that economic agents form expectations about aggregate outcomes that are consistent with the underlying model. For instance, standard New Keynesian models predict increases in both unemployment and inflation in response to a negative supply-side shock, but an increase in unemployment and a decrease in inflation in response to a negative demand-side shock. Households are assumed to recognize these relationships and form their expectations accordingly. In this paper, we study empirically people's beliefs about the response of inflation and unemployment to different types of macroeconomic shocks.

Characterizing households' subjective models in the context of macroeconomic shocks not only sheds light on a crucial assumption underlying macroeconomic models, but is also important to better understand how households form expectations about unemployment and inflation. Recent evidence indicates that these expectations matter for households' consumption and financial decisions, and thereby shape aggregate outcomes (Armona et al., 2018; Bailey et al., 2019, 2018; Coibion et al., 2019a,c; D'Acunto et al., 2019a; Kuchler and Zafar, 2019). Understanding households' beliefs about macroeconomic shocks is also essential to assess the effects of government and central bank policies, as the impact of such policies depends on households' forward-looking expectations of aggregate dynamics. Indeed, some of these policies, such as monetary forward guidance, aim to influence people's behavior primarily by shifting macroeconomic expectations (Coibion et al., 2018b).

Measuring households' beliefs about changes in inflation and unemployment in response to macroeconomic shocks poses several challenges. First, it is difficult to find clean exogenous variation in beliefs about the likelihood of macroeconomic shocks in the real world. Second, people's beliefs should be measured in a way that enables a comparison with benchmarks from the theoretical and empirical literature. Third, beliefs should be measured using survey questions that are understandable to individuals without an economics education.

We propose an approach to measure beliefs about the effects of different macroeconomic shocks, which we apply to a representative sample of 2,200 individuals of the US population. We use vignettes in which we ask our respondents to predict future unemployment and inflation under different hypothetical macroeconomic shocks. We focus on

four different exogenous shocks that are among the most commonly studied in macroeconomics: an oil supply shock, a monetary policy shock, a government spending shock, and an income tax shock. Our approach allows us to fix the respondents' beliefs about the sources of the shocks and to ensure that the shocks are perceived as exogenous. This enables us to compare our respondents' beliefs to benchmarks from theoretical models and empirical evidence. Moreover, we conduct the same survey with a sample of more than 1,000 experts from academia, policy institutions and the private sector. This provides us with another benchmark that does not rely on any particular model or piece of empirical evidence. In addition, the expert survey allows us to compare economists' beliefs about the functioning of the US economy with the predictions of standard models and empirical evidence.

For each vignette, we elicit the respondents' expectations about the unemployment rate and the inflation rate twice: First, under a hypothetical scenario in which the shock variable of interest, such as the oil price, will not change over the next 12 months; second, under a "rise-scenario" or a "fall-scenario" to which the respondents are randomly allocated. In the "rise-scenario", the shock variable of interest increases relative to the baseline scenario. For example, in the oil price vignette, we tell our respondents that the oil price will be on average $30 higher over the next 12 months or, in the income tax vignette, that tax rates increase by 1 percentage point. In the "fall-scenario", the shock variable of interest decreases relative to the baseline scenario. By taking the difference in the predictions of unemployment and inflation between the rise/fall and the baseline scenario we measure each respondent's beliefs about effects of the shock, while taking out constant differences in expectations across individuals.

We first present the predictions from the experts. Experts think that both inflation and unemployment positively respond to an increase in oil prices. Experts forecast that unemployment responds positively to increases in taxes and to decreases in government spending, while inflation responds negatively. Finally, experts think that inflation drops in response to interest rate hikes, while unemployment increases. The experts' forecasts are all precisely estimated and largely consistent with predictions from DSGE models and empirical evidence from VARs. Moreover, there is relatively little disagreement among the experts.

We next compare the expert predictions to those of respondents from the representa-

tive online panel. The respondents in this sample hold very similar beliefs as the experts in the oil vignette, both for unemployment and inflation. While households think that inflation positively responds to changes in government spending, they on average do not think that unemployment significantly responds to changes in government spending. The most striking deviation from the experts' forecasts is that households think that increases in the federal funds target rate would increase inflation, as would a rise in income tax rates. Across the vignettes, households' predictions about the unemployment rate are largely aligned with those of experts, while their inflation forecasts are less in line with those of experts. These findings suggest that the assumption that households form expectations in line with equilibrium relationships of standard macroeconomic models may serve as a reasonable approximation in some contexts, but not always. Finally, there is a large degree of heterogeneity in household beliefs, consistent with models of learning in which households may disagree about the structural parameters of the economy.

What is driving the deviations of household predictions from the benchmarks? We include a series of tailored questions that allow us to study the mechanisms behind our findings. We begin by measuring our respondents' beliefs about the propagation mechanisms of macroeconomic shocks. Holding more accurate beliefs about propagation channels positively correlates with making predictions that are qualitatively aligned with those of experts, but the effects are relatively moderate. Moreover, respondents' evaluation of how relevant aggregate unemployment or inflation are to their own economic situation is uncorrelated with how close their predictions are to the benchmarks. Furthermore, deviations of households' predictions from those of experts about the inflation response to interest rate and tax shocks are not purely driven by households who believe in greater importance of a cost channel, suggesting that our findings do not operate through households holding a supply-side view of the economy. Among respondents' demographic characteristics, higher education and age are associated with predictions that are more in line with those of experts, providing some support for learning over the life-cycle and cognitive constraints (D'Acunto et al., 2019b,c,d).

Research from psychology and economics emphasizes that individuals revert to simple heuristics in complex decision environments that are plagued by a large degree of uncertainty (Gigerenzer and Todd, 1999; Lacetera et al., 2012). Building on these insights, we propose a behavioral channel based on respondents' affective representation of macroe-

conomic variables, which we term the "good-bad heuristic" (GBH). The GBH posits that individuals perceiving two variables as both good (or both bad) predict that these variables move in the same direction, while, if they perceive one variable as good and the other as bad, they predict movement in opposing directions. For each variable of interest, we measure whether respondents consider higher values of the variable as good or bad for their own household and for the US economy. We find that a substantial fraction of deviations from benchmarks can be explained by the GBH, consistent with an important role for sentiment in expectation formation (Kamdar, 2018). These findings call for more theoretical and empirical work featuring a role for economic agents' affective judgment in shaping their macroeconomic expectations.

Finally, we illustrate the potential implications of our findings in a simple quantitative exercise. We show that households' partial-equilibrium responses to monetary policy shocks are substantially altered if their inflation and unemployment expectations are in line with our evidence. This indicates that heterogeneity in beliefs can have important implications for the transmission mechanism of policies. Moreover, our findings highlight the importance of adequate communication of specific policy measures to the public (Blinder et al., 2008; Coibion et al., 2019b; Haldane and McMahon, 2018).

We build on a growing literature studying the formation of macroeconomic expectations of experts, households and firms (Acosta and Afrouzi, 2019; Afrouzi, 2019; Armantier et al., 2016, 2015; Bachmann et al., 2019, 2015; Binder and Rodrigue, 2018; Binder and Makridis, 2018; Bordalo et al., 2018, 2019; Cavallo et al., 2017; Coibion and Gorodnichenko, 2012, 2015a,b; Fuster et al., 2012, 2010, 2019; Goldfayn-Frank and Wohlfart, 2019; Malmendier and Nagel, 2011, 2016). We contribute to this research effort by providing the first direct evidence on households' beliefs about the effects of macroeconomic shocks.

Our paper is related to work studying beliefs about the relationship between different macroeconomic variables. Carvalho and Nechio (2014), Dräger et al. (2016) and Kuchler and Zafar (2019) use observational data to examine how households' beliefs about unemployment, inflation and interest rates are correlated. A series of experiments have studied cross-learning, i.e. how respondents update their expectations about unemployment in response to information about inflation (Coibion et al., 2019a, 2018a, 2019c), and how they update their expectations about inflation in response to learning about the likelihood

of a recession (Roth and Wohlfart, 2019). While the exogenous variation in information in these experiments allows for causal identification, the interpretation of cross-learning is difficult as beliefs about the source of changes in inflation and GDP growth are unrestricted. Our approach directly measures households' beliefs about the effects of different exogenous macroeconomic shocks, and therefore allows us to characterize how well beliefs about the response of inflation and unemployment align with theoretical and empirical benchmarks.[1]

We methodologically contribute to the literature on macroeconomic expectation formation by proposing the use of hypothetical vignettes to characterize people's subjective beliefs about the responses of unemployment and inflation to different macroeconomic shocks. A series of recent papers uses hypothetical vignettes to study belief formation and behavior in contexts that are difficult to study in a real-world setting, such as in the area of education and human capital (Delavande and Zafar, 2018; Wiswall and Zafar, 2017), and recently also for the study of the consumption response to changes in current or expected future economic resources (Christelis et al., 2017; Fuster et al., 2018). We contribute to this literature by applying this approach to study household beliefs about changes in unemployment and inflation in response to commonly modelled shocks. We also test for the role of effort by providing a random subset of respondents with prediction incentives, leveraging expert responses as an external benchmark. We find that incentives only moderately affect the predictions. Finally, our evidence on the predictions from experts and the representative sample contributes to a literature studying differences in forecasts between experts and the general population (DellaVigna and Pope, 2018a,b).

The rest of this paper is structured as follows. Section 2 provides an overview of the samples of households and experts and the survey design. Section 3 presents our main evidence on experts' and households' predictions in the different vignettes and compares them to benchmarks from standard DSGE models and VARs. Section 4 discusses a series of mechanisms that could be driving heterogeneity in household predictions. Section 5 provides evidence on the robustness of our results. Section 6 discusses the implications of our findings for the transmission of shocks, macroeconomic modeling and policymaking. Section 7 concludes.

[1]This also relates to a small literature providing causal evidence on misspecified mental models in tightly controlled abstract prediction tasks, and on sources of model misspecification (Graeber, 2019).

2 Data and Design

2.1 Samples

Representative samples We collect a sample of 2,200 respondents that is representative of the US population in terms of education, gender, age, region, and total household income in two waves. We conducted the first wave of the online survey in February and March 2019 in collaboration with the market research company Research Now SSI, which is commonly used in economics. In July 2019, we collected the second wave in collaboration with a widely used online panel provider, Lucid.

Table A2 provides summary statistics for the pooled sample and compares it to the general population using data from the 2017 American Community Survey (ACS). Our sample matches the distributions of education, gender, age, region and total household income very closely. 55 percent of our respondents are female, compared to 51 percent in the ACS. The average age in our sample is 46.4, while it is 47.4 in the ACS. 32 percent of the respondents in our sample have at least a bachelor's degree compared to 30 percent in the ACS. The median income in our sample is \$62,500 compared to \$65,700 in the ACS.

Expert samples We recruited a sample of approximately 1,100 experts. In the first wave, which took place in February and March 2019, we invited economists who were authors or discussants in at least one of a series of leading conferences on macroeconomics.[2] In addition, we invited Ph.D. students, experts from several policy institutions, as well as several experts working in the broader areas of expectation formation and macroeconomic forecasting. In total, 179 experts completed the first wave of the survey.[3] We also included our own module in the World Economic Survey (WES) conducted by the ifo Institute (experts Wave 2). Their sample comprises about 1,800 economic experts from around the world who make forecasts about the economic prospects of their country on a quarterly basis (Boumans and Garnitz, 2017). The relevant survey round was conducted in July 2019 and 908 experts completed at least one hypothetical vignette.[4]

As shown in Table A3, 16% percent of the experts that participated in Wave 1 are from policy institutions, such as the IMF and central banks, and 83 percent of the experts

[2]For details on the conferences considered, see Appendix E.

[3]Demographic data on the 179 experts were collected directly from their CVs. For respondents from PhD programs and policy institutions we asked a few questions on demographics at the end of the survey.

[4]743 out of the 908 experts responded to all questions in the survey.

are from academic institutions. Moreover, only 26 percent of the experts are female. 21 percent are Full Professor, while 18 percent are PhD students.

In Wave 2 of our expert survey, 16 percent of experts are from policy institutions, 56% from academia, 16 percent work in a bank or a company, while the remaining 12 percent have another type of employer. The field of study of most of the experts is economics (84 percent). 65 percent of the experts have a PhD, and they predominantly come from North America or Western Europe (50 percent).[5]

2.2 Structure of the survey

Representative sample The structure of our survey with the representative sample is summarized in Figure 1. First, our respondents complete a series of demographic questions. Second, they receive brief non-technical definitions of the unemployment rate and the inflation rate to establish a common-ground definition of the two terms at the start of the survey. We also inform the respondents about the current values of the unemployment rate and the inflation rate in the US to ensure comparable information sets about the current state of the economy across respondents. Third, in the main part of the survey participants make predictions about unemployment and inflation under two hypothetical vignettes. Fourth, we ask additional questions to better understand the mechanisms driving our respondents' predictions. Finally, we collect data on some additional demographics. The full set of experimental instructions for Wave 1 and Wave 2 of the experiments can be found under the following link: `https://www.dropbox.com/s/tovmpbqhp0xm9pq/instructions_subjectivemodels.pdf?dl=0`.

Expert survey The expert survey consists of a subset of the survey given to the general population (see Figure 1). After being introduced to the question format, experts directly proceed to the prediction task in two randomly selected vignettes. At the end, experts respond to a few demographic questions.[6,7]

[5]In Wave 2 of our survey we have slightly different descriptives due to the specific demographic questions posed by the World Economic Survey.

[6]In Wave 1, PhD students and experts from policy institutions responded to the demographic questions. We obtained demographic information for academics from their websites.

[7]We do not include the definitions of inflation and unemployment, but still provide the experts with the most recent values of both variables to ensure common information sets about the current state of the economy.

2.3 Hypothetical vignettes

To measure our respondents' beliefs about the effects of different macroeconomic shocks, we use hypothetical vignettes in which we introduce our respondents to different scenarios and ask them to predict future unemployment and inflation. This approach allows us to fix people's beliefs about the source and the parameters of the shock and, importantly, to ensure that our respondents understand that the shocks are truly exogenous. The vignettes focus on four different exogenous shocks which are among the most commonly studied in macroeconomics: an oil supply shock, a government spending shock, a monetary policy shock, and a tax shock. Our participants are randomly assigned to make predictions for two out of four different hypothetical vignettes that are presented in random order.[8]

Each vignette follows the same structure (summarized in Figure 1). All start with a short introduction to familiarize respondents with the setting of the vignette and the variable that will be subject to the shock (a). For example, in the income tax vignette they are informed about the average tax rate and the amount that the median household currently pays in taxes on labor income. Then, respondents are presented with a baseline scenario in which they are asked to imagine that the variable of interest does not change (b). In this scenario, we elicit people's expectations about the unemployment rate in 12 months from now and the inflation rate over the next 12 months. Our respondents are then randomly assigned to either a "rise-scenario" or a "fall-scenario" (c). In the "rise-scenario" ("fall-scenario"), respondents predict unemployment and inflation under a hypothetical increase (decrease) of the shock variable of interest. Eliciting beliefs under both a baseline as well as a rise/fall scenario allows us to measure beliefs about the effects of shocks by taking differences across the scenarios. Thereby, constant individual-level differences in expectations about the future level of unemployment and inflation are cancelled out. Moreover, this leverages the behavioral phenomenon of "coherent arbitrariness" according to which statements about levels are often arbitrary and susceptible to ephemeral influences such as framing or anchoring, whereas differences between stated levels are comparatively coherent, robust, and reliable (Ariely et al., 2003). At the end of each vignette, respondents are asked how confident they are about their unemployment

[8]In Wave 2 of the expert survey we did not fully randomize the order of vignettes as this is not feasible in the World Economic Survey.

and inflation predictions for the vignette on five-point scales (d).

Respondents indicate the expected unemployment and inflation rates on two sliders that range from 0% to 10% for unemployment and from -2% to 8% for inflation. The default position of the sliders is the current value of the respective rate. The sliders ease the task for our respondents and reduce noise and cognitive strain. Finally, to account for potential order effects, we cross-randomize whether respondents first receive the questions on the inflation rate or the questions on the unemployment rate. For each participant, the order of the inflation and unemployment questions is identical across all vignettes. In what follows, we provide details on the vignettes.

Oil supply shock In the introduction to the oil vignette, respondents learn about the current average price of one barrel of crude oil. Then, in the baseline scenario, our respondents are told to imagine that the average price of crude oil stays constant over the next 12 months. Thereafter, they are randomly assigned to either an "oil price rise scenario" or an "oil price fall scenario". Specifically, respondents in the "oil price rise scenario" receive the following instructions:

> Imagine the average price of crude oil unexpectedly rises due to problems with the local production technology in the Middle East. On average, the price will be $30 higher for the next 12 months than the current price. That is, the price will be on average $84 for the next 12 months.

As is the case for all other vignettes, instructions for the fall-scenario are analogous to the rise-scenario.

Government spending shock This vignette first provides respondents with information on the size of yearly government spending in the US and its usual growth rate. In the baseline scenario, our respondents are told to imagine that federal government spending grows as usual over the next 12 months. In the "rise-scenario", our respondents receive the following instructions:

> Imagine federal government spending unexpectedly grows to a larger extent than usual over the next 12 months due to a newly announced spending program on defense. In particular, total government spending grows by 2.4

> percentage points more than the usual growth that took place in the previous years.
>
> The government announces: The change is temporary and occurs despite no changes in the government's assessment of national security or economic conditions. Moreover, federal taxes do not change in response to the spending program.

Monetary policy shock We familiarize respondents with the federal funds target rate and its current value. The baseline scenario asks our respondents to imagine that the Federal Open Market Committee announces that it will keep the federal funds target rate constant. In the subsequent "fall-scenario" our respondents receive the following instructions:

> Imagine the federal funds target rate is unexpectedly 0.5 percentage points lower. That is, in its next meeting, the Federal Open Market Committee announces that it is reducing the rate from 2.5% to 2%.
>
> Imagine the committee announces it does so with no changes in their assessment of the economic conditions.

Tax shock After a brief explanation of federal income taxes in the US, the baseline scenario tells our respondents to imagine that income tax rates stay constant for all US citizens over the next 12 months. In the subsequent "rise-scenario", our respondents receive the following instructions:

> Imagine that income tax rates are unexpectedly 1 percentage point higher for all households in the US over the next 12 months. This means that the typical US household would pay about $400 more in taxes.
>
> The government announces: The tax change is temporary and occurs despite no changes in the government's assessment of the economic conditions. Moreover, government spending does not change in response to the tax increase.

Endogenous policy response Our main object of interest is people's beliefs about the effects of the shocks on the unemployment and inflation rate while accounting for any

endogenous responses to the shocks by policymakers. In Wave 1 of our experiments we did not explicitly mention that respondents should take into account endogenous government and central bank responses when responding to the survey. To make this point more explicit, we gave respondents in Wave 2 of both the household and the expert survey the following instructions:

> In all your responses to the following questions, please think about what policymakers (the government and the central bank) would do under the different scenarios. Please account for the actions of policymakers that you would expect under the different scenarios and include them in your predictions.

Additional instructions In our instructions in both Wave 1 and Wave 2, we clearly state that respondents should make predictions about the US economy. Nevertheless, in Wave 2 we include an additional instruction to make sure that respondents do not instead consider a hypothetical economy when they respond to our questions. Moreover, to make sure that the respondents do not just interpret our questions as a test of their knowledge of economics, we tell them that we are interested in their actual views on what would happen under the different scenarios:

> In all of the following questions, please give us your best guess about how the unemployment rate and the inflation rate in the US economy would actually develop under the scenarios considered. This may or may not be in line with theoretical findings and evidence from economics. We are only interested in your own views and opinions on the US economy. All of the hypothetical scenarios and all of your predictions deal with the US economy.

Despite these differences in instructions across waves there are barely any differences in responses across the two waves, neither in the household nor in the expert survey. We therefore focus on the pooled sample in our main analysis.

Incentives To study the role of effort for the predictions in the vignettes, we provide a random subset of respondents with monetary incentives in Wave 1 of the household survey if their responses are close to the expert forecasts. Since incentivizing respondents with expert forecasts measures households' second-order beliefs about experts, we also measure perceived objectivity and accuracy of experts.

Discussion of vignette design Since we work with a general population sample, we face a trade-off between the precision of the vignette and the ease of understanding it. To avoid cognitive overload among the general population sample, we make the vignettes as simple to understand as possible. At the same time, we are careful to make it clear that the shocks are exogenous to the economy, which makes our estimates comparable to theoretical models and empirical evidence. For instance, we attribute the oil supply shock to changes in the local production technology in the Middle East. Similarly, in the interest rate scenario, we explicitly state that the change in interest rates occurs with no changes in the Fed's assessment of the economic conditions. Moreover, we also fix people's beliefs about the duration of the shocks by clarifying that the changes in taxation and government spending only last for one year.[9] For the government spending and taxation shocks, we clarify that the temporary nature of the shock is common knowledge by using the wording "the government announces". Finally, many of our design choices are motivated by common modeling assumptions in DSGE models and by empirical evidence from VARs in order to ensure comparability of our survey responses to these external benchmarks. For example, empirical evidence on government spending shocks often focuses on defense spending (e.g. Auerbach and Gorodnichenko (2012); Nakamura and Steinsson (2014)) as this type of spending does not affect the economy's productivity and does not directly redistribute resources across the income distribution.

2.4 Additional variables

We elicit several additional variables in the representative survey that allow us to study determinants and predictors of beliefs about the effects of macroeconomic shocks. In Wave 1 of the experiment we collect beliefs about the propagation mechanisms of the shocks and three standard questions measuring financial literacy (Lusardi and Mitchell, 2014). In Wave 2 of the experiment we collect additional variables to capture respondents' evaluations of how important the inflation rate and the unemployment rate are for their own economic situation, whether people perceive the shock variables, inflation, and unemployment as good or bad, several questions testing the numeracy of respondents, as well as beliefs about supply-side mechanisms operating in the context of the tax shock

[9] We do not fix beliefs about the duration of the change in interest rates under the monetary policy shock, since the interest rate should react endogenously to changes in inflation and unemployment in response to the shock through the Taylor rule.

and the interest rate shock.

In both waves of the survey, we also collect data on gender, age, income, education, economic education, political affiliation as well as financial assets, real assets, and outstanding debt. Moreover, participants report whether they generally follow news on the national economy.

3 Results

3.1 Benchmarks from the empirical and the DSGE literature

We compile a set of quantitative benchmarks for each shock from the theoretical and empirical literature. This enables us to compare the forecasts of experts and the general population with how the macroeconomic literature conventionally assesses the effect of each shock on inflation and unemployment. The details of each set of calculations and the sources used are contained in Appendix C. Table 1 reports the benchmark changes for the unemployment and inflation rate from the theoretical and empirical literature. The benchmarks are also depicted in Figure 2 and we discuss their magnitude compared to the expert forecasts in the next subsection.

