The term structure of judgement: interpreting survey disagreement*

Federica Brenna[a] and Žymantas Budrys[b]
[a]Bank of Lithuania, Vilnius University
[b]CEFER - Bank of Lithuania, Vilnius University

January 2024
Most recent version

Abstract

Consensus forecasts by professionals are highly accurate, yet hide large heterogeneity. We develop a framework to extract the judgement component from survey forecasts and analyse the extent to which it contributes to respondents' disagreement. For the average respondent, we find a substantial contribution of judgement about the current quarter, which often steers unconditional forecasts towards the realisation, thereby improving accuracy. We identify the structural components of judgement exploiting stochastic volatility and give an economic interpretation to expected future shocks. For individual respondents, about a third of disagreement is due to differences in coefficients or models used, and the remainder is due to different assessments about future shocks; the latter mostly concerns the size of shocks, while there is a general agreement on their source.

Keywords: Expectations Formation, Judgement, Identification via Stochastic Volatility, Survey of Professional Forecasters
JEL Codes: C32, C33, C51, D84, E37

*We are grateful, for useful comments and suggestions, to Ana B. Galvão, Gergely Á. Gánics, Leland E. Farmer, Michele Lenza, Daniel J. Lewis, Malte Knüppel, Elmar Mertens, Boriss Siliverstovs, and to seminar and conference participants at the National Bank of Belgium, International Institute of Forecasters, IAAE 2023, Baltic Central Bank Researcher Meeting 2023, $34^{th} EC^2$ Conference. All remaining errors are our own. The views expressed are those of the authors and do not necessarily reflect those of the Bank of Lithuania or the ESCB.

Contact: FBrenna@lb.lt, ZBudrys@lb.lt

1 Introduction

Forecasting economic variables is a complex task: even professional forecasters, who are well-informed and skilled agents, often make mistakes. Nevertheless, their projections are among the most accurate and highly regarded by policy makers and market participants. The average forecast is the most widely reported quantity in each release of the Survey of Professional Forecasters (SPF), although it masks a large degree of heterogeneity in the responses. Where does this heterogeneity come from and what can it tell us about the forecasting process of the respondents? While we cannot directly observe this process, we can use publicly available information to begin to form an idea: Stark (2013) and ECB (2019) provide insights from surveys conducted among panellists of the US and euro area SPFs, respectively. Two main results are of interest for this paper: first, respondents rely heavily on models, with time series (univariate or multivariate) being the most common; second, most of them apply a component of judgement to their model results, including judgement about economic relationships (see also Andre et al., 2022). We use these two pieces of information to build a framework which allows us to back-engineer the structural shocks expected by each professional forecaster at each forecast horizon.[1] In other words, we extract the combination of structural shocks that is consistent with the forecast paths of all the variables considered and then analyse them in an economically meaningful way.

To do this, we assume that each quarter SPF panellists run a generic vector autoregression (VAR) model to produce forecasts of several macroeconomic variables between the current quarter and four quarters ahead, and adjust their model results by adding a judgement component in the form of conditions on future shocks. The rationale for these assumptions lies in the surveys of panellists mentioned above, but also in the characteristics of VAR models: flexible, general specifications that allow the nesting of simpler models (univariate autoregressive process, random walk...) as well as some more complex ones (time series representations of dynamic stochastic general equilibrium (DSGE) models). The choice to model the judgement component, a well-established feature of survey forecasts, as expected future shocks is in line with widely used econometric methods to construct conditional forecasts, as described by Waggoner and Zha (1999) and, more recently, in terms of structural scenarios, by Antolín-Díaz et al. (2021).

Our methodology derives from these assumptions: for several macroeconomic aggregates,

[1]The framework and results of this paper build on companion work by Brenna (2022), which introduces the structural interpretation of judgement in a frequentist approach, with a bivariate model. We extend the model and results by moving to a Bayesian setting, introducing more variables and shocks, and providing further insights into the structural interpretation of judgement and disagreement.

we collect individual survey forecasts from US SPF respondents, from the current quarter to four quarters ahead, and feed them together with the observed data into a Bayesian VAR. We allow the volatilities of the shocks in the VAR to be time-varying and use this feature to identify the structural shocks. Identification refers to both current and expected future shocks, meaning that we are able to decompose the entire path of each respondent's forecast in a structural way.

We consider results for the average SPF respondent as well as for individual panellists; namely, we analyse impulse response functions (IRFs), historical shock decompositions (including a decomposition between current shocks and judgement), and disagreement about shock decompositions across individual respondents.

We include six variables in our main specification and label five of the six shocks, both by relating them to shock series estimated in the literature and by observing IRFs dynamics. These five shocks behave as, and correspond to, well-known macroeconomic disturbances: unanticipated demand and supply, anticipated demand, cost-push, and financial shock.

For the average respondent, we find that a significant amount of judgement is built into the forecasts and that it contributes to forecast accuracy, particularly at shorter horizons and in times of economic turmoil. For individual models, we find that most of the disagreement among respondents is due to differences in judgement, with only 15-30% of the total disagreement due to differences in coefficients. We find that while respondents disagree on the size of the shocks affecting the variables, they tend to agree on which shock has the largest impact, with the exception of forecasts of the term spread, whose variance is attributed to both an anticipated demand shock and an interest rate shock. Our results help to shed light on the expectation formation process of professional forecasters in an empirical setting. Moreover, our framework can serve as a powerful tool for policymakers to identify, in real time, the type of shocks that professionals expect to affect macroeconomic aggregates and the extent to which they agree on the size and nature of these shocks.

Our paper fits into the growing literature on expectation formation, as well as the large literature analysing disagreement among forecasters and the more specific literature focusing on professional forecasters.

The first strand includes a large number of papers investigating the rationality of expectations and their deviation from the full information paradigm. Seminal works include Mankiw et al. (2003), Sims (2003), and Woodford (2013), with two papers by Coibion and Gorodnichenko (2012, 2015) offering important advances for the modelling of expectations with information rigidities. A few of the later models include, for example, those with learning or sticky information (Born et al., 2020; Farmer et al., 2021; Del Negro et al., 2022), inattentive or heterogeneous forecasters (Dovern & Hartmann, 2017; Giacomini et al., 2020), and overreaction to news (Bordalo et al., 2020; Kohlhas & Walther, 2021). All of the above literature introduces structural assumptions in order to model a specific non-FIRE mechanism. In this paper, we take a different approach by not taking a specific stance on the microfoundations behind the processes of expectation formation, but by modelling expectations in a flexible, reduced form. We then identify structural drivers of forecasts, rather than looking at what types of irrational behavior are consistent with specific types of forecast revisions. In other words, while we acknowledge that forecasters can be irrational and are certainly heterogeneous in their priors and judgments, we do not try to assign their responses to a particular model, but simply use them to extract implied structural shocks.

Several papers have studied disagreement among forecasters, both from a purely empirical perspective and from a more structural one. Examples of the former include Dovern (2015), Andrade et al. (2016), and Clements (2022), while some of the latter are Ricco et al. (2016), Born et al. (2020), Kuang et al. (2020), Falck et al. (2021), and Herbst and Winkler (2021). We contribute to this literature by offering a deeper look into the structural interpretation of disagreement through time: on which shocks do forecasters disagree the most, and in which periods?

Surveys of professionals, mainly from the euro area and the United States, are the subjects of many studies: a comprehensive review is provided by Clements et al. (2022); some of the literature focuses on the accuracy of SPF forecasts, either alone (Andrade & Le Bihan, 2013; Dovern & Hartmann, 2017) or in combination with other models (Bańbura et al., 2021; Galvão et al., 2021), with a particular emphasis on the uncertainty, or density, dimension (Abel et al., 2016; Aastveit et al., 2018; Casey, 2021). We focus here on information contained in individual point forecasts, consider several variables, and the full-term structure of forecasts (from nowcast to one-year-ahead), therefore providing an in-depth analysis of this specific survey.

The rest of the paper is organised as follows. Section 2 describes the empirical framework and identification method, Section 3 introduces the data used and the estimation procedure, Section 4 discusses the results for average forecasters and Section 5 for individuals. Section 6 concludes.

2 An empirical framework for survey forecasts

In this section, we describe the model structure that we assume for SPF respondents and introduce our approach in accounting for judgement and disagreement among forecasters.

We assume that forecasts are produced by SPF participants using a VAR model:

$$y_t = c + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + A_0^{-1} e_t \qquad\qquad e_t \sim \mathcal{N}(0, I_n) \tag{1}$$

where y_t is a vector of N observables, c is a vector of constants, β_i, $i = 1, ..., p$ are $N \times N$ matrices of lagged coefficients, p is the number of lags, A_0 is a matrix of structural impact coefficients and e_t is a vector of structural form disturbances. For simplicity, we omit the constant and lags above one in the rest of the model description.

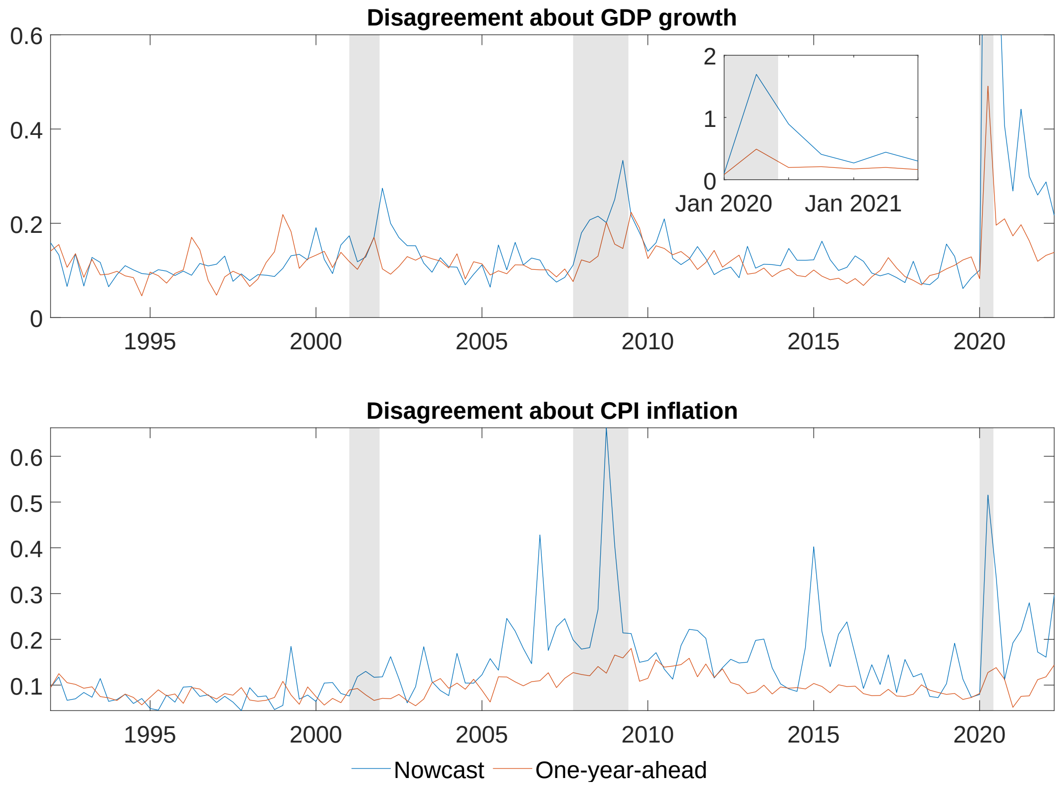
On the basis of a loss function of mean squared forecast error, the forecast implied by the model for h periods ahead is:

$$y_{t+h|t} = \beta^h y_t \tag{2}$$

We call this an unconditional forecast, namely a forecast where future errors are zero on average and the information set at time t is optimally used given the assumed loss function.

We know from observing the data (see Figure 1, which shows the disagreement calculated as the standard deviation of point forecasts across individuals) and the literature (e.g. Mankiw, Reis, & Wolfers, 2003; Andrade, Crump, Eusepi, & Moench, 2016; Glas, 2020) that (professional) forecasters often disagree on the exact value of the most likely outcome. The presence of disagreement is not compatible with a forecast produced as in equation 2, which, if coming from the same model, using the same available data, and assuming the same priors, would return the same outcome for each forecaster. A first explanation for disagreement could be that forecasters might run different models and have diverse priors (Patton & Timmermann, 2010; Giacomini et al., 2020). Second, while they are likely to observe the same information, being specialised economic agents (Farmer et al., 2021), they might attribute varying importance to different pieces of information (Kohlhas & Walther, 2021). Third, as we argue in this paper, they could *both* use different models *and* adjust their model forecasts based on different subjective views on the meaning of the information received.

Figure 1: Disagreement for Real GDP growth and CPI inflation



Note: The figure shows the disagreement for real GDP growth and CPI inflation from the Fed SPF forecasts at the 0 and 4 quarter horizons. Disagreement is calculated as the standard deviation of the individual point forecasts, excluding the two smallest and largest values. Shaded bars are recessions as defined by the NBER.

2.1 Intuition for the model with conditional forecasts

We model this latter possibility by claiming that each agent produces *conditional* forecasts in which they assume that future shocks are not equal to zero, but instead reflect their beliefs about future economic developments. It follows that the h periods ahead forecast is:

$$y_{t+h|t} = \underbrace{\beta^h y_t}_{\text{unconditional forecast}} + \underbrace{A_0^{-1} e_{t+h|t} + \beta A_0^{-1} e_{t+h-1|t} + \ldots + \beta^{h-1} A_0^{-1} e_{t+1|t}}_{\text{future assumptions}} \tag{3}$$

where the first term is the unconditional (or model-based) forecast, and the other terms are the dynamic effects of future assumptions, $e_{t+h|t}$, hereafter "judgement". Judgement reflects the forecaster's opinion about the most likely scenario for the path of certain variables or shocks.

Our motivation for this specification reflects three main aspects. First, the fact that experts produce their forecasts by collecting available information and running models (Stark, 2013), but they can also rely on their past experience and complement their answers with a judgement component (see findings of Croushore & Stark, 2019; ECB, 2019; Andre et al., 2022). Second, while we cannot deduce the exact model used by the panellists, our specification represents a parsimonious and general method of modelling forecasts, capable of nesting different cases: simpler univariate models, different VAR specifications, VAR representations of DSGE models, etc. The parsimony stems from the fact that we do not need to introduce additional coefficients or constrain relationships, but simply recover those coefficients that best fit each respondent's forecasts for all h horizons. Third, this specification is consistent with the econometric methods that forecasters and policymakers can use to supplement their statistical models when making conditional forecasts in real life.

Waggoner and Zha (1999) developed the main formal contribution on how to construct scenarios for conditional forecasts. Their approach was then improved to give the statistician many options for analysing alternative scenarios or expressing her beliefs about future economic developments.[2] All conditional forecasting methods are based on the idea of constraining future shocks within a VAR framework.

We assume that the judgement is distributed independently of the information set inherent in our assumed model specification, y_t, as well as the realised shocks, e_{t+j}:

$$\begin{aligned} &\mathbf{E}\left(e_{t+h|t}|y_{t-s}, s \geq 0\right) = \mathbf{0}_{N\times 1} \\ &\mathbf{E}\left(e_{t+h|t}e'_{t+j}|y_{t-s}, s \geq 0\right) = \mathbf{0}_N \quad \forall h, j \end{aligned} \qquad \mathbf{E}\left(e_{t+h|t}e'_{t+j|t}|y_{t-s}, s \geq 0\right) = \begin{cases} \mathbf{I}_N & \forall h = j \\ \mathbf{0}_N & \forall h \neq s \end{cases} \tag{4}$$

This assumption stems partly from computational convenience and partly from our primary interest in estimating structural (realised and future) shocks, which by definition must be independent.

2.2 Baseline model with conditional forecasts

In this paper, we want to "back-engineer" the most likely combination of shocks assumed by the SPF respondents by estimating a specification that takes into account the structural

[2] Several papers explored different ways of imposing subjectivity, e.g. "hard" versus "soft" conditions to reflect the uncertainty surrounding the assumptions. Antolín-Díaz et al. (2021) drew attention to the need to distinguish between assumptions arising from structural and reduced-form shocks in order to provide more interpretable scenarios. The latest contribution and overview on this topic is provided by Chan et al. (2023).

dynamics of forecasts imposed in the VAR.[3] In the following, we rewrite the model in compact form and extend it to include time-varying volatility in structural shocks.

$$\begin{aligned} & Y_{t+h|t} = C + BY_{t+h-1|t-1} + F\left(I_{h+1} \otimes A_0^{-1}\right)\Lambda_t E_{t+h|t} \\ & E_{t+h|t} \sim \mathcal{N}(0_{N(h+1)\times 1}, I_{N(h+1)}) \end{aligned} \tag{5}$$

As above, we define N as the number of variables and h as the number of forecast horizons. In the system, there are $N(h+1)$ structural shocks that explain the forecast structure, collected in the vector $E_{t+h|t}$.

$$E'_{t+h|t} = \begin{bmatrix} e'_{t+h|t} & e'_{t+h-1|t} & \dots & e'_{t+1|t} & e'_t \end{bmatrix}' \tag{6}$$

The state vector, $Y_{t+h|t}$, is of size $N(h+p)$ and collects conditional forecasts $(y_{t+h|t})$, nowcasts $(y_{t+1|t})$, data (y_t) and lags (y_{t-p+1}):

$$Y'_{t+h|t} = \begin{bmatrix} y'_{t+h|t} & y'_{t+h-1|t} & \dots & y'_{t+1|t} & y'_t & \dots & y'_{t-p+1} \end{bmatrix}' \tag{7}$$

In order to construct the matrices C, B and F so they conform to the state space representation, we first define $B_{*,h}$:

$$B_{*,h} = B_*^h \qquad \text{where} \qquad \underset{Nm\times Nm}{B_*} = \begin{bmatrix} \underset{N\times Np}{\beta} & 0_{N\times N(m-p)} \\ I_{N(m-1)} & 0_{N(m-1)\times N} \end{bmatrix} \tag{8}$$

where m is the maximum between the forecast horizons, $h+1$, and the chosen lag length, p. The matrices then capture the dynamic responses in the following way:

$$\begin{aligned} B &= \begin{bmatrix} 0_{N(h+1)\times Nh} & B_{*,h+1}^{[1:N(h+1),1:Np]} \\ 0_{N(p-1)\times Nh} & I_{N(p-1)\times Np} \end{bmatrix} \\ F &= \begin{bmatrix} I_{N(h+1)\times N} & B_{*,1}^{[1:N(h+1):,1:N]} & B_{*,2}^{[1:N(h+1),1:N]} & \dots & B_{*,h}^{[1:N(h+1),1:N]} \\ 0_{N(p-1)\times N} & 0_{N(p-1)\times N} & 0_{N(p-1)\times N} & \dots & 0_{N(p-1)\times N} \end{bmatrix} \end{aligned} \tag{9}$$

where superscript values in square brackets represent corresponding rows and columns extracted from the matrix. The constant collects the deterministic component over forecast

[3] In the case of no conditional forecasts, i.e. $h = 0$, the model specification collapses to a standard structural VAR.

horizons in a companion form:

$$C' = \begin{bmatrix} c'_{*,h} & \dots & c'_{*,0} & 0'_{N(p-1)\times 1} \end{bmatrix}' \qquad \text{where} \quad \underset{N\times 1}{c_{*,h}} = \sum_{j=0}^{h} B_{*,j}^{[1:N,:]} \begin{bmatrix} \underset{N\times 1}{c} \\ 0_{N(m-1)\times 1} \end{bmatrix} \tag{10}$$

Λ_t collects expected and current scaling factors for the stochastic volatility:

$$\Lambda_t = diag(e^{\lambda_{1,t+h|t}}, ..., e^{\lambda_{N,t+h|t}}, e^{\lambda_{1,t+h-1|t}}, ..., e^{\lambda_{N,t+h-1|t}}, ..., e^{\lambda_{1,t}}, ..., e^{\lambda_{N,t}}) \tag{11}$$

We run the model described above for the average forecaster and for each individual respondent. By estimating the above specification, we obtain coefficients and structural shocks that are consistent with the full term structure of forecasts up to the h we consider. Despite the large amount of information included in the estimation (for each variable, data, and forecasts from horizon 0 to horizon h), the number of parameters does not increase with the number of horizons. The additional information provided by the forecasts allows us to estimate the parameters of the model more precisely while maintaining a parsimonious specification. To further verify this result, we perform a Monte Carlo simulation in Appendix B. The exercise with fully observed forecasts over five periods confirms that the posterior standard deviation for the parameters of the model is a factor of 2.15 smaller than estimates using the model without additional forecast information.

To better appreciate the structural nature of the model, note that in our application we study up to 6 variables with a forecasting horizon of 5 periods, resulting in a state vector of size 36. For the 4 lag specification, the coefficient matrix, β and the vector of constants c contain 150 elements to be estimated, in line with a small VAR system. Alternatively, if we did not impose the forecast structure, the VAR would become large with a coefficient vector of length 5220. In this case, one would have to resort to Bayesian shrinkage, alternative model specifications, or other computational methods for the estimation (see, for example Bańbura et al., 2010; Carriero et al., 2016; Carriero et al., 2022).

The representation described in equation 5 may appear rather cumbersome. In particular, conformable matrices are non-linear functions of the underlying parameters, in order to capture dynamic effects. Dealing with non-linear parameters can lead to inefficiencies in estimation using Bayesian methods, as the conditional distribution may become non-standard. However, our specification is flexible enough to avoid these problems. By constructing conditional forecasts within the VAR framework, we can take advantage of the iterative procedure, which results in linear parameters:

$$y_{t+h|t} = c + \beta y_{t+h-1|t} + A_0^{-1}\Lambda_{t+h|t}e_{t+h|t}$$

As a result, the conditional distributions become standard. We discuss this point in more detail in Appendix A.

Controlling for the full term structure of forecasts also alleviates concerns about invertibility. The assumption of invertibility is related to the informational sufficiency of the VAR system to successfully recover the underlying structural shocks (Lippi & Reichlin, 1994; Fernández-Villaverde et al., 2007; Leeper et al., 2013; Forni & Gambetti, 2014; Forni et al., 2019). In a small VAR setting, the usual remedy is to extend the system either to include forward-looking variables, such as stock market prices, or information encapsulated in some latent factors, or, as we do, to capture agents' expectations by including survey variables.

2.3 Identification of structural shocks

Our choice to introduce stochastic volatility into the specification helps to address the common VAR problem of identifying structural shocks. We want to be able to interpret the shocks we extract from the observed forecasts in economic terms. Exploiting the time variation in the volatility of shocks to identify structural shocks is an established technique (dating back to Rigobon, 2003), which has gained popularity in macroeconomic settings in recent years. Lewis (2021) generalises the method to identification via time-varying volatility. For the present analysis, we rely on a recent contribution by Bertsche and Braun (2022) and an extension to a Bayesian setting by Chan et al. (2021), and set the law of motion of stochastic volatility to:[4]

$$\log \sigma_{i,t} \equiv \lambda_{i,t} = \rho_i \lambda_{i,t-1} + u_{i,t} \qquad\qquad u_{i,t} \sim \mathcal{N}(0, \sigma^2_{u,i}) \tag{12}$$

The setting assumes that log-volatilities follow an independent autoregressive process with a long-run mean restricted to be zero. Under this specification and assuming that $\rho_i \neq 0$ and $|\rho_i| < 1$, Bertsche and Braun (2022) suggest that the structural A_0 is full and identified up to sign changes and column permutations.

[4] To complement the estimation of the structural impact, the study of Chan et al. (2021) focuses on the feature of variable order-invariance in stochastic volatility VARs, that was previously neglected due to the imposed triangular parameterisation of the covariance.

To complete our model specification, we need to define the law of motion for the expected volatility. For the sake of simplicity, we assume that the expected volatility follows an unconditional path, that is, $\lambda_{i,t+j|t} = \rho_i^j \lambda_{i,t}$ or $u_{i,t+j|t} = 0$, $\forall j = 0, ..., h$, implying that respondents do not assume any conditionality about future uncertainty.[5] This assumption allows us to simplify the number of latent variables to one per each observable.

Identification through heteroskedasticity offers some advantages in addressing the questions of our study. First, stochastic volatility is considered an intrinsic feature of macroeconomic data and accounting for it greatly improves forecasting performance (see Cogley & Sargent, 2005; Primiceri, 2005; Clark, 2011; D'Agostino et al., 2013; Clark & Ravazzolo, 2015, to name a few). We believe that respondents are exploiting this feature when conducting their forecasts, considering that the forecast uncertainty, as inferred from probability distributions in surveys, tends to vary over time and across individuals (Boero et al., 2015; Glas, 2020). In addition, enriching the VAR setting with stochastic volatility seems a viable solution to capture extreme economic developments in light of the COVID pandemic (Carriero et al., 2021; Lenza & Primiceri, 2022). Second, contrary to many "traditional" identification methods (such as zero, sign and long-run restrictions), this one does not call for imposing ex-ante or external information on structural coefficients; instead, it limits identification to the statistical features of structural shocks. This aspect is particularly relevant for us, as we want to interpret forecasts produced by economic agents who may have imposed different identification constraints (if any) on their models. The use of this methodology means that we do not have to assume a specific identification strategy of the agents, but only exploit the information contained in the observed forecasts.

2.3.1 External instruments and labelling of structural shocks

The downside of identification through heteroskedasticity is that structural shocks are not automatically given economic meaning, but need to be labelled post-estimation. Several papers propose solutions to this issue, such as Lütkepohl et al. (2021) who test for traditional restrictions that become overidentifying in the presence of heteroskedasticity, or Brunnermeier et al. (2021), who label shocks based on the signs of impulse responses. Here, we label shocks both based on impulse responses and exploiting structural shocks from the literature, similarly to Bertsche and Braun (2022) and Schlaak et al. (2023). More specifically, we relate external shocks from other papers (that may have been estimated from a

[5] The restriction is set for pure convenience and can be easily relaxed to allow for conditionality, though at the cost of reducing computational efficiency and introducing complexity regarding expected volatilities. We do not explore the relaxation of this assumption in this paper.

model, collected using high-frequency information, or otherwise constructed using narrative evidence) to our structural shocks and check whether the latter can be valid instruments for the former. If both standard conditions for instrumental variables (*relevance* and *exogeneity*) are satisfied, we can label our shocks to reflect the structural interpretation of the external shock series.

We assume observed external shocks w_t are linearly related to shock estimates, $\hat{\varepsilon}_t = [\exp(\hat{\lambda}_{1,t})\hat{e}_{1,t}, ..., \exp(\hat{\lambda}_{N,t})\hat{e}_{N,t}]$, as arising from our model with heteroskedasticity identification:

$$w_t = \tau + \psi\hat{\varepsilon}_t + o_t \qquad o_t \sim \mathcal{N}(0, \sigma_o^2) \qquad \hat{\varepsilon}_t \sim p(\varepsilon_t, \Sigma_{\varepsilon,t}) \tag{13}$$

where $p(m, v)$ and $N(m, v)$ present arbitrary and normal distributions with mean m and variance v; o_t is an i.i.d. measurement error. We account for the estimation errors in our structural shocks by explicitly modelling them using the posterior distribution $p(\varepsilon_t, \Sigma_{\varepsilon,t})$ from our VAR results.[6] ε_t is the vector containing smoothed posterior mean estimates of the shock series; $\Sigma_{\varepsilon,t}$ is a diagonal matrix containing heteroskedasticity estimates.

One can infer the structural interpretation of smoothed shocks, ε_t, by observing whether they satisfy the two conditions for a valid instrument. In the frequentist framework, the conditions are relevance, i.e. $\psi_k \neq 0$, and exogeneity $\psi_i = 0$ for all $i \neq k$. To choose the candidate's shock, we select the one with the largest correlation in absolute value and for which the zero value is not in the 90% credible set of coefficient ψ_i. In addition to ensuring that the candidate shock is relevant, we also explore whether the instrument is exogenous to other structural shocks on the basis of a Bayesian model comparison via a likelihood-ratio test. Section D in the Appendix provides an in-depth description and technical details of the procedure.

We relate our shocks to more than 100 proxies collected from over 40 studies; we summarise sources in Table F.1 in the Appendix.

3 Data and estimation

In this section, we describe the data used and the estimation procedure.

[6]If the estimation errors are not accounted for, estimates are biased in line with the classical implication from models with errors in independent variables.

3.1 Data

We use data from the US SPF and Real-Time Data Set for Macroeconomists (RTDSM), both provided by the Federal Reserve Bank of Philadelphia.

The SPF is a quarterly survey of experts conducted each quarter since 1968 (since 1990 by the Philadelphia Fed). Respondents to the survey include forecasting firms, financial institutions, and research centres, among others.[7] These are asked to provide point and probability forecasts for a set of macroeconomic variables for several horizons. For our analysis, we use point forecasts between horizon zero (nowcast) and four (one year ahead). At the time of filling out the survey, the panellists have access to the Bureau of Economic Analysis's advance report of the national income and product accounts, which contains the first estimate for the GDP of the previous quarter. After the first data release, forecasters have about one week to send their responses to the Philadelphia Fed, which elaborates and publishes them (for more details on the survey timing, see Croushore & Stark, 2019).

Forecasts for most variables used in this study are given initially in levels. Still, where needed, we transform them into differences or log-differences, mainly for two reasons: first, to reduce the effect of revisions and second, to account for the various rebasing of data series that occurred throughout the sample.[8] We use real-time data to ensure they resemble the information set available to respondents at each point in time. Revisions mainly concern GDP and its components; inflation measures are usually revised to a smaller extent (primarily due to seasonal adjustment) while interest rates, yields and spreads are observed metrics, hence not affected by revisions.

We collect two types of responses: individual and average across individual responses.[9] Our main specification runs from 1984Q2 to 2022Q2, but we also investigate additional samples for robustness. Since panellists have entered and exited the survey throughout the years and do not provide responses in every survey, we need to deal with several missing observations. As a starting point, we exclude all respondents who have responded to fewer than thirty surveys for both GDP and CPI inflation over the period considered. We deal with missing data for the remaining respondents using the Kalman smoother.

For the baseline specification, we estimate our model for 63 forecasters. Table 1 summarises the variables used in the main specification, their transformations, the date when

[7] While responses are anonymised, a partial list of respondents is available in each survey's report, see Survey of Professional Forecasters.

[8] As highlighted by Howrey (1996) and later by Croushore and Stark (2001), forecasts of the rate of growth of real GNP are less affected by data revisions than forecasts of the level.

[9] SPF also provides the median across responses. For robustness, we also estimate the baseline specification using this alternative aggregate measure, and we find that the differences are negligible.

Table 1: Data and transformations

Data Series	Transformation	Available from	Avg Periods	Avg Resp.
Real GDP	log-diff	1968:Q4	61	25 (29)
Investment[11]	log-diff	1981:Q3	57	23 (27)
Term Spread[12]	level	1992:Q1	52	22 (27)
AAA-10y spread	level	1992:Q1	44	18 (23)
CPI Inflation	log	1981:Q3	60	25 (29)
T-bill	diff	1981:Q3	58	24 (28)

Note: The table summarises variables used in the baseline specification, their transformation and the availability of individual responses. "Available from" is the date when forecast information became available in the SPF dataset; "Avg Periods" indicates the average number of quarters in which each respondent reported the forecast for a variable; "Avg Resp." indicates the average number of respondents at each time point in the sample, with the average from 1992q1 in brackets.

forecast information became available in the SPF, the average number of periods for which there are responses over the sample, and the average number of panellists who reported a forecast for a given variable.[10]

3.2 Bayesian estimation and priors

For estimation, we use Bayesian inference because it is well suited to our specification, including the large VAR system, the presence of latent variables and missing observations. In a nutshell, our priors are mostly uninformative but proper. Below we describe them in more detail.

$$vec(\beta_{avg}) \sim \mathcal{N}(\beta_0, \Sigma_\beta(\kappa)) \qquad\qquad vec(\beta_{indiv}) \sim \mathcal{N}(\widehat{\beta}_{avg}, 3 \cdot I)$$

For the coefficient vector, $vec(\beta)$, we set a rather uninformative Minnesota prior in the specification with average SPF forecasts (Litterman, 1986). Although we also experiment with variations of Minnesota priors and different adaptive shrinkage levels à la Chan (2021),

[10] Appendix Figure F.1 shows the number of respondents who provided their forecasts for a particular variable over time. The figure shows that the number of respondents persistently increased in 1990 after the takeover of the survey by the Philadelphia Fed from ASA/NBER.

[11] Note that the information on gross private domestic investment is not available in the survey in aggregate form, but only for its components: real nonresidential fixed investment (RNRESIN), real residential fixed investment (RRESINV) and real change in private inventories (RCBI). We sum the three series to obtain aggregate investment.

[12] We define the term spread or the slope of the term structure of Treasury securities as the difference between the nominal yield on a 10-year Treasury bond and the nominal rate on a 3-month Treasury bill (SPR_TBOND_TBILL).

we find that results are not very sensitive to them.[13] For the estimation of individual specifications, we apply a slight level of pooling by specifying the mean of the prior to be the posterior mean estimate of the aggregate specification, in resemblance to panel VARs (Zellner & Hong, 1992; Jarociński, 2010).[14]

$$\forall i = 1, ..., N \quad a_{avg,0,i,i} \sim \mathcal{N}(\widehat{\sigma}_{AR,i}^{-1}, 40) \qquad \forall i \neq j \quad a_{avg,0,i,j} \sim \mathcal{N}(0, 40)$$
$$\forall i, j \quad a_{indiv,0,i,j} \sim \mathcal{N}(\widehat{a}_{avg,0,i,j}, 4)$$

The prior for the structural impact matrix, A_0, is set such that the diagonal elements, $a_{0,i,i}$, are centered around the inverse of the residual standard deviation of independent AR(4) processes. The off-diagonal elements, $a_{0,i,j}$, are shrunk towards zero. In line with the priors on the βs, the prior belief is that every variable follows an independent random walk. For the estimation of individual specifications, the priors for the structural parameters $a_{indiv,0,i,j}$ are the same across forecasters, to ensure comparability. We set the expected value equal to the posterior mean estimate from the aggregate specification. The prior reduces the parameter space, allowing to neglect permutations that arise when using heteroskedasticity identification, an issue we return to in Section 5.1.

$$\sigma_{u,i}^2 \sim \mathcal{IG}(3/2, S_{u,i}) \qquad S_{u,i} \sim \mathcal{G}(1.6/2, 1)$$
$$\rho_i \sim \mathcal{N}(0.9, 0.09)\mathbb{1}_{(-1<\rho<1)} \qquad \lambda_{1,i}|\rho_i, \sigma_{u,i}^2 \sim \mathcal{N}\left(0, \frac{\sigma_{u,i}^2}{1-\rho_i^2}\right)$$

Priors for the parameters that govern the law of motion of the stochastic volatility, ρ and σ_u^2, are independent across different processes and conditionally conjugate. For the latter parameter we use a non-standard hierarchical setup that allows for a 'fatter' tail for values close to zero and which does not rule out homoskedasticity a priori.[15] The prior for the initial state of the log standard deviation, λ_1, is hierarchical and set to the long-run values of an AR(1) process with mean value of zero and variance determined by hyper-parameters

[13] We set hyperparameters for the variance matrix $\Sigma_\beta(\kappa)$ such that κ_1 that governs the shrinkage on variables' own lags is equal to unity, κ_2, the shrinkage on other lags, is equal to unity and κ_4 on intercepts is equal to 100. In a hierarchical set-up à la Chan (2021), we find that an 'aggressive' cross-variable shrinkage is prescribed, i.e. $\kappa_1 = 0.134$ and $\kappa_2 = 0.004$.

[14] Irrespective of the prior, the most significant amount of pooling for coefficients should arise from the likelihood, as the same observed data but not forecasts is used across respondents.

[15] Our specification provides a computationally convenient prior adhering to remarks by Gelman (2006) and Kastner and Frühwirth-Schnatter (2014) that a standard prior of inverse gamma tends to be over informative a posteriori.

ρ_i and $\sigma^2_{u,i}$.

$$Y_0 \sim \mathcal{N}(\bar{Y}_0, 5 \cdot \Sigma_{Y,0})$$

Priors for the initial conditions Y_0 are centred around observed data prior to the estimation period; the variance matrix is a scaled matrix, $\Sigma_{Y,0}$ which has on the diagonal the long-run variances of independent AR(4) processes over the entire sample. For the estimation of individual specifications, the priors for the initial conditions are the same across forecasters, to ensure comparability.

We estimate the model using the Markov Chain Monte Carlo (MCMC) algorithm of Gibbs sampling, with detailed explanations in Appendix A. The algorithm's steps are similar to previous efforts to estimate models with stochastic volatility (most notably Cogley & Sargent, 2005). However, we deviate by, first, relying on the recent contribution by Chan et al. (2021) to account for the full parameterisation of the structural matrix and, second, by fully modelling SPF forecasts within the assumed VAR specification - an addition that introduces a few complexities that we leave for the Appendix. We apply heavy thinning by taking every 30th draw after the burn-in sample to ensure satisfactory mixing. In the end, we generate 10^6 draws, of which 7×10^5 are for burn-in, leaving 10^4 draws to approximate posterior distributions.[16]

Lastly, as discussed in the previous subsection, individual information on the SPF forecast can be scarce depending on the respondent. For that reason, we cast our model in a state-space representation to interpolate series using the simulation smoother of Durbin and Koopman (2002).

We use four lags in the VAR system.

4 Results for the average respondent

We first present results for the average respondent specification, mainly for two reasons: first, the "average" or "consensus" forecast is the most widely reported quantity, hence the most scrutinised; therefore, it is relevant to understand how this hypothetical respondent works. Second, results for the average respondent are easier to interpret and a useful starting point to then move to individual results. The labelling of shocks is done based on the impulse responses of this first specification.

[16] MCMC convergence and efficiency are satisfactory, see inefficiency factors and Geweke (1992) convergence diagnostic statistics in Appendix Figure C.1.

4.1 Impulse responses and shocks' labelling

In this section, we describe impulse responses and label structural shocks. As mentioned, our identification does not allow to map shocks into their economic meaning directly. We approach this issue from two sides: first, we look at the correlation with external shocks, as described in Section 2.3.1.[17] Second, we analyse the sign and dynamics of impulse responses and relate them to regularities arising from structural models. We interpret shocks rather broadly, as is common in the VAR literature. Economies are subject to a myriad of shocks; for that reason, we do not expect every shock in our analysis to have an exact underlying structural foundation. Figure 2 depicts our baseline model's impulse response functions. Each column includes a shock and each row a variable, so each panel represents the response of a variable to a shock, such that variables on the diagonal are normalised to increase by one unit. The black line represents the posterior mean and the blue lines are the posterior 90% credible sets.

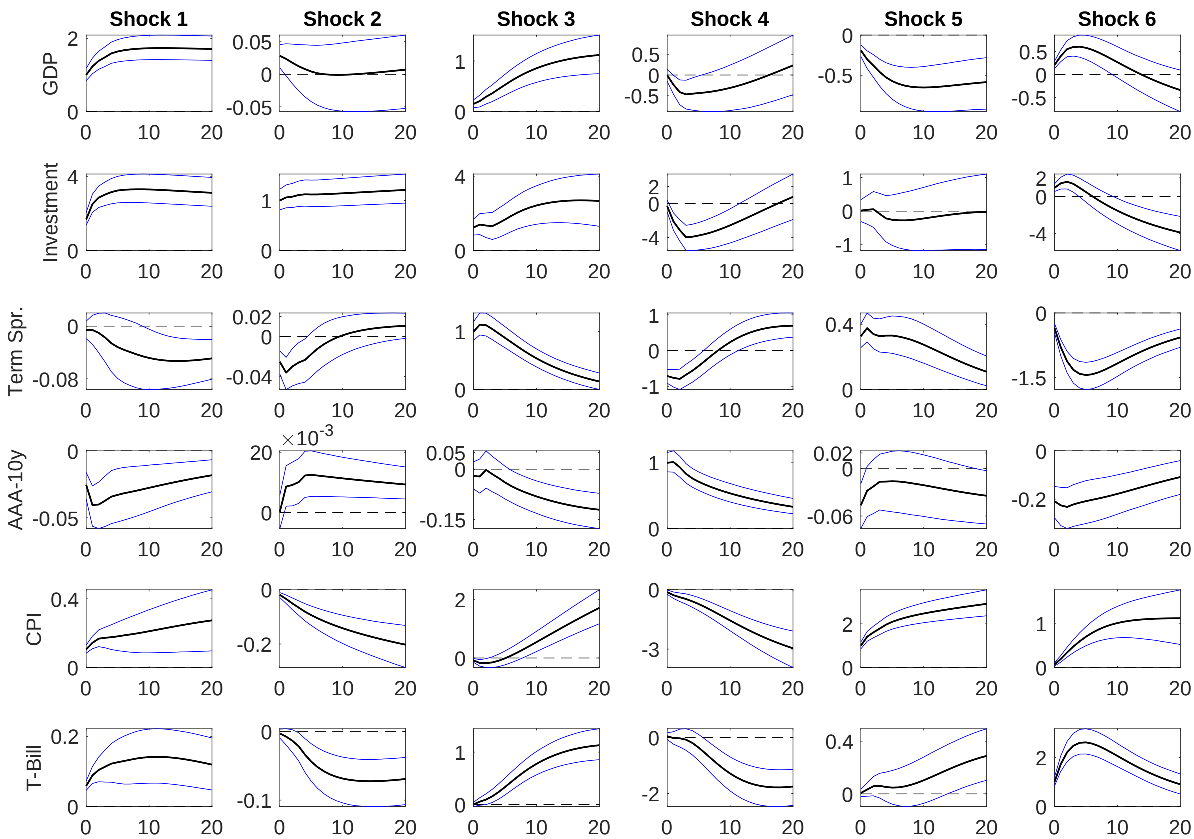
Unanticipated demand shock. The first shock (Shock 1 in the figure), causes a comovement between GDP, investment and prices, in line with a positive aggregate demand shock.[18] The monetary authority responds to an increase in inflation by tightening - a feature that distinguishes this shock from a monetary policy surprise. The delayed decrease in the term spread reflects dissipating demand effects, as expected short-term interest rates are lowered. The boost in demand ensures more favourable financing conditions for firms - the credit spread falls.