To ensure comparability, we first calculate the relative size of the shock in each paper relative to the corresponding shock in the vignettes. For instance, most studies on government spending consider a shock with a magnitude of 1 percent of GDP while the spending change in our vignette is approximately 0.5 percent of GDP. We then rescale the estimated responses of output and inflation in the papers until quarter 4 accordingly. Since most papers focus on output as the main variable of real activity, we translate the responses into changes in the unemployment rate using Okun's Law. For the theoretical benchmarks, when possible, we consider as an immediate benchmark the most comparable shock in a model that is widely accepted as a standard medium-size New Keynesian DSGE model. Our main choice is Smets and Wouters (2007) and its extension to unemployment dynamics in Galí et al. (2011). As theoretical benchmarks for government spending and tax rate shocks we also draw on the multipliers computed by Zubairy (2014). For the oil price shock benchmarks, we take into account the changing relevance of exogenous shocks in the world supply of oil for the US market.[10]

[10] As discussed by Baumeister and Kilian (2016), while the price spikes of the early 1970s shaped the

3.2 Expert predictions

We next characterize the experts' beliefs about the effects of macroeconomic shocks pooling responses from Waves 1 and 2 as we do not find any qualitative differences in predictions across waves (Table A14). The expert forecasts mainly serve two purposes: (i) to provide a benchmark for the representative sample that does not rely on any particular model or piece of evidence; and ii) to document the average point beliefs and disagreement within the profession and compare them to the literature benchmarks. In the analysis, we estimate the following equations for each of the four vignettes separately:

$$\Delta u_i = \beta_1 \text{Rise}_i + \beta_2 \text{Fall}_i + \varepsilon_i \tag{1}$$

$$\Delta \pi_i = \gamma_1 \text{Rise}_i + \gamma_2 \text{Fall}_i + \eta_i \tag{2}$$

Δu_i denotes the difference in unemployment predictions between the rise/fall scenario and the baseline scenario ($u_{i,rise/fall} - u_{i,baseline}$). Correspondingly, $\Delta \pi_i$ denotes the difference in inflation predictions between the rise/fall scenario and the baseline scenario ($\pi_{i,rise/fall} - \pi_{i,baseline}$). We employ robust standard errors throughout the paper and, wherever applicable, we cluster standard errors at the respondent level. Results are summarized in Panel A of Table 2 and in Figure 2.

Oil supply shock Experts view the oil supply shock as a supply-side shock that moves unemployment and inflation in the same direction (Columns 1 and 2 in Table 2, Panel A). In particular, they predict an increase in unemployment of 0.24 percentage points and a rise in inflation of 0.45 percentage points in the scenario where the oil price increases by \$30. In the scenario in which the oil price decreases by \$30, they predict that the unemployment rate would be lower by 0.13 percentage points, and that the inflation rate would be lower by 0.33 percentage points. The absolute size of these expected changes is

way economists conventionally think about oil price shocks, the many developments in the structure of the US economy, including falling oil imports and the rise of domestic extraction, imply that the drivers of oil prices and their transmission to the rest of the economy have shifted. For this reason, we choose as an empirical benchmark the VAR estimate by Blanchard and Galí (2010) for the post-1984 period, which excludes the age of large oil price changes due to world supply shocks. Moreover, we base the theoretical benchmarks for the impact of oil supply shocks on the results of Bodenstein et al. (2011) and Balke and Brown (2018), where the former treats the US solely as an importer while the latter also includes domestic production.

qualitatively in line but quantitatively below the theoretical and empirical benchmarks.

Government spending shock Experts perceive the government spending shock as a demand-side shock that leads to changes in unemployment and inflation in opposite directions (Columns 3 and 4 in Table 2, Panel A). The experts predict a 0.31 percentage point lower unemployment rate and a 0.3 percentage point higher inflation rate in the rise-scenario. In the fall scenario, they predict that the unemployment rate would be higher by 0.3 percentage points and that the inflation rate would be lower by 0.22 percentage points.

The benchmark relationship for unemployment from the literature has the same sign as perceived by the experts, although it is smaller in size than the average expert forecast. This difference is not large and the confidence intervals of the expert beliefs include the upper bound of the benchmark. However, the difference may appear larger in light of the fact that the experts were surveyed at what was likely a peak of the business cycle. As shown by Auerbach and Gorodnichenko (2012), fiscal multipliers are empirically much closer to zero during business cycle peaks, while they tend to be largest during recessions. With respect to inflation, the benchmark from the DSGE literature is in line with expert predictions qualitatively, although it is also smaller in size.

Interest rate shock Our experts predict that unemployment would be higher by 0.29 percentage points, while inflation would be lower by 0.15 percentage points in response to an unexpected increase in the interest rate. In the fall scenario, experts predict that unemployment would be lower by 0.19 percentage points, and that inflation would be higher by 0.16 percentage points. This implies that the experts view monetary policy shocks as demand-side shocks.

The predictions for the change in unemployment after a monetary contraction are very close to the benchmark of 0.2 to 0.4 percentage points, while those for the expansion are just below the lower bound. Inflation predictions are also very close to the 0.15-0.20 percentage point change from the literature benchmark. Once again, considering the point of the business cycle in which the survey was conducted provides a broader perspective. As shown by Barnichon and Matthes (2016), when the labor market is tight, the unemployment response to monetary policy shocks may be muted while that of inflation is amplified. Taking this into account, the expert predictions appear even closer

to literature benchmarks.

Tax shock On average, experts think that income taxes act as a demand-side shock moving unemployment and inflation in opposite directions (Columns 7 and 8 in Table 2, Panel A). A 0.22 percentage point higher unemployment rate and a 0.11 percentage point lower inflation rate are expected under the rise-scenario. For the fall-scenario, experts predict a 0.24 percentage point lower unemployment rate and a 0.21 percentage point higher inflation rate.

The unemployment prediction is above the theoretical benchmark of 0.06 percentage points but is inside the 0.2 to 0.6 range from the empirical evidence. The empirical studies we reviewed focus on tax revenues rather than income tax rates, and may therefore not be closely comparable. Experts believe that the responses of macro variables are stronger for government spending shocks than for tax shocks.[11] This view is consistent with theoretical work such as that of Zubairy (2014).

Measuring disagreement Finally, we study the extent of disagreement among experts in the different scenarios. The variances of the predicted changes in unemployment and inflation, Δu and $\Delta \pi$, are relatively small, highlighting that there is relatively little disagreement among the experts (see Table A6).

Result 1. *Overall, experts' predictions are broadly consistent with the empirical and theoretical consensus for all shocks. The magnitudes of expert forecasts are broadly in the range of quantitative estimates from DSGE models and VAR evidence.*

Heterogeneity in the expert sample Using the rich demographic background data from the World Economic Survey (Wave 2 of the expert survey), we also shed light on correlates of making predictions that are qualitatively consistent with the theoretical and empirical benchmarks. Overall, we find relatively little heterogeneity in expert forecasts by gender, age, and location of the experts (Figure A.8). However, respondents who work for policy institutions and those with a PhD are more likely to make predictions consistent with the benchmarks (see Table A4). Moreover, the predictions of experts based in the US are very similar to predictions from experts not based in the US (Table

[11] Given that the government shock is approximately equal to 0.3% of GDP and the tax shock is close to 0.5% (see Appendix C for the calculation of these sizes), the difference would be even larger if the two shocks were scaled to the same magnitude.

A5). We find no significant relationship between confidence and whether the forecast is qualitatively aligned with the literature benchmarks (Table A13).

3.3 Predictions from the representative sample

We continue with the forecasts from the general population. Panel B of Table 2 and Figure 2 display the predictions of the inflation and the unemployment rate under the different hypothetical scenarios. Throughout this section we report pooled results from Wave 1 and Wave 2 as we find barely any differences in responses across the waves (Table A14).

Oil supply shock Respondents make qualitatively and quantitatively similar predictions to the experts: On average, they predict the unemployment rate to be 0.45 percentage points higher and the inflation rate to be 0.67 percentage points higher in the scenario where the oil price rises by $30. In the oil price fall-scenario, they expect the unemployment rate to be 0.21 percentage points lower and the inflation rate to be 0.33 percentage points lower.

Government spending shock Households' beliefs about how inflation responds to government spending changes is qualitatively consistent with expert beliefs. Households think that inflation would be lower by 0.26 percentage points in response to an exogenous reduction in government spending, and that it would be higher by 0.13 percentage points in response to an increase in government spending. Households on average think that unemployment responds to neither an increase nor a decrease in government spending.

Interest rate shock While respondents think that unemployment would be 0.17 percentage points higher in response to a rise in interest rates, they expect it to remain unchanged in response to a decrease in interest rates. Respondents expect a 0.15 percentage point *lower* inflation rate in response to a fall in the federal funds target rate, which is at odds with the expert forecasts and theory. Even more striking is that households think that an increase in the federal funds target rate would lead to a 0.19 percentage point *higher* inflation rate, again contrary to the benchmarks.

Tax shock Respondents think higher taxes would lead to a 0.3 percentage point higher unemployment rate, and that lower taxes would result in a 0.25 percentage point lower unemployment rate. However, they expect exogenous tax changes to influence inflation in the opposite direction of the expert forecasts. Specifically, they predict that a tax hike would result in a 0.21 percentage point higher inflation rate, while they forecast a 0.12 percentage point lower inflation rate in response to a tax cut.

Disagreement There is more disagreement among households than among experts. The standard deviations of the predicted changes in unemployment and inflation, Δu and $\Delta\pi$, are about twice as large as those among experts (see Table A6).

Result 2. *For the oil supply shock and the government spending shock the general population expects responses of both unemployment and inflation that are qualitatively aligned with benchmarks from expert predictions and the empirical and theoretical literature. By contrast, households predict movements in inflation in the opposite directions compared to the benchmarks in response to the interest rate and the tax shocks, while their predictions about unemployment are qualitatively aligned with the benchmarks. Finally, there is substantial heterogeneity in predictions in the representative panel.*

4 Mechanisms: heterogeneity of forecasts

In the vignettes, only 48% of all predictions from the general population are qualitatively in line with expert and literature benchmarks (see Figure 3). In this section we study different mechanisms that could be driving heterogeneity in the representative sample. Understanding the determinants of forecast heterogeneity among households is critical for informing models of expectation formation and for the policy implications of our findings. We define making benchmark-consistent predictions as a dummy variable that takes value one if a respondent believes that a shock affects a given variable (unemployment or inflation) directionally in line with the benchmarks.[12] Figure 4 depicts regression coefficients of different potential determinants of the fraction of benchmark-consistent forecasts that a respondent makes.

[12]We have at least one theoretical or empirical benchmark in all cases except for the effects of income tax shocks on inflation. In this case we rely on the conventional view of income tax shocks as demand-side shocks.

Beliefs about propagation mechanisms A potential determinant of people's beliefs about the effects of macroeconomic shocks is their understanding of the underlying propagation mechanisms. For instance, do households understand how an oil supply shock influences demand and supply? Do they understand how demand and supply affect inflation and unemployment? To better understand our respondents' subjective models, we measure their beliefs about the propagation mechanisms through which aggregate shocks unfold. For instance, in a broad class of canonical models, a contractionary oil supply shock typically (i) increases firms' production costs and (ii) reduces households' purchasing power. This respectively shifts (iii) the supply curve and (iv) the demand curve to the left which, in turn, affects (v) inflation and (vi) unemployment. In the interest rate shock, it is important how the financing costs of firms (i.e. the costs of borrowing money) and the interest rates that households earn on savings or pay on loans are usually affected. In the income tax vignette, a pure demand-side shock, tax changes lead to a change in purchasing power which ultimately affects inflation and unemployment through aggregate demand.

In Wave 1 of the survey we elicit people's beliefs about the causal direction of 14 relationships that should be relevant to the transmission of the four shocks we focus on. For example, to measure the perceived relation between the oil price and firms' production costs, we ask our respondents, "How do firms' production costs usually react if the oil price increases?". The answer categories are: "they increase"; "they decrease"; "neither of the above". We focus on questions for which there is agreement in the literature on the sign of the relationship.[13]

There is substantial heterogeneity in respondents' beliefs about propagation mechanisms behind all four macroeconomic shocks (Figure A.2). We calculate a score indicating the fraction of these 14 questions the respondent answered correctly. On average, respondents answer two thirds of the questions on propagation mechanisms in line with the benchmarks. Holding above-median correct beliefs about propagation mechanisms increases the fraction of benchmark-consistent forecasts by 10 percentage points (Figure 4). These results hold in all vignettes, even after controlling for financial literacy and a

[13]Having a benchmark allows us to cross-randomize a subset our respondents to receiving a bonus payment of \$0.50 if a randomly selected question is answered correctly and, hence, to mitigate survey fatigue or inattention. The incentives, however, do not affect whether respondents make forecasts that are consistent with the benchmarks. We ask all questions to all participants (randomly ordered), irrespective of which two vignettes they actually face.

range of demographics (Table A7).[14]

Financial literacy Figure 4 shows that respondents with above median financial literacy are 7 percentage points more likely to make benchmark-consistent predictions in the vignettes. However, once we condition on people's knowledge of propagation mechanisms, financial literacy is not significantly related to making benchmark-consistent macroeconomic forecasts except for predictions in the oil vignette, suggesting that specific knowledge of propagation mechanisms seems more important than a basic understanding of more general financial concepts (Table A7).

Supply-side view of the economy If households believe that higher income taxes or interest rates mainly affect inflation through higher costs for firms, this could explain why they perceive interest rate or tax hikes as inflationary. In the context of monetary policy, this view of the shock transmission would be consistent with the widely studied "cost channel" (Barth and Ramey, 2002). To measure respondents' perceived relevance of supply-side mechanisms operating in the interest rate and the taxation shock scenarios, in Wave 2 we ask our respondents to what extent they agree with the following three statements: (i) Firms tend to pass on increases in production costs to consumers in the form of higher prices; (ii) Increases in income taxes tend to increase firms' production costs; (iii) Increases in interest rates tend to increase firms' production costs. We categorize individuals who agree to statements i) and ii)/iii) as believing in a supply-side effect of increases in income taxes/interest rates. There is substantial heterogeneity in people's perceptions of the role of supply-side mechanisms, with 68 percent (77 percent) of respondents believing in a cost channel going from income taxes (interest rates) to prices (Figure A.4). People who believe in a cost channel predict a somewhat stronger (i.e. more positive) inflation response to income tax changes, but not to interest rate hikes (Figure 4 and Table A8). The likelihood of making benchmark-consistent forecasts decreases by 12 percentage points for the income tax vignette if people have a supply-side view, but there is no such effect in the interest rate vignette (Panel B of Table A8). Therefore, respondents' predictions of a positive co-movement of unemployment and inflation after an interest rate change are not driven by a strong supply-side view of the world (i.e. the

[14] In unreported regressions we find no strong differences in the role of beliefs about specific propagation mechanisms in shaping forecasts across vignettes.

cost-channel). It thus seems that structural uncertainty about the importance of supply-side factors cannot explain our results. However, there are potentially other dimensions of structural uncertainty that may affect households' beliefs, which should be tested in future research.[15]

Rational inattention Do household predictions deviate from benchmarks because holding biased beliefs about the functioning of the macroeconomy is not very costly to households? To examine whether rational inattention is driving deviations from expert predictions, we measure our respondents' beliefs about the relevance of unemployment and inflation for their own economic situation.[16] Specifically, we ask respondents about their agreement with the following two statements: (i) The US inflation rate is relevant for my own economic situation; (ii) The US unemployment rate is relevant for my own economic situation. Figure A.3 shows that there is substantial heterogeneity in the extent to which respondents perceive the unemployment rate and the inflation rate to be relevant for them personally. A majority of respondents (66 percent) considers the unemployment rate to be relevant for their own situation, and the fraction is even higher for inflation (87 percent). However, these measures of perceived relevance are uncorrelated with making predictions that are closer to benchmarks (Figure 4), suggesting that rational inattention is not the main driver of deviations from benchmarks.

Numeracy Can a lack of numeracy explain deviations of household predictions from benchmarks? We measure our respondents' numeracy using seven questions that have been applied previously in the context of macroeconomic expectations and are collected in the Survey of Consumer Expectations (Armantier et al., 2017). We find no significant correlation of a dummy indicating above median numeracy with the deviation of predictions from benchmarks (Figure 4).

[15]For instance, the perceived persistence of the monetary policy action crucially determines households' expectation of inflation dynamics. If households interpret an interest rate rise as permanent, they will also revise upward their expectation of the long-run inflation rate. Furthermore, in a broad class of New-Keynesian models, a permanent monetary tightening can increase inflation even in the short run, an outcome known as the Neo-Fisherian effect (Cochrane, 2016). While this effect has in the past been considered a puzzle of New-Keynesian theory, recent work provides support for its empirical existence (Uribe, 2019).

[16]Macroeconomic models of rational inattention are usually concerned with inattention to the current state of economy (Maćkowiak and Wiederholt, 2015; Reis, 2006; Sims, 2003). If rational inattention was driving benchmark-inconsistent predictions among our respondents, this would need to be inattention to the underlying structure of the economy.

Good-bad heuristic Research from psychology emphasizes that individuals revert to simple heuristics in complex decision environments in which there is much uncertainty (Gigerenzer and Todd, 1999). In light of this evidence, we consider whether a simple heuristic, namely that good things only lead to good things and bad things only lead to bad things, can explain the heterogeneity in predictions in the representative sample. We refer to this as the good-bad heuristic (GBH). It postulates that households perceiving two variables as both good or both bad (symmetric affective evaluation) are more likely to predict a positive co-movement between them, while predicting a movement in opposing directions if they perceive one variable as good and the other one as bad (asymmetric affective evaluation). Affective evaluations provide a particularly promising basis for heuristic extrapolation because emotional responses are known to be automatic, fast, and effortless and they typically exert a strong influence on behavior and reasoning (Kahneman, 2003; Loewenstein, 2000; Pham, 2007).

To test this hypothesis, for each variable of interest (i.e. the four shock variables, unemployment, and inflation), we measure whether respondents consider higher values as good or bad for the US economy and for their own household on 7-point scales, ranging from very bad (-3) to very good (3).[17,18] Then, we derive the directional prediction that follows from the GBH for each individual forecast. If a respondent evaluates the two variables underlying a forecast (e.g. government spending and inflation) symmetrically (asymetrically), the GBH implies a predicted change of the outcome variable in the same (opposite) direction as the change in the shock variable. If at least one variable is evaluated neutrally (neither good nor bad), no change is predicted. For example, if a respondent perceives both higher government spending and higher inflation as bad, the GBH predicts that she expects that inflation will increase in response to an exogenous increase in government spending. Finally, we construct a dummy that takes value one whenever the predicted change suggested by the GBH is in line with the benchmarks, that is, whenever following the GBH would result in a benchmark-consistent forecast. This dummy is used in our analyses.

We uncover a striking explanatory power of the good-bad heuristic. On average, forecast consistency with benchmarks increases by 21 percentage points when the GBH makes

[17] Our evidence is related to small-scale survey evidence from psychology studying psychology students' understanding of the macroeconomy (Leiser and Aroch, 2009).

[18] Figure A.5 highlights strong heterogeneity in affective evaluations of the different macroeconomic variables.

a consistent prediction (Figure 4, table 3). This amounts to a 51% increase in making forecasts that are consistent with the benchmarks. Moreover, the good-bad heuristic turns out to be a powerful explanatory variable in each vignette. Panel A of Table 3 shows the disaggregated results for each vignette.

The above analysis leverages two distinct sources of variation: (i) Average affective evaluations differ across vignettes and forecasts. (ii) Within each vignette and forecast, the affective evaluations differ across individuals. Both sources of variation explain roughly half of the overall effect size, as shown by the inclusion of question fixed effects (Panel B of Table 3). The joint estimate is robust to the inclusion of a large battery of individual-level demographic controls. Since we have variation in forecast consistency and the GBH on the individual level, we can even include individual fixed effects. This even increases the estimate to 22 percentage points, which underscores the robustness of the results on the good-bad heuristic.

How are macroeconomic variables affectively encoded by different groups? Females, more numerate respondents, people with above-median age, and people with more debt are substantially less likely to evaluate higher values of inflation, unemployment, oil prices, the federal funds rate and the income taxes as good (Table A11). Holdings of real and financial assets are largely unrelated to affective evaluations. Individuals with higher incomes view higher values of unemployment, inflation, the oil price or the fed funds rate as less negative, perhaps because they are less exposed to adverse macroeconomic shocks. Democrats are more likely to evaluate higher government spending and higher taxes as good.

Who uses the good-bad heuristic? We create a dummy variable taking value one if a respondent's predictions are consistent with the predictions of the good-bad heuristic. Most demographic variables have little predictive power for which respondents use the GBH. However, people who have an above-median age are more likely to make a prediction consistent with the good-bad heuristic (Table A11).

Our findings on the good-bad heuristic are related to a literature on extrapolation, which has shown that individuals extrapolate recent price changes (Bordalo et al., 2018, 2019) or extrapolate from their own situation to the macroeconomy (Kuchler and Zafar, 2019). The patterns in our data highlight a particular form of heuristic extrapolation across macroeconomic variables that depends on how the variables are affectively encoded,

consistent with an important role for sentiment in expectation formation (Kamdar, 2018). Our results thus call for a more systematic measurement of how households affectively represent macroeconomic variables to better understand the formation of economic expectations.

Political affiliation People's beliefs about how the economy is affected by the different shocks could also be driven by their political ideology. However, we find no significant heterogeneity in responses to the vignettes by people's political affiliation, even in the government spending vignette (see Figure 4 and Table A12).

Demographics Figure 4 highlights that there are substantial differences in the consistency of predictions with benchmarks across demographic groups. People with higher net wealth, older respondents, and college-educated respondents make a significantly higher fraction of forecasts consistent with benchmarks, in line with roles for cognitive limitations and learning over the life-cycle.

Confidence Confidence in predictions of the unemployment rate and the inflation rate among respondents from the representative sample is uncorrelated with the consistency of responses with benchmarks (Table A13 and Figure 4).

Result 3. *Taken together, our evidence suggests a quantitatively important role for the good-bad heuristic compared to all other potential determinants of household beliefs. Knowledge about propagation mechanisms and perceived importance of supply-side mechanisms are also correlated with making predictions that are consistent with benchmarks, but numeracy, perceived relevance of macroeconomic variables for the personal situation, financial literacy, standard demographics as well as confidence are all at best weakly correlated with the respondents' predictions.*

5 Robustness

In this section we discuss a series of robustness checks.