In relation to external instruments, we find that our unanticipated demand shock resembles exogenous tax cuts, see Table 2. The time series of estimated shock is strongly correlated to the exogenous tax shocks of Romer and Romer (2010), the unanticipated tax increases of Mertens and Ravn (2012) and the personal income tax increases as described by Mertens and Ravn (2013). To be a valid instrument, the external shock series has to be exogenous to other shocks, $\psi_i = 0 \; \forall i = 2, ..., 6$. The posterior model probabilities, $P(M_r|y)$, indicate that the restricted model is preferred with almost certainty across all candidate shocks. The same conclusion is upheld for the likelihood ratio test, the frequentist alternative for hypothesis testing. Both findings provide rigorous evidence for the first shock representing unanticipated demand shocks, such as tax decreases.

[17] For brevity, we present selected results focused on a particular shock that satisfy both relevance and exogeneity conditions. The full table, relating all estimated shocks to external proxies, is presented in the Appendix with clarifying details, see Table F.1.

[18] A positive comovement between output and prices is often used as a defining feature of an aggregate demand shock, such as preference/taste shocks, for structural identification when using sign restrictions (Canova & Paustian, 2011; Foroni et al., 2018; Furlanetto et al., 2019, for example).

Figure 2: Impulse responses, average model



Note: The figure shows impulse response functions for the baseline model. Each sub-panel shows the response of a variable (in the rows) to a shock (in the columns), normalised to increase the variable on the diagonal by one unit. The solid black line is the posterior mean as a point estimate and the blue lines are the posterior 90% credible sets.

Table 2: External shocks related to unanticipated demand shock

	RR10exo	MR12unc	MR2013TPI
ψ_1	**-0.024*** **	**-0.02*** **	**-0.023*** **
ψ_2	0.001	0.001	0.001
ψ_3	-0.019**	-0.013*	-0.016*
ψ_4	-0.011	-0.006	-0.004
ψ_5	-0.004	-0.005	-0.005
ψ_6	-0.007	-0.002	-0.002
Candidate	1	1	1
$P(M_r\|y)$	1	1	1
$\log_{10} BF$	7.964	8.827	8.364
LRT	0.139	0.488	0.475

Note: The table presents the posterior mean of coefficients, ψ_i, obtained from regressing shocks from the literature (column names) on our shock estimates, see equation 13. Asterisks denote different levels of high probability density intervals when the zero value is not included (***=99%, **=95%, *=90%). "Candidate" is the shock with the highest absolute correlation, "$P(M_r|y)$" is the posterior probability of the restricted model (i.e. the model including only the most relevant shock) to be preferred, "$\log_{10} BF$" is the logarithm of Bayes' factor in favour of the restricted model, and "LRT" is the p-value from the likelihood ratio test.

Unanticipated supply shock. The second shock is more akin to an aggregate (positive) supply disturbance: quantities (both GDP and investment) rise, while prices decrease. Interest rates are set higher to combat disinflation, so the treasury bills' yield slowly decreases. The term spread goes down, due to expected decreases in the short-term rates, but may as well reflect the decrease in the term premium, as noted by Rudebusch and Swanson (2012) for positive technology shocks. The credit spread somewhat rises, unusually for a standard positive supply shock which normally would see a decrease in risk. Nevertheless, these dynamics may also be reflective of a mechanism of the extensive margin of borrowers: given an increase in total investments, more agents will borrow money, including both "good" and "bad" borrowers, therefore causing an increase in credit spread (see, e.g., Justiniano et al., 2011). When relating our shock series to previous vast research efforts on technology shocks, we find rather mixed evidence, possibly reinforcing the conclusion of Ramey (2016) that there is immense complexity and lack of consensus when identifying technology shocks.

Anticipated demand shock. The third shock stands for the anticipated demand shock. The shock shares similarities with the unanticipated demand shock, such that GDP and inflation rise. However, the overall dynamics of variables reflect an announced or anticipated demand shock. The responses of aggregates are more delayed, with increases happening after a few quarters. Meanwhile, the forward-looking variables react on impact, indicating an effect on expectations. The term spread has an immediate positive response: long-term treasury yield rises, indicating that future rises in short-term rates are expected to materialise. The corporate spread drops on impact, reflecting more favourable credit conditions for firms due to expected future demand. Another indicator of anticipation lies in the increase in investments, relatively larger than the one in GDP, that could stem from an accumulation of inventories and capital in view to fulfil the anticipated demand increase.

Financial shock. We define the fourth shock as a shock to financial conditions. As indicated by an increase in the spread between AAA-rated corporate and 10-year government bonds, the worsened financing conditions lead to a delayed decrease of slower-moving variables, real GDP, investment and prices. A monetary easing accommodates the bust in the economy, as the yield on treasury bills is adjusted downwards. The term spread contemporaneously decreases, likely due to an expected monetary policy easing in response to the financial shock, but reverts once the economy recovers. The impulse responses for this shock closely resemble those labelled by Brunnermeier et al. (2021) as a "non-bank financial shock", a shock to the excess bond premia by Gilchrist and Zakrajšek (2012) and credit supply shocks of Bassett et al. (2014). Indeed, we find that our series for this structural shock are correlated to innovations of the latter two studies, see Table 3.[19]

Table 3: External shocks related to financial shock

	BCDZ14	GZ12	NB09	NB09FMT	NB09MMT
ψ_1	-0.003	-0.005	-0.006	0.002	-0.001
ψ_2	0.001	0	0.001	-0.001	-0.001
ψ_3	-0.001	0.006	0.014*	0	0
ψ_4	**0.038****	**0.09*****	**0.064*****	**0.013****	**0.016*****
ψ_5	0.006	-0.004	-0.001	-0.001	0
ψ_6	0.007	0.01	-0.004	0.002	0.002
Candidate	4	4	4	4	4
$P(M_r\|y)$	1	1	1	1	1
$\log_{10} BF$	9.035	9.013	8.731	11.506	11.473
LRT	0.775	0.39	0.249	0.781	0.665

See note for table 2.

The shock has an alternative structural rationale to portray uncertainty shocks, as suggested by Bloom (2009). An exogenous variation in possible worse future outcomes encourages households and firms to postpone consumption and investment, which may possibly have macroeconomic effects. The dynamics of our impulse responses align with these mech-

[19]The historical decomposition also confirms our narrative, see Figure F.6 in the Appendix. Particularly the contribution of a financial shock to the corporate spread spikes at periods of the "dot-com" bubble in the 2000s and the Great Financial Crisis in 2007 - instances of tight financial conditions.

anisms, whereas our structural shocks correlate with those found in Bloom's study.

Cost-push shock. The fifth innovation represents a cost-push shock - a shock that arises due to unwarranted aggregate variation in production input prices, such as commodity prices. An exogenous increase in commodity prices leads to a contemporaneous pass-through onto consumer prices and diminishing economic activity. Note that our interpretation may be hindered by the fact that the VAR specification does not incorporate any commodity price index, as well as the dynamic responses for these shocks are not dissimilar from the one we named "supply" (with an inverted sign), with a negative co-movement between GDP and CPI inflation. Despite that, we find that our shock series are valid instruments to oil supply news shocks by Känzig (2021), "pure" oil price expectation shocks by Baumeister (2023), oil price shocks by Hamilton (2003) an instrument constructed by Caldara et al. (2019), see Table 4.[20]

Table 4: External shocks related to cost-push shock

	DK21s	HAM03b	BH2022E	CCI19inst
ψ_1	-0.003	0.001	0	0.006
ψ_2	-0.002	-0.004*	-0.002	0.001
ψ_3	-0.005	0.001	0.004	-0.003
ψ_4	-0.001	0.014	-0.009	0.009
ψ_5	**0.013*** **	**0.013*** **	**0.018*** **	**-0.01** **
ψ_6	0.011	0.001	0.005	0.009
Candidate	5	5	5	5
$P(M_r\|y)$	1	1	1	1
$\log_{10} BF$	8.553	8.746	9.084	8.766
LRT	0.443	0.509	0.815	0.712

See note for table 2.

We also confirm that the implied contributions of these shocks to historical variations of consumer price inflation fit the historical narrative of notable events related to drastic oil price developments, including the recent increase in input prices arising due to disruptions in global value chains and the Ukraine/Russia war (see Figure F.5 in the Appendix).

Interest rate shock. Finally, the last shock takes the name from the variable it affects the most, namely the T-bill rate. Around 50% of short-run variation in the T-bill is allocated to this variable.[21] This shock has a large positive impact on the T-bill and a negative impact

[20] Differently to other proxies in the table, the sign of coefficient for the instrument by Caldara et al. (2019) is negative one. This happens because authors construct the instrument as leading to a disruption of oil supply, therefore representing shocks of an opposite sign.

[21] Figure F.4 in the Appendix summarises results for the forecast variance decomposition across all variables and shocks.

on the AAA spread. The term spread reacts negatively with a delay. Real variables move consistently with a demand-side shock: higher GDP and investment, together with higher CPI inflation, although investment starts to decrease only after one year. We are cautious about providing a structural interpretation for this shock, as we believe it may represent a convoluted response of multiple innovations. It resembles yet another demand shock in our system, which reflects the reallocation of demand from investment onto other components of GDP. We also find that the shock series relates to monetary policy shocks as identified by Jarociński and Karadi (2020) and Miranda-Agrippino and Ricco (2021), although the impulse responses are at odds with the conventional dynamics for monetary surprises.[22]

Figure G.1 in the appendix shows impulse responses from alternative specifications: excluding the COVID period, beginning the sample in 1976, or using median SPF responses as opposed to the average across individuals, or the specification with two lags.[23] While most results are qualitatively and quantitatively similar to the baseline model, one notable difference is in the impulse responses for the sub-sample excluding the COVID-19 pandemic. This specification features a stronger treasury bill response to an unanticipated demand shock, accompanied by a larger term spread and a more muted reaction of CPI. This difference probably reflects the monetary policy stance during the pandemic, which was less accommodative than in the previous part of the sample, given the extraordinary drop in GDP.

4.2 Historical shock decomposition

In this section, we look at historical shock decomposition for nowcasts and one-year-ahead forecasts, visible in figures 3 and 4, for real GDP growth and CPI inflation, respectively.[24] In a conventional VAR setting, the historical decomposition corresponds to a vector moving-average process divided into a deterministic component (DC in the legend) and stochastic

[22]We suspect that these puzzling results may be due to the lack of variation arising from monetary surprises, which are considered to reflect only a small share of business-cycle variation (Ramey, 2016). Identification using heteroskedasticity may fail to distinguish this shock without any additional source of information, unlike in the studies exploiting high-frequency information.

[23]Specification with two lags is suggested by various information criteria, e.g. Akaike or Schwarz, and likelihood ratio tests.

[24]Figures F.7 and F.8 in the Appendix present results for all variables in the VAR system.

structural contributions:[25]

$$y_t = \underbrace{\left(\sum_{j=0}^{t-1} \beta^j\right) c + \beta^t y_0}_{\text{deterministic comp.}} + \underbrace{\sum_{j=0}^{t} \beta^j A_0^{-1} \varepsilon_{t-j}}_{\text{stochastic comp.}}$$

For the sake of brevity and clarity, we use structural shocks scaled by the stochastic volatility in the rest of the paper, so that elements of vector $\varepsilon_{t+h|t}$ are $\varepsilon_{i,t+h|t} = \exp(\lambda_{i,t+h|t}) e_{i,t+h|t}$. Our specification and its use of both forecasts and data allow us to decompose forecasts into contributions of *observed* shocks, referring to the history of innovations, and *expected* shocks, which reflect judgement about the current (unknown) quarter and future shocks:[26]

$$y_{t+h|t} = \overbrace{\underbrace{\left(\sum_{j=0}^{t-1} \beta^{j+h}\right) c + \beta^{t+h} y_0}_{\text{deterministic comp.}} + \underbrace{\sum_{j=0}^{t} \beta^{j+h} A_0^{-1} \varepsilon_{t-j}}_{\text{stochastic comp.}}}^{\text{unconditional forecast}} + \overbrace{\underbrace{\beta^{h-1} A_0^{-1} \varepsilon_{t+1|t}}_{\text{nowcast judg.}} + \underbrace{\sum_{l=2}^{h} \beta^{h-l} A_0^{-1} \varepsilon_{t+l|t}}_{\text{future judg.}}}^{\text{judgement}}$$

The panels on the left of the two figures show the decomposition of each forecast (in deviation from its long-run mean) into each structural shock, named in the legend with the same numbering of Figure 2.[27] These are sums of observed and future shocks. The panels on the right show the decomposition of each forecast into initial conditions, observed shocks, judgement about the current quarter or nowcast, and judgement about future shocks.

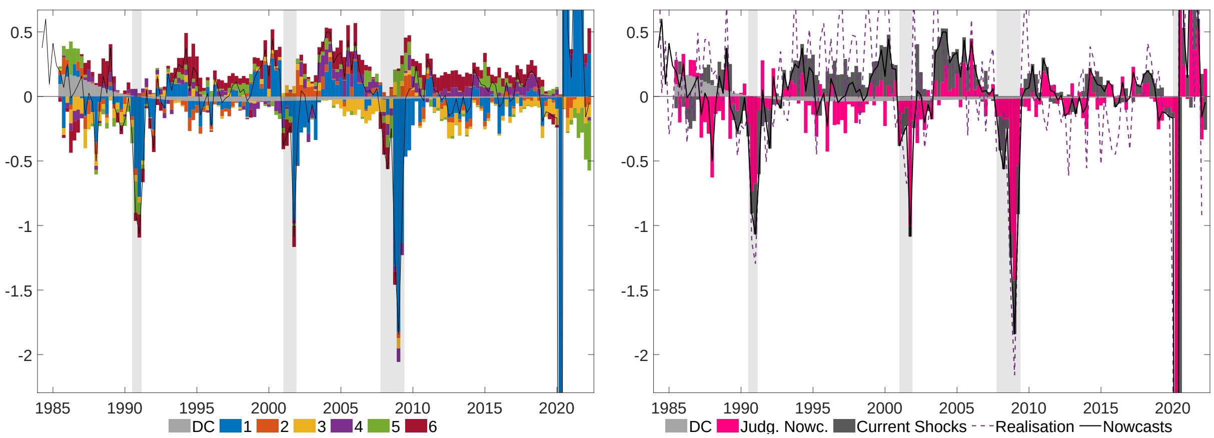
Looking at the left panels, results suggest that unanticipated demand shocks were the main drivers of the forecast for real GDP during the periods of the Great Financial Crisis (GFC) and the COVID pandemic. However, there is one notable difference between the two instances: for the GFC, forecasters expected the growth to be hindered for a few years, whereas the pandemic bust was considered to be temporary, with a sharp rebound forecast in the upcoming four quarters. Inflation in the past two years was thought to be affected by cost-push shocks caused by supply-side bottlenecks and the war in Ukraine. These shocks also had a negative effect on the forecast for GDP.

[25] Figure F.6 in the Appendix shows "traditional" historical shock decompositions of data.

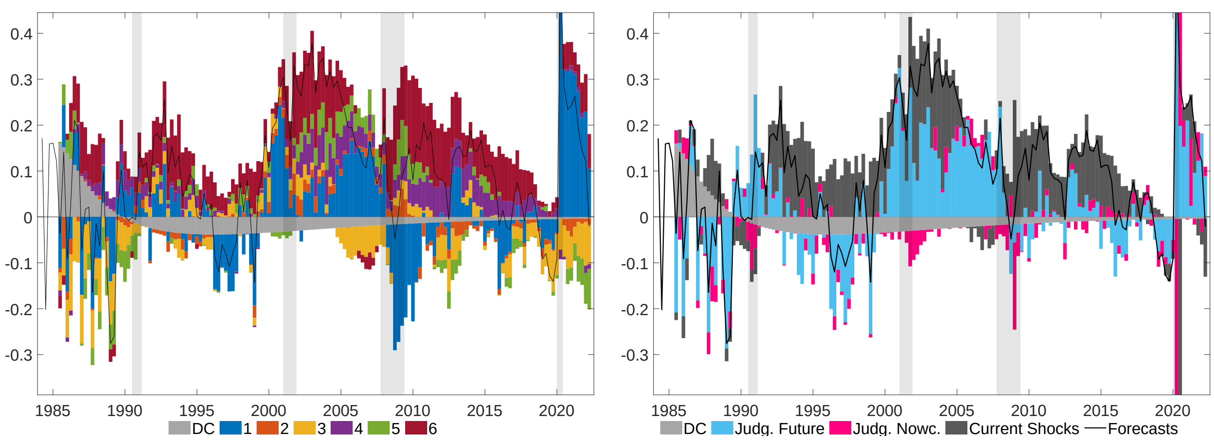
[26] Note that our concept of nowcast is slightly different than the conventional one: we define nowcast as the forecast for quarter $t+1$, given information up to quarter t. Any information at higher than quarterly frequency will be included in the judgement component of the nowcast: this arises from the fact that we do not employ mixed-frequency estimation techniques to capture nowcasting rigorously but, instead, we rely on a simplification allowing us to distinguish SPF nowcasts and forecasts.

[27] Shock decompositions are built by simulating the model multiple times, each time setting one of the shocks equal to zero.

Figure 3: Historical shock decomposition and judgement decomposition of Real GDP q-o-q growth rate



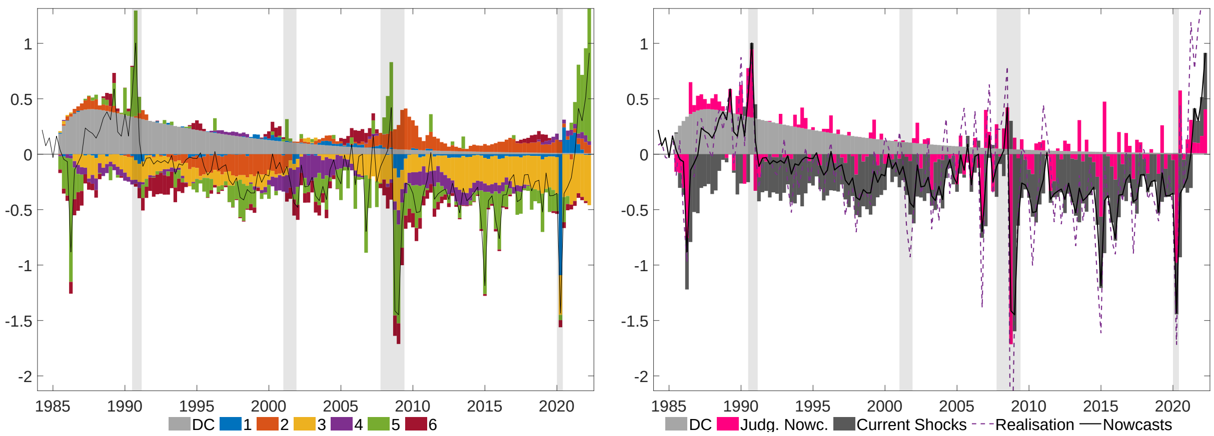
(a) Nowcast



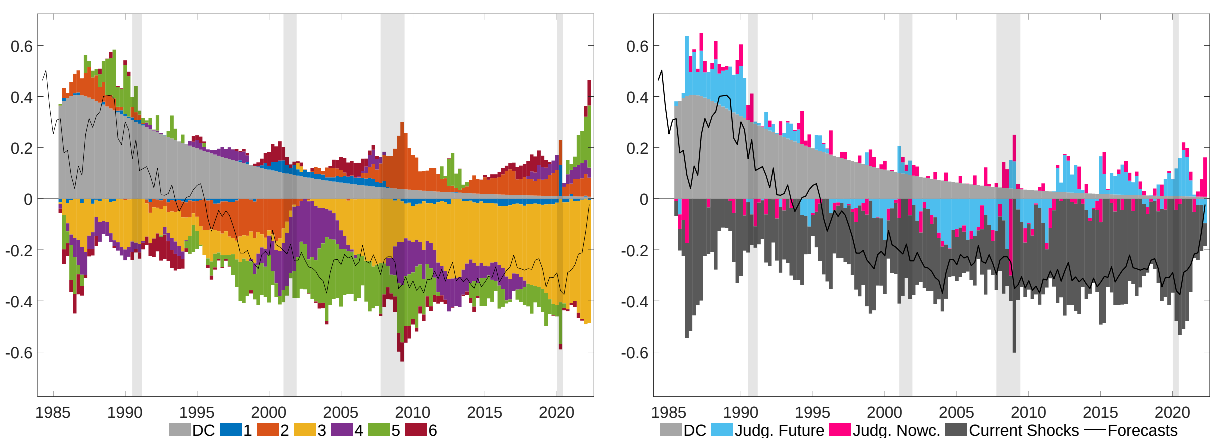
(b) One-year-ahead

Note: The left panels show the historical shock decomposition for the nowcast and one-year-ahead forecast in terms of deviation from its long-run mean. The right panels show the decomposition of the nowcast and one-year-ahead forecast into deterministic conditions, observed shocks, judgement about nowcasts and judgement about expected shocks in future horizons. We use the posterior mean of the historical decomposition as our point estimate. Shaded areas represent NBER recession periods.

Figure 4: Historical shock decomposition and judgement decomposition of CPI inflation



(a) Nowcast



(b) One-year-ahead

See note for Figure 3 above.

The right panels can tell us yet another story: to what extent have forecasters adjusted their model forecasts by adding non-zero future shocks, and how did these expected shocks affect overall forecasts? A first point to note is that forecasters exercise their subjective views about the economy widely: a non-negligible share of observed predictions is due to judgement.

We distinguish here between judgement added in the current quarter, or judgement about nowcasts (pink bars) and judgement about future shocks, or referring to the following quarters (blue bars). In addition to subjective assumptions, judgement about nowcasts can incorporate all information which becomes available during the survey quarter up to the day of its submission. In our system, this applies to financial yields that are observed daily. However, nowcasts can also reflect external information beyond variables in the system, for example, industrial production, energy prices, electricity consumption etc. - indicators that are observed more timely. We find that the contribution of judgement about current macroeconomic developments can be sizeable; particularly, a large piece of variation of nowcasts is explained by this component.

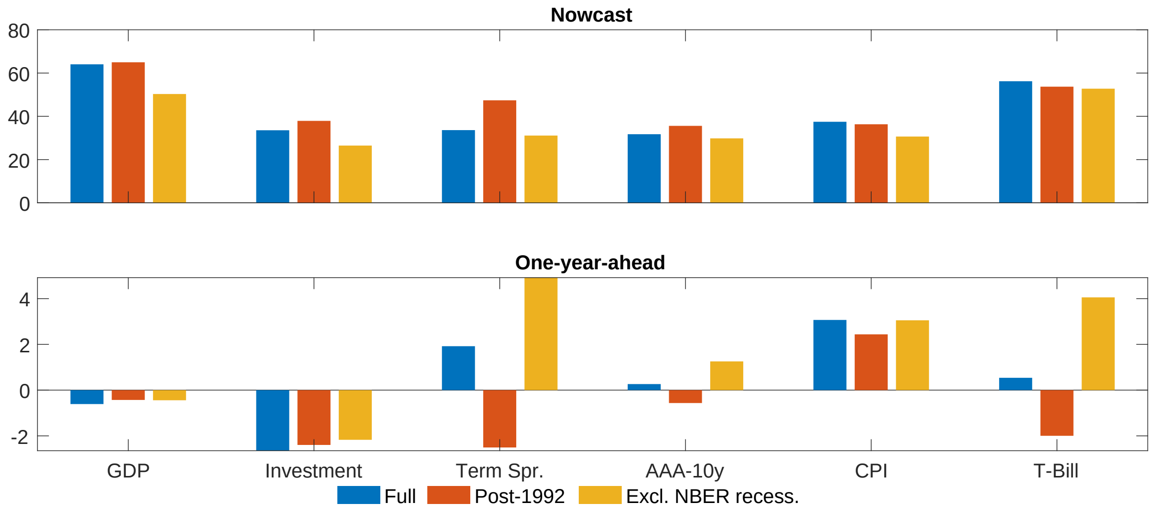
When nowcasts are compared to the realisation for that quarter (dashed purple line), we observe that this type of judgement correctly anticipates the realised value.[28] Figure 5 presents percentage gains in terms of root mean squared errors (RMSE) for SPF forecasts versus unconditional forecasts. RMSEs for nowcasts are consistently smaller with judgement than without it. This holds across all variables and different samples. The most significant contributions appear in recession periods, suggesting that VAR predictions without additional information would have greatly failed to predict the current economic stance. In contrast, professional forecasters do "get it right".

The contribution of nowcast judgement is smaller for four-step-ahead forecasts, by construction, as a result of stationary autoregressive dynamics in our VAR system. However, the higher horizon forecasts do include a non-negligible component of expected future shocks (blue bars) in addition to the prescriptions by the VAR model. In contrast to nowcasts, the overall contribution of judgment does not lead to an improvement in the four-step-ahead horizon accuracy: differences in RMSEs between SPF and unconditional forecasts are negligible (see the bottom panel of Figure 5).

In this section, we confirm the well-known result that SPF performs better than model-consistent unconditional forecasts at shorter horizons. In our framework, this higher accu-

[28] We define *realisation* the value recorded in the SPF, often reflecting the first-release value. We also find qualitatively and quantitatively similar estimates when using realisations revised five months after the initial value or currently observed values.

Figure 5: Forecast performance gains of SPF versus unconditional forecasts



Note: The figure shows the percentage gains in terms of root mean squared error (RMSE) for the SPF forecasts compared to model-consistent unconditional forecasts: $100(1 - RMSE_{SPF}/RMSE_{UC})$. The top panel shows the evaluation for nowcasts, while the bottom panel shows the evaluation for four-step-ahead forecasts. The different colours show results using different sub-periods for the forecast error: "Full" is the full sample; "Post-1992" is the sample from 1992q1; "Excl. NBER recess." is the sample excluding NBER recessions. In the appendix, Figure F.10 includes results for all forecast horizons.

racy is mainly due to subjective judgement, which we find to be pervasive at each horizon and which is a source of disagreement among respondents. In the next section, we analyse this disagreement further: we estimate our model for individual respondents to understand the extent to which different structural sources of judgement can rationalise the disagreement.

5 Results for individual respondents

In this section we present the results of estimating our model specification for each forecaster in order to analyse the underlying drivers of disagreement across respondents.

Recall the specification of a conditional forecast for each individual i:

$$y_{t+h|t,i} = \left(\sum_{j=0}^{h-1} \beta_i^j\right) c_i + \beta_i^h y_t + A_{0,i}^{-1}\varepsilon_{t+h|t,i} + \beta_i A_{0,i}^{-1}\varepsilon_{t+h-1|t,i} + \ldots + \beta_i^{h-1} A_{0,i}^{-1}\varepsilon_{t+1|t,i} \quad (14)$$

The equation highlights different sources that explain the heterogeneity of forecasts across respondents. In our setting, the dispersion may arise from different judgement about current and future economic developments, as reflected in the structural assumptions $\varepsilon_{t+1|t,i}$

for each forecaster. It is also plausible to expect that agents use different models or priors for the underlying structural parameters, c_i, $A_{0,i}$ and β_i. As a result, even though agents observe the same information, as summarised in y_t, they can still produce very different prescriptions about the future. The SPF also records respondents' expected revisions to past data, which we take into account when estimating individual specifications. While this is yet another source of disagreement, expected revisions are rarely recorded and account for a negligible proportion of total disagreement.

Before presenting the results in more detail, we briefly discuss the issue of shock permutation, a feature of identification through heteroskedasticity.

5.1 Permutations and cross-sectional comparison

As mentioned in subsection 2.3, identification through heteroskedasticity allows us to identify the structural A_0 matrix up to sign changes and column permutation (Lewis, 2021; Bertsche & Braun, 2022). This can become an issue when multiple models need to be compared, as is the case for our individual estimations. As we run a separate estimation for each respondent, we need to ensure that the order of the shocks is the same.

To do this, we first normalise the parameter space by imposing weakly informative priors on the matrix of structural coefficients, A_0, as mentioned in Section 3.2. Following Hamilton et al. (2007), we find that the prior ensures parameter identification and reliable inference by selecting the appropriate neighborhood of the posterior mode. In addition, the prior allows permutations to be neglected when comparing results between individual forecasters.

Second, we use a rigorous classification procedure to ensure that the comparison is robust. We rely on the idea that current structural shocks across respondents should reflect similar information, since they are determined by the same observed data. Therefore, the structural shocks associated with the observed data should have a common structure across agents, where each structural shock for an agent should be explained by only one factor. If we can determine to which factor the shock is most related, we can label that shock as similar across agents and determine the ordering. In addition, by observing whether the correlation with the factor is positive or negative, we can also determine the sign of the permutation. For a description of the exact procedure used to reorder shocks consistently across agents, we refer the reader to Section E in the Appendix.

5.2 Heterogeneity of parameters

Figure 6 shows impulse response functions across individuals, which give an indication about the heterogeneity of structural parameters. The black line is the cross-sectional median of the individual posterior estimates, while the dark and light grey areas represent the 68% and 90% percentile bands, respectively. The red line is the posterior mean coming from the average model. Individual impulse responses are roughly similar to average ones; the dynamics follow the same patterns, in line with the structural narrative we provided for each shock in Section 4.1.

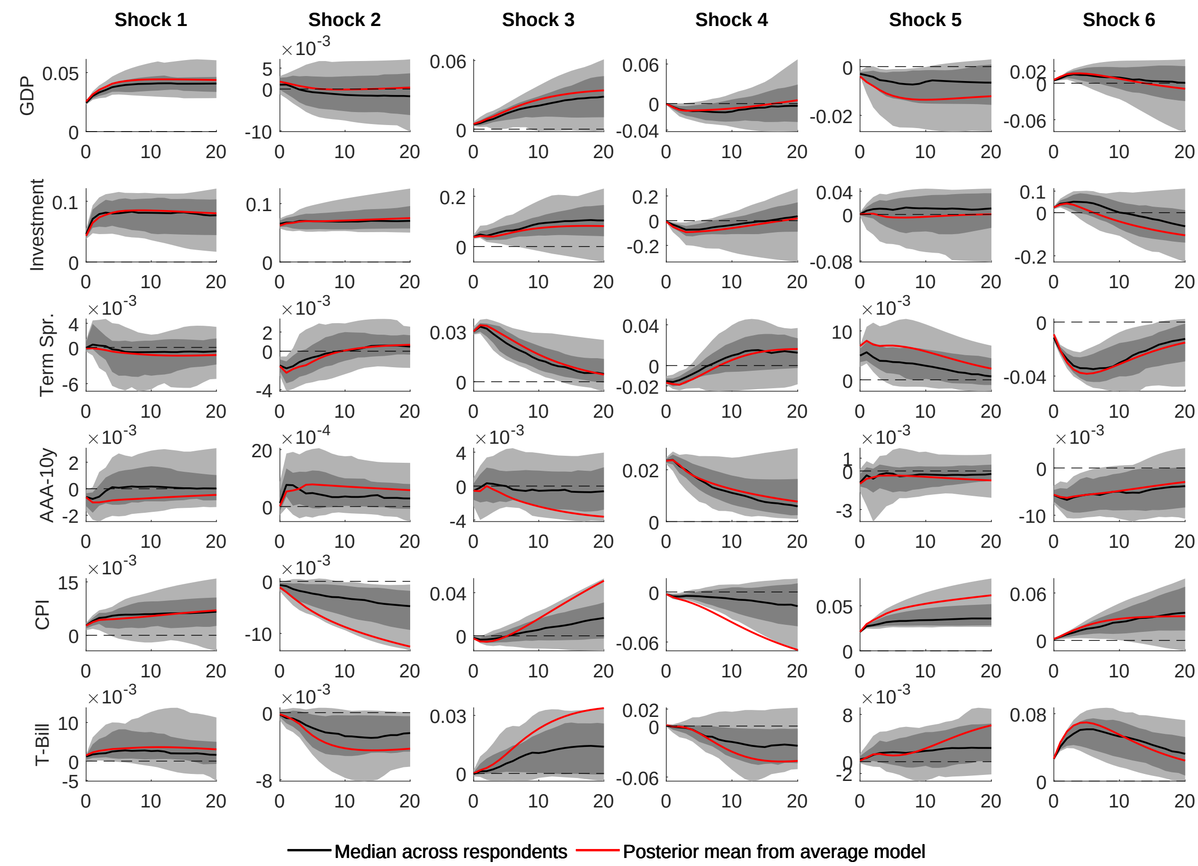
The percentile bands are quite wide. Our reading of the result is that there is evidence of differences in structural parameters across respondents. It is plausible to expect that forecasters use different tools, information and methods (which may change over time) to produce forecasts, leading to differences in the dynamics of the responses. As explained in Section 2, we have to make some assumptions when specifying our model for the individual respondent, both about the exact specification and the observed information set. While these assumptions could bias our results towards relatively homogeneous dynamics, as they somehow restrict respondents' heterogeneity, we find that this is not the case when looking at the cross-section of respondents.

Nevertheless, we are cautious about interpreting these results as definitive, but rather as indicative. The data in the SPF are subject to attrition and sometimes infrequent responses from forecasters: the last two columns of Table 1 give an indication of the number of observations, i.e., the average number of quarters in which each respondent reported the forecast for a variable and the average number of respondents at each point in the sample. While we deal with these issues rigorously by using interpolation within estimation, differences in sample size between individuals still introduce additional noise that can affect the structural parameters.

5.3 Analysis of disagreement

What is the relative role of each structural shock in explaining overall disagreement across respondents? We answer this question by analysing the cross-sectional dispersion across respondents, decomposed into items reflecting structural components of the forecaster's judgement and heterogeneity of unconditional forecast arising from heterogeneous model coefficients. We calculate disagreement as the standard deviation of the individual point forecasts. We exclude the two smallest and largest values to minimise the effect of outliers, however, our results change only slightly when we consider all respondents, or we exclude

Figure 6: Impulse responses, individual models



Note: The figure shows impulse response functions for the individual models. Each sub-panel shows the response of a variable (in the rows) to a shock (in the columns), normalised to increase the variable on the diagonal by one standard deviation. The dark and light grey areas are the 68% and 90% percentile bands, respectively.

only the smallest and largest value.

Based on equation 14, we deconstruct structural assumptions into contributions associated with each structural shock, $\varepsilon_{t+j|t,i,l}$, $l = 1, ..., N$:

$$y_{t+h|t,i} = \underbrace{\left(\sum_{j=0}^{h-1} \beta_i^j\right) c_i + \beta_i^h y_t}_{y_{h,t,i}^{(uf)} \text{ (uncond. forecast)}} + \underbrace{\sum_{j=1}^{h} \psi_{h-j,i,1} \varepsilon_{t+j|t,i,1}}_{\tilde{\varepsilon}_{h,t,i}^{(1)} \text{ (shock 1 judgement)}} + ... + \underbrace{\sum_{j=1}^{h} \psi_{h-j,i,N} \varepsilon_{t+j|t,i,N}}_{\tilde{\varepsilon}_{h,t,i}^{(N)} \text{ (shock N judgement)}} \tag{15}$$

where the coefficient in front of each structural shock assumption at each future horizon $j = 1, ..., h$ represents the impulse response to structural shocks, such that $\psi_{h-j,i,q}$ is the q^{th} column of matrix $\beta_i^{h-j} A_{0,i}^{-1}$. Note that all parameters in the equation represent posterior mean estimates, whereas, for shocks, we use a smoothed posterior mean as our point estimates. We define disagreement as the cross-sectional variance across respondents and, following Herbst and Winkler (2021), decompose it into *empirical* cross-sectional covariances for each component of equation 15:

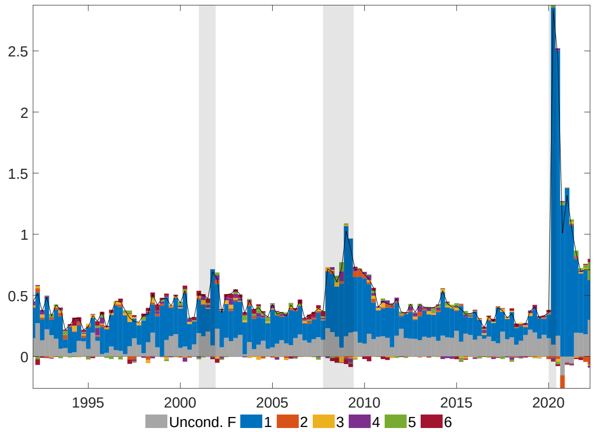
$$\widehat{\mathbb{V}}_i(y_{t+h|t,i}) = \widehat{\mathbb{C}ov}_i(y_{t+h|t,i}, y_{h,t,i}^{(uf)}) + \widehat{\mathbb{C}ov}_i(y_{t+h|t,i}, \tilde{\varepsilon}_{h,t,i}^{(1)}) + ... + \widehat{\mathbb{C}ov}_i(y_{t+h|t,i}, \tilde{\varepsilon}_{h,t,i}^{(N)}) \tag{16}$$

We present the cross-sectional dispersion across forecasters as the empirical standard deviation for better readability in our plots. Therefore, we divide both sides of equation 16 by the standard deviation of $y_{t+h|t,i}$.

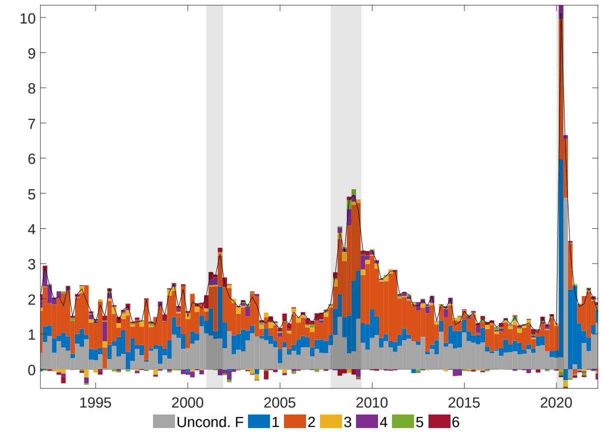
Figure 7 presents historical decompositions of one-year-ahead disagreement for all variables. Grey bars represent disagreement about unconditional forecasts, in other words, disagreement arising from differences in coefficients among respondents. Coloured bars represent disagreement about the contribution of each shock, expressed as the covariance between contribution and variable, as explained above.

A first point to note is that disagreement for each variable comes mostly from differences in judgement, as opposed to heterogeneity in model coefficients. More precisely, the latter contributes to between 15% and 30% of total disagreement, depending on the variable. Second, recessions are periods when disagreement spikes for some variables, with the COVID pandemic an unprecedented example in our sample for GDP, investment, and the AAA spread. Disagreement about CPI inflation increases during COVID, too, to a similar extent to the period following the GFC. Term spread and T-bill have less discernible patterns of disagreement, with the one of the latter decreasing markedly in the aftermath of the GFC, then rising again to about half the pre-crisis levels.