Misperceived endogeneity of the interest rate shock In the interest rate vignette we stress that the change in the Fed funds rate does not occur due to a change in the

Fed's assessment of economic conditions. However, there may still be a concern that respondents in the representative sample believe that higher interest rates indicate that the Fed is reacting to a rise in inflation, and therefore predict higher inflation. Since we anchor people on the current level of inflation, this could only be the case if respondents think that the Fed is reacting to a change in its outlook for future inflation. To address this concern, in Wave 2 we elicit subjective beliefs about how the Fed usually adjusts interest rates to (i) an unexpected increase in the outlook for inflation and (ii) an unexpected increase in the outlook for unemployment. To do so we prompt our respondents to "imagine that the FOMC changes their outlook for inflation (unemployment) over the next 12 months due to data revisions, while there is no change in the outlook for unemployment (inflation). Specifically, the Fed believes that the inflation (unemployment) rate will be 0.25 percentage points higher than their initial estimate." Thereafter, we measure people's beliefs about how the Fed would adjust the federal funds rate. Figure A.6 shows that there is substantial heterogeneity in beliefs on how the Fed would adjust interest rates in response to unexpected changes in its outlook for inflation or for unemployment. If our results were driven by respondents attributing a higher fed funds rate to a change in the Fed's outlook for inflation, we would expect stronger predicted increases in inflation in response to a positive interest rate shock for respondents who believe that the Fed more strongly raises interest rates in response to a higher outlook for the inflation rate. However, there is no significant heterogeneity along this dimension and, if anything, the patterns go in the opposite direction of what would be predicted by this potential confound (Table A17).

Incentives for vignettes To examine the role of effort and attention in responses to the hypothetical vignettes, we provide a random subset of respondents with monetary incentives in Wave 1 of the household survey. We inform these respondents that we asked economic experts the same questions and that for one randomly selected question they can earn an additional \$0.50 if their response is at most 0.2 percentage points away from the average expert response.[19]

Incentives moderately increase the fraction of benchmark-consistent predictions of inflation by 4 percentage points (Table A16 Column 1), while the predictions regarding

[19]\$0.50 corresponds to approximately one third of the show-up fee for respondents and is thus a relatively sizeable amount for respondents.

unemployment are completely unaffected (Column 2). In a joint test, no effect of incentives on consistency of predictions with the benchmarks can be detected (Column 4), even though incentivized respondents spend roughly 40 seconds longer in the vignettes – a 25% increase in response time (Column 6). Since incentivizing respondents with expert forecasts measures households' second-order beliefs about experts, we also interact the incentive dummy with a measure of perceived objectivity and accuracy of experts. The effect of incentives does not significantly vary with this measure of trust in experts (Panel B of Table A16).

Order effects To account for potential order effects, we randomize both the order of vignettes as well as the order in which unemployment or inflation forecasts are elicited. There are no significant order effects for the sample of experts from Wave 1 (Table A15).[20] Figure A.7 shows our main results for consumers pooling across Waves 1 and 2, separately for i) all forecasts, ii) forecasts under the first vignette faced by each respondent, iii) forecasts for the first variable (either unemployment or inflation) in both vignettes faced by a respondent. The figure highlights that the responses are very similar, indicating a limited relevance of order effects, even though a highly powered formal test shows statistical evidence of some small order effects (Table A15).

Attention to the survey Figure A.7 also displays forecasts separately iv) for a restricted sample excluding respondents in the upper and lower 10% tails of the survey time distribution, and (v) for a restricted sample excluding the 20% of respondents with the largest absolute difference in predictions in the baseline scenarios across the two vignettes to which they responded.[21] Our figure highlights very similar patterns for those two different samples, suggesting that a lack of attention does not account for the patterns observed in the household survey.

Wave effects We conducted both the household and expert surveys in two separate waves, where the second wave made it more explicit i) that respondents should account for endogenous policy responses, ii) that the survey is about the respondents' own opinions

[20]The randomization of order was not feasible in the World Economic Survey (experts Wave 2).

[21]Given that the baseline scenarios ask respondents to assume no change in the shock variable of interest, large differences in predictions between the two baseline scenarios each respondent faced could indicate inattention or random response behavior.

and not a test of their economic knowledge, and iii) that the questions are about the US economy, as discussed in Section 2. There are no significant differences across waves in our household survey or in our expert survey (Table A14). This suggests that the small changes in wording and the timing of the data collection across waves do not strongly affect the predictions.

6 Implications

In this section, we discuss the implications of our findings for macroeconomic models and fiscal and monetary policy.

Quantitative relevance of households' expectations What are the implications of our findings on households' expectations for their own optimal consumption decisions within a theoretical framework? To provide a tentative answer, we present a simple proof-of-concept exercise to highlight their potential quantitative importance. We embed the empirical beliefs about the effects of a monetary policy contraction into a partial-equilibrium version of a canonical medium-size DSGE model.

We combine the results from the monetary policy vignette with the aggregate dynamics from the model of Galí et al. (2011). First, we solve the model and compute the Impulse Response Functions (IRFs) from a contractionary monetary policy shock. We then construct "First-Order Conditions" (FOC) IRFs for consumption and labor supply which are derived by combining the households' optimization problem with the expectations for inflation and unemployment from the vignette. Specifically, we compute counterfactual IRFs by solving the households' first-order conditions for consumption and labor supply assuming that households expect inflation and unemployment to react to the shock as indicated by the mean prediction from the survey respondents. Expectations about all other endogenous variables are assumed to follow the path of the actual General Equilibrium (GE) IRFs.[22] These FOC IRFs can be interpreted as the representative household's planned consumption and labor supply decisions in response to the monetary contraction, formulated at the time the shock realizes, if the household holds beliefs about the macroeconomic effects of the shock in line with our survey findings. In other words, the exercise examines the contemporaneous response of expectations for the household's choice vari-

[22] Appendix D contains the details of this exercise.

ables, but it does not account for any general equilibrium feedback in the succeeding periods.

The lower panel of Figure 5 compares the GE IRFs for consumption and labor supply with the FOC IRFs derived by imposing the empirical expectations of inflation and unemployment from the vignettes (upper panel). Expecting positive inflation, the real interest rate faced by households falls despite the rise in the nominal rate. Households' optimal consumption decision, underpinned by the Euler equation, thus implies a rise in consumption rather than the fall consistent with the GE requirement. The expected rise in the unemployment rate is lower than in the model, leading to a smaller fall in wages (not shown) and thus a lower reduction in households' labor supply.

This simple illustrative exercise suggests that for the monetary shock the deviation of households' predictions from the equilibrium dynamics determined by the rational expectations restrictions may have important qualitative and quantitative implications.

Modeling the expectation formation mechanism A recent research effort tries to incorporate more realistic assumptions on the expectation formation mechanism into macroeconomic models. What are the implications of our findings for this literature?

First, standard models assume that economic agents agree on the true model of the economy. This also holds for models of imperfect information, in which agents may disagree about the current state of the economy but agree on its structure (Mankiw and Reis, 2002; Wiederholt, 2015; Woodford, 2003). These assumptions are at odds with the substantial disagreement in household predictions about the effects of all shocks in our survey. This finding lends support to a class of models in which households are uncertain and may disagree about structural parameters of the economy (Bhandari et al., 2019; Evans and Honkapohja, 2012; Milani, 2007; Orphanides and Williams, 2005).

Second, while there is substantial disagreement among our respondents across shocks, average predictions about the inflation response to government spending shocks and oil supply shocks and about the unemployment response to all shocks are qualitatively aligned with benchmarks. This suggests that the assumption that households hold beliefs about the effects of macroeconomic shocks in line with standard models may be a reasonable first approximation to the average household's beliefs in these contexts.

Third, households predict movements in inflation in response to monetary policy

shocks and tax shocks in the opposite direction compared to those predicted by standard models. Which models could generate such beliefs? Behavioral macroeconomic models featuring rational inattention (Maćkowiak and Wiederholt, 2015; Reis, 2006; Sims, 2003), k-level thinking (Farhi and Werning, 2017), no higher-order beliefs (Angeletos and Lian, 2017), or myopia (Gabaix, 2018) all entail that agents' expectations of future macroeconomic fluctuations are somewhat muted in magnitude but directionally aligned with the model's equilibrium, which contrasts with our evidence. Our findings on the explanatory power of the good-bad heuristic call for models in which agents' affective judgments shape their macroeconomic expectations (Kamdar, 2018). Extrapolation across macroeconomic variables according to affective evaluation parallels the extrapolation of recent price trends or personal circumstances, which has been documented by the literature (Bordalo et al., 2018, 2019; Kuchler and Zafar, 2019). Our findings on the GBH are also consistent with recent evidence showing that exogenously higher inflation expectations are associated with a more pessimistic outlook about general economic conditions, and lead firms to *decrease* investment expenditure (Coibion et al., 2019c) and households to *reduce* their spending on durable goods (Coibion et al., 2019a). Similarly, Binder (2019) finds that when unemployment falls, many consumers revise their inflation forecasts downward.

Fiscal and monetary policy-making Our findings also have several implications for policymakers. For instance, on average households expect unemployment to fall in response to a tax cut, but not after a government spending stimulus. This could lead to differential consumption responses to the announcement of different types of fiscal policies. Our quantitative exercise above indicates that our findings matter for the transmission of monetary policy. Moreover, deviations of household beliefs from expert views about changes in inflation in response to government and central bank policies highlight the importance of communication. Different communication strategies could crucially influence the effectiveness of fiscal and monetary policy (Blinder et al., 2008; Coibion et al., 2019b; Haldane and McMahon, 2018; Hansen et al., 2017, 2019). Finally, our finding of substantial heterogeneity in households' beliefs about macroeconomic relationships implies a large degree of variation in the effectiveness of monetary policy and fiscal policy in shifting expectations and behavior for different subpopulations of interest.

7 Conclusion

Drawing on a sample of experts and a large representative sample of the US population, we provide evidence on beliefs about the unemployment and inflation responses to four different macroeconomic shocks: an oil supply shock, a monetary policy shock, a government spending shock, and a tax shock.

We establish a series of novel results: First, the expert predictions are both qualitatively and quantitatively aligned with the predictions from standard models and there is relatively little disagreement among experts. Second, we find substantial heterogeneity in responses to the vignettes in the representative panel. While for the oil supply shock and the government spending shock experts and households expect rather similar responses of unemployment and inflation, households' predictions substantially deviate from those of experts in the interest rate and the tax shock vignettes. In general, households' predictions about unemployment are largely in line with experts' predictions, while households' predictions about inflation in several cases deviate from our benchmarks. Third, we show that a large fraction of deviations of household predictions from expert predictions can be explained by the use of a simple heuristic that we label the "good-bad heuristic". According to this heuristic, households who perceive two variables as both good or both bad predict that these two variables co-move in the same direction. Our findings call for a more systematic measurement of people's mental representations of macroeconomic variables to better understand the formation of their economic expectations.

References

Acosta, Miguel and Hassan Afrouzi, "Rationally Confused: On the Aggregate Implications of Information Provision Policies," 2019.

Afrouzi, Hassan, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," *Working Paper*, 2019.

Angeletos, George-Marios and Chen Lian, "Dampening General Equilibrium: From Micro to Macro," *Working Paper*, 2017.

Ariely, Dan, George Loewenstein, and Drazen Prelec, ""Coherent Arbitrariness": Stable Demand Curves without Stable Preferences," *The Quarterly Journal of Economics*, 2003, *118* (1), 73–105.

Armantier, Olivier, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar, "An Overview of the Survey of Consumer Expectations," *Economic Policy Review*, 2017, *23* (2).

_ , **Scott Nelson, Giorgio Topa, Wilbert van der Klaauw, and Basit Zafar**, "The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment," *Review of Economics and Statistics*, 2016, *98* (3), 503–523.

_ , **Wändi Bruine de Bruin, Giorgio Topa, Wilbert Klaauw, and Basit Zafar**, "Inflation Expectations and Behavior: Do Survey Respondents Act on their Beliefs?," *International Economic Review*, 2015, *56* (2), 505–536.

Armona, Luis, Andreas Fuster, and Basit Zafar, "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment," *The Review of Economic Studies*, 2018, *86* (4), 1371–1410.

Auerbach, Alan J. and Yuriy Gorodnichenko, "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 2012, *4* (2), 1–27.

Auerbach, Alan J, Yuriy Gorodnichenko, and Daniel Murphy, "Local Fiscal Multipliers And Fiscal Spillovers in the United States," *NBER Working Paper Series*, 2019.

Bachmann, Rüdiger, Kai Carstensen, Stefan Lautenbacher, and Martin Schneider, "Uncertainty and Change: Survey Evidence of Firms' Subjective Beliefs," *Working Paper*, 2019.

_ , **Tim O Berg, and Eric R Sims**, "Inflation Expectations and Readiness to Spend: Cross-Sectional Evidence," *American Economic Journal: Economic Policy*, 2015, *7*, 1–35.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel, "House Price Beliefs and Mortgage Leverage Choice," *Review of Economic Studies*, 2019.

_ , **Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel**, "The Economic Effects of Social Networks: Evidence from the Housing Market," *Journal of Political Economy*, 2018, *126* (6), 2224–2276.

Balke, Nathan S. and Stephen P.A. Brown, "Oil Supply Shocks and the U.S. Economy: An Estimated DSGE Model," *Energy Policy*, 2018, *116* (February), 357–372.

Barnichon, Regis and Christian Matthes, "Gaussian Mixture Approximations of Impulse Responses and the Nonlinear Effects of Monetary Shocks," *Federal Reserve Bank of Richmond Working Paper Series*, 2016, (16-08).

Barth, Marvin J III and Valerie Ramey, "The Cost Channel of Monetary Transmission," in Ben Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2001*, Cambridge, MA: MIT Press, 2002, pp. 199–240.

Basso, Henrique S. and Omar Rachedi, "The Young, the Old, and the Government: Demographics and Fiscal Multipliers," *Working Paper*, 2019.

Baumeister, Christiane and Lutz Kilian, "Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us," *Journal of Economic Perspectives*, February 2016, *30* (1), 139–60.

Bhandari, A., J. Borovicka, and P. Ho, "Survey Data and Subjective Beliefs in Business Cycle Models," *Federal Reserve Bank of Richmon Working Paper Series*, 2019, (19-14).

Binder, Carola, "Long-Run Inflation Expectations in the Shrinking Upper Tail," *Working Paper*, 2019.

_ **and Alex Rodrigue**, "Household Informedness and Long-Run Inflation Expectations: Experimental Evidence," *Southern Economic Journal*, 2018, *85* (2), 580–598.

_ **and Christos Makridis**, "Stuck in the Seventies: Gas Prices and Macroeconomic Expectations," *Working Paper*, 2018.

Blanchard, Olivier and Jordi Galí, "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s so Different from the 1970s? Chapter," in "International Dimensions of Monetary Policy" 2010.

Blinder, Alan S, Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen, "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence," *Journal of Economic Literature*, 2008, *46* (4), 910–45.

Bodenstein, Martin, Christopher J. Erceg, and Luca Guerrieri, "Oil shocks and External Adjustment," *Journal of International Economics*, 2011, *83* (2), 168–184.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Diagnostic Expectations and Credit Cycles," *The Journal of Finance*, 2018, *73* (1), 199–227.

_ , _ , **Yueran Ma, and Andrei Shleifer**, "Overreaction in Macroeconomic Expectations," *Working Paper*, 2019.

Boumans, Dorine and Johanna Garnitz, "Ifo World Economic Survey Database–An International Economic Expert Survey," *Jahrbücher für Nationalökonomie und Statistik*, 2017, *237* (1), 71–80.

Carvalho, Carlos and Fernanda Nechio, "Do People Understand Monetary Policy?," *Journal of Monetary Economics*, 2014, *66*, 108–123.

Cavallo, Alberto, Guillermo Cruces, and Ricardo Perez-Truglia, "Inflation Expectations, Learning and Supermarket Prices: Evidence from Field Experiments," *American Economic Journal: Macroeconomics*, 2017, *9* (3), 1–35.

Christelis, Dimitris, Dimitris Georgarakos, Tullio Jappelli, Luigi Pistaferri, and Maarten Van Rooij, "Asymmetric Consumption Effects of Transitory Income Shocks," *Economic Journal*, 2017.

Cochrane, J., "Do Higher Interest Rates Raise or Lower Inflation?," *Working Paper*, 2016.

Coibion, Olivier and Yuriy Gorodnichenko, "What Can Survey Forecasts Tell us about Information Rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

_ **and** _ , "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *The American Economic Review*, 2015, *105* (8), 2644–2678.

_ **and** _ , "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, 2015, *7* (1), 197–232.

_ **, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij**, "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial," *National Bureau of Economic Research Working Paper*, 2019.

_ **, Yuriy Gorodnichenko, and Michael Weber**, "Monetary Policy Communications and their Effects on Household Inflation Expectations," *Working Paper*, 2019.

_ **,** _ **, and Saten Kumar**, "How do firms form their expectations? new survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

_ **,** _ **, and Tiziano Ropele**, "Inflation expectations and firm decisions: New causal evidence," *Quarterly Journal of Economics*, 2019.

_ **,** _ **, Saten Kumar, and Mathieu Pedemonte**, "Inflation Expectations as a Policy Tool?," *Working Paper*, 2018.

D'Acunto, Francesco, Daniel Hoang, and Michael Weber, "Managing Households' Expectations with Simple Economic Policies," *Working Paper*, 2019.

_ **,** _ **, Maritta Paloviita, and Michael Weber**, "Cognitive Abilities and Inflation Expectations," *American Economic Review Papers and Proceedings*, 2019.

_ **,** _ **,** _ **, and** _ , "Human Frictions to the Transmission of Economic Policy," *Working Paper*, 2019.

_ **,** _ **,** _ **, and** _ , "IQ, Expectations, and Choice," *Working Paper*, 2019.

Delavande, Adeline and Basit Zafar, "University Choice: The Role of Expected earnings, Non-pecuniary Outcomes and Financial Constraints," *Journal of Political Economy*, 2018.

DellaVigna, Stefano and Devin Pope, "Predicting Experimental Results: Who Knows What?," *Journal of Political Economy*, 2018, *126* (6), 2410?2456.

_ **and** _ , "What Motivates Effort? Evidence and Expert Forecasts," *Review of Economic Studies*, 2018, *85* (2), 1029–1069.

Dräger, Lena, Michael J Lamla, and Damjan Pfajfar, "Are Survey Expectations Theory-consistent? The Role of Central Bank Communication and News," *European Economic Review*, 2016, *85*, 84–111.

Evans, George W and Seppo Honkapohja, *Learning and Expectations in Macroeconomics*, Princeton University Press, 2012.

Farhi, Emmanuel and Iván Werning, "Monetary Policy, Bounded Rationality and Incomplete Markets," *Working Paper*, 2017.

Fuster, Andreas, Benjamin Hebert, and David Laibson, "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing," *NBER Macroeconomics Annual*, 2012, *26* (1), 1–48.

_ , **David Laibson, and Brock Mendel**, "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 2010, *24* (4), 67–84.

_ , **Greg Kaplan, and Basit Zafar**, "What Would You Do With $500? Spending Responses to Gains, Losses, News and Loans," *Working Paper*, 2018.

_ , **Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar**, "Expectations with Endogenous Information Acquisition: An Experimental Investigation," *Working Paper*, 2019.

Gabaix, Xavier, "A Behavioral New Keynesian Model," *Working Paper*, 2018.

Galí, Jordi, Frank Smets, and Rafael Wouters, "Unemployment in an Estimated New Keynesian Model," *NBER Macroeconomics Annual*, 2011, *26* (1), 329360.

Gigerenzer, Gerd and Peter M Todd, *Simple heuristics that make us smart*, Oxford University Press, USA, 1999.

Goldfayn-Frank, Olga and Johannes Wohlfart, "Expectation Formation in a New Environment: Evidence from the German Reunification," *Journal of Monetary Economics*, 2019.

Graeber, Thomas, "Inattentive Inference," *Working Paper*, 2019.

Haldane, Andrew and Michael McMahon, "Central bank communications and the general public," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 578–83.

Hansen, Stephen, Michael McMahon, and Andrea Prat, "Transparency and deliberation within the FOMC: a computational linguistics approach," *The Quarterly Journal of Economics*, 2017, *133* (2), 801–870.

_ , _ , **and Matthew Tong**, "The long-run information effect of central bank communication," 2019.

Kahneman, Daniel, "Maps of Bounded Rationality: Economicst Psychology for Behavioral Economics," *The American Economic Review*, 2003, *93* (5), 1449–1475.

Kamdar, Rupal, "The Inattentive Consumer: Sentiment and Expectations," 2018.

Kuchler, Theresa and Basit Zafar, "Personal Experiences and Expectations about Aggregate Outcomes," *Journal of Finance*, 2019.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor, "Heuristic thinking and limited attention in the car market," *American Economic Review*, 2012, *102* (5), 2206–36.

Leiser, David and Ronen Aroch, "Lay understanding of macroeconomic causation: The good-begets-good heuristic," *Applied Psychology*, 2009, *58* (3), 370–384.

Loewenstein, George, "Emotions in Economic Theory and Economic Behavior," *American Economic Reivew*, 2000, *90* (2), 426–432.

Lusardi, Annamaria and Olivia S Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 2014, *52* (1), 5–44.

Maćkowiak, Bartosz and Mirko Wiederholt, "Business Cycle Dynamics under Rational Inattention," *The Review of Economic Studies*, 2015, *82* (4), 1502–1532.

Malmendier, Ulrike and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk-taking?," *The Quarterly Journal of Economics*, 2011, *126* (1), 373–416.

_ **and** _ , "Learning from Inflation Experiences," *The Quarterly Journal of Economics*, 2016, *131* (1), 53–87.

Mankiw, N Gregory and Ricardo Reis, "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 2002, *117* (4), 1295–1328.

Milani, Fabio, "Expectations, Learning and Macroeconomic Persistence," *Journal of Monetary Economics*, 2007, *54* (7), 2065–2082.

Nakamura, Emi and Jòn Steinsson, "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, March 2014, *104* (3), 753–92.

Orphanides, Athanasios and John C Williams, "The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning, and Expectations," *Journal of Economic Dynamics and Control*, 2005, *29* (11), 1927–1950.

Pham, Michel Tuan, "Emotion and rationality: A critical review and interpretation of empirical evidence," *Review of General Psychology*, 2007, *11* (2), 155–178.

Reis, Ricardo, "Inattentive Consumers," *Journal of Monetary Economics*, 2006, *53* (8), 1761–1800.

Roth, Christopher and Johannes Wohlfart, "How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?," *Review of Economics and Statistics, forthcoming*, 2019.

Sims, Christopher A, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

Uribe, M., "The Neo-Fisher Effect: Econometric Evidence from Empirical and Optimizing Models," *Working Paper*, 2019.

Wiederholt, Mirko, "Empirical Properties of Inflation Expectations and the Zero Lower Bound," *Working Paper*, 2015.

Wiswall, Matthew and Basit Zafar, "Preference for the Workplace, Investment in Human Capital, and Gender," *The Quarterly Journal of Economics*, 2017, *133* (1), 457–507.

Woodford, Michael, "Imperfect Common Knowledge and The Effects of Monetary Policy," in Philippe Aghion, Roman Frydman, Joseph E. Stiglitz, and Michael Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, Princeton, NJ: Princeton Univ. Press, 2003.