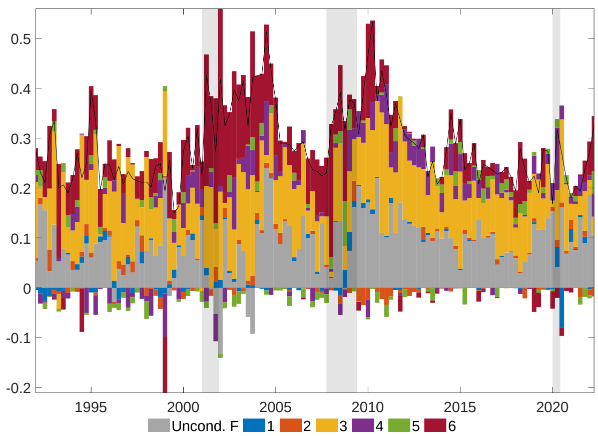
Figure 7: Historical decomposition of one-year-ahead disagreement



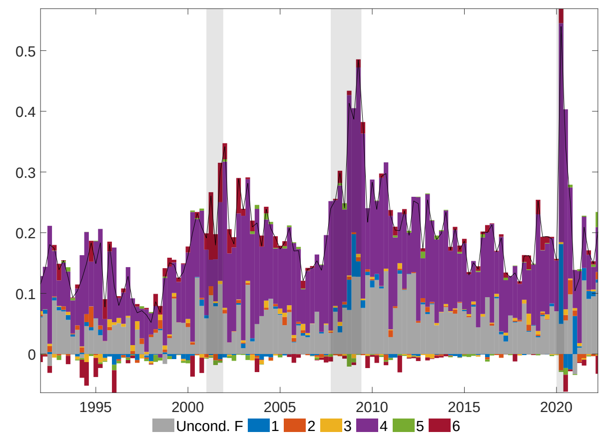
(a) Real GDP, y-o-y growth rate



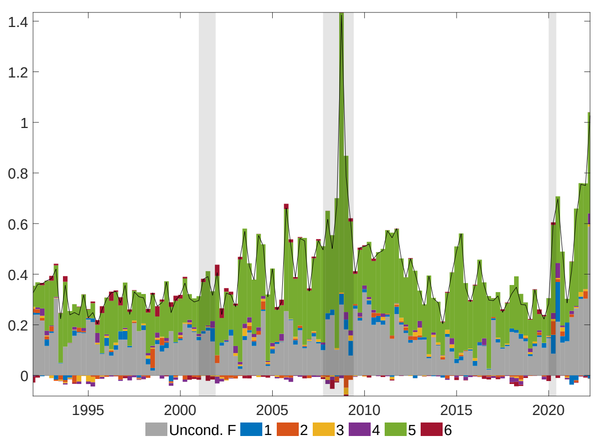
(b) Real Investment, y-o-y growth rate



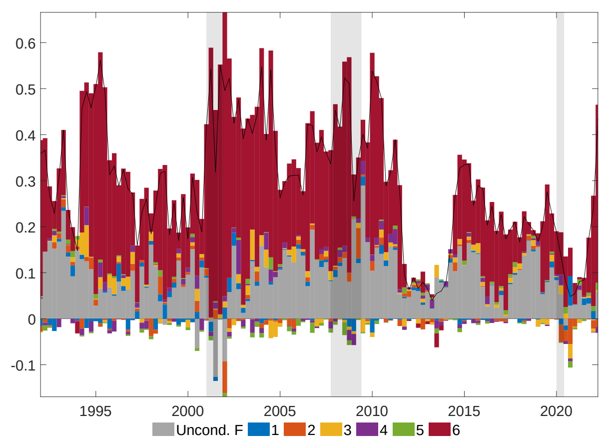
(c) Term spread



(d) AAA-10y spread



(e) CPI, y-o-y log-growth rate



(f) T-bill

Note: The figure shows the historical decomposition of the one-year-ahead disagreement, calculated as the standard deviation of the individual point forecasts, excluding the two smallest and largest values. Shaded areas represent NBER recession periods. Numbers in the legends correspond to the shock numbering of Figure 2.

Third, forecasters seem to agree on which shock contributes the most to a variable and mostly disagree on its size: this is particularly striking for GDP and CPI, where most disagreement is attributed to unanticipated demand and cost-push shock (1 and 5), respectively. Other variables incorporate disagreement about multiple shocks, namely: for investment, supply and to a smaller extent unanticipated demand (2 and 1); for the term spread, anticipated demand and interest rate, to a smaller extent the financial shock (3,6 and 4); for the AAA-spread, financial and to a smaller extent unanticipated demand and interest rate; for the T-bill, interest rate and to a smaller extent anticipated demand. The unanticipated demand shock is incorporated in most variables' disagreements during the COVID pandemic: forecasters agreed this shock would affect GDP, investment, the AAA spread and CPI inflation in the first quarter of 2021.

The fact that respondents, for the most part, agree on the type of shock affecting the volatility of each variable may be surprising but can be rationalised by looking at another exercise, namely the one-year-ahead forecast error variance decomposition for the average respondent. In the Appendix, we present the decomposition in the upper panel of Figure F.4. The figure confirms the importance of predominantly one shock for each variable forecast's variance, with the exception of the term spread. Our intuition for this finding is that the underlying model specifications across forecasters, on average, agree that the short-term forecast variation of one variable is mainly driven by one structural force. As a consequence, disagreement about expected developments is also guided by one structural innovation. To fix ideas, one can establish an analytical correspondence between disagreement and forecast variance decomposition on the basis of our framework with no coefficient heterogeneity. In such a model, all the variation in a specific forecast would come from the variation in the shocks. Equation 16 without heterogeneity in coefficients asymptotically reflects a standard forecast error variance decomposition.

A final, less crucial, point concerns the composition of the dataset on which disagreement is calculated. Because of missing observations in the survey, in our estimation we interpolate data and forecasts via the simulation smoother. This allows to "fill in the gaps" whenever a respondent did not return the survey for some or all the variables. Without additional forecast information, the procedure of smoothing presumes zero expected future shocks, in other words, assumes the respondent is performing an unconditional forecast. So the smoother recovers an unconditional forecast in case there is *no* forecast information available, neither across different horizons nor variables. Alternatively, if the forecast information is available for *some* variables, the smoother ensures that some combination of judgement generates conditional forecasts. Since the interpolated forecasts are unconditional (i.e. have zero judgement), taking them into account when calculating disagreement would give more weight to coefficients and less weight to judgement in an artificial way. For this reason, we deem it more correct to focus on disagreement based only on observed forecasts without including the smoothed observations (see Figure 7). For comparison, Figure F.9 in the Appendix shows results using interpolated data. While the main messages are robust when using the alternative method, there are specific periods where disagreement in unconditional forecasts spikes, for example, for the term spread and the T-bill during the 2001 crisis.

6 Conclusions

We develop a novel framework to structurally interpret the term structure of professional forecasts and to analyse their heterogeneity. The framework allows forecasts to be incorporated into a VAR model in a parsimonious way, i.e. without increasing the number of parameters to be estimated, and it can account for heterogeneity across respondents due to both different model coefficients and different expected future shocks. In our main specification, we consider a six-variable VAR model and identify five shocks through stochastic volatility. The shocks are identified post-estimation, both by looking at the impulse responses and by analysing their correlation with standard shocks in the literature. We discuss the importance of these shocks for the SPF forecasts at the average and individual level. For the average respondent, we find a substantial contribution of judgement about the current quarter, which often biases unconditional forecasts towards the realisation, thereby improving accuracy. We decompose the individual forecast variance into a component due to differences in unconditional forecasts (from different coefficients or models) and components due to differences in judgement. We find that the former has a relatively lower importance, accounting for about one third of the observed forecast, while the remaining disagreement is due to different subjective future shocks. Respondents tend to agree on the most relevant shock for the forecast of a variable and disagree mainly on its magnitude. Our findings help to shed light on the expectation formation process of professional forecasters in an empirical setting. Moreover, our framework can serve as a powerful tool for policymakers to identify in real time which shocks professionals expect to affect macroeconomic aggregates and to what extent they agree on the size and nature of these shocks.

References

Aastveit, K. A., Ravazzolo, F., & van Dijk, H. K. (2018). Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics*, *36*(1), 131–145. https://doi.org/10.1080/07350015.2015.1137760

Abel, J., Rich, R., Song, J., & Tracy, J. (2016). The measurement and behavior of uncertainty: Evidence from the ECB survey of professional forecasters. *Journal of Applied Econometrics*, *31*(3), 533–550. https://doi.org/10.1002/jae.2430

Andrade, P., Crump, R. K., Eusepi, S., & Moench, E. (2016). Fundamental disagreement. *Journal of Monetary Economics*, *83*, 106–128. https://doi.org/10.1016/j.jmoneco.2016.08.007

Andrade, P., & Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, *60*(8), 967–982.

Andre, P., Pizzinelli, C., Roth, C., & Wohlfart, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies*, rdac008. https://doi.org/10.1093/restud/rdac008

Antolín-Díaz, J., Petrella, I., & Rubio-Ramírez, J. F. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics*, *117*, 798–815. https://doi.org/10.1016/j.jmoneco.2020.06.001

Bańbura, M., Brenna, F., Paredes, J., & Ravazzolo, F. (2021). Combining bayesian VARs with survey density forecasts: Does it pay off? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3838719

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92. https://doi.org/10.1002/jae.1137

Bassett, W. F., Chosak, M. B., Driscoll, J. C., & Zakrajšek, E. (2014). Changes in bank lending standards and the macroeconomy. *Journal of Monetary Economics*, *62*, 23–40. https://doi.org/10.1016/j.jmoneco.2013.12.005

Baumeister, C. (2023, January 1). Measuring market expectations. In R. Bachmann, G. Topa, & W. van der Klaauw (Eds.), *Handbook of economic expectations* (pp. 413–441). Academic Press. https://doi.org/10.1016/B978-0-12-822927-9.00022-7

Bertsche, D., & Braun, R. (2022). Identification of structural vector autoregressions by stochastic volatility. *Journal of Business & Economic Statistics*, *40*(1), 328–341. https://doi.org/10.1080/07350015.2020.1813588

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685. https://doi.org/10.3982/ECTA6248

Boero, G., Smith, J., & Wallis, K. F. (2015). The measurement and characteristics of professional forecasters' uncertainty. *Journal of Applied Econometrics*, *30*(7), 1029–1046. https://doi.org/10.1002/jae.2400

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, *110*(9), 2748–2782. https://doi.org/10.1257/aer.20181219

Born, B., Dovern, J., & Enders, Z. (2020). Expectation dispersion, uncertainty, and the reaction to news. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3765299

Brenna, F. (2022). *Behind the scenes of expectations: Interpreting survey forecasts*. https://fedbre.github.io/federicabrenna/JMP.pdf

Brunnermeier, M., Palia, D., Sastry, K. A., & Sims, C. A. (2021). Feedbacks: Financial markets and economic activity. *American Economic Review*, *111*(6), 1845–1879. https://doi.org/10.1257/aer.20180733

Caldara, D., Cavallo, M., & Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, *103*, 1–20. https://doi.org/10.1016/j.jmoneco.2018.08.004

Canova, F., & Paustian, M. (2011, July). *Business cycle measurement with some theory* (No. 1203). Department of Economics and Business, Universitat Pompeu Fabra. Retrieved May 13, 2021, from https://ideas.repec.org/p/upf/upfgen/1203.html

Carriero, A., Chan, J., Clark, T. E., & Marcellino, M. (2022). Corrigendum to "large bayesian vector autoregressions with stochastic volatility and non-conjugate priors" [j. econometrics 212 (1) (2019) 137–154]. *Journal of Econometrics*, *227*(2), 506–512. https://doi.org/10.1016/j.jeconom.2021.11.010

Carriero, A., Clark, T. E., & Marcellino, M. (2016). Common drifting volatility in large bayesian VARs. *Journal of Business & Economic Statistics*, *34*(3), 375–390. https://doi.org/10.1080/07350015.2015.1040116

Carriero, A., Clark, T. E., Marcellino, M., & Mertens, E. (2021, March 25). *Addressing COVID-19 outliers in BVARs with stochastic volatility* (DP15964). Retrieved February 20, 2023, from https://cepr.org/publications/dp15964

Casey, E. (2021). Are professional forecasters overconfident? *International Journal of Forecasting*, *37*(2), 716–732. https://doi.org/10.1016/j.ijforecast.2020.09.002

Chan, J. C. C. (2021). Minnesota-type adaptive hierarchical priors for large bayesian VARs. *International Journal of Forecasting*, *37*(3), 1212–1226. https://doi.org/10.1016/j.ijforecast.2021.01.002

Chan, J. C. C., Koop, G., & Yu, X. (2021). Large order-invariant bayesian VARs with stochastic volatility, 34.

Chan, J. C., Pettenuzzo, D., Poon, A., & Zhu, D. (2023). Conditional forecasts in large bayesian VARs with multiple soft and hard constraints. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4358152

Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, *29*(3), 327–341. https://doi.org/10.1198/jbes.2010.09248

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility: Forecasting performance and specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575. https://doi.org/10.1002/jae.2379

Clements, M. P. (2022). Forecaster efficiency, accuracy, and disagreement: Evidence using individual-level survey data. *Journal of Money, Credit and Banking*, *54*(2-3), 537–568.

Clements, M. P., Rich, R. W., & Tracy, J. S. (2022). Surveys of professionals. *Federal Reserve Bank of Cleveland, Working Paper*. https://doi.org/https://doi.org/10.26509/frbc-wp-202213

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302. https://doi.org/10.1016/j.red.2004.10.009

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, *120*(1), 116–159. https://doi.org/10.1086/665662

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–2678. https://doi.org/10.1257/aer.20110306

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, *105*(1), 111–130. https://doi.org/10.1016/S0304-4076(01)00072-0

Croushore, D., & Stark, T. (2019). Fifty years of the survey of professional forecasters. *Economic Insights*, *4*(4), 1–11.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101. https://doi.org/10.1002/jae.1257

Del Negro, M., Casarin, R., & Bassetti, F. (2022). A bayesian approach to inference on probabilistic surveys. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4164182

Dovern, J. (2015). A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data. *European Economic Review*, *80*, 16–35. https://doi.org/10.1016/j.euroecorev.2015.08.009

Dovern, J., & Hartmann, M. (2017). Forecast performance, disagreement, and heterogeneous signal-to-noise ratios. *Empirical Economics*, *53*(1). https://doi.org/10.1007/s00181-016-1137-x

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–615. Retrieved August 12, 2022, from http://www.jstor.org/stable/4140605

ECB. (2019). The ECB survey of professional forecasters forecast processes and methodologies: Results of the 2018 special survey. *The ECB Survey of Professional Forecasters (SPF)*. Retrieved March 10, 2023, from https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/pdf/ecb.spf201902_specialsurvey~7275f9e7e6.en.pdf

Falck, E., Hoffmann, M., & Hürtgen, P. (2021). Disagreement about inflation expectations and monetary policy transmission. *Journal of Monetary Economics*, *118*, 15–31.

Farmer, L., Nakamura, E., & Steinsson, J. (2021). *Learning about the long run* (tech. rep.). National Bureau of Economic Research.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., & Watson, M. W. (2007). ABCs (and ds) of understanding VARs. *American Economic Review*, *97*(3), 1021–1026. https://doi.org/10.1257/aer.97.3.1021

Forni, M., & Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics*, *66*, 124–136. https://doi.org/10.1016/j.jmoneco.2014.04.005

Forni, M., Gambetti, L., & Sala, L. (2019). Structural VARs and noninvertible macroeconomic models. *Journal of Applied Econometrics*, *34*(2), 221–246. https://doi.org/10.1002/jae.2665

Foroni, C., Furlanetto, F., & Lepetit, A. (2018). Labor supply factors and economic fluctuations. *International Economic Review*, *59*(3), 1491–1510. https://doi.org/10.1111/iere.12311

Furlanetto, F., Ravazzolo, F., & Sarferaz, S. (2019). Identification of financial factors in economic fluctuations. *The Economic Journal*, *129*(617), 311–337. https://doi.org/10.1111/ecoj.12520

Galvão, A. B., Garratt, A., & Mitchell, J. (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting*, *37*(3), 1247–1260. https://doi.org/10.1016/j.ijforecast.2021.02.007

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics* (4th ed.). Clarendon Press.

Giacomini, R., Skreta, V., & Turen, J. (2020). Heterogeneity, inattention, and bayesian updates. *American Economic Journal: Macroeconomics*, *12*(1), 282–309. https://doi.org/10.1257/mac.20180235

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, *102*(4), 1692–1720. https://doi.org/10.1257/aer.102.4.1692

Glas, A. (2020). Five dimensions of the uncertainty–disagreement linkage. *International Journal of Forecasting*, *36*(2), 607–627. https://doi.org/10.1016/j.ijforecast.2019.07.010

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*(2), 363–398. https://doi.org/10.1016/S0304-4076(02)00207-5

Hamilton, J. D., Waggoner, D. F., & Zha, T. (2007). Normalization in econometrics. *Econometric Reviews*, *26*(2), 221–252. https://doi.org/10.1080/07474930701220329

Herbst, E., & Winkler, F. (2021, July 1). *The factor structure of disagreement* (SSRN Scholarly Paper No. 3899646). Social Science Research Network. Rochester, NY. https://doi.org/10.17016/FEDS.2021.046

Howrey, E. P. (1996). Forecasting gnp with noisy data: A case study. *Journal of Economic and Social Measurement*, *22*(3), 181–200. https://api.semanticscholar.org/CorpusID:152830609

Jarociński, M. (2010). Responses to monetary policy shocks in the east and the west of europe: A comparison. *Journal of Applied Econometrics*, *25*(5), 833–868. https://doi.org/10.1002/jae.1082

Jarociński, M., & Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, *12*(2), 1–43. https://doi.org/10.1257/mac.20180090

Justiniano, A., Primiceri, G. E., & Tambalotti, A. (2011). Investment shocks and the relative price of investment. *Review of Economic Dynamics*, *14*(1), 102–121. https://doi.org/10.1016/j.red.2010.08.004

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, *111*(4), 1092–1125. https://doi.org/10.1257/aer.20190964

Kastner, G., & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423. https://doi.org/10.1016/j.csda.2013.01.002

Kohlhas, A. N., & Walther, A. (2021). Asymmetric attention. *American Economic Review*, *111*(9), 2879–2925. https://doi.org/10.1257/aer.20191432

Kuang, P., Tang, L., Zhang, R., & Zhang, T. (2020). Forecast disagreement about long-run macroeconomic relationships. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3614426

Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica*, *81*(3), 1115–1145. https://doi.org/10.3982/ECTA8337

Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*, *37*(4), 688–699. https://doi.org/10.1002/jae.2895

Lewis, D. J. (2021). Identifying shocks via time-varying volatility. *The Review of Economic Studies*, *88*(6), 3086–3124. https://doi.org/10.1093/restud/rdab009

Lippi, M., & Reichlin, L. (1994). VAR analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, *63*(1), 307–325. https://doi.org/10.1016/0304-4076(93)01570-C

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38. https://doi.org/10.2307/1391384

Lütkepohl, H., Meitz, M., Netšunajev, A., & Saikkonen, P. (2021). Testing identification via heteroskedasticity in structural vector autoregressive models. *The Econometrics Journal*, *24*(1), 1–22. https://doi.org/10.1093/ectj/utaa008

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, *18*, 209–248. https://doi.org/10.1086/ma.18.3585256

Mertens, K., & Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks. *American Economic Journal: Economic Policy*, *4*(2), 145–181. https://doi.org/10.1257/pol.4.2.145

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, *103*(4), 1212–47. https://doi.org/10.1257/aer.103.4.1212

Miranda-Agrippino, S., & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107. https://doi.org/10.1257/mac.20180124

Patton, A. J., & Timmermann, A. (2010). Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, *57*(7), 803–820. https://doi.org/10.1016/j.jmoneco.2010.07.001

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, *72*(3), 821–852. https://doi.org/10.1111/j.1467-937X.2005.00353.x

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In *Handbook of macroeconomics* (pp. 71–162). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.03.003

Ricco, G., Callegari, G., & Cimadomo, J. (2016). Signals from the government: Policy disagreement and the transmission of fiscal shocks. *Journal of Monetary Economics*, *82*, 107–118. https://doi.org/10.1016/j.jmoneco.2016.07.004

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics*, *85*(4), 777–792.

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, *100*(3), 763–801. https://doi.org/10.1257/aer.100.3.763

Rudebusch, G. D., & Swanson, E. T. (2012). The bond premium in a DSGE model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, *4*(1), 105–143. https://doi.org/10.1257/mac.4.1.105

Schlaak, T., Rieth, M., & Podstawski, M. (2023). Monetary policy, external instruments, and heteroskedasticity. *Quantitative Economics*, *14*(1), 161–200. https://doi.org/10.3982/QE1511

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/https://doi.org/10.1016/S0304-3932(03)00029-1

Stark, T. (2013). SPF panelists' forecasting methods: A note on the aggregate results of a november 2009 special survey. *Federal Reserve Bank of Philadelphia*. https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-special-survey-on-forecast-methods.pdf?la=en&hash=DA9492A3DE5E3BF70D40F807B7278C83

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics*, *81*(4), 639–651. https://doi.org/10.1162/003465399558508

Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. *Annual Review of Economics*, *5*(1), 303–346. https://doi.org/10.1146/annurev-economics-080511-110857

Zellner, A., & Hong, C. (1992). Forecasting international growth rates using bayesian shrinkage and other procedures. In P. K. Goel & N. S. Iyengar (Eds.), *Bayesian analysis in statistics and econometrics* (pp. 327–352). Springer. https://doi.org/10.1007/978-1-4612-2944-5_22

A Estimation algorithm

In this section, we summarise the steps necessary to estimate the model. As mentioned in the main text, we rely on Bayesian inference, requiring Gibbs sampling across different parameters and latent variables to approximate marginal posterior distributions. The algorithm is in line with previous efforts to estimate models with stochastic volatility (most notably Cogley & Sargent, 2005). However, we deviate by first, relying on a recent contribution by Chan et al. (2021) to account for the full parameterisation of the structural matrix and second, by fully modelling SPF forecasts within the assumed VAR specification - an addition that introduces a few complexities.

The aim of the algorithm is to obtain draws for parameters β[29], $A_0, \rho, \sigma_u^2, S_{\sigma_{u,i}^2}$, initial conditions y_0 and matrices of latent variables $y = [Y_{1+h|1}, ..., Y_{T+h|T}]$ and $\lambda = [\lambda_1, ..., \lambda_T]$. Cycling through the following sampling steps provides the posterior estimates.

A.1 Interpolating missing observations and drawing initial conditions

$p(y, y_0|y^{obs}, \lambda, \beta, A_0, \bar{y_0}, V_{\bar{y_0}})$

This is the interpolation step for missing data y, given observed data, y^{obs}. The normality of the conditional distribution is preserved given the parameters, other latent variables and normally distributed priors for initial conditions. We cast our model in a state-space representation, see equation 5, such that the interpolated series can then be sampled following the simulation smoother of Durbin and Koopman (2002) and Jarociński (2015).[30]

A.2 Drawing coefficients

$p(\beta|y, y_0, \lambda, A_0)$

We assume that agents produce conditional forecasts according to the following specification:

$$y_{t+h|t} = (1 + \beta + ... + \beta^{h-1})c + \beta^h y_t + A_0^{-1}\Lambda_{t+h|t}e_{t+h|t} + \beta A_0^{-1}\Lambda_{t+h-1|t}e_{t+h-1|t} + ... + \beta^{h-1}A_0^{-1}\Lambda_{t+1|t}e_{t+1|t} \quad \text{(A.1)}$$

[29]For the sake of brevity, we include the constant, c, in the coefficient vector, β.

[30]We also tried to use the novel approach proposed by Chan et al. (2023), but it did not lead to any efficiency gains in our application.

We describe the one lag case, to give an intuition of how we construct the algorithm. This can easily be extended to a case with more than one lag by writing the expression in companion form. Note that coefficients appear in a non-linear form: they both scale the current information set and capture the dynamic effects of judgment or shocks expected in future horizons. For this reason, the conditional distribution of coefficients for this specification does not have a well-known form from which one can easily draw. To obtain draws, one could rely on the random-walk Metropolis-Hastings algorithm with a Gaussian as the importance distribution. However, this comes at the cost of lower sampling efficiency. Instead, we acknowledge that under our assumptions of independent judgement (see equation 4), the above specification adheres to the iterative dynamics that are usual for VAR frameworks.

This assumption allows us to cast our specification in the following form:

$$y_{t+h|t} = c + \beta y_{t+h-1|t} + A_0^{-1} \Lambda_{t+h|t} e_{t+h|t}, \tag{A.2}$$

Given the assumption of independence, we can iterate the previous equation backwards:

$$\begin{aligned} y_{t+h-1|t} &= c + \beta y_{t+h-2|t} + A_0^{-1} \Lambda_{t+h-1|t} e_{t+h-1|t}, \\ &\dots \\ y_{t+1|t} &= c + \beta y_t + A_0^{-1} \Lambda_{t+1|t} e_{t+1|t} \end{aligned}$$

Compared to equation A.1, equation A.2 (and subsequent ones) have nothing in the information set to predict further horizons, conditional on the forecasts of shorter horizons. Furthermore, the coefficients enter in the linear form. Both points allow the likelihood to be expressed in a standard way with a known form for the conditional distribution of coefficients.

We preserve these relationships across forecast horizons, in addition to the autoregressive process of the observed data in a VAR system by an adequate construction of matrices representing dependent and independent variables.

$$Y' = \begin{bmatrix} y_{t+h|t} & y_{t+h-1|t} & \dots & y_{t+1|t} & y_t \end{bmatrix}' \quad X' = \begin{bmatrix} y_{t+h-1|t} & y_{t+h-2|t} & \dots & y_t & y_{t-1} \end{bmatrix}' \tag{A.3}$$

For the Gibbs step, standard sampling techniques for Bayesian VAR regressions can be used. In this case, we prefer to use the algorithm developed by Chan et al. (2021) and adapted from Carriero et al. (2019), Carriero et al. (2022) for the case where the structural coefficient matrix A_0 is full. This algorithm samples the reduced form parameters 'row by row', reducing computational complexity and increasing efficiency.

A.3 Drawing the structural impact matrix

$p(A_0|y, y_0, \lambda, \bar{A}_0, V_{\bar{A}_0})$

To draw the impact matrix A_0, we apply the algorithm of Chan et al. (2021) for our case with conditional forecasts in the vector autoregression model. To briefly summarise the derivation, note that the matrix of forecast errors, U, is expressed as:

$$U = F_1(I_{h+1} \otimes A_0^{-1})E \tag{A.4}$$

where U is $Nh^* \times T$ matrix of forecast errors; $F_1 = F^{[1:Nh^*,:]}$ and where $h^* \equiv h + 1$ for the sake of brevity. Note that Chan et al. (2021) propose to take the draws of A_0 row by row, so we rearrange the equation to align the conditional distribution:

$$U^* a_i = vec(E'^{[i:N:Nh^*]}) \tag{A.5}$$

where subscripts and superscripts i denote the row and column of a matrix. Superscript $[i : N : Nh^*]$ stands for every Nth column of the matrix starting from column i. The matrix U^* represents the rearrangement of $U'F_1'^{-1}$ s.t.

$$U^* = \begin{bmatrix} (U'F_1'^{-1})^1 & \dots & (U'F_1'^{-1})^N \\ (U'F_1'^{-1})^{N+1} & \dots & (U'F_1'^{-1})^{2N} \\ \dots & \dots & \dots \\ (U'F_1'^{-1})^{N(h^*-1)+1} & \dots & (U'F_1'^{-1})^{h^*N} \end{bmatrix} \tag{A.6}$$

The vector $vec(E'^{i:N:Nh^*})$ follows a normal distribution with mean zero and variance-covariance matrix of dimensions $Th^* \times Th^*$ s.t.

$$diag(e^{2\lambda_{1,i}}, ..., e^{2\lambda_{T,i}}, e^{2\lambda_{2|1,i}}, ..., e^{2\lambda_{T+1|T,i}}, e^{2\lambda_{m-1|1,i}}, ..., e^{2\lambda_{T+m|T,i}}) \tag{A.7}$$

It can be shown that the conditional of $p(a_i|a_{-i}, y, y_0, \lambda, \bar{A}_0, V_{\bar{A}_0})$ is a non-standard distribution but can be efficiently drawn from using the algorithm suggested by Villani (2009) as a combination of draws from Gaussian and absolute normal distributions. We refer the reader to Chan et al. (2021) for more details.

A.4 Drawing the log-volatilities

$p(\lambda|y, y_0, A_0, \bar{\lambda_1}, V_{\bar{\lambda_1}}, \rho, \sigma_u^2)$

The algorithm to draw log-volatilities λ is rather standard apart from the fact that we have to account for the expected volatility. The system of equations, respectively, observation and transition equations, can be summarised as follows:

$$\log(e^2_{t+j|t,i} + \bar{c}) = 2\lambda_{t+j|t,i} + \log(\epsilon_{t+j|t,i}) \qquad \forall j = 0, ..., h \tag{A.8}$$

$$\lambda_{t+1,i} = \rho_i \lambda_{t,i} + u_{t+1,i} \tag{A.9}$$

where $\bar{c}$ is an offset constant to robustify the estimation due to possible computational errors. For the sake of simplicity, we assume that the expected volatility follows an unconditional path, s.t. $\lambda_{t+j|t,i} = \rho_i^j \lambda_{j,i}$ or $u_{t+j|t,i} = 0$, $\forall j = 1, ..., h$. This allows us to simplify the system to one latent variable and $h + 1$ observations.

$$\log(e^2_{t+j|t,i} + \bar{c}) = 2\rho_i^j \lambda_{t,i} + \log(\epsilon_{t+j|t,i}) \qquad \forall j = 0, ..., h \tag{A.10}$$

$$\lambda_{t+1,i} = \rho_i \lambda_{t,i} + u_{t+1,i} \tag{A.11}$$

To draw latent log-volatilities, we combine the auxiliary mixture sampler of Kim et al. (1998) with the precision sampler of Chan and Jeliazkov (2009), ensuring high pace and little computational burden.

A.5 Drawing the variance for AR process of log-volatilities

$p(\sigma_u^2|\lambda, \rho, v_{\bar{\sigma_u^2}}, S_{\sigma_u^2})$

Since we assume that log-volatilities, λ_t, are conditionally independent, elements of σ_u^2 can be drawn one-by-one from the conditional:

$$p(\sigma^2_{u,i}|\lambda_i, \rho_i, v_{\bar{\sigma^2_{u,i}}}, S_{\sigma^2_{u,i}}) \sim \mathcal{IG}\left(v_{\bar{\sigma^2_{u,i}}} + \frac{T}{2}, S_i\right) \tag{A.12}$$

$$S_i = S_{\sigma^2_{u,i}} + 0.5 \sum_{t=1}^{T} (\lambda_{i,t} - \rho_i \lambda_{i,t-1})^2 \tag{A.13}$$

A.6 Drawing the hierarchical parameter for the variance

$p(S_{\sigma_u^2}|\sigma_u^2, v_{\bar{\sigma_u^2}}, \theta_1, \theta_2)$

The conditional distribution is of known-form, whereas individual elements are independent.

$$p(S_{\sigma^2_{u,i}}|\sigma^2_{u,i}, v_{\sigma^{\bar{2}}_u}, \theta_1, \theta_2) \sim \mathcal{G}\left(\theta_1 + v_{\sigma^{\bar{2}}_{u,i}}, \theta_2 + (\sigma^2_{u,i})^{-1}\right) \tag{A.14}$$

A.7 Drawing the autoregressive coefficient for AR process of log-volatilities

$p(\rho|\lambda, \sigma^2_u, \bar{\rho}, V_{\bar{\rho}})$

This is a nonstandard conditional distribution. We follow Chan and Hsiao (2014) to implement an independence-chain Metropolis Hastings with a proposal distribution of the following form:

$$p(\rho|\lambda, \sigma^2_u) \sim \mathbb{1}_{(-1<\rho<1)}\mathcal{N}(\tilde{\rho}, V_{\tilde{\rho}}) \tag{A.15}$$

$$V_{\tilde{\rho}_i} = \left(\frac{1}{V_{\bar{\rho}_i}} + \frac{\sum \lambda^2_{i,t-1}}{\sigma^2_{u,i}}\right)^{-1} \qquad \tilde{\rho}_i = \left(\frac{\sum \lambda_{i,t}\lambda_{i,t-1}}{\sigma^2_{u,i}} + \frac{\bar{\rho}_i}{V_{\bar{\rho}_i}}\right) V_{\tilde{\rho}_i} \tag{A.16}$$

The candidate draw, ρ^*_i, is accepted with the probability $\min(1, g(\rho^*_i)/g(\rho^{j-1}_i)$, where $g(\rho_i) = (1-\rho^2_i)^{0.5}\exp(-\frac{1}{2\sigma^2_{u,i}}(1-\rho^2_i)\lambda^2_{1,i})$. Otherwise, the previous draw is kept, $\rho^*_i = \rho^{j-1}_i$.

B Efficiency gains: Monte Carlo exercise

We conduct a Monte Carlo exercise to measure the extent to which our model including the term structure of conditional forecasts provides more efficient estimates than the model without.

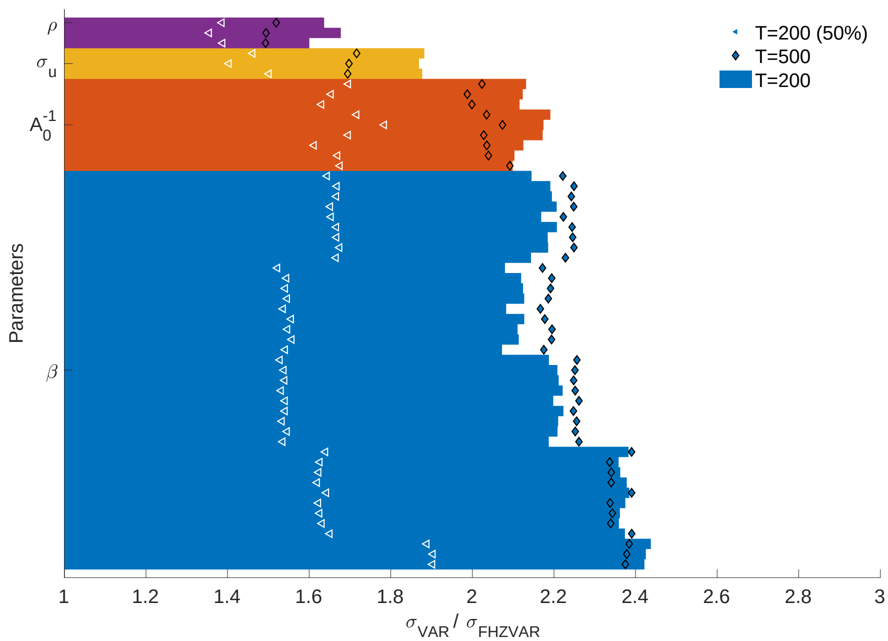
For this purpose, we generate $S = 1000$ simulations following our model's data generating process (DGP), as mentioned in Section 2. To mimic our application with SPF information, we generate forecasts for 5 periods ahead ($h = 5$) and use a VAR system with four lags ($p = 4$). The underlying parameters are different in every simulation. Coefficients, β, are obtained as follows:[31] the constants are drawn independently from a uniform distribution with boundaries -10 and 10, $U(-10, 10)$; coefficients for own first lags are from $U(0, 1)$; off-diagonal elements are from $U(-0.2, 0.2)$; elements for other lags, $j > 1$, are from an independent normal distribution with mean 0 and standard deviation $0.1/j$. The matrix of structural parameters, A_0, is obtained by taking independent draws from $U(0.5, 1.5)$, for diagonal elements and $\mathcal{N}(0, 1)$ for the off-diagonal ones. Stochastic volatility processes are calibrated such that ρ_i are drawn from $U(0.8, 0.98)$ and $\sigma^2_{u,i}$ are from $U(0.02, 0.2)$. For every simulation, we estimate our model, denoted $FHZVAR$, that includes the term structure of conditional forecasts and the VAR specification without any forecasts. The specification for priors in both models is aligned. We conduct this exercise for cases with different sample sizes ($T = 200$ and $T = 500$) and several variables ($N = 3$ and $N = 6$). To investigate the influence of missing data, we additionally conduct an experiment where around 50% of forecast information are randomly missing ($T = 200$ 50%).

We summarise efficiency gains by computing the ratio of posterior standard deviations across models, averaged across simulations s: $\frac{1}{S}\sum_s (\sigma_{VAR,s}/\sigma_{FHZVAR,s})$. A value above one indicates that our proposed model provides more precise estimates. Figure B.1 presents results for the specification with $N = 3$ in the upper panel and $N = 6$ in the lower panel. The results are as follows. First, the inclusion of additional information in the conditional forecasts leads to more accurate parameter estimates: for both the three-variable and the six-variable model, the improvements are on average by a factor of 2.15. Second, the improvement in precision is similar regardless of the sample size. The ratio of standard deviations is comparable for the two sample sizes $T = 200$ and $T = 500$, apart from small discrepancies that may result from a lower importance of priors in the larger sample. As expected, when forecast information is missing the efficiency advantage of the model with forecasts decreases to a factor of 1.58 on average. Finally, the increase in precision is evident

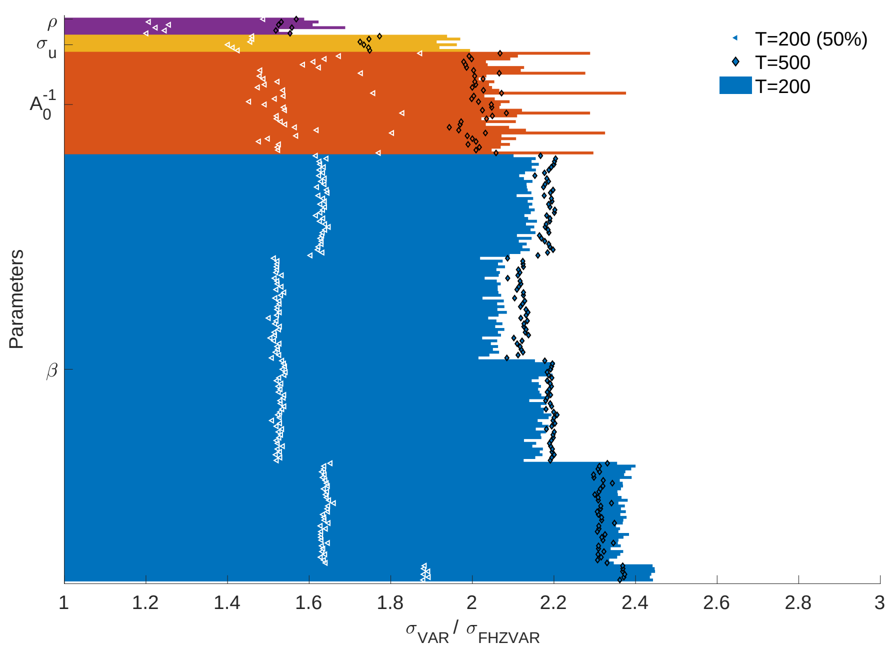
[31] We follow the Monte Carlo study by Chan et al. (2021).

for all parameters, to varying degrees for different types. In particular, efficiency is most improved for the coefficient parameters, β and the structural impact parameters in the A_0 matrix.