Zubairy, Sarah, "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model," *International Economic Review*, 2014, *55* (1), 169–195.

Main Figures

Figure 1: Overview of the survey structure and the structure of the vignettes

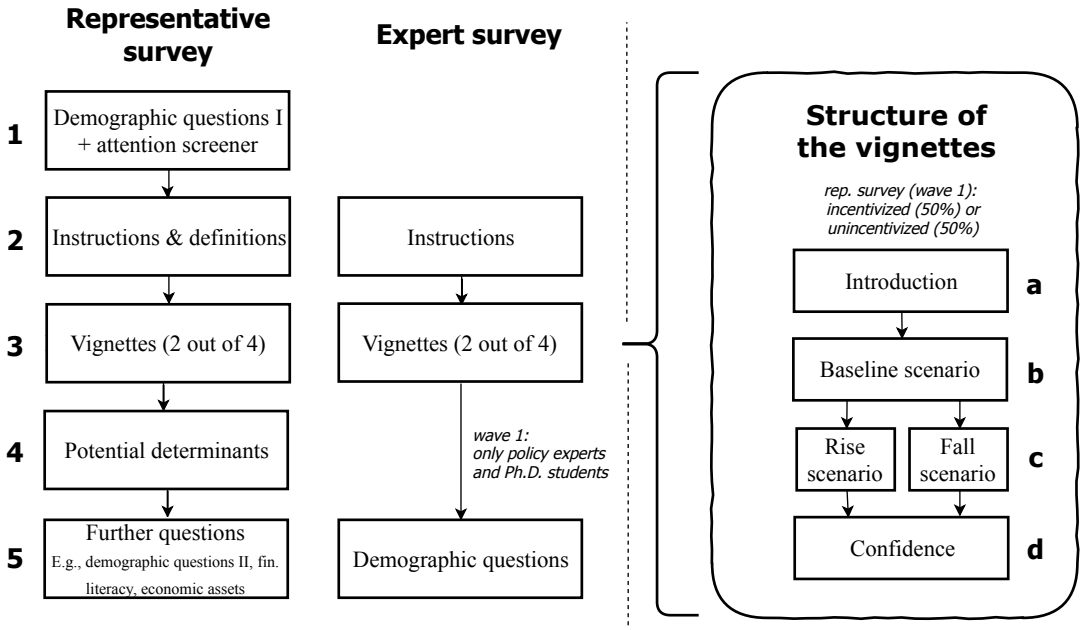
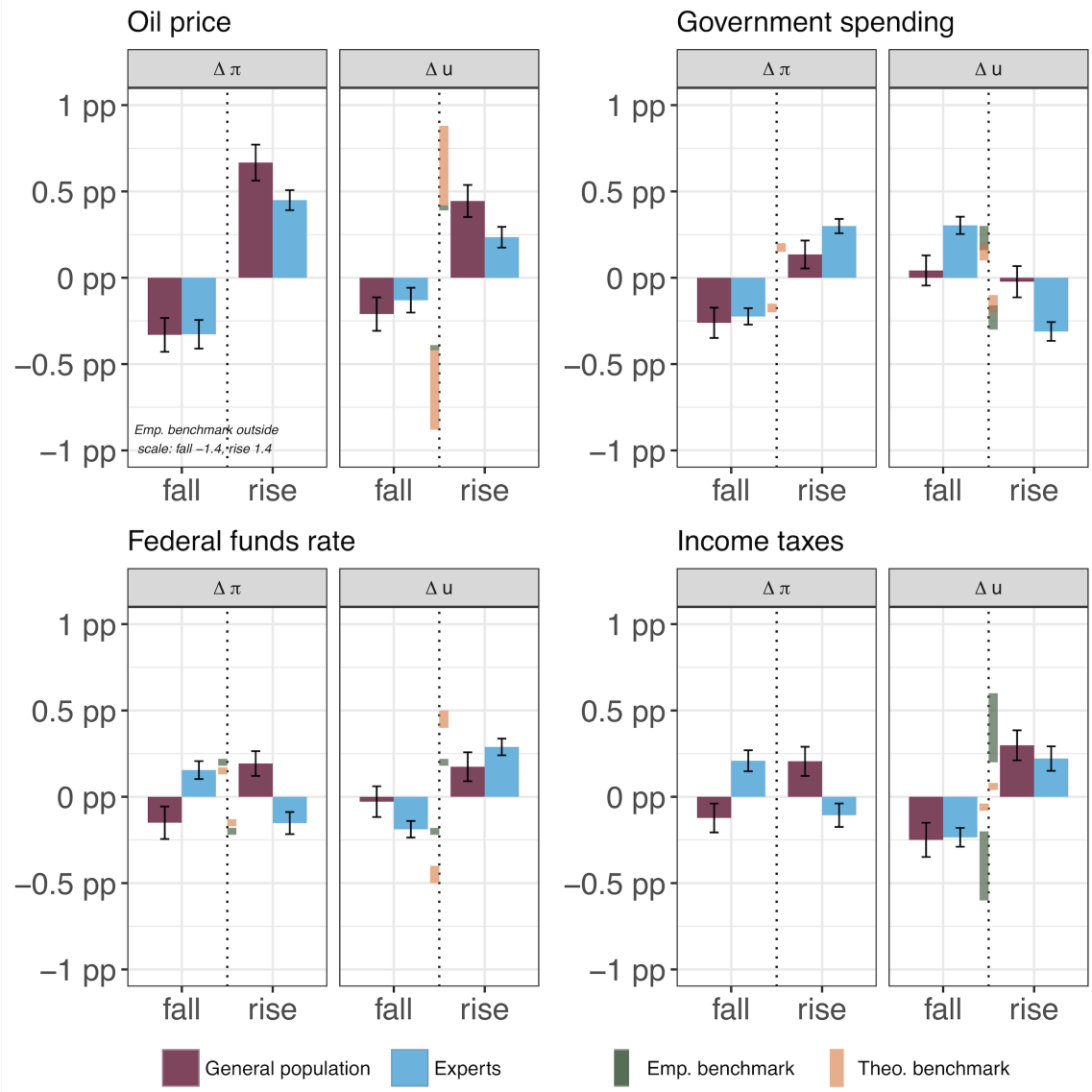
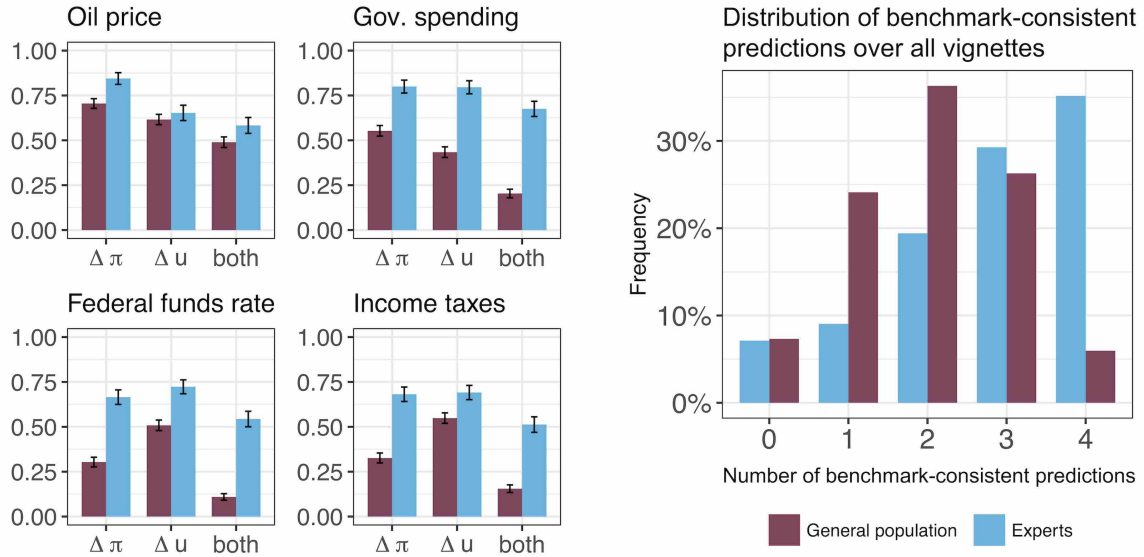


Figure 2: Beliefs about the effects of different shocks (with 95% CI)



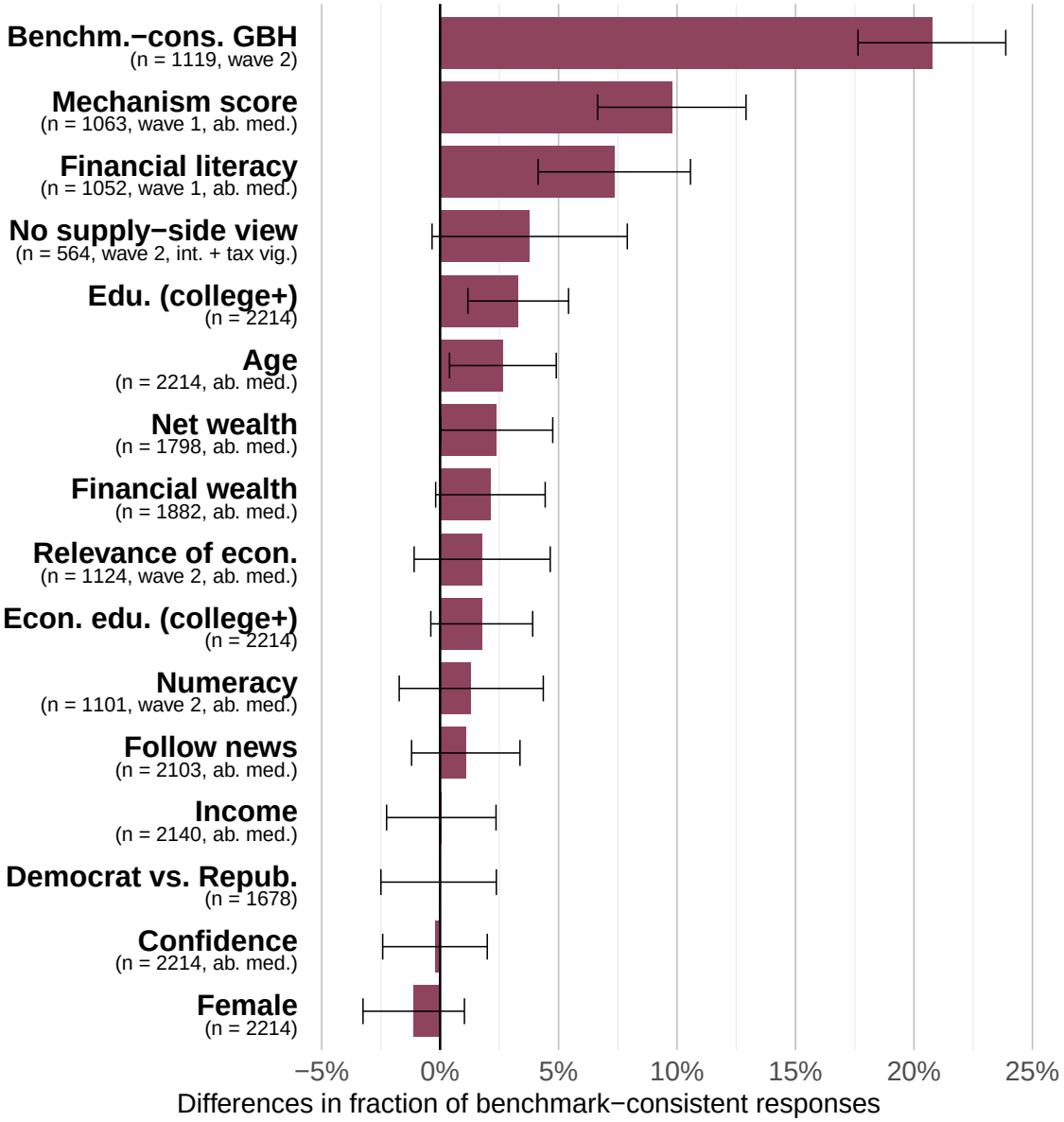
Notes: This figure provides an overview of beliefs about the effects of macroeconomic shocks on the unemployment rate and the inflation rate. It plots beliefs for the "rise" and "fall" scenarios for each of the different vignettes separately. The red bars show the mean responses in the representative sample, while the blue bars show the responses from our expert survey. Error bars display 95% confidence intervals using robust standard errors. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta \pi$ denotes the expected change in the inflation rate compared to the baseline scenario.

Figure 3: Distribution of benchmark-consistent forecasts among experts and the general population



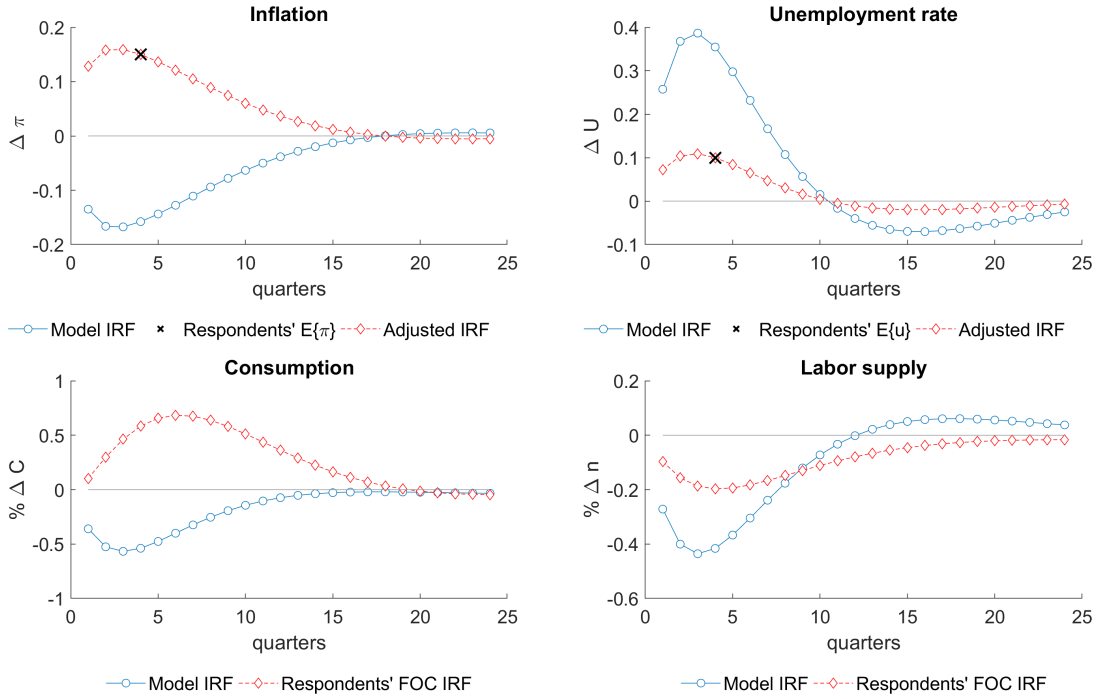
Notes: This figure shows which fraction of the predictions from experts or the general population are directionally consistent with the theoretical and empirical benchmarks. The left panel depicts the fraction of benchmark-consistent responses (and their 95% confidence intervals using robust standard errors) for each vignette and for inflation and unemployment separately. The third column, "both" shows how many respondents make benchmark-consistent predictions for both the unemployment and the inflation rate. The right panel presents the frequency distribution of total benchmark-consistent responses over all vignettes. It ranges from 0 to 4 because respondents predict four changes in total (2 vignettes $\times$ 2 macroeconomic variables (unemployment, inflation)).

Figure 4: Correlates of benchmark-consistent responses of the general population (with 95% CI)



Notes: This figure presents the effect of various covariates on the likelihood of making predictions that are consistent with the benchmarks using the representative online panel. A prediction is viewed as benchmark-consistent if it is directionally in line with the benchmarks. The effects are estimated in bivariate regressions. The fraction of benchmark-consistent predictions is regressed on each binary covariate. 95% confidence intervals are displayed. Where applicable clustered standard errors at the respondent level are used. Otherwise, robust standard errors are used. The underlying sample size is reported in parentheses. If a covariate is available only in a specific wave, this is reported. "ab. med." denotes an indicator variable for above-median values. "Benchm.-cons. GBH" denotes an indicator taking value 1 if the respondents' good-bad heuristic makes the benchmark-consistent prediction. For more details, see section 4. "No supply-side view" denotes an indicator taking value 1 if a respondent does not perceive the shock as a supply-side shock. This analysis relies on responses to the interest rate and income tax rate scenarios only. See Section 4 for more details.

Figure 5: Impulse Response Functions to a contractionary monetary policy shock (interest rate rise): rational expectations model and "first-order conditions" responses using empirical expectations.



Notes: The blue cirlces report the general-equilibrium rational-expectations IRFs from a monetary policy shock from Galí et al. (2011), rescaled to be consistent with the size of the shock in the vignette. In the upper panels, the red diamonds plot the IRFs of the model for unemployment and inflation "chained" to be consistent with the households' mean expectations from the vignette (black crosses). In the lower panels, the red diamonds report the "First-Order Conditions" IRFs for consumption and labor supply computed using the household's first-order conditions and the model's IRFs for the necessary endogenous variables, and the adjusted IRFs for inflation and unemployment from the upper panel. The FOC IRFs can be interpreted as the household's optimal plan for future consumption and labor supply, formulated once the shock occurs, if expectations for inflation and unemployment were aligned with those from the vignette. See Section 6 and Appendix D for more details.

Main Tables

Table 1: Benchmarks for the sign and size of the effects of different shocks

Shock		**Unemployment Response**		**Inflation Response**	
		Sign	Value (p.p.)	Sign	Value (p.p.)
Oil price rise	Theory	+	0.42 to 0.88		
(55% higher price)	Empirical	+	0.42	+	1.4
Government spending rise	Theory	−	−0.1 to −0.2	+	0.15 to 0.2
(2.4% higher growth rate)	Empirical	−	−0.16 to −0.3		
Interest rate rise	Theory	+	0.4 to 0.5	−	−0.15
(0.5 b.p. higher rate)	Empirical	+	0.2	−	−0.2
Tax rate rise	Theory	+	0.06		
(1 p.p. higher rates)	Empirical	+	0.2 to 0.6		

Notes: The table reports the benchmarks for changes in the unemployment rate and the inflation rate four quarters after the respective shock from the theoretical and empirical literature. The values are adjusted to be comparable to the size of the shocks in our survey. Empty fields indicate that – to the best of our knowledge – there is no robust and rigorous evidence on the effect of a given shock on the respective outcome variable of interest. Appendix C contains details on the papers surveyed for this exercise and the calculations used to the derive the values.

Table 2: Beliefs about the effects of different shocks

Panel A: Experts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.327^{***}	-0.130^{***}	-0.224^{***}	0.303^{***}	0.155^{***}	-0.188^{***}	0.209^{***}	-0.235^{***}
	(0.042)	(0.037)	(0.024)	(0.026)	(0.027)	(0.025)	(0.031)	(0.028)
rise	0.449^{***}	0.235^{***}	0.299^{***}	-0.311^{***}	-0.152^{***}	0.289^{***}	-0.107^{***}	0.221^{***}
	(0.030)	(0.030)	(0.021)	(0.028)	(0.033)	(0.025)	(0.035)	(0.036)
p-values from additional tests								
(i) fall≠rise	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
(ii) |fall|≠|rise|	0.858	0.596	0.088	0.566	0.106	0.055	0.657	0.331
Observations	482	481	474	475	517	513	515	521
R^2	0.333	0.120	0.373	0.352	0.096	0.270	0.093	0.164

Panel B: General Population

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.331^{***}	-0.210^{***}	-0.261^{***}	0.042	-0.150^{***}	-0.028	-0.122^{***}	-0.250^{***}
	(0.050)	(0.049)	(0.045)	(0.045)	(0.048)	(0.045)	(0.043)	(0.051)
rise	0.667^{***}	0.445^{***}	0.135^{***}	-0.023	0.193^{***}	0.174^{***}	0.206^{***}	0.298^{***}
	(0.053)	(0.047)	(0.041)	(0.046)	(0.037)	(0.043)	(0.043)	(0.044)
p-values from additional tests								
(i) fall≠rise	<0.001	<0.001	<0.001	0.31	<0.001	0.001	<0.001	<0.001
(ii) |fall|≠|rise|	0.007	0.643	0.089	0.694	0.054	0.418	0.257	0.634
(iii) ≠ exp.: fall	0.954	0.192	0.467	<0.001	<0.001	0.002	<0.001	0.794
(iv) ≠ exp.: rise	<0.001	<0.001	<0.001	<0.001	<0.001	0.02	<0.001	0.18
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.159	0.085	0.042	0.001	0.029	0.014	0.027	0.056

Notes: This table provides an overview of beliefs about the effect of the different shocks on the unemployment rate and the inflation rate. It displays beliefs for the "rise" and "fall" scenarios for each of the different vignettes separately. Panel A provides evidence from the expert sample. Panel B displays responses from the representative online panel. Δu denotes the predicted change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the predicted change in the inflation rate compared to the baseline scenario. Additionally, p-values from the following regressions are reported: (i) test whether there is a difference between rise and fall predictions, (ii) test whether there is a difference in the absolute size of rise and fall predictions, (iii) test whether there is a difference in fall predictions between experts and the general population, (iv) repeats this test for rise predictions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3: Evidence for the Good-Bad Heuristic

Panel A: Effect on forecast consistency with benchmarks for each vignette

	pooled over all questions				with question FE			
	oil✓	gov.✓	fed.✓	tax✓	oil✓	gov.✓	fed.✓	tax✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
benchm. c. GBH	0.130***	0.147***	0.155***	0.217***	0.132***	0.115***	0.048	0.084**
	(0.039)	(0.031)	(0.032)	(0.031)	(0.038)	(0.035)	(0.039)	(0.039)
Constant	0.571***	0.440***	0.357***	0.365***	0.608***	0.485***	0.306***	0.310***
	(0.035)	(0.018)	(0.018)	(0.018)	(0.036)	(0.030)	(0.020)	(0.020)
Question FE?	No	No	No	No	Yes	Yes	Yes	Yes
Obs.	1,110	1,118	1,114	1,134	1,110	1,118	1,114	1,134
R^2	0.014	0.021	0.024	0.045	0.020	0.024	0.048	0.074

Panel B: Effect on forecast consistency with benchmarks pooled over all vignettes

	pooled over all questions			with question FE		
	all✓	all✓	all✓	all✓	all✓	all✓
	(1)	(2)	(3)	(4)	(5)	(6)
benchm. c. GBH	0.208***	0.204***	0.221***	0.096***	0.099***	0.096***
	(0.016)	(0.018)	(0.021)	(0.019)	(0.022)	(0.026)
Constant	0.408***	0.602***	0.834***	0.515***	0.710***	0.894***
	(0.010)	(0.145)	(0.016)	(0.025)	(0.144)	(0.036)
Question FE?	No	No	No	Yes	Yes	Yes
Controls?	No	Yes	-	No	Yes	-
Ind. FE?	No	No	Yes	No	No	Yes
Obs.	4,476	3,580	4,476	4,476	3,580	4,476
R^2	0.043	0.044	0.279	0.078	0.073	0.309

Panel C: Effect on forecasts pooled over all vignettes

	pooled over all questions		with question FE	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
fall	−0.215***	−0.040	−0.289***	−0.066
	(0.059)	(0.075)	(0.068)	(0.083)
rise	0.158***	0.048	0.088	0.018
	(0.061)	(0.070)	(0.068)	(0.080)
fall×GBH pos.	−0.050	−0.142*	−0.060	−0.155*
	(0.072)	(0.086)	(0.072)	(0.086)
rise×GBH pos.	0.176**	0.278***	0.163**	0.271***
	(0.071)	(0.079)	(0.071)	(0.080)
Question FE?	No	No	Yes	Yes
Obs.	2,238	2,238	2,238	2,238
R^2	0.054	0.034	0.056	0.037

Notes: This table reports two tests of the Good-Bad Heuristic (GBH), conducted with Wave 2 from the general population sample. A prediction is viewed as benchmark-consistent if it is directionally in line with the benchmarks. In panels A and B, an indicator for a benchmark-consistent forecast is regressed on the dummy *benchm. c. GBH* that takes value one if a participant's GBH makes the benchmark-consistent forecast. Thus, the coefficients can be interpreted as the effect of a benchmark-consistent GBH on the probability of a benchmark-consistent forecast. Panel A conducts the analysis vignette by vignette. For instance, the first column *oil✓* runs the analysis for the forecasts in the oil vignette. Panel B pools over all vignettes. Panel C summarizes the effect of the GBH on raw forecasts. Δu and $\Delta\pi$ denote the expected changes in the unemployment rate and the inflation rate compared to the baseline scenario. *GBH pos.* is a dummy that takes value 1 if the GBH predicts a positive change in the outcome variable. Question FE indicate fixed effects for each vignette-rate forecast. Controls (Panel B) include age, education, log income, net wealth (inverse hyperbolic sine transformation), log financial wealth, gender, numeracy, news consumption, the perceived relevance of macroeconomic conditions for the personal situation, and political affiliation. Standard errors clustered at the respondent level are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Online Appendix: Subjective Models of the Macroeconomy: Evidence from Experts and a Representative Sample

Peter Andre[1] Carlo Pizzinelli[2]

Christopher Roth[3] Johannes Wohlfart[4]

Summary of the online Appendix

Section A provides additional figures. Figure A.1 provides an overview of the raw data in the different scenarios. Figure A.2 shows descriptive evidence regarding people's beliefs about propagation mechanisms. Figure A.3 shows descriptive evidence regarding respondents' perceptions of the relevance of inflation and unemployment for their own household. Figure A.4 displays descriptives regarding people's perceptions of the supply-side channels through which the interest rate shock and the income taxation shock could operate. Figure A.5 provides descriptives on the "good-bad heuristic". Figure A.6 provides descriptive evidence regarding subjective interest rate rules. Figure A.7 graphically displays the household predictions for different subsamples of respondents. Figure A.8 displays experts' forecasts for different subgroups.

Section B provides additional tables. Table A1 shows an overview of the four surveys that we conducted. Tables A2 and A3 provide summary statistics for the covariates of the representative online panel and the expert sample respectively. Tables A4 and A5 shed light on heterogeneity in expert predictions. Table A6 displays the variance in responses across the vignettes separately for consumers and experts. Table A7 analyzes whether benchmark-consistent forecasts are associated with higher understanding of the shocks' propagation mechanisms. Table A8 shows heterogeneous results by people's perception of the supply-side effects of the macroeconomic shocks. Table A9 tests for the relevance of rational inattention. Table A10 correlates demographics with a dummy for making

[1]Peter Andre, University of Bonn, p.andre@uni-bonn.de

[2]Carlo Pizzinelli, IMF, cpizzinelli@imf.org

[3]Christopher Roth, Department of Economics, University of Warwick, briq, CESifo, CAGE, email: Christopher.Roth@warwick.ac.uk

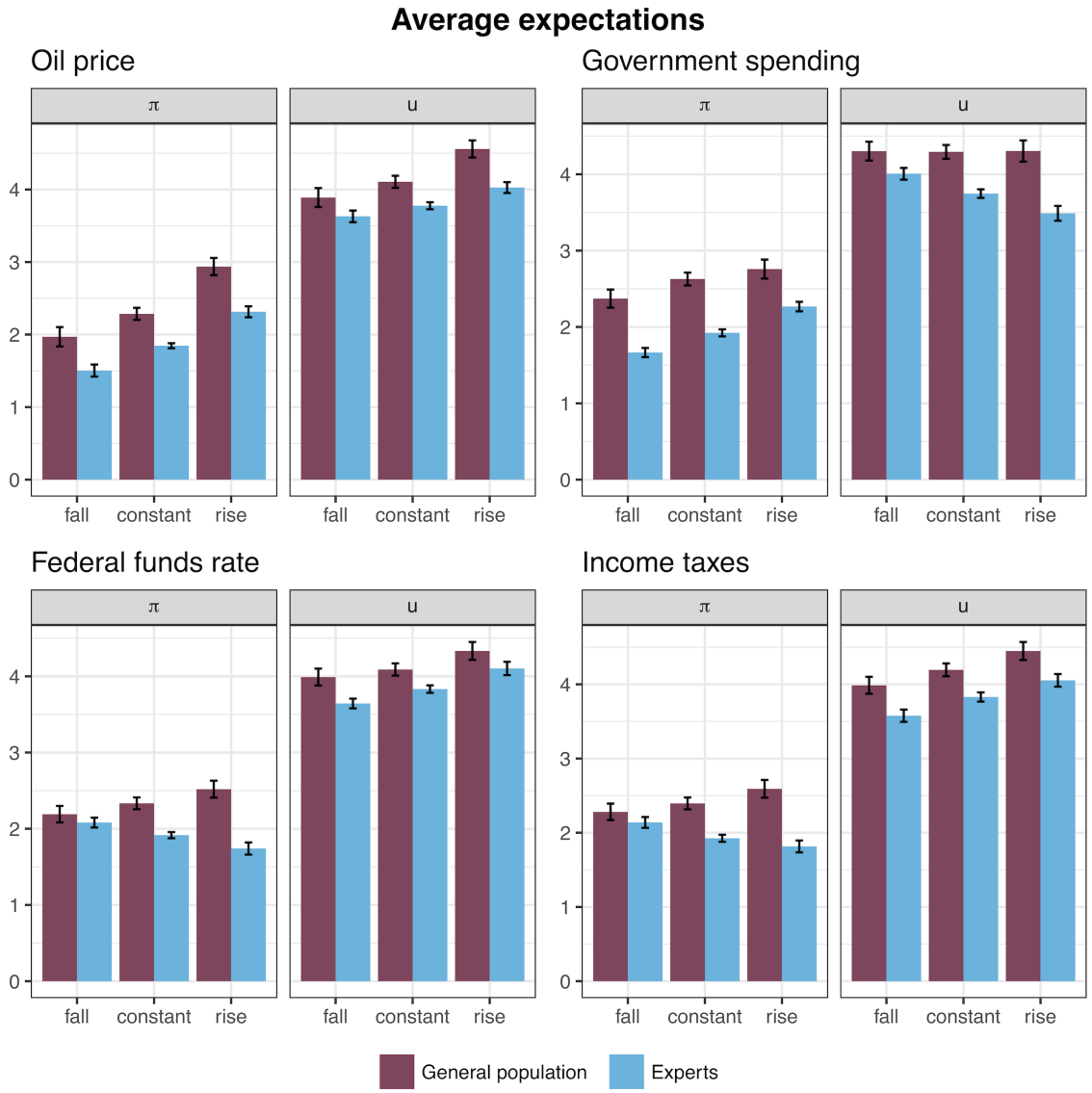
[4]Johannes Wohlfart, Department of Economics, University of Copenhagen, e-mail: johannes.wohlfart@gmx.de

predictions that are consistent with the good-bad heuristic. Table A11 displays correlates of the affective evaluations of different macroeconomic variables. Table A12 examines whether people's political affiliation significantly affects responses to the vignettes. Table A13 correlates measured confidence with the consistency of predictions with benchmarks. Table A14 tests for the relevance of wave effects. Table A15 displays tests for order effects. Table A16 displays the effect of incentives. Table A17 displays heterogeneous responses to the vignettes by people's perception of how interest rates would respond to changes in expected inflation and unemployment.

Section C provides details on the empirical and theoretical literature used to derive the benchmarks for changes in unemployment and inflation in response to shocks. Section E provides additional details on recruitment in Wave 1 of the expert survey.

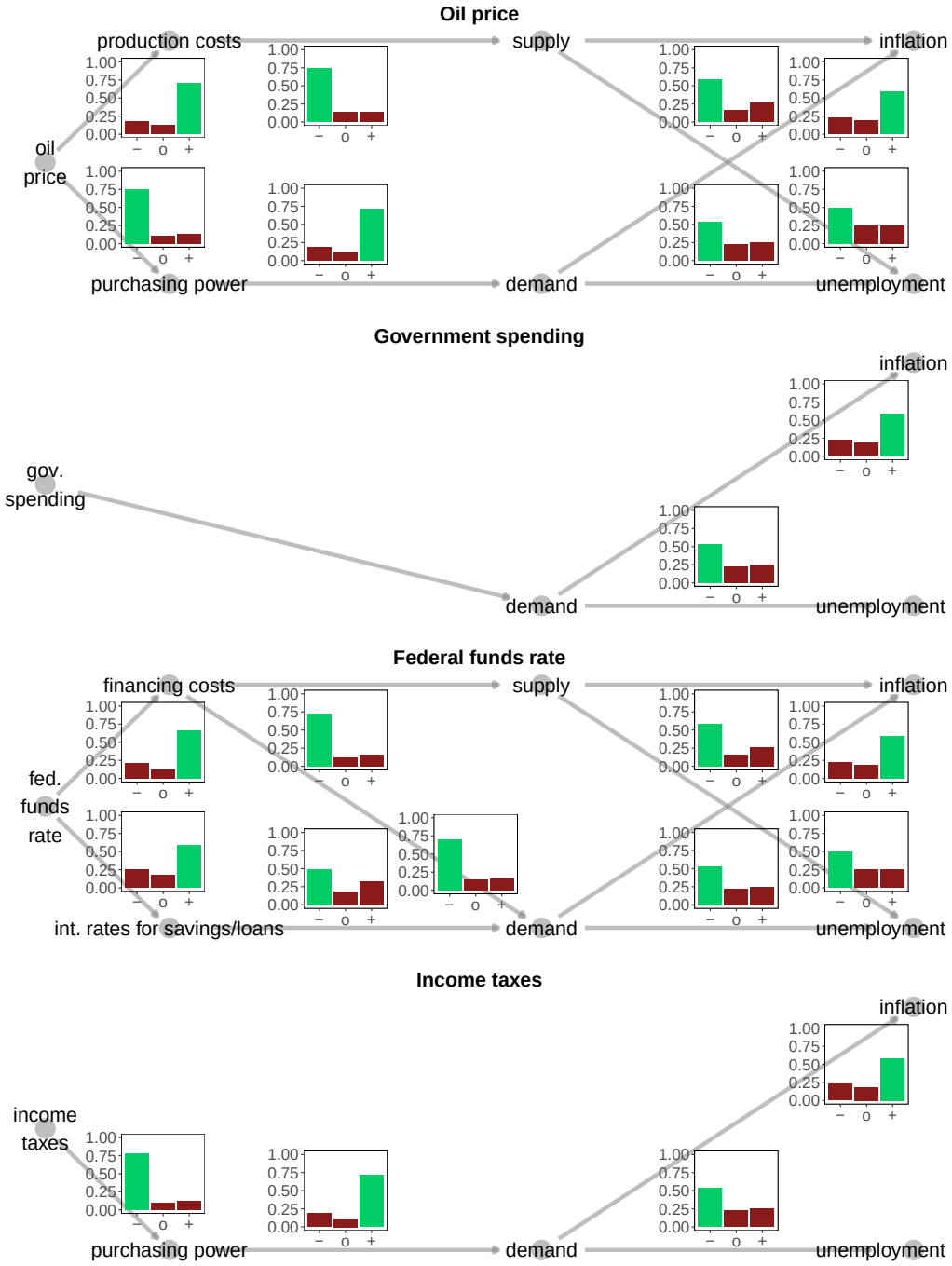
A Additional figures

Figure A.1: Overview of raw data: Predictions



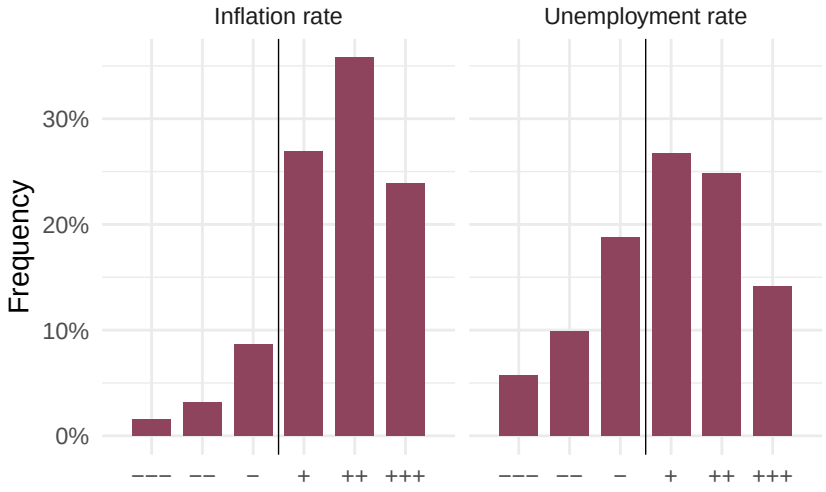
Notes: This figure plots mean predicted levels of unemployment and inflation by experts and respondents from the representative samples across the different vignettes and scenarios.

Figure A.2: Beliefs about propagation mechanisms



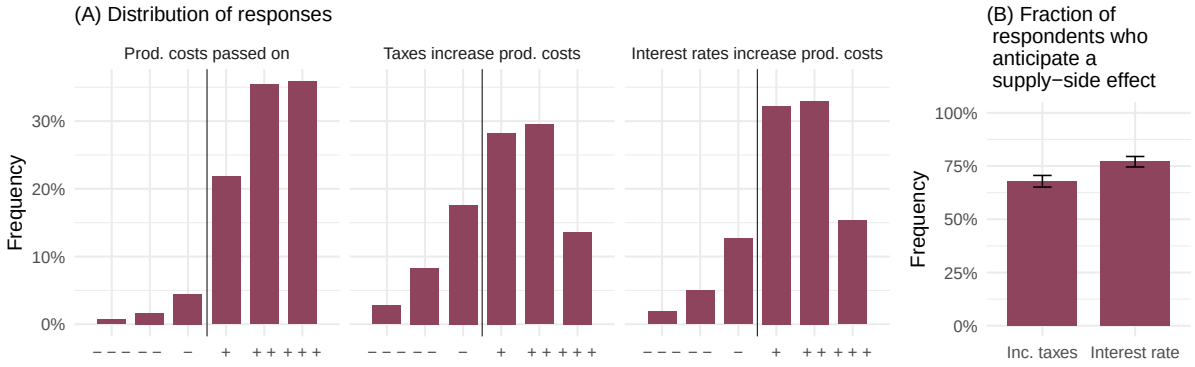
Notes: This figure uses data from the representative online panel and shows the frequency of people's responses to the propagation mechanism questions related to all four vignettes. "+" indicates that people perceive a positive association between two variables; "−" indicates that people perceive a negative association between two variables. "o" indicates that people perceive neither a positive nor a negative association. The green bars show the fractions of correct responses, while the red bars show the fractions of incorrect responses for each of the questions separately.

Figure A.3: Descriptive statistics: Rational inattention



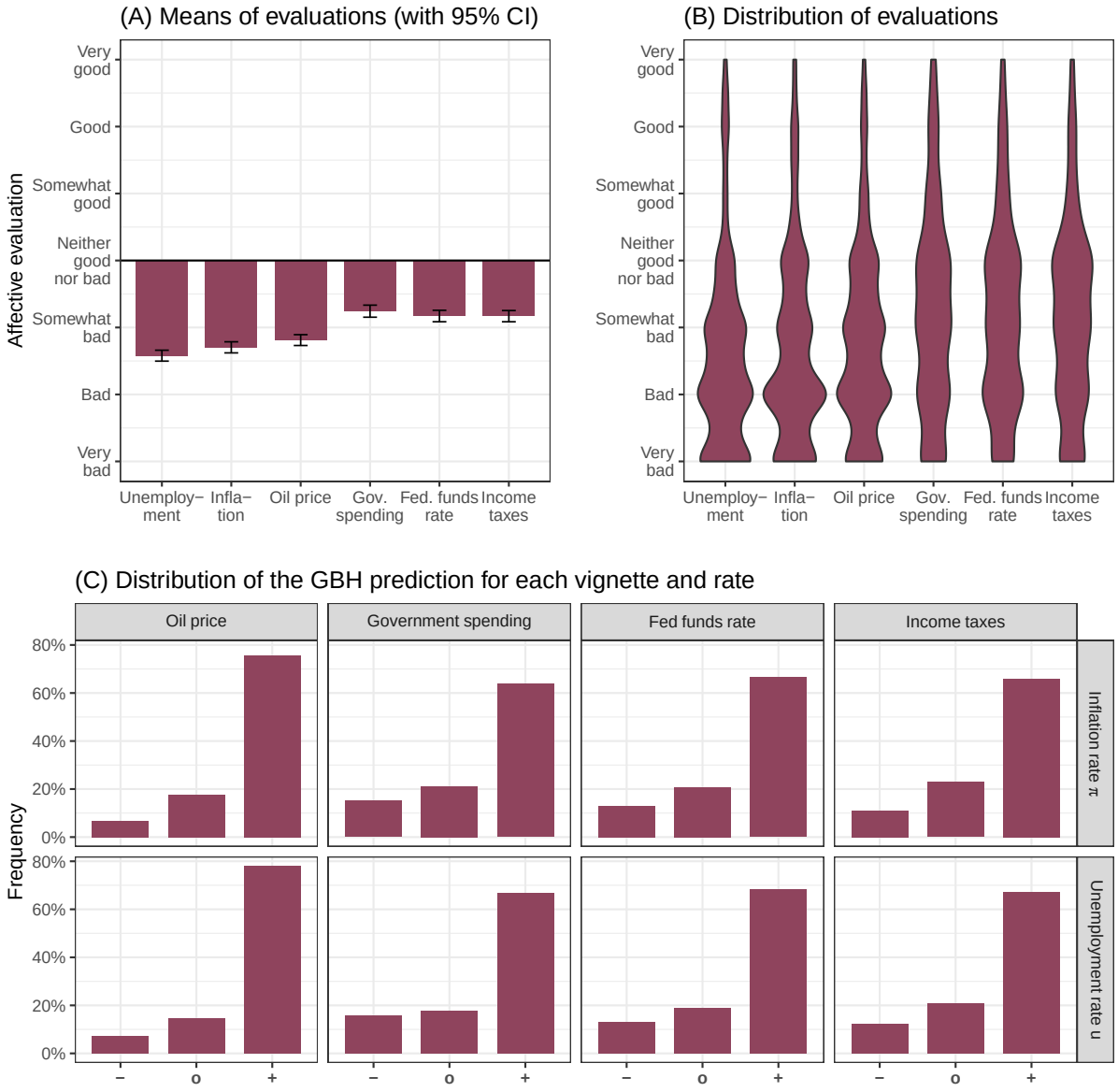
Notes: The figure presents the distribution of responses to the two rational inattention questions. The questions measure agreement with the statements: (i) The US inflation rate is relevant for my own economic situation. (ii) The US unemployment rate is relevant for my own economic situation. The answers are coded as follows: +++: strongly agree, ++: agree, +: somewhat agree, - somewhat disagree, - -: disagree, - - -: strongly disagree.

Figure A.4: Descriptive statistics: Supply-side view of the economy



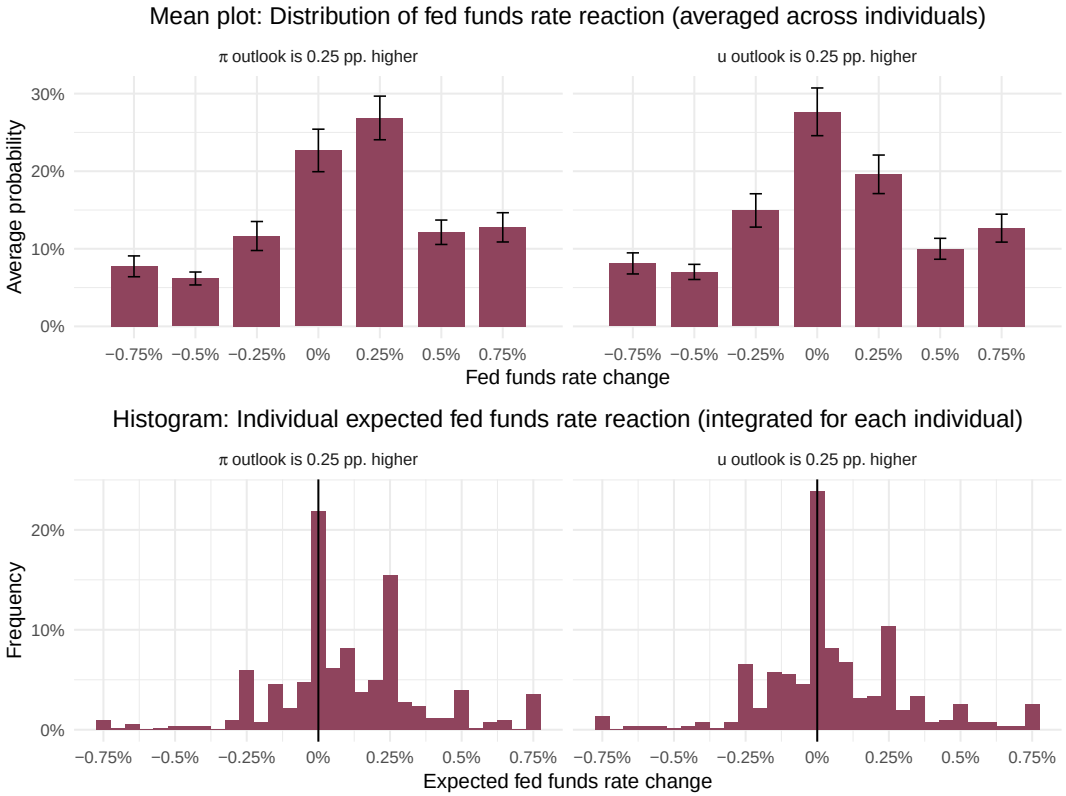
Notes: Panel A presents the distribution of responses to the three supply-side questions. The questions measure agreement with the statements: (i) Firms tend to pass on increases in production costs to consumers in the form of higher prices. (ii) Increases in income taxes tend to increase firms' production costs. (iii) Increases in interest rates tend to increase firms' production costs. The answers are coded as follows: + + +: strongly agree, + +: agree, +: somewhat agree, - somewhat disagree, - -: disagree, - - -: strongly disagree. Panel B displays the fraction of respondents that believe in a supply-side effect, i.e. agree with statements (i) and (ii) for the income taxes case and statements (i) and (iii) for the interest rate case.