Figure B.1: Efficiency gains



(a) N=3

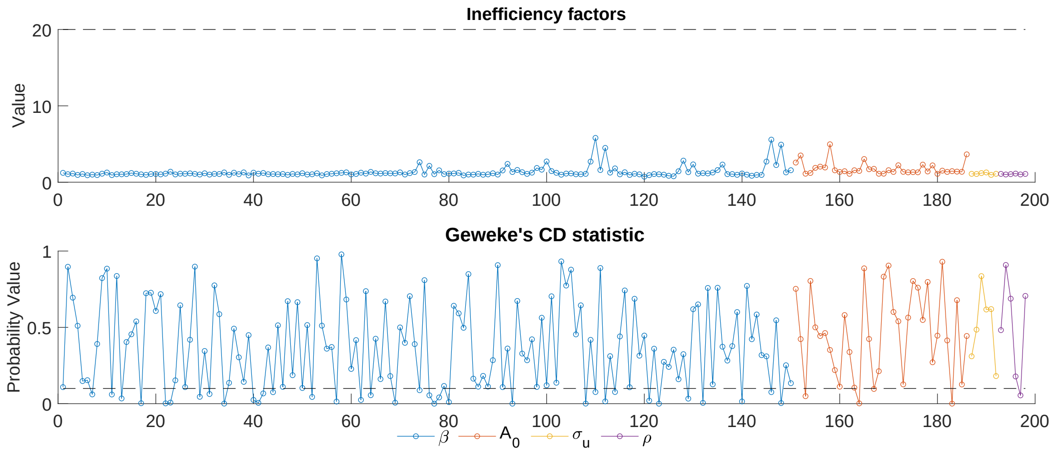


(b) N=6

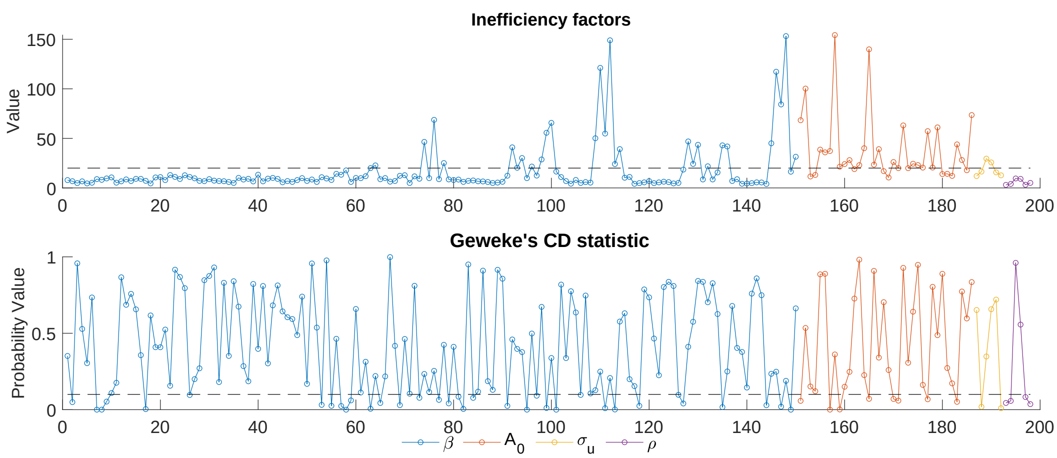
Note: The figure shows the ratio of posterior standard deviations between models, averaged over Monte Carlo simulations. Values above one indicate that the model with conditional forecasts is more efficient than the one without. The top panel shows results for the specification with three endogenous variables ($N = 3$), the bottom panel with $N = 6$. Bar charts show the specification with a sample size of 200 periods, while diamonds are for a sample of 500 and white triangles for a sample of 200, but 50% of the forecast information is missing. On the y-axis, different colours are used to indicate different types of parameters in the model.

C MCMC convergence and efficiency

Figure C.1: Convergence and mixing statistics



(a) Baseline



(b) No thinning

Note: The figure shows the inefficiency factors and convergence diagnostic statistics of Geweke (1992) for different parameters of the model, shown in different colours. The statistics are computed using a single chain and only the retained draws. The top panel shows information for the baseline specification with thinning, while the bottom panel shows results without thinning. For the inefficiency factors, the dashed black line is set at a value of 20, the threshold considered to ensure satisfactory MCMC mixing. For the Geweke's CD statistic, the dashed line is at the 10% significance level.

D External instruments and labelling structural shocks

To provide a structural interpretation of the shocks, we relate shocks w_t from the literature (which may have been estimated from a model, collected using high frequency information, or otherwise constructed using narrative evidence), to shock estimates, $\hat{\varepsilon}_t = [\exp(\hat{h}_{1,t})\hat{e}_{1,t}, ..., \exp(\hat{h}_{N,t})\hat{e}_{N,t}]$, from our model with heteroskedasticity identification:

$$w_t = \psi\hat{\varepsilon}_t + o_t \qquad\qquad o_t \sim \mathcal{N}(0, \sigma_o^2) \tag{D.1}$$

$$\hat{\varepsilon}_t \sim p(\varepsilon_t, \Sigma_{\varepsilon,t}) \tag{D.2}$$

where $p(m, v)$ and $\mathcal{N}(m, v)$ represent an arbitrary and a normal distribution with mean m and variance v; o_t is an i.i.d. measurement error. We account for the estimation errors in our structural shocks, $\hat{\varepsilon}_t$, by explicitly modelling them using the posterior distribution $p(\varepsilon_t, \Sigma_{\varepsilon,t})$ from the VAR results.[32] ε_t is the mean estimate of the smoothed posterior of the shock series; $\Sigma_{\varepsilon,t}$ is a diagonal matrix containing heteroskedasticity estimates. The latter parameters are considered to be known in the specification to infer parameters ψ and σ_o^2.

The specification has a natural explanation in terms of Bayesian updating. Estimates from the VAR with stochastic volatility provide a posterior $p(\varepsilon_t, \Sigma_{\varepsilon,t})$; instruments w_t, however, may or may not provide additional information about unobserved shocks $\hat{\varepsilon}_t$, which can be summarised in a Bayesian updating framework:

$$p(\hat{\varepsilon}_t|w_t, \psi, \sigma_o^2, \varepsilon_t, \Sigma_{\varepsilon,t}) = \frac{p(w_t|\hat{\varepsilon}_t, \psi, \sigma_o^2)p(\hat{\varepsilon}_t|\varepsilon_t, \Sigma_{\varepsilon,t})}{p(w_t)} \tag{D.3}$$

Consequently, one can infer the structural interpretation of smoothed shocks, ε_t, by observing whether they satisfy the two well-known conditions for a valid instrument. In the frequentist framework, the conditions include relevance, i.e. $\psi_k \neq 0$, and exogeneity ($\psi_i = 0$ for all $i \neq k$). In a related study Schlaak et al. (2023), the authors suggest testing for validity using the likelihood-ratio test for the unrestricted and restricted models when equation D.1 is directly incorporated into the VAR system. Instead, we propose a post-estimation procedure for the test to provide a structural interpretation, similarly to Bertsche and Braun (2022), but extended to a Bayesian setting.

[32] If the estimation errors are not accounted for, estimates are biased in line with the classical implication from models with errors in independent variables.

The posterior kernel can be summarised as follows:

$$p(\psi, \sigma_o | w_t) \propto p(w_t | \psi, \sigma_o, \varepsilon_t, \Sigma_{\varepsilon,t}) p(\psi) p(\sigma_o) \tag{D.4}$$

The likelihood can be obtained after marginalising the latent state:

$$\begin{aligned} p(w_t | \psi, \sigma_o, \varepsilon_t, \Sigma_{\varepsilon,t}) &= \int p(w_t | \psi, \sigma_o, \hat{\varepsilon}_t) p(\hat{\varepsilon}_t | \varepsilon_t, \Sigma_{\varepsilon,t}) d\hat{\varepsilon}_t & \text{(D.5)} \\ &= \mathbf{E}_{\hat{\varepsilon}_t} p(w_t | \psi, \sigma_o, \hat{\varepsilon}_t) & \text{(D.6)} \\ &\approx N(\psi \varepsilon_t, \sigma_o + \psi \Sigma_{\varepsilon,t} \psi') & \text{(D.7)} \end{aligned}$$

Therefore, the likelihood can be approximated using all draws from the posterior to obtain the expected value through Monte Carlo integration; see equation D.6. This would consider the posterior's intricacies but may be computationally intensive. The alternative is to assume that the normal distribution describes the posterior well, allowing for a convenient form of the likelihood, see equation D.7. In our application, we find that approximation errors are negligible.

Additionally, our specification includes a constant, τ, that was omitted from equation D.1 for the sake of brevity. We include it to capture a non-zero mean of the instrument. We find that the assumption is innocuous.

Priors are set to be proper but relatively uninformative:

$$\tau, \psi \sim \mathcal{N}(0_k, 0.25 \cdot I_k) \qquad \sigma_o \sim IG(3, 3) \tag{D.8}$$

We produce posterior draws using a single-block Metropolis-Hastings algorithm. Given the small system, the procedure is efficient and converges relatively fast.

The procedure allows determining which candidate shock from the vector ε_t is correlated with the suspected out-of-system shock w_t. To choose the candidate's shock, we select the one with the largest correlation in absolute value, and for which the zero value is not in the 90% credible set of coefficient ψ_i. In addition to ensuring that the candidate shock is relevant, we also explore whether the instrument is exogenous to other structural shocks. For that purpose, we compute the marginal likelihood following Chib and Jeliazkov (2001) and conduct a Bayesian model comparison.

E Permutations and cross-sectional comparison

Identification using heteroskedasticity ensures that the structural impact matrix A_0 is unique only up to column permutations and sign changes (Lewis, 2021; Bertsche & Braun, 2022). This presents some challenges when comparing different models. In our case, we want to compare estimates from individual-level information. However, the distinct sequence of draws from Gibbs sampling can lead to the posterior estimates of the structural impact matrix being permuted across respondents. For this reason, we introduce a procedure to detect permutations, which allows the results to be transformed and leads to an adequate cross-sectional comparison.

To identify permutations, we rely on the idea that current structural shocks should reflect rather similar information across respondents, since they are determined by the same observed data. This contrasts with judgement shocks, which differ across forecasters and lead to the observation of disagreement. For this reason, structural shocks associated with observed data should have a common factor structure across agents, where each structural shock for an agent should be explained by only one factor.[33] If we can determine to which factor the shock is most related, we can label that shock as similar across agents and determine the ordering. In addition, by observing whether the correlation with the factor is positive or negative, we can also determine the sign of the permutation.

In order to do this, we estimate a static factor model for structural shocks concatenated across agents:

$$\hat{E}_t = WZ_t + O_t \qquad\qquad O_t \sim \mathcal{N}(0, \sigma_o^2 I) \tag{E.1}$$

where $\hat{E}_t$ presents the stacked N number of estimated current structural errors across K number of agents and for the aggregate specification at time t:

$$\left[\hat{e}'_{agg,1,t|t}, ..., \hat{e}'_{agg,N,t|t}, \hat{e}'_{1,1,t|t}, ..., \hat{e}'_{1,N,t|t}, \hat{e}'_{2,1,t|t}, ..., \hat{e}'_{2,N,t|t}, \hat{e}'_{3,1,t|t}, ..., , \hat{e}'_{K,N,t|t}\right]' \tag{E.2}$$

W represents the factor loadings of size $N(K+1) \times l$, where l is the number of factors. Z_t is a vector of length l that represents latent factors. O_t is a vector of measurement errors, arising either due to sampling errors, as structural shocks are estimated, or idiosyncratic differences in shocks.

[33] Our conjecture does not exclude the possibility that expected shocks or judgment have a common factor structure across respondents, which could theoretically be explained by a common signal received (Herbst & Winkler, 2021; Fisher et al., n.d.). Instead, we expect the common factor to be "stronger" or to explain a larger proportion of the variation in the structural shocks associated with the observed data.

We estimate the factor model using the EM algorithm to account for missing observations across respondents due to different sample sizes. For the estimation, we use the same number of factors as the number of variables ($l = N$). We find factors to robustly explain around 87% of the variation in structural errors. The result confirms that current structural errors share a common component among respondents due to the same observed data. After the estimation, we apply the specific rotation to interpret each factor as a common structural shock. Particularly, we constrain the loading matrix, W, such that only one factor explains most of the variation in one *aggregate* structural shock:

$$W = \begin{bmatrix} I_N \\ \bar{W} \end{bmatrix} \tag{E.3}$$

We also find that results are robust to alternative rotations: e.g. "Quartimax" rotation, that ensures only one factor explains most of the variation in one estimated structural shock.

Then we label structural shocks for each individual by determining which factors explain most of the variation. More particularly, we select the factor that loads the most onto the shock q, s.t. $w_{q,max} = \max(w_{q,1}^2, w_{q,2}^2, ..., w_{q,N}^2)$. The sign is determined by obtaining the sign of the loading corresponding to the selected factor, $sign(w_{q,max})$.

We find the procedure is robust across different specifications or sampling chains. As a result, the routine allows us to determine column and sign permutations, which we apply to provide an adequate comparison of impulse responses in Section 5.2 and estimate the decomposition of disagreement among respondents in Section 5.3.

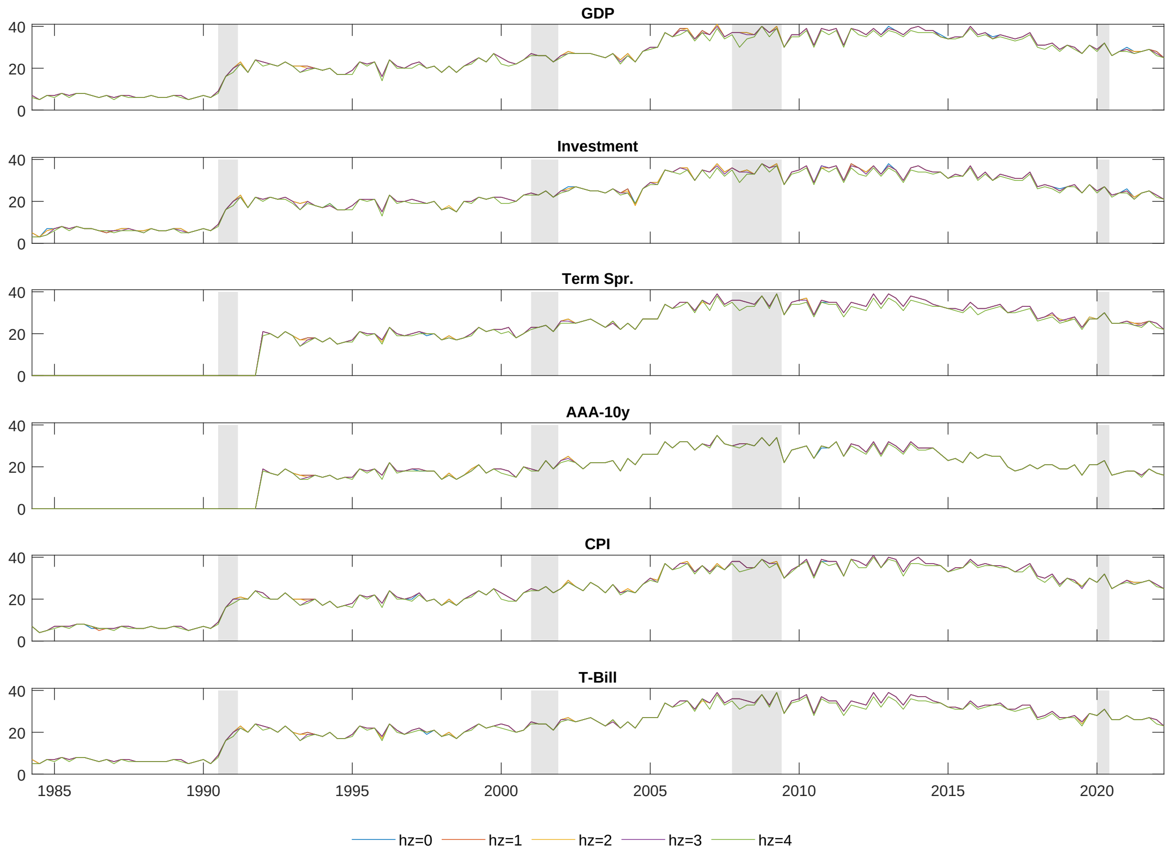
For the sake of further robustness, we also explore a different way of identifying the rotations. We look for which individual estimated shock j for individual i is most explained by the shocks from the aggregate specification by running the following regression line:

$$\hat{e}_{i,j,t|t} = \beta_1 \hat{e}_{agg,1,t|t} + ... + \beta_N \hat{e}_{agg,N,t|t} + o_{i,j,t} \tag{E.4}$$

Then we determine permutations by aligning each individual shock with an aggregate component whose coefficient, β, is the largest in absolute value. Since all shocks in our permutation procedures are normalised to having a variance of unity, selecting the largest coefficient is equivalent to selecting the factor explaining most of the variation. The sign of the permutation is also set by the coefficient: $sign(\beta)$. We find that this alternative routine provides almost identical results to our chosen factor procedure explained above.

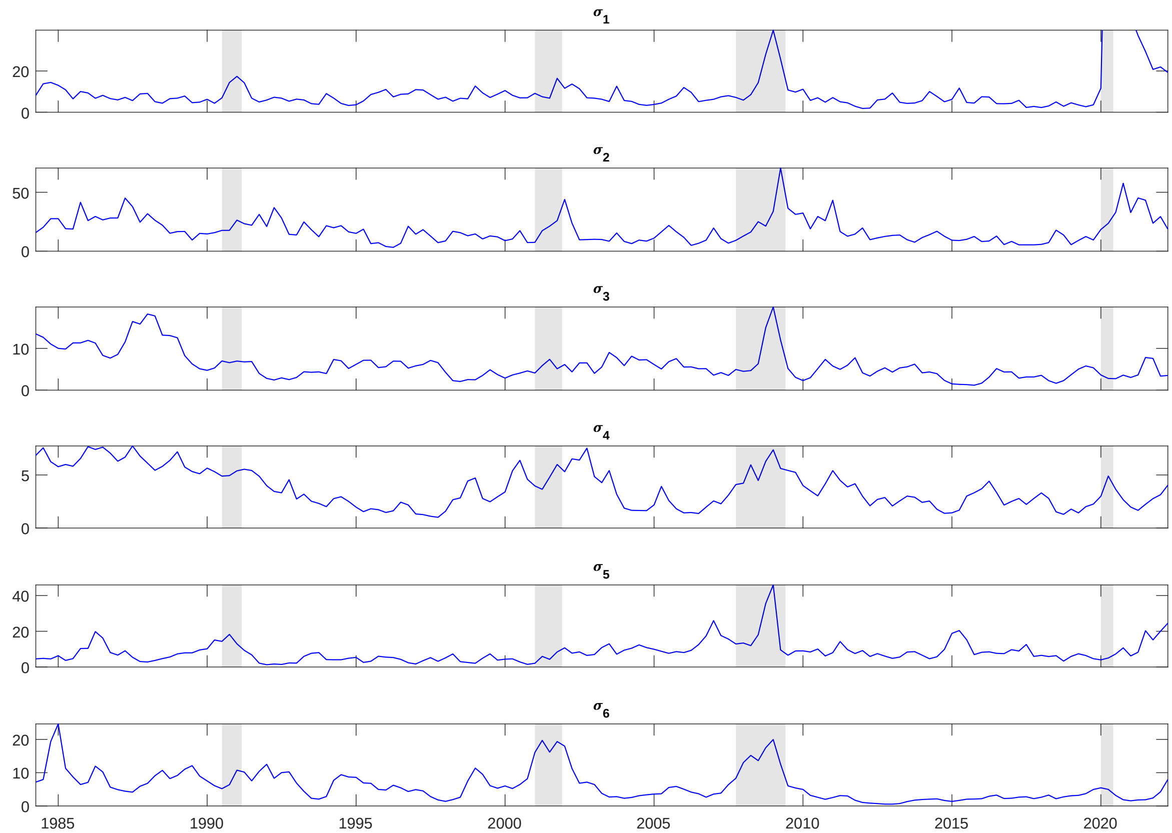
F Additional results

Figure F.1: Number of respondents over time



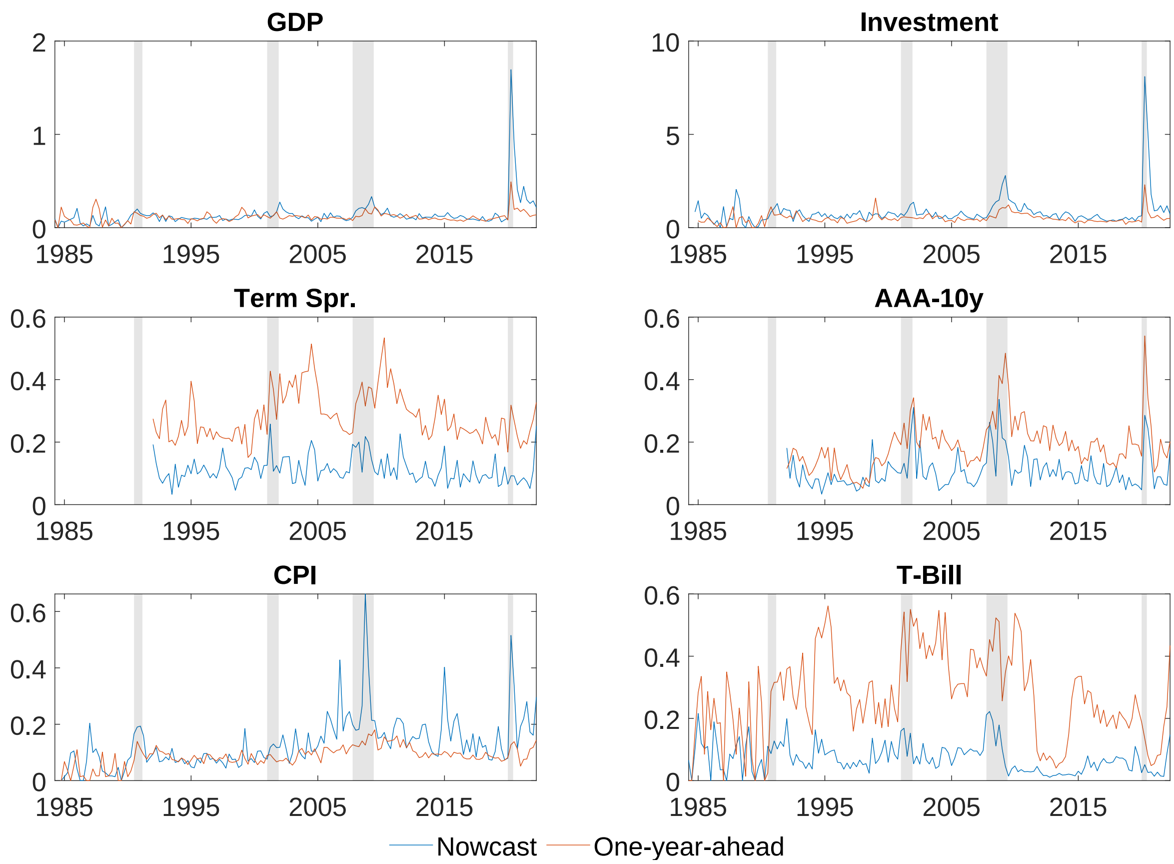
Note: The figure shows the number of respondents over time that provided their forecasts for a given variable. Different lines in each panel correspond to a different forecast horizons.