Figure A.5: Descriptive statistics: Good-Bad Heuristic



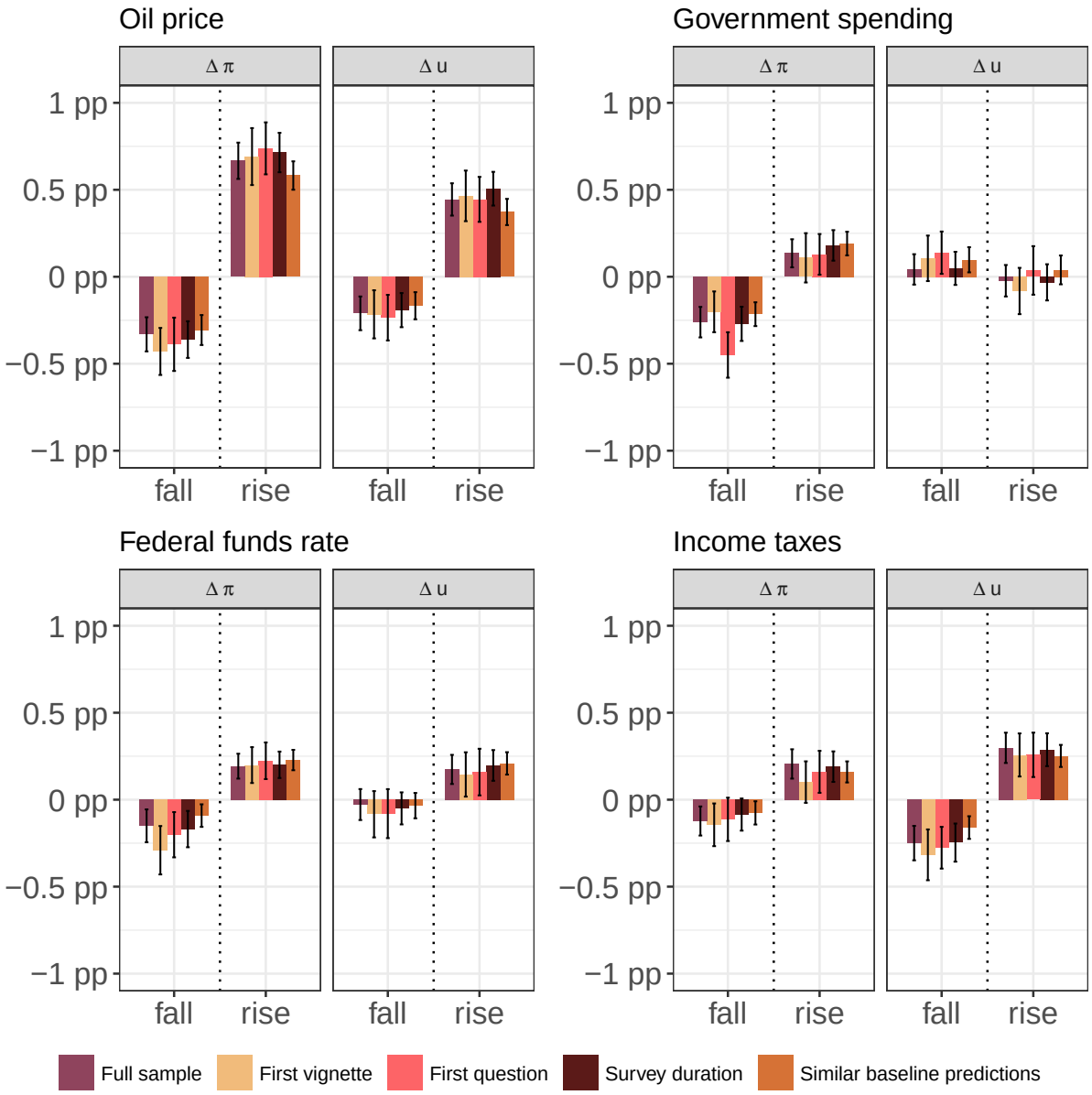
Notes: Panel A presents the average affective evaluation of the different variables (with 95% CI) and panel B its distribution. Panel C displays how many respondents' GBH predicts an increase (+), a decrease (–), or no change (o) of unemployment or inflation across vignettes. For a discussion, see Section 4.

Figure A.6: Descriptive statistics for the subjective interest rate rules



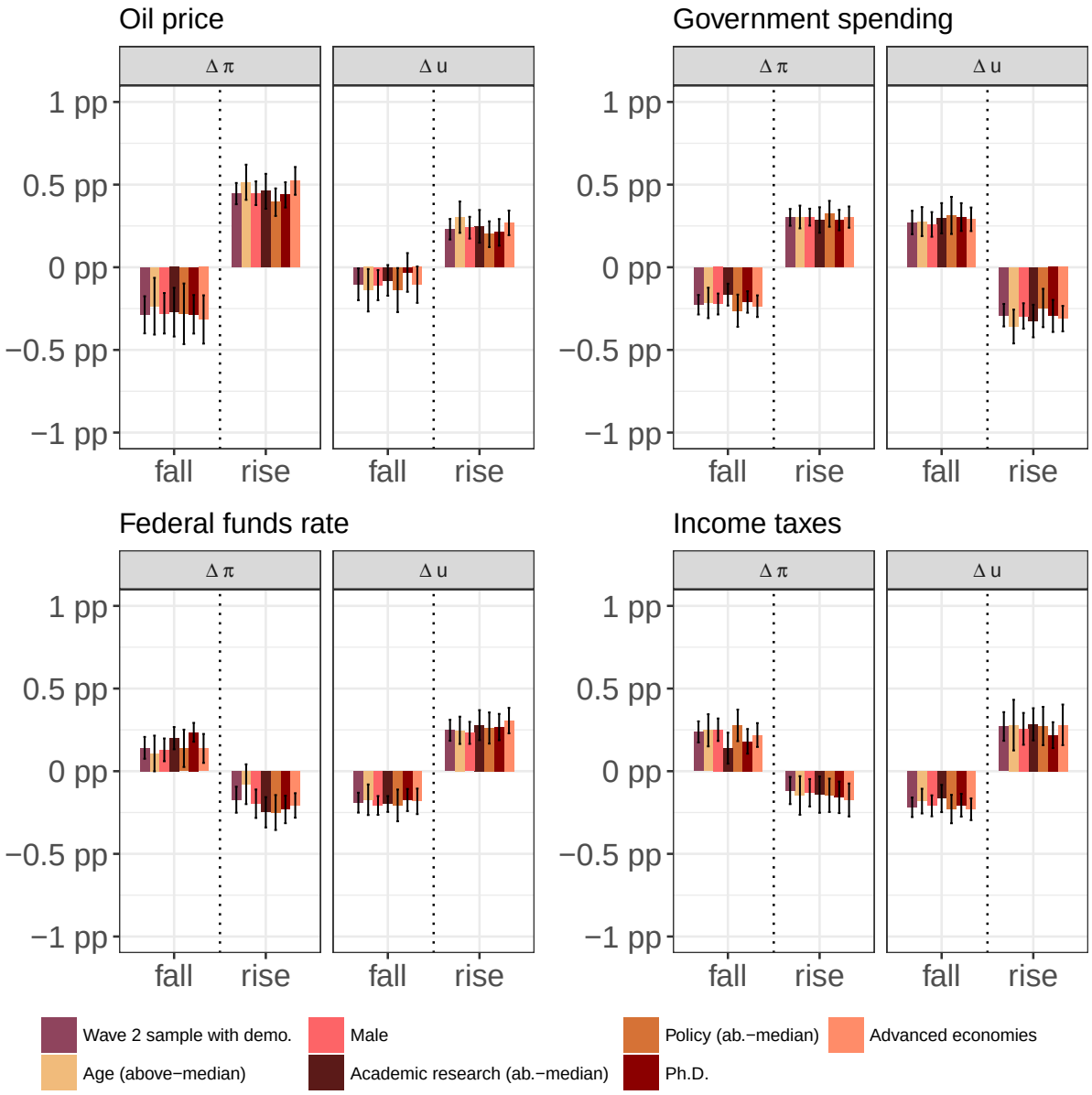
Notes: This figure analyzes the distribution of responses to the subjective interest rate rule questions in Wave 2 of the general population sample. Respondents were asked to estimate the likelihood of different federal funds target rate changes in response to a 0.25 pp. increase in the Fed's outlook for the inflation rate or the unemployment rate. For each possible federal funds target rate reaction, the "Mean plot" summarizes the average probability assigned to this event (averaged across individuals). The histogram plots the distribution of individual-level expected changes in the federal funds target rate in response to increases in the Fed's outlook for inflation or unemployment (integrated for each individual).

Figure A.7: Procedural robustness of results (with 95% CI)



Notes: This figure provides an overview of procedural robustness checks that repeat the main analysis for different subsamples. It plots predicted changes in the unemployment rate and the inflation rate for the "rise" and "fall" scenarios for each of the different vignettes separately. Error bars show 95% confidence intervals using robust standard errors. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta \pi$ denotes the expected change in the inflation rate compared to the baseline scenario. "Full sample" denotes the full sample and, thus, replicates the results of figure 2. "First vignette" contains only the responses to the first vignette, while "First question" focuses only on responses to the first forecast question (in both vignettes). "Survey duration" excludes both 10% tails in the survey duration distribution, and "Similar baseline prediction" excludes the 20% respondents with the largest absolute difference in baseline predictions across the two vignettes they responded to.

Figure A.8: Heterogeneity in experts' forecasts (with 95% CI)



Notes: This figure provides an overview of forecast heterogeneity for the expert wave 2 sample for which data on background characteristics are available ($n = 596$). It repeats the main analysis for different subsamples and plots expected changes in the unemployment rate and the inflation rate for the "rise" and "fall" scenarios for each of the different vignettes separately. Error bars show the 95% confidence intervals. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta \pi$ denotes the expected change in the inflation rate compared to the baseline scenario. "Wave 2 sample with demo." denotes the full sample that satisfies the conditions mentioned above. "Age (above-median)" contains only respondents with above-median age. "Male" contains only male respondents. "Academic research (ab.-median)" focuses on respondents that spend an above-median percentage of their working time on academic research, while "Policy (ab.-median)" restricts the sample to those who do an above-median amount of policy work. "Ph.D." contains only respondents with a Ph.D., and "Advanced economies" contains only respondents that are registered at the WES to make forecasts about an advanced economy (as classified by the IMF).

B Additional tables

Table A1: Overview of experiments

Experiment	**Sample**	**Treatments Arms**	**Mechanism Questions**
Consumers Wave 1 (February/March 2019) (N=1,063)	Representative online panel (in terms of age, income, region, gender, and education) in collaboration with Research Now	Hypothetical vignettes on oil supply shock, government spending shock, interest rate shock and tax shock	Beliefs about propagation mechanisms, financial literacy
Consumers Wave 2 (July 2019) (N=1,151)	Representative online panel (in terms of age, income, region, gender, and education) in collaboration with Lucid	Hypothetical vignettes on oil supply shock, government spending shock, interest rate shock and tax shock	Good-bad heuristic, rational inattention, numeracy, beliefs about supply-side mechanisms, subjective interest rate rule
Experts Wave 1 (February/March 2019) (N=179)	Experts recruited via email invitation (for details see Section E)	Hypothetical vignettes on oil supply shock, government spending shock, interest rate shock and tax shock	None
Experts Wave 2 (July 2019) (N=908)	Experts recruited via the ifo World Economic Survey	Hypothetical vignettes on oil supply shock, government spending shock, interest rate shock and tax shock	None

Notes: This table provides an overview of the different conducted experiments.

Table A2: Summary statistics: Covariates in the general population sample

Variable	**ACS (2017)**	**Rep. online panel**
Female	51%	55%
Age	47.43	46.41
At least bachelor's degree	30%	32%
Household net income (median)	65700	62500
Northeast	18%	21%
Midwest	21%	22%
South	38%	41%
West	24%	16%

Notes: This table compares the distributions of individual characteristics in our sample with those in the American Community Survey (ACS) 2017.

Table A3: Summary statistics: Covariates in the expert sample

Variable	**Wave 1**	**Wave 2 (WES)**
Female	26%	14%
Age (median)		52
Policy institution	16%	16%
Academia	83%	56%
Bank or company	0%	16%
Full professor	21%	
PhD student	18%	
Field of study: economics		84%
Field of study: business		7%
Ph.D.		65%
Region: Western Europe		42%
Region: Eastern Europe		12%
Region: CIS		7%
Region: North America		8%
Region: Latin America		10%
Region: Africa		7%
Region: Near East		2%
Region: Asia		10%
Region: Oceania		2%

Notes: This table provides an overview of the covariates in the expert sample. Different covariates were collected in the two waves. Moreover, demographic data are not available for all respondents.

Table A4: Heterogeneity among experts

	$\Delta\pi$✓	Δu✓	all✓
	(1)	(2)	(3)
age (ab.-median)	−0.031	0.028	−0.002
	(0.029)	(0.030)	(0.024)
female	−0.043	0.001	−0.021
	(0.049)	(0.044)	(0.035)
policy (ab.-median)	0.066**	0.057*	0.062**
	(0.030)	(0.030)	(0.024)
research (ab.-median)	0.006	0.028	0.017
	(0.034)	(0.034)	(0.028)
Ph.D.	0.089**	0.013	0.051*
	(0.035)	(0.035)	(0.029)
advanced economy	0.006	0.059*	0.033
	(0.031)	(0.031)	(0.025)
avg. confidence (ab.-median)	−0.037	−0.039	−0.038
	(0.031)	(0.033)	(0.026)
Observations	525	525	525
R^2	0.036	0.018	0.031

Notes: This table displays predictors of benchmark-consistent forecasts among experts from the wave 2 expert sample that responded to all forecast questions and for whom data on background characteristics is available (n=525). A forecast change is viewed as benchmark-consistent if it is directionally in line with the median response and the macroeconomic benchmark estimates. For each expert, $\Delta\pi$✓measures the fraction of benchmark-consistent inflation forecasts (out of 2), Δu✓the fraction of benchmark-consistent unemployment forecasts (out of 2), and *all*✓the overall fraction of benchmark-consistent forecasts (out of 4). Thus, the coefficients can be interpreted as the effects on the probability of a benchmark-consistent forecast. The predictors include above-median dummies in age, time spent on policy work, time spent on research work, and average expressed confidence as well as indicators for female respondents, respondents with a Ph.D. and respondents from advanced economies (IMF classification). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A5: Do US experts make more forecasts that are more consistent with models?

	$\Delta\pi$✓	Δu✓	all✓
	(1)	(2)	(3)
US	−0.071	−0.043	−0.057
	(0.044)	(0.041)	(0.035)
Constant	0.758***	0.717***	0.738***
	(0.012)	(0.012)	(0.010)
Observations	862	862	862
R^2	0.003	0.001	0.003

Notes: This Table uses data from all experts from Wave 1 and Wave 2 that responded to all forecast questions and for whom we know whether they are based in the US or not. This table analyzes whether experts from the US make more benchmark-consistent forecasts in the hypothetical vignettes about the US economy. A predicted change is viewed as benchmark-consistent if it is directionally in line with the median response and the macroeconomic benchmark estimates. For each expert, $\Delta\pi$✓measures the fraction of benchmark-consistent inflation forecasts (out of 2), Δu✓the fraction of benchmark-consistent unemployment forecasts (out of 2), and *all*✓the overall fraction of benchmark-consistent forecasts (out of 4). Thus, the coefficients can be interpreted as the effect on the probability of a benchmark-consistent forecast. An expert is classified as "US" if he or she works at a US institution (Wave 1, $n = 33$) or is registered to make regular forecasts about the US economy in the ifo World economic survey (Wave 2, $n = 52$). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A6: Disagreement in predicted changes in unemployment and inflation

	$\Delta\pi$			Δu		
	$\sigma_{experts}$	$\sigma_{gen.\ pop.}$	p	$\sigma_{experts}$	$\sigma_{gen.\ pop.}$	p
oil price – rise	0.28	0.74	<0.01	0.27	0.64	<0.01
oil price – fall	0.32	0.71	<0.01	0.28	0.69	<0.01
gov. spend. – rise	0.22	0.54	<0.01	0.27	0.61	<0.01
gov. spend. – fall	0.20	0.61	<0.01	0.24	0.63	<0.01
fed. funds rate – rise	0.31	0.52	<0.01	0.27	0.55	<0.01
fed. funds rate – fall	0.28	0.59	<0.01	0.25	0.63	<0.01
inc. taxes – rise	0.29	0.52	<0.01	0.26	0.55	<0.01
inc. taxes – fall	0.25	0.58	<0.01	0.27	0.56	<0.01
weighted mean	0.27	0.60		0.26	0.61	

Notes: This table reports the standard deviations of expert forecasts and of forecasts from the general population as well as p-values from a Levene's test of equality of variance (trimmed, median-based, bootstrapped) for each rise or fall scenario. The data is trimmed by 0.05 percentage points on both tails to reduce the influence of outliers. The last row presents the average within-scenario standard deviation, weighted by the differential number of respondents across scenarios.

Table A7: Understanding of propagation mechanisms and benchmark-consistent forecasts

Panel A: For each vignette

	oil price			gov. spending			fed. funds rate			income taxes		
	$\Delta\pi$✓	Δu✓	both✓	$\Delta\pi$✓	Δu✓	both✓	$\Delta\pi$✓	Δu✓	both✓	$\Delta\pi$✓	Δu✓	both✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
MS	0.120***	0.079***	0.090***	0.058**	0.057**	0.039**	−0.008	0.070***	0.030**	0.031	0.052*	0.038**
	(0.022)	(0.024)	(0.022)	(0.026)	(0.025)	(0.019)	(0.024)	(0.024)	(0.014)	(0.024)	(0.027)	(0.017)
fin. lit.	0.050**	0.060**	0.083***	−0.012	0.013	0.013	0.001	−0.027	−0.018	0.003	−0.011	−0.006
	(0.023)	(0.025)	(0.025)	(0.026)	(0.025)	(0.020)	(0.024)	(0.026)	(0.017)	(0.024)	(0.026)	(0.018)
Constant	0.712***	0.596***	0.470***	0.553***	0.423***	0.212***	0.293***	0.492***	0.113***	0.333***	0.514***	0.156***
	(0.019)	(0.021)	(0.021)	(0.022)	(0.022)	(0.018)	(0.020)	(0.021)	(0.014)	(0.020)	(0.022)	(0.016)
Obs.	521	521	521	508	508	508	543	543	543	532	532	532
R^2	0.120	0.066	0.097	0.012	0.019	0.015	0.000	0.014	0.006	0.004	0.008	0.008

Panel B: Pooled over all vignettes

	$\Delta\pi$✓	Δu✓	both✓	all✓	$\Delta\pi$✓	Δu✓	both✓	all✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MS	0.047***	0.067***	0.049***	0.057***	0.040***	0.061***	0.050***	0.051***
	(0.012)	(0.013)	(0.010)	(0.009)	(0.015)	(0.015)	(0.011)	(0.011)
fin. lit.	0.010	0.008	0.017	0.009	0.009	−0.000	0.013	0.005
	(0.013)	(0.013)	(0.011)	(0.010)	(0.015)	(0.016)	(0.013)	(0.012)
Constant	0.468***	0.506***	0.235***	0.487***	0.402**	0.421**	0.111	0.411***
	(0.011)	(0.011)	(0.009)	(0.008)	(0.180)	(0.196)	(0.160)	(0.135)
Add. ctrl?	–	–	–	–	✓	✓	✓	✓
Obs.	1,052	1,052	1,052	1,052	870	870	870	870
R^2	0.022	0.036	0.037	0.055	0.022	0.043	0.042	0.052

Notes: This table analyzes the relationship between responding correctly to the propagation mechanism questions and making benchmark-consistent vignette forecasts. A vignette forecast is viewed as benchmark-consistent if it is directionally in line with the median expert response. MS is the standardized number of correctly answered mechanism questions. $fin.\ lit.$ is the standardized number of correct answers in the standard three-item financial literacy test. The coefficients can be read as the percentage point increase in benchmark-consistent forecasts associated with a 1 standard deviation increase in MS (or $fin.\ lit.$).

Panel A conducts the analyze for each vignette separately. The outcome variables ($\Delta\pi$✓, Δu✓, both✓) are binary and indicate whether a vignette forecast (for the inflation rate, the unemployment rate, or both rates jointly) is benchmark-consistent. Panel B performs an analysis pooled over all vignettes. Here, $\Delta\pi$✓measures the fraction of benchmark-consistent inflation forecasts (out of 2), Δu✓the fraction of benchmark-consistent unemployment forecasts (out of 2), *both*✓the fraction of vignettes in which both forecasts are benchmark-consistent (out of 2), and *all*✓the overall fraction of benchmark-consistent forecasts (out of 4). Controls (Panel B) include age, education, log income, net wealth (inverse hyperbolic sine transformation), log financial wealth, gender and news consumption. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A8: Effect of having a supply-side view on vignette forecasts

Panel A: The effect of having a supply-side view on forecasts

	fed. funds rate		income taxes		both vignettes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)
fall	−0.195	0.102	−0.027	−0.154	−0.088	−0.061
	(0.142)	(0.151)	(0.104)	(0.123)	(0.086)	(0.098)
rise	0.197*	0.152	0.162	0.297**	0.178*	0.232**
	(0.104)	(0.135)	(0.168)	(0.143)	(0.099)	(0.106)
fall×int supply	0.001	−0.148				
	(0.163)	(0.169)				
rise×int supply	−0.010	0.034				
	(0.126)	(0.151)				
fall×tax supply			−0.217	−0.271*		
			(0.135)	(0.159)		
rise×tax supply			0.070	0.058		
			(0.183)	(0.162)		
fall×supply					−0.130	−0.165
					(0.102)	(0.115)
rise×supply					0.031	0.034
					(0.110)	(0.119)
Obs.	557	557	571	571	1,128	1,128
R^2	0.031	0.016	0.031	0.072	0.030	0.036

Panel B: The effect of having a supply-side view on the consistency of forecasts with the benchmarks

	fed. funds rate			income taxes			both vignettes		
	$\Delta\pi$✓	Δu✓	all✓	$\Delta\pi$✓	Δu✓	all✓	$\Delta\pi$✓	Δu✓	all✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
int supply	−0.008	0.003	−0.003						
	(0.048)	(0.051)	(0.031)						
tax sup.				−0.116***	−0.001	−0.058**			
				(0.043)	(0.044)	(0.029)			
supply							−0.070**	−0.006	−0.038*
							(0.032)	(0.034)	(0.021)
Constant	0.317***	0.520***	0.419***	0.401***	0.577***	0.489***	0.367***	0.554***	0.461***
	(0.042)	(0.045)	(0.027)	(0.036)	(0.037)	(0.024)	(0.027)	(0.029)	(0.018)
Obs.	557	557	557	571	571	571	1,128	1,128	1,128
R^2	0.000	0.000	0.000	0.013	0.000	0.007	0.004	0.000	0.003

Notes: This table reports regressions that test whether having a supply-side view of the economy affects vignette forecasts. *int/tax supply* are dummies taking value one if the respondent thinks that the shock (fed. funds rate / income taxes) increases production costs *and* that firms pass this on to consumers. Columns 5 and 6 in panel A as well as 7 to 9 in panel B report estimates with data pooled over both vignettes and standard errors that are clustered on the respondent level. Here, *supply* indicates a supply-side view for the vignette at hand.

Panel A analyzes the effect of having a supply-side view on the predicted changes of the inflation and unemployment rate. Panel B analyzes the effect on making benchmark-consistent forecasts. A vignette forecast is viewed as benchmark-consistent if it is in line with the median expert response or the VAR/DSGE benchmark. The outcome variables $\Delta\pi$✓and Δu✓are binary and indicate whether a vignette forecast (for the inflation rate, the unemployment rate) is benchmark-consistent. *all*✓ takes value zero if no forecast in a vignette is benchmark-consistent, 0.5 if exactly one forecast is benchmark-consistent, and 1 if both forecasts are benchmark-consistent. Thus, the coefficients can be interpreted as the effect of having a supply-side view on the probability of making a benchmark-consistent forecast. Robust standard errors clustered at the respondent level are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A9: Effect of rational (in)attention on benchmark-consistent forecasts

	Only inflation forecast			Only unemployment forecast			Pooled		
	$\Delta\pi$✓	$\Delta\pi$✓	$\Delta\pi$✓	Δu✓	Δu✓	Δu✓	all✓	all✓	all✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
π relevant	0.035*			0.010			0.022		
	(0.021)			(0.022)			(0.015)		
u relevant		0.030			−0.022			0.004	
		(0.021)			(0.022)			(0.015)	
index			0.055***			−0.019			0.018
			(0.021)			(0.022)			(0.015)
Constant	0.452***	0.461***	0.444***	0.540***	0.554***	0.556***	0.496***	0.507***	0.500***
	(0.017)	(0.013)	(0.015)	(0.017)	(0.014)	(0.016)	(0.011)	(0.009)	(0.011)
Obs.	1,124	1,124	1,124	1,124	1,124	1,124	1,124	1,124	1,124
R^2	0.002	0.002	0.006	0.000	0.001	0.001	0.002	0.000	0.001

Notes: This table reports regressions of making benchmark-consistent forecasts on three measures of respondents' perceived relevance of macroeconomic variables for their own situation. A vignette forecast is viewed as benchmark-consistent if it is directionally in line with the median expert response. For each individual, $\Delta\pi$✓measures the fraction of benchmark-consistent inflation forecasts (out of 2), Δu✓the fraction of benchmark-consistent unemployment forecasts (out of 2), and *all*✓the overall fraction of benchmark-consistent forecasts (out of 4). π *relevant* is a dummy that takes value 1 if a respondent agrees or strongly agrees with the statement "The US inflation rate is relevant for my own economic situation.". u *relevant* is an analogous variable for the unemployment rate. *index* is a combined binary measure that takes value one for respondents with above-median agreement to the two statements. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A10: Who uses the Good-Bad Heuristic?

	Follows GBH
age (above-median)	0.058**
	(0.023)
female	0.003
	(0.020)
college+	−0.002
	(0.021)
income (above-median)	−0.031
	(0.022)
net wealth (above-median)	0.016
	(0.022)
numeracy (above-median)	0.030
	(0.021)
Constant	0.466***
	(0.023)
Observations	895
R^2	0.015

Notes: This table analyzes who uses the Good-Bad Heuristic (GBH). The outcome variable is the fraction of an individual's forecasts that are in line with his/her GBH (out of four forecasts in total). Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A11: Heterogeneity in affective evaluations

	inflation	unemployment	oil price	gov. spending	fed. funds rate	income taxes
	(1)	(2)	(3)	(4)	(5)	(6)
female	−0.053**	−0.063***	−0.019	−0.096***	−0.011	−0.060**
	(0.023)	(0.021)	(0.023)	(0.030)	(0.028)	(0.027)
age	−0.091***	−0.078***	−0.087***	−0.065**	−0.096***	−0.104***
	(0.019)	(0.016)	(0.019)	(0.031)	(0.028)	(0.024)
college+	−0.006	−0.001	0.023	0.047	0.049*	0.032
	(0.024)	(0.020)	(0.025)	(0.032)	(0.029)	(0.031)
econ. coll.+	0.020	0.024	0.030	0.050	0.049	0.025
	(0.024)	(0.020)	(0.024)	(0.032)	(0.030)	(0.031)
numeracy	−0.096***	−0.089***	−0.071***	0.029	−0.046	−0.044*
	(0.019)	(0.017)	(0.021)	(0.031)	(0.028)	(0.027)
log income	0.050***	0.067***	0.060***	0.021	0.067***	0.016
	(0.016)	(0.014)	(0.017)	(0.022)	(0.019)	(0.020)
log real est. wealth	−0.002	−0.003	−0.003	−0.003	−0.003	−0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
log debt	−0.008***	−0.005**	−0.011***	−0.003	−0.010***	−0.009***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
log fin. wealth	−0.005*	−0.005*	−0.004	−0.002	0.002	−0.004
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
rat. inatt.	0.004	−0.017	−0.028	0.014	0.009	−0.060**
	(0.021)	(0.018)	(0.021)	(0.028)	(0.027)	(0.026)
democrat	0.006	0.011	0.002	0.104***	0.018	0.064**
	(0.021)	(0.019)	(0.022)	(0.029)	(0.027)	(0.026)
avg. conf.	0.089***	0.082***	0.108***	0.050*	0.068**	0.082***
	(0.021)	(0.020)	(0.022)	(0.029)	(0.028)	(0.027)
Constant	−0.284*	−0.506***	−0.396**	−0.014	−0.487**	0.122
	(0.167)	(0.139)	(0.174)	(0.222)	(0.199)	(0.214)
Observations	895	895	895	895	895	895
R^2	0.127	0.136	0.140	0.066	0.084	0.097

Notes: This table studies heterogeneity in affective evaluations of the six variables (inflation, unemployment, oil price, government spending, federal funds rate and income tax rates). The dependent variables are indicators that take value 1 if a variable is evaluated positively (that is, the average rating on the two scales from -3 (very bad) to 3 (very good) for the two subitems (personal and economy-wide evaluation) is strictly positive). Explanatory variables include indicators for being female, above-median age, college education (or higher), economics education on college level (or higher), an above-median numeracy score, an above-median rational inattention score, being a Democrat (versus Republican or independent), and an above-median average forecast confidence. Moreover, log income, log real estate wealth, log debt, and log financial wealth are included, where 1 is added to include zeros. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A12: Political heterogeneity in forecasts

Panel A: Political heterogeneity in forecasts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.362^{***}	-0.117	-0.253^{***}	0.143^{**}	-0.139^{*}	-0.054	-0.160^{**}	-0.190^{**}
	(0.072)	(0.083)	(0.069)	(0.066)	(0.082)	(0.075)	(0.067)	(0.079)
rise	0.614^{***}	0.448^{***}	0.147^{**}	-0.026	0.083	0.166^{**}	0.232^{***}	0.381^{***}
	(0.089)	(0.084)	(0.070)	(0.073)	(0.066)	(0.073)	(0.057)	(0.070)
fall×democrat	0.119	-0.108	0.006	-0.130	0.001	0.087	0.025	-0.150
	(0.109)	(0.110)	(0.110)	(0.104)	(0.115)	(0.106)	(0.101)	(0.124)
rise×democrat	0.118	-0.001	-0.114	0.006	0.208^{**}	0.089	-0.062	-0.201^{**}
	(0.125)	(0.110)	(0.095)	(0.103)	(0.089)	(0.097)	(0.089)	(0.095)
Joint F-test does not detect a significant effect of democrat. p = 0.211								
Observations	839	839	827	827	834	834	856	856
R^2	0.161	0.086	0.033	0.005	0.030	0.023	0.032	0.060

Panel B: Political heterogeneity in consistency of forecasts with the benchmarks

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$✓	Δu✓	$\Delta\pi$✓	Δu✓	$\Delta\pi$✓	Δu✓	$\Delta\pi$✓	Δu✓
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
democrat	-0.033	0.012	0.001	0.013	-0.046	-0.004	0.015	-0.006
	(0.031)	(0.034)	(0.035)	(0.035)	(0.032)	(0.035)	(0.032)	(0.034)
Constant	0.728^{***}	0.608^{***}	0.543^{***}	0.432^{***}	0.332^{***}	0.522^{***}	0.314^{***}	0.549^{***}
	(0.023)	(0.025)	(0.025)	(0.024)	(0.023)	(0.025)	(0.023)	(0.024)
Joint F-test does not detect a significant effect of democrat. p = 0.823								
Observations	839	839	827	827	834	834	856	856
R^2	0.001	0.000	0.000	0.000	0.002	0.000	0.000	0.000

Notes: This table provides an overview of political heterogeneity in forecasts in the representative sample. Respondents for whom data on political affiliation is missing or who respond that they are Independent are excluded. Panel A displays predicted changes in unemployment and inflation for each of the different vignettes separately. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the expected change in the inflation rate compared to the baseline scenario. Panel B displays differences in making benchmark-consistent forecasts. The outcome variables ($\Delta\pi$✓, Δu✓, both✓) are binary and indicate whether a vignette forecast (for the inflation rate, the unemployment rate, or both rates jointly) is benchmark-consistent. A vignette forecast is viewed as benchmark-consistent if it is directionally in line with the median expert response. Both panels report a joint F-test that results from Seemingly Unrelated Regressions (SUR) with respondent-level clustered standard errors and tests for an overall zero effect of *democrat*. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A13: Confidence and making benchmark-consistent forecasts

Panel A: Experts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
confidence	0.026	0.042	−0.023	0.035	−0.017	0.027	−0.011	0.050*
	(0.023)	(0.030)	(0.025)	(0.026)	(0.030)	(0.026)	(0.029)	(0.027)
Constant	0.846***	0.660***	0.799***	0.797***	0.665***	0.726***	0.681***	0.700***
	(0.016)	(0.022)	(0.019)	(0.018)	(0.021)	(0.020)	(0.021)	(0.020)
Observations	481	478	474	472	516	509	514	513
R^2	0.003	0.004	0.002	0.004	0.001	0.002	0.000	0.007

Panel B: General population

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$	$\Delta\pi\checkmark$	$\Delta u\checkmark$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
confidence	0.002	−0.027	−0.024	−0.011	0.019	0.037**	0.036**	0.024
	(0.016)	(0.017)	(0.016)	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)
Constant	0.705***	0.618***	0.554***	0.435***	0.303***	0.507***	0.325***	0.547***
	(0.014)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.000	0.003	0.002	0.000	0.001	0.005	0.005	0.002

Panel C: Determinants of confidence in general population

	avg. confidence	avg. confidence: oil price	avg. confidence: gov. spending	avg. confidence: fed. funds rate	avg. confidence: income taxes
	(1)	(2)	(3)	(4)	(5)
female	−0.381***	−0.334***	−0.440***	−0.442***	−0.320***
	(0.037)	(0.053)	(0.055)	(0.054)	(0.056)
age	−0.002*	−0.000	−0.002	0.000	−0.005***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
std. ln(inc.)	0.055***	0.073***	0.048	0.077***	0.022
	(0.019)	(0.028)	(0.030)	(0.028)	(0.029)
college	0.004	−0.024	0.033	−0.023	0.024
	(0.038)	(0.055)	(0.058)	(0.056)	(0.058)
Constant	0.351***	0.250***	0.409***	0.283***	0.474***
	(0.064)	(0.094)	(0.096)	(0.093)	(0.100)
Observations	2,140	1,064	1,055	1,075	1,086
R^2	0.057	0.046	0.068	0.070	0.039

Notes: This table assesses the role of confidence in the predictions across vignettes. A forecast is classified as benchmark-consistent if it follows the same qualitative direction as the median expert forecast. Confidence is measured on a 5-point scale reaching from -2 to 2. Panel A shows how confidence affects making benchmark-consistent predictions among experts. Panel B shows how confidence affects making benchmark-consistent forecasts in the representative sample. Panel C displays the determinants of confidence (averaged over confidence in inflation and unemployment forecasts) in the representative sample separately for each shock. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A14: Robustness: Wave effects

Panel A: Experts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.502^{***}	-0.168^{**}	-0.283^{***}	0.337^{***}	0.165^{**}	-0.156^{**}	0.198^{***}	-0.228^{***}
	(0.091)	(0.077)	(0.032)	(0.055)	(0.068)	(0.074)	(0.072)	(0.045)
rise	0.496^{***}	0.294^{***}	0.369^{***}	-0.336^{***}	-0.255^{***}	0.364^{***}	-0.181^{***}	0.260^{***}
	(0.057)	(0.056)	(0.067)	(0.062)	(0.054)	(0.044)	(0.054)	(0.046)
fall×wave 2	0.216^{**}	0.047	0.075^{*}	-0.043	-0.012	-0.038	0.013	-0.008
	(0.103)	(0.087)	(0.043)	(0.062)	(0.074)	(0.078)	(0.080)	(0.055)
rise×wave 2	-0.059	-0.075	-0.082	0.029	0.124^{*}	-0.092^{*}	0.091	-0.047
	(0.066)	(0.067)	(0.070)	(0.069)	(0.066)	(0.053)	(0.067)	(0.063)
Joint F-test does not detect a significant effect of wave. p = 0.305								
Observations	482	481	474	475	517	513	515	521
R^2	0.341	0.122	0.378	0.353	0.100	0.274	0.095	0.164

Panel B: General Population

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.319^{***}	-0.244^{***}	-0.224^{***}	0.075	-0.121^{*}	-0.038	-0.058	-0.170^{***}
	(0.069)	(0.068)	(0.058)	(0.061)	(0.067)	(0.062)	(0.050)	(0.057)
rise	0.773^{***}	0.419^{***}	0.100	-0.048	0.184^{***}	0.161^{***}	0.191^{***}	0.263^{***}
	(0.078)	(0.062)	(0.062)	(0.059)	(0.044)	(0.061)	(0.052)	(0.059)
fall×wave 2	-0.022	0.062	-0.072	-0.062	-0.057	0.019	-0.118	-0.147
	(0.100)	(0.098)	(0.089)	(0.089)	(0.096)	(0.091)	(0.083)	(0.098)
rise×wave 2	-0.212^{**}	0.051	0.065	0.047	0.017	0.025	0.029	0.069
	(0.106)	(0.095)	(0.083)	(0.091)	(0.073)	(0.086)	(0.086)	(0.089)
Joint F-test does not detect a significant effect of wave. p = 0.620								
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.162	0.086	0.043	0.002	0.029	0.014	0.029	0.058

Notes: This table provides an overview of differences in forecasts across survey waves. Panel A uses data from the expert sample. Panel B uses data from the general population sample. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the expected change in the inflation rate compared to the baseline scenario. *wave* 2 is an indicator taking value one if a respondent participates in the second wave of the data collection. Both panels report a joint F-test that results from SUR regressions with respondent-level clustered standard errors and tests for an overall zero effect of *wave* 2. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A15: Robustness: Vignette and question order effects

Panel A: Order effects on forecasts of experts

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.670^{***}	-0.109	-0.323^{***}	0.384^{***}	0.150	-0.264^{**}	0.227^{**}	-0.270^{***}
	(0.108)	(0.088)	(0.053)	(0.084)	(0.113)	(0.123)	(0.112)	(0.087)
rise	0.588^{***}	0.115	0.364^{***}	-0.160^{*}	-0.335^{***}	0.331^{***}	-0.223^{***}	0.219^{***}
	(0.095)	(0.105)	(0.093)	(0.085)	(0.118)	(0.096)	(0.065)	(0.062)
fall × u. first	0.133	-0.070	-0.038	-0.024	-0.043	0.139	-0.137	0.067
	(0.195)	(0.159)	(0.063)	(0.117)	(0.129)	(0.133)	(0.167)	(0.096)
rise × u. first	-0.148	0.093	-0.059	-0.176	0.079	0.061	0.290^{**}	-0.096
	(0.116)	(0.107)	(0.140)	(0.139)	(0.121)	(0.099)	(0.123)	(0.114)
fall × 2nd vig.	0.196	-0.045	0.113^{*}	-0.067	0.079	0.071	0.104	0.013
	(0.198)	(0.164)	(0.064)	(0.116)	(0.130)	(0.134)	(0.162)	(0.092)
rise × 2nd vig.	-0.029	0.277^{***}	0.063	-0.224^{*}	0.106	0.010	-0.174	0.143
	(0.118)	(0.107)	(0.147)	(0.134)	(0.116)	(0.097)	(0.109)	(0.106)
Joint F-tests.								
u. first	p = 0.521							
2nd vig.	p = 0.122							
Observations	95	95	85	85	85	85	90	90
R^2	0.513	0.272	0.519	0.473	0.236	0.342	0.222	0.412

Panel B: Order effects on forecasts of general population

	oil price		gov. spending		fed. funds rate		income taxes	
	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
fall	-0.487^{***}	-0.192^{**}	-0.391^{***}	0.001	-0.325^{***}	-0.025	-0.136^{*}	-0.293^{***}
	(0.092)	(0.081)	(0.075)	(0.080)	(0.079)	(0.069)	(0.081)	(0.100)
rise	0.761^{***}	0.464^{***}	0.103	-0.131^{*}	0.231^{***}	0.159^{**}	0.059	0.293^{***}
	(0.097)	(0.091)	(0.088)	(0.080)	(0.065)	(0.068)	(0.079)	(0.075)
fall × u. first	0.117	-0.049	0.361^{***}	0.200^{**}	0.071	-0.122	-0.015	-0.044
	(0.099)	(0.099)	(0.089)	(0.089)	(0.097)	(0.091)	(0.086)	(0.102)
rise × u. first	-0.142	0.001	0.013	0.108	-0.063	-0.028	0.092	-0.078
	(0.107)	(0.095)	(0.082)	(0.093)	(0.073)	(0.085)	(0.085)	(0.089)
fall × 2nd vig.	0.192^{*}	0.011	-0.113	-0.122	0.292^{***}	0.130	0.045	0.139
	(0.100)	(0.099)	(0.088)	(0.089)	(0.096)	(0.091)	(0.085)	(0.100)
rise × 2nd vig.	-0.047	-0.042	0.049	0.111	-0.015	0.058	0.201^{**}	0.082
	(0.106)	(0.095)	(0.085)	(0.092)	(0.073)	(0.085)	(0.085)	(0.089)
Joint F-tests.								
u. first	p = 0.007							
2nd vig.	p = 0.066							
Observations	1,099	1,099	1,085	1,085	1,123	1,123	1,121	1,121
R^2	0.164	0.086	0.059	0.010	0.040	0.018	0.033	0.059

Notes: This table provides an overview of vignette and question order effects on the forecasts. Panel A uses data from Wave 1 of the expert sample. Panel B uses data from the general population sample. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the expected change in the inflation rate compared to the baseline scenario. *u. first* is an indicator taking value one if a respondent first answered the unemployment forecast question and then the inflation forecast question. *2nd vig.* is an indicator taking value one if the forecast corresponds to the second vignette that a respondent faced. Both panels report joint F-tests that result from Seemingly Unrelated Regressions (SUR) with respondent-level clustered standard errors and test for overall zero effects of *u. first* and *2nd vig.* Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A16: Robustness: Incentive effects

Panel A: Incentives

	$\Delta\pi\checkmark$	$\Delta u\checkmark$	both✓	all✓	time instructions	time vignettes
	(1)	(2)	(3)	(4)	(5)	(6)
incentives	0.044**	−0.000	0.038**	0.022	−0.537	38.589***
	(0.022)	(0.023)	(0.019)	(0.016)	(10.361)	(13.236)
Constant	0.447***	0.508***	0.216***	0.477***	112.689***	165.001***
	(0.015)	(0.017)	(0.013)	(0.011)	(9.261)	(6.490)
Observations	1,063	1,063	1,063	1,063	1,063	1,063
R^2	0.004	0.000	0.004	0.002	0.000	0.008

Panel B: Incentives crossed with subjective perception of expert accuracy

	$\Delta\pi\checkmark$	$\Delta u\checkmark$	both✓	all ✓	time instructions	time vignettes
	(1)	(2)	(3)	(4)	(5)	(6)
incentives	0.040*	0.001	0.038**	0.020	−1.128	38.692***
	(0.022)	(0.023)	(0.019)	(0.016)	(10.555)	(13.029)
exp. acc.	0.006	−0.015	0.012	−0.005	7.261	5.663
	(0.015)	(0.017)	(0.013)	(0.012)	(7.068)	(4.976)
incentives × exp. acc.	−0.017	−0.007	−0.029	−0.012	0.246	7.697
	(0.022)	(0.024)	(0.018)	(0.017)	(9.717)	(17.501)
Constant	0.449***	0.506***	0.217***	0.477***	113.222***	165.118***
	(0.015)	(0.017)	(0.013)	(0.011)	(9.502)	(6.577)
Observations	1,049	1,049	1,049	1,049	1,049	1,049
R^2	0.004	0.003	0.006	0.003	0.002	0.010

Panel C: Incentives for the mechanism questions

	total mechanism score	time mechanism questions
	(1)	(2)
incentives	0.027	13.637
	(0.066)	(14.873)
Constant	−0.019	234.530***
	(0.055)	(11.090)
Observations	1,063	1,063
R^2	0.000	0.001

Notes: This table provides an overview of the effect of monetary incentives on the response behavior of the general population in Wave 1. A forecast is classified as benchmark-consistent if it follows the same qualitative direction as the median expert forecast. Panel A displays the effect on the benchmark-consistency of forecasts and response times. *incentives* constitutes a binary variable that takes value one for incentivized respondets. For each individual, $\Delta\pi\checkmark$ measures the fraction of benchmark-consistent inflation forecasts (out of 2), $\Delta u\checkmark$ the fraction of benchmark-consistent unemployment forecasts (out of 2), *both✓* the fraction of vignettes in which both forecasts are benchmark-consistent (out of 2), and *all✓* the overall fraction of benchmark-consistent forecasts (out of 4). Thus, the coefficients can be interpreted as the effect of incentives on the probability of a benchmark-consistent forecast. Columns 5 and 6 show effects on the time spent reading the instructions and the total time spent on the vignettes. Panel B examines heterogeneity according to the respondents' perceived accuracy of experts (*exp. acc.*, standardized) to rule out that incentives might be ineffective merely because expert forecasts are perceived as inaccurate. Panel C studies the effect of incentives on response accuracy to the mechanism propagation questions. *total mechanism score* denotes the standardized number of correctly answered mechanism questions, *time mechanism questions* measures the time spent for answering the mechanism questions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A17: Misperceived endogeneity of interest rate shock

Panel A: Binary monetary policy reaction

	fed. funds rate			
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
fall	-0.358^{***}	-0.046	-0.407^{***}	-0.054
	(0.114)	(0.087)	(0.127)	(0.098)
rise	0.164	0.249^{***}	0.155	0.190
	(0.100)	(0.092)	(0.118)	(0.124)
fall$\times 1(\alpha > 0)$	0.257^{*}		0.216	0.017
	(0.150)		(0.147)	(0.148)
rise$\times 1(\alpha > 0)$	0.046		0.039	0.125
	(0.128)		(0.131)	(0.128)
fall$\times 1(\beta > 0)$		0.071	0.145	0.066
		(0.131)	(0.145)	(0.146)
rise$\times 1(\beta > 0)$		-0.113	0.028	-0.144
		(0.127)	(0.126)	(0.122)
Obs.	503	503	503	503
R^2	0.039	0.020	0.041	0.022

Panel B: Expected monetary policy reaction

	fed. funds rate			
	$\Delta\pi$	Δu	$\Delta\pi$	Δu
	(1)	(2)	(3)	(4)
fall	-0.304^{***}	-0.014	-0.309^{***}	-0.060
	(0.080)	(0.066)	(0.081)	(0.071)
rise	0.178^{***}	0.198^{***}	0.176^{**}	0.183^{***}
	(0.068)	(0.067)	(0.070)	(0.071)
fall$\times\alpha/4$	0.860^{**}		0.761^{**}	0.526^{*}
	(0.355)		(0.334)	(0.270)
rise$\times\alpha/4$	0.135		0.093	0.194
	(0.230)		(0.234)	(0.249)
fall$\times\beta/4$		0.061	0.234	-0.146
		(0.285)	(0.304)	(0.266)
rise$\times\beta/4$		-0.048	0.135	-0.114
		(0.231)	(0.262)	(0.241)
Obs.	503	503	503	503
R^2	0.052	0.018	0.054	0.026

Notes: This table reports regressions that test for misperception of the interest rate shock as an endogenous reaction of the Fed to a changed outlook in inflation. α denotes the coefficient on π^e in the Fed's linear forward-looking interest rate rule, and β denotes the coefficient on u^e. Δu denotes the expected change in the unemployment rate compared to the baseline scenario. $\Delta\pi$ denotes the expected change in the inflation rate compared to the baseline scenario. Panel A regresses both variables on $1(\alpha > 0)$ – a dummy taking value one if the respondent believes that the Fed would increase the federal funds target rate in response to an unexpected increase in the outlook for future inflation – and $1(\beta > 0)$ – a dummy taking value one if the respondent believes that the Fed would increase the federal funds target rate in response to an unexpected increase in the outlook for future unemployment. Panel B uses α and β which are the respondent's estimates of the coefficients in the forward-looking interest rate rule. They are divided by 4 because the inflation and unemployment outlook change by 0.25 percentage points (rather than 1 pp.) in the survey questions. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

C Details on the derivation of the theoretical and empirical benchmarks

In this section, we provide details on the sources, the assumptions, and calculations used to turn the empirical and theoretical evidence on each shock into comparable multipliers, as outlined in section 3.1. We use an Okun's Law coefficient of -0.4, based on Ball et al. (2017), which implies a 0.4 percentage point rise in unemployment associated to a 1 percent fall in output over the course of a year. Below, Δy indicates a percent fall in output over four quarters, and $\Delta\pi$ and Δu are the respective four quarter changes of inflation and the unemployment rate in percentage points.[5] In each case, the following five key steps are involved: 1) identifying the size of the shock in the source paper(s), 2) identifying the size of the response of the variables of interest in the source paper(s), 3) determining the size of the shock in the vignettes, 4) rescaling the shocks from the source papers to be of the same size as those from the vignettes, 5) translating output changes into unemployment changes when needed. All calculations contain a small degree of approximation.

Oil price Blanchard and Galí (2010) show that since 1984, a date conventionally considered as the beginning of the Great Moderation, the response of the US economy to oil price fluctuations has become milder. We thus derive our benchmark from the authors' post-1984 VAR results. As shown in Table 1, the benchmark unemployment rate change for an oil price rise of $30 is 0.4 to 0.45 percentage points. For inflation, we derive an empirical benchmark rise of 1.25 to 1.5 percentage points.

We choose two papers as theoretical references: Bodenstein et al. (2011) and Balke and Brown (2018). Both papers model the effect of shocks to oil supply outside the US. While the former paper models the US as a purely oil-importing country, the latter treats the US as both oil-producing and oil-importing, providing us with a theoretical benchmark effect ranging from 0.35 to 0.8 percentage points (see Table 1). Neither of these papers studies the impact of oil supply shocks on domestic inflation.

[5] In the case of government and tax shocks in the model of Galí et al. (2011), the responses of output and unemployment exhibited very low persistence, likely due to the specification of the shock process itself. We therefore opted for using the average change over for four quarters rather than the change in the fourth quarter only.

Oil price - Empirical Source: Blanchard and Galí (2010), Figure 1, Panel B (i.e. post-84). 1) Shock is 10% change in price. 2) $\Delta y = -0.2$, $\Delta \pi = 0.25$. 3) Size of shock in vignette 55 percent (Wave 2) or 56 percent (Wave 1) so we approximately multiply the original shocks by 5.5. 4) $\Delta y = -1.1$, $\Delta \pi = 1.4$. 5) Okun's Law: $\Delta u = 0.425$.

Oil price - Theory Source: Bodenstein et al. (2011), Figure 2. 1) Shock is 8% change in price 2) $\Delta y = -0.15$. 3) Size of shock in vignette 55 percent (Wave 2) 56 percent (Wave 1) so we approximately multiply the original shocks by 7. 4) $\Delta y = -1.05$ 5) Okun's Law: $\Delta u = 0.42$.

Source: Balke and Brown (2018), Figure 3. 1) Shock is 2.5% change in price 2) $\Delta y = -0.1$. 3) Size of shock in vignette 55 percent (Wave 2) 56 percent (Wave 1) so we approximately multiply the original shocks by 22. 4) $\Delta y = -2.2$ 5) Okun's Law: $\Delta u = 0.88$.

Government spending Regarding government spending, the growing body of works focusing narrowly on defense spending shocks (Auerbach et al., 2019; Basso and Rachedi, 2019; Nakamura and Steinsson, 2014) would in theory constitute the optimal comparison for the vignette. However, these studies compute fiscal multipliers at the local level (e.g., metro area or state), which are not necessarily applicable to the national level. We therefore refer to studies that examine the impact of spending at the national level and that utilize the same methodologies (i.e., VAR models) as the papers we consider for the other shocks. For the effect of government spending increases on unemployment, we compute an empirical reference range of -0.1 to -0.2 percentage points (Auerbach and Gorodnichenko, 2012; Blanchard and Perotti, 2002; Ramey, 2011). No results are available for the effect on inflation. On the theoretical side, we interpret the exogenous spending shock in Smets and Wouters (2007) and Galí et al. (2011) as a government spending shock. A third source is the government spending shock in Zubairy (2014).[6] The theoretical reference range of values for the change in unemployment after a rise in spending of 0.5 percent of GDP, reported in Table 1, is between -0.1 to -0.2 percentage points, while the benchmark rise in inflation is 0.15 to 0.2 percentage points.

[6]Note that we do not use this paper as a benchmark for the response of inflation. Although inflation dynamics resulting from fiscal policy are embedded in the model, they are not discussed in detail by the author.

Government spending - Empirical Source: Blanchard and Perotti (2002), Ramey (2011) and sources therein, Auerbach and Gorodnichenko (2012). 1) Shock is 1% of GDP 2) $\Delta y = 0.8$ to 1.5. 3) Size of shock in vignette is 2.4% of 4.2 trillion of government spending. US 2018 GDP is 20.89 trillion according to the Bureau of Economic Analysis, so the shock is about 2.4% of 20% of GDP, which is 0.5% of GDP. So we divide the original shock by 2. 4) $\Delta y = 0.4$ to 0.75. 5) Okun's Law: $\Delta u = -0.16$ to -0.3.

Government spending - Theory Source: Galí et al. (2011), Figure 3. 1) Size of shock is 0.47, with exogenous spending formulated in percent of output, so it can be interpreted as 0.5% of GDP. 2) $\Delta u = -0.1$, $\Delta \pi = 0.2$. 3) The shock in the vignette is very similar in size, so there is no need to scale it. 4) $\Delta u = -0.1$, $\Delta \pi = 0.2$.

Source: Smets and Wouters (2007), Figure 2. 1) Size of shock is 0.5, with exogenous spending formulated in percent of output, so it can be interpreted as 0.5% of GDP. 2) $\Delta y = 0.3$, $\Delta \pi = 0.15$. 3) The shock in the vignette is very similar in size, so there is no need to scale it. 4) $\Delta y = 0.3$, $\Delta \pi = 0.15$. 5) Okun's Law: $\Delta u = -0.12$.

Source: Zubairy (2014), Table 2. 1) Size of shock is 1% of GDP. 2) $\Delta y = 1$. 3) Divide by 2 to make it comparable to the vignette. 4) $\Delta y = 0.5$. 5) Okun's Law: $\Delta u = -0.2$.

Monetary policy Arias et al. (2019) gives an empirical benchmark effect of 0.2 percentage points on unemployment and 0.2 percentage points on inflation for our federal funds rate rise by 50 basis points. This is largely in line with a large and consistent body of VAR evidence since the late 1990's (Bernanke and Mihov, 1998; Bernanke et al., 2005; Christiano et al., 1999; Primiceri, 2005; Romer and Romer, 2004; Stock and Watson, 2001; Uhlig, 2005). As a theoretical reference, we again use Smets and Wouters (2007) and Galí et al. (2011) and arrive at a benchmark of 0.4 to 0.5 percentage points for unemployment and a benchmark of -0.15 percentage points for inflation.

Monetary policy - Empirical Source: Arias et al. (2019) Figure 5 (i.e. estimation on full post-WWII sample, imposing a zero restriction on the systematic response of monetary policy to commodity prices). 1) Shock size is 0.25 percentage points. 2) $\Delta y = -0.25$, $\Delta \pi = -0.1$. 3) To make the shock comparable to the vignette, we multiply by 2. 4) $\Delta y = -0.5$, $\Delta \pi = -0.2$. 5) Okun's Law: $\Delta u = 0.2$.

Monetary policy - Theory Source: Galí et al. (2011), Figure 3. 1) Size of shock is 0.15 percentage points 2) $\Delta u = -0.15$, $\Delta \pi = -0.05$. 3) We approximately multiply by 3.3 to make it comparable to the vignette. 4) $\Delta u = 0.5$, $\Delta \pi = -0.16$.

Source: Smets and Wouters (2007), Figure 2. 1) Size of shock is 0.175. 2) $\Delta y = -0.35$, $\Delta \pi = -0.05$. 3) We approximately multiply by 3 to make it comparable to the vignette. 4) $\Delta y = -1$, $\Delta \pi = -0.15$. 5) Okun's Law: $\Delta u = 0.4$.

Income tax rate The empirical benchmark for the unemployment change in response to the increase in the income tax rate by 1 percentage point on average ranges between 0.2 and 0.6 percentage points (Blanchard and Perotti, 2002; Favero and Giavazzi, 2012; Mertens and Ravn, 2012, 2014; Perotti, 2012; Romer and Romer, 2010). To our knowledge, the only paper modeling the impact of labor income tax rate fluctuations in a New Keynesian model is Zubairy (2014). For the theoretical benchmark of the effect on unemployment, we derive a value of 0.06.[7]

Tax rate change - Empirical Source: Blanchard and Perotti (2002), Romer and Romer (2010), Favero and Giavazzi (2012), Mertens and Ravn (2012, 2014), and Perotti (2012). 1) Shock size is a 1 percent of GDP increase in tax revenue. 2) Range of empirical output multipliers at 4 to 6 quarters is 1 to 3 percent of GDP. 3) The shock size in the vignette is approximately 0.5 percent of GDP. So we divide by 2 to make the shock comparable to the vignette. 4) $\Delta y = 0.5$ to 1.5. 5) Okun's Law: 0.2 to 0.6.

Tax rate change - Theory Source: Zubairy (2014), Table 2. 1) Size of shock is 1 percent of GDP. 2) $\Delta y = 0.32$. 3) Divide by 2 to make it comparable to the vignette. 4) $\Delta y = 0.15$. 5) Okun's Law: $\Delta u = -0.06$.

[7]Once again, we do not use this paper as a benchmark for the response of inflation. Although inflation dynamics resulting from fiscal policy are embedded in the model, they are not discussed in detail by the author.

D Details on the First-Order Conditions Impulse Response Functions

This exercise examines how the representative household's expectations of its optimal consumption and labor supply decisions differ from the rational-expectations general-equilibrium solution when its expectations for inflation and unemployment in response to a monetary shock resemble those we observe in the survey. The FOC IRFs that we compute should be interpreted neither as an equilibrium dynamic nor as the household's actual response to a monetary shock. Instead, their interpretation is that of an intended path for future consumption and labor supply formulated once the household observes the monetary shock. The implicit assumption is that the household expects all the endogenous variables to react as in the rational-expectations model except for inflation and unemployment (and the real wage markup, which is directly pinned down by unemployment).

The steps of the exercise are as follows.

1. Solve the Galí et al. (2011) model through a first-order perturbation method (e.g., with the Dynare toolbox) and compute the IRFs to the monetary policy shock.

2. Rescale all the IRFs to make them consistent with the size of the shock in the vignette. The standard deviation of the shock in the model is 0.21 percentage points while in the vignette it is 0.5. As the model is linear, this simply changes the magnitude of the deviations from the steady state implied by the IRFs.

3. Rebase the IRFs of inflation and unemployment to have the same value as the mean household expectation from the vignette at a horizon of four quarters, assuming that the overall path shape stays the same. In other words, we "force" the IRFs of inflation and unemployment to pass through the black cross marks in the upper panels of Figure 5.

4. Use the relevant equations to recompute the household's optimal consumption and labor supply from time 0 (when the shock hits) until the final period of the IRFs. These equations are the household's first-order conditions for consumption (Euler equation) and labor supply and the model equations linking unemployment and inflation to the relevant state variables for the household's decision (wage markup

equation). The resulting values constitute the First-Order Conditions IRFs (lower panel of Figure 5).

The details of the full model are in Galí et al. (2011). We use the same parameter values as the estimated modes of the posterior distribution from the original paper. To make it clear that the relevant equations are solved as expectations at time 0 (i.e. when the shock is realized), we denote them with an e superscript. For instance, r_t^e represents the time-0 expected nominal interest rate for time t under rational expectations. Meanwhile, we use a tilde to denote expectations that follow those we observe in the survey or the expectations of the households' choice variables (e.g., $\tilde{\pi}_t^e$ and $\tilde{c}_t^e$ for inflation and consumption, respectively).

The Euler equation (with habits) is used to solve for the consumption IRFs. In practical terms, we compute c_t using the household's expectations rather than the rational expectations IRF for inflation. Note that for the nominal interest rate we keep the rational-expectations IRF as in the model because we did not measure households' expectations for it. The resulting equation is

$$\tilde{c}_t^e = \frac{h}{1+h}\tilde{c}_{t-1}^e + \left(1 - \frac{h}{1+h}\right)\tilde{c}_{t+1}^e - \frac{1-h}{1+h}\left(r_t^e - \tilde{\pi}_{t+1}^e\right), \tag{3}$$

where h is a habit parameter (adjusted for trend growth).

The relevant equations for computing labor supply factor in the expected inflation through the consumption-labor preference smoother z_t and the effect of the unemployment rate on the real wage markup. Galí et al. (2011) describe z_t as a "consumption externality": when consumption rises, the marginal disutility from work also falls.

$$\tilde{z}_t^e = (1-\nu)\tilde{z}_{t-1}^e + \nu\left(\frac{1}{1-h}\tilde{c}_t^e - \frac{h}{1-h}\tilde{c}_{t-1}^e\right) \tag{4}$$

The wage markup (μ_t^w) is determined by wages and unemployment.

$$\tilde{u}_t^e = \frac{\tilde{\mu^w}_t^e}{\phi}. \tag{5}$$

Labor supply (n_t) is pinned down by the usual first-order condition equalizing the real wage received by the household and marginal rate of substitution between consumption

and labor supply, which in the model is formulated as follows:

$$\tilde{\mu^{w}}_{t}^{e} = w_t^e - \phi \tilde{n}_t^e - \tilde{z}_t^e. \tag{6}$$

E Details on wave 1 of the expert survey

We compiled a list of participants of the following conferences:

- SITE Macroeconomics of Uncertainty and Volatility (2018, 2017, 2016)
- SITE Macroeconomics and Inequality (2018)
- Cowles Macro Conference (2018, 2017, 2016)
- NBER Annual Conference on Macroeconomics (2018, 2017, 2016)
- ifo Conference on "Macroeconomics and Survey Data" (2018, 2017, 2016)
- Venice Summer Institute on Expectation Formation (2018)
- Workshop on Subjective Expectations NY Fed (2016)

We also recruited a sample of graduate students in macroeconomics from the following institutions:

- University of Bonn
- Goethe University Frankfurt
- University of Oxford

Finally, we also recruited a sample of economists from the following policy institutions:

- The Federal Reserve Board, Washington D.C.
- The International Monetary Fund, Washington D.C.
- Bank for International Settlements, Basel
- Deutsche Bundesbank, Frankfurt
- European Central Bank, Frankfurt
- ifo centre, Munich

Below is a list of the institutions that our experts (from Wave 1) have as one of their main institutions: Kellogg School of Management, Northwestern University, University of Cologne, Haverford College, University of Minnesota, Ross School of Business, University of Michigan, Federal Reserve Bank of Boston, University of Amsterdam, Boston University, Questrom School of Business, Federal Reserve Bank of St. Louis, Goethe University Frankfurt, LMU Munich, University of Notre Dame, University of California San Diego, University of Oxford, Temple University, International Monetary Fund, University of Toronto, Carleton University, Yale University, Federal Reserve Board, University of Copenhagen, University of Bologna, Georgia Institute of Technology Atlanta, Statistics Norway, Deutsche Bundesbank, Frankfurt School of Finance & Management, Johns Hopkins University, Baltimore, Brandeis University, Federal Reserve Bank of Cleveland, Bank of England, MIT Sloan School of Management, Rand Corporation, University of Copenhagen, International Monetary Fund, Swiss National Bank, Boston College, University of Reading, UNC Kenan-Flagler Business School, Bonn Graduate School of Economics, Institute for Employment Research Friedrich-Alexander University (FAU) Erlangen-Nuremberg, College of Business Clemson University, ifo Institute Munich, Stockholm University, Banque de France, University of Nantes, Uppsala University, World Bank, University of St.Gallen, Austrian Institute of Economic Research, Copenhagen Business School, Federal Reserve Bank of Minneapolis, NYU Stern School of Busines, University of Bonn, Mannheim University, University of Manchester, University College London, University of Lausanne, Arizona State University, University of Birmingham, Federal Reserve Bank of Chicago, European Central Bank, Bank for International Settlements, Basel, University of Maryland, Amsterdam School of Economics, Columbia University, Christian Albrechts University at Kiel, Princeton University, Stockholm School of Economics, University of Chicago Booth School of Business, University of Warwick, Leibniz University Hannover, University of Heidelberg, University of Copenhagen, Northwestern University, New York University, Federal Reserve Bank of Minneapolis, Indiana University, Karlsruhe Institute of Technology.

References for the Appendix

Arias, Jonas E., Dario Caldara, and Juan F. Rubio-Ramírez, "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," *Journal of Monetary Economics*, 2019, *101*, 1–13.

Auerbach, Alan J. and Yuriy Gorodnichenko, "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 2012, *4* (2), 1–27.

Balke, Nathan S. and Stephen P.A. Brown, "Oil Supply Shocks and the U.S. Economy: An Estimated DSGE Model," *Energy Policy*, 2018, *116* (February), 357–372.

Ball, L, D Leigh, and P Loungani, "Okun's Law: Fit at 50?," *Journal of Money, Credit and Banking*, 2017, *49* (7), 1413–1441.

Bernanke, Ben S. and Ilian Mihov, "Measuring Monetary Policy Expectations," *Quarterly Journal of Economics*, 1998, *113* (3), 869–902.

_ **, Jean Boivin, and Piotr Eliasz**, "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 2005, *120* (1), 387–422.

Blanchard, Olivier and Jordi Galí, "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s so Different from the 1970s? Chapter," in "International Dimensions of Monetary Policy" 2010.

_ **and Roberto Perotti**, "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, 2002, *117* (4), 1329–1368.

Bodenstein, Martin, Christopher J. Erceg, and Luca Guerrieri, "Oil shocks and External Adjustment," *Journal of International Economics*, 2011, *83* (2), 168–184.

Christiano, L J, M Eichenbaum, and C Evans, "Monetary policy shocks: what have we learned and to what end? Handbook of Macroeconomics," *Handbook of Macroeconomics*, 1999, *A*.

Favero, Carlo and Francesco Giavazzi, "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs," *American Economic Journal: Economic Policy*, 2012, *4* (2), 69–94.

Galí, Jordi, Frank Smets, and Rafael Wouters, "Unemployment in an Estimated New Keynesian Model," *NBER Macroeconomics Annual*, 2011, *26* (1), 329360.

Mertens, Karel and Morten O Ravn, "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks," *American Economic Journal: Economic Policy*, may 2012, *4* (2), 145–181.

_ **and Morten O. Ravn**, "A reconciliation of SVAR and narrative estimates of tax multipliers," *Journal of Monetary Economics*, dec 2014, *68*, S1–S19.

Perotti, Roberto, "The Effects of Tax Shocks on Output: Not so Large, but Not Small Either," *American Economic Journal: Economic Policy*, 2012, *4* (2), 214–237.

Primiceri, Giorgio E., "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 2005, *72*, 821–852.

Ramey, Valerie A, "Can government spending stimulate the economy ?," *Journal of Economic Literature*, 2011, *49* (3), 673–685.

Romer, Christina D and David H Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–1084.

_ and _ , "The Macroeconomic Effects of Tax Changes : Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 2010, *100* (3), 763–801.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

Stock, James H. and Mark W Watson, "Vector Autoregressions," *Journal of Economic Perspectives*, 2001, *15*, 101–116.

Uhlig, Harald, "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 2005, *52* (2), 381–419.

Zubairy, Sarah, "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model," *International Economic Review*, 2014, *55* (1), 169–195.