Figure F.2: Estimated stochastic volatility



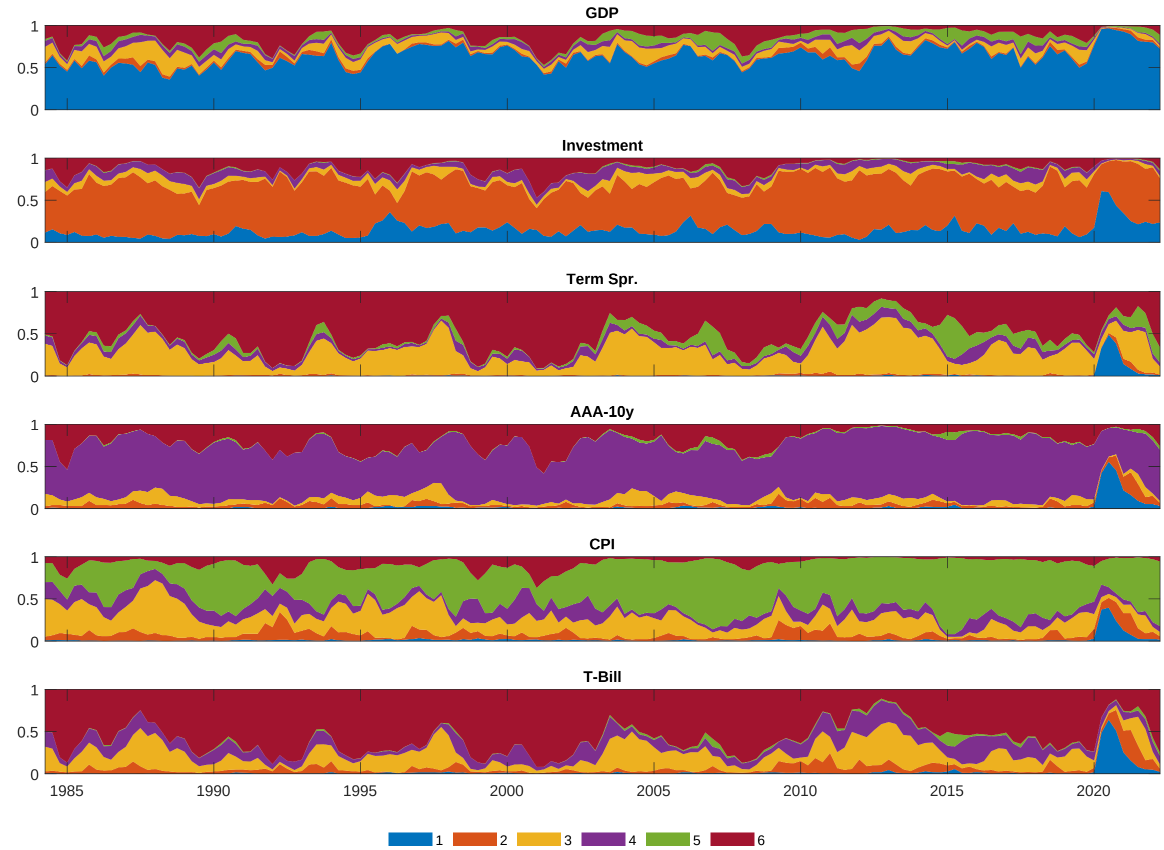
Note: The figure shows the posterior mean of stochastic volatility over the sample for each structural shock.

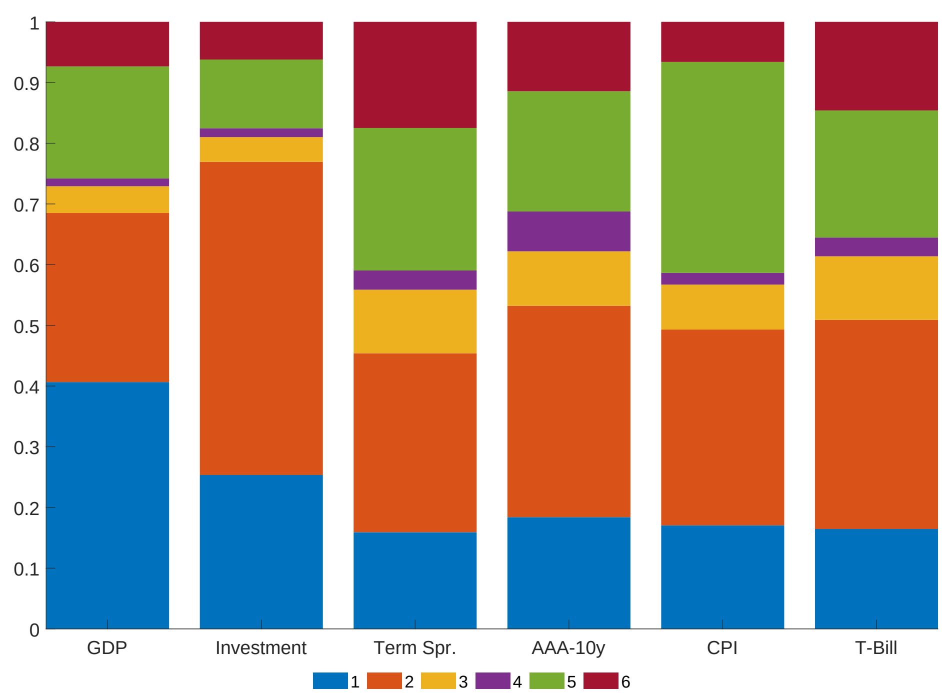
Figure F.3: Disagreement across respondents



Note: The figure shows the disagreement for all variables of our main specification from Fed SPF forecasts, at the 0 and 4 quarters horizons. Disagreement is calculated as the standard deviation of point forecasts across individuals, excluding the two smallest and largest values. Shaded bars are recessions as defined by the NBER.

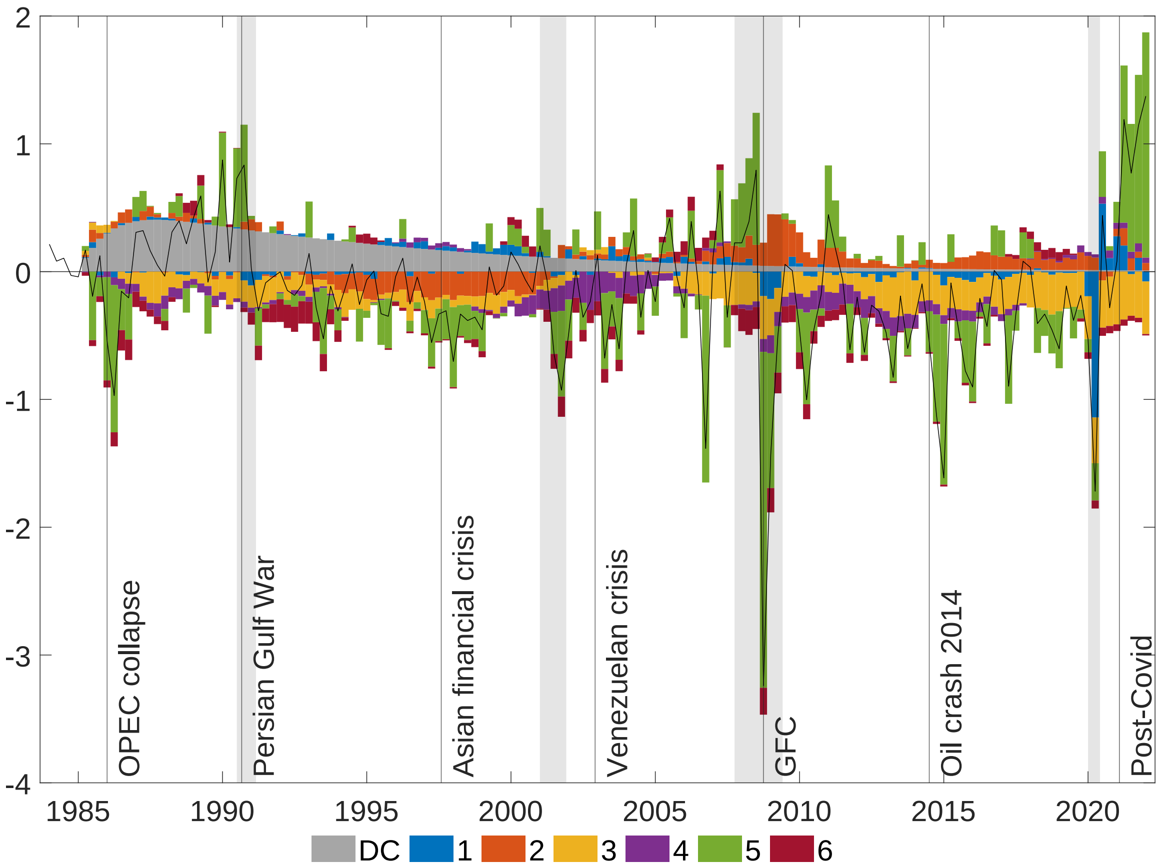
Figure F.4: Forecast error variance decomposition





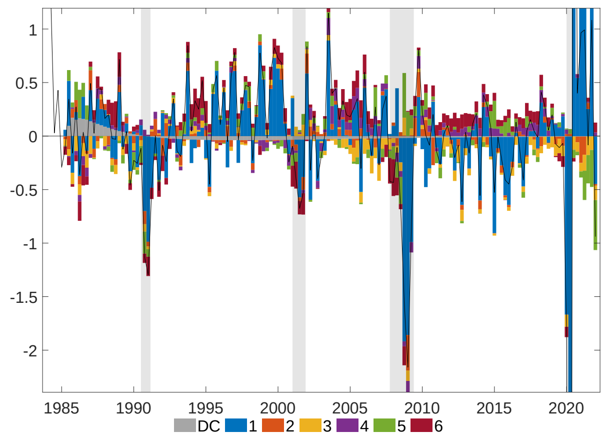
Note: The top panel shows the forecast error variance decomposition for the one-year-ahead forecasts. The bottom panel shows the unconditional forecast error variance decomposition. The numbers in the legend correspond to the shock numbering of Figure 2.

Figure F.5: Historical decomposition of CPI inflation and Oil supply events

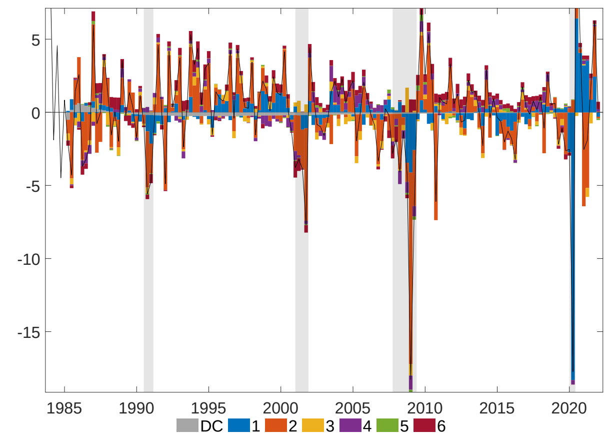


Note: The figure shows the historical shock decomposition for the q-o-q CPI inflation rate. The vertical lines indicate relevant events for the oil market, taken from Känzig (2021). The numbers in the legend correspond to the shock numbering of Figure 2.

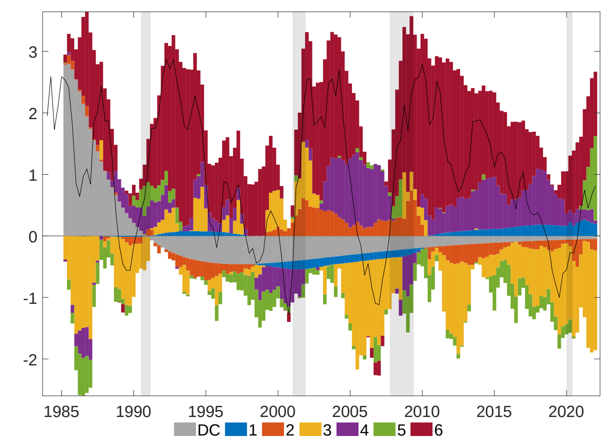
Figure F.6: Historical shock decomposition for the observed data



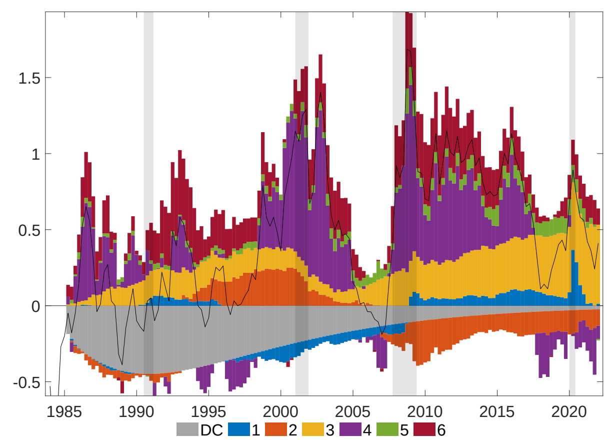
(a) Real GDP, q-o-q growth rate



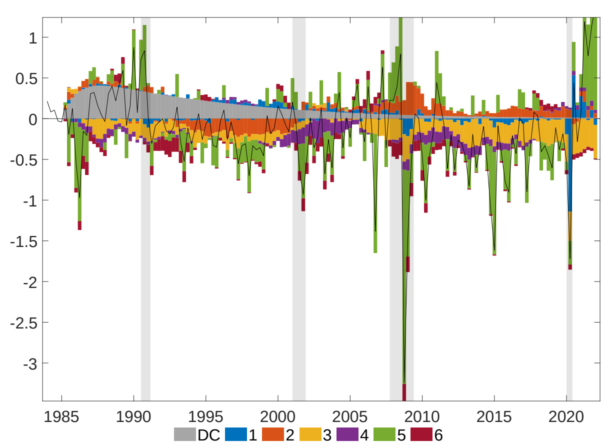
(b) Real Investment, q-o-q growth rate



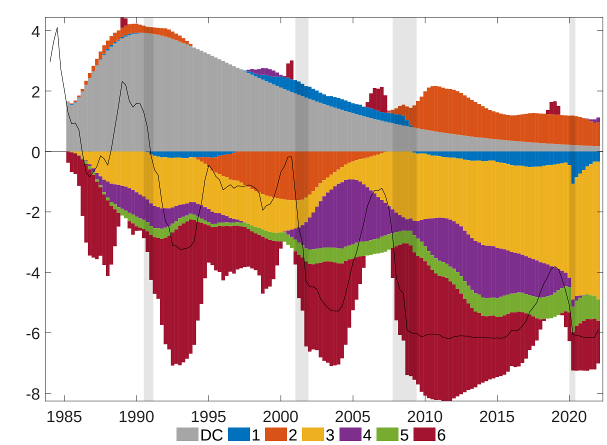
(c) Term spread



(d) AAA-10y spread



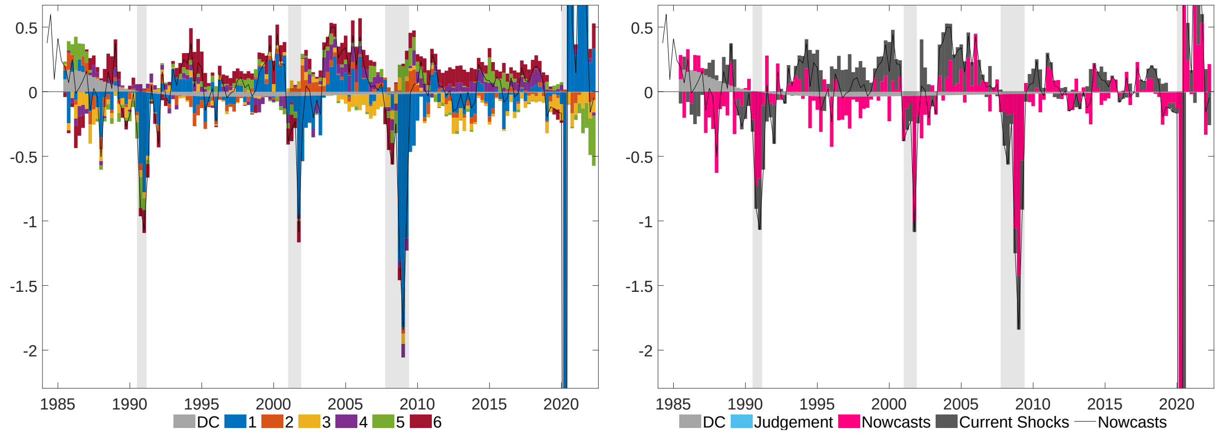
(e) CPI, q-o-q log-growth rate



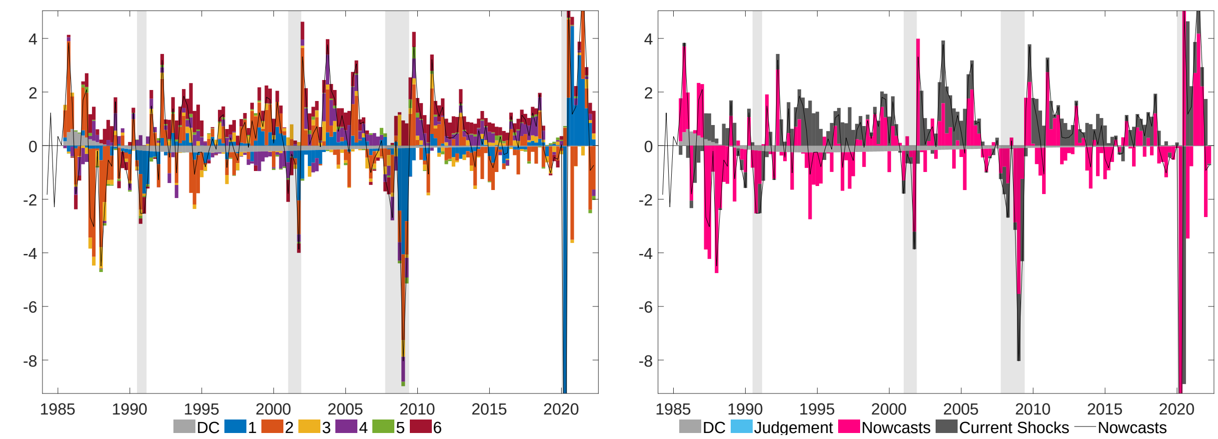
(f) T-bill

Note: The figure shows the historical shock decomposition for the observed data in deviation from its long-run mean. We use the posterior mean of the historical decomposition as our point estimate. The shaded areas represent NBER recession periods. The numbers in the legends correspond to the shock numbering of Figure 2.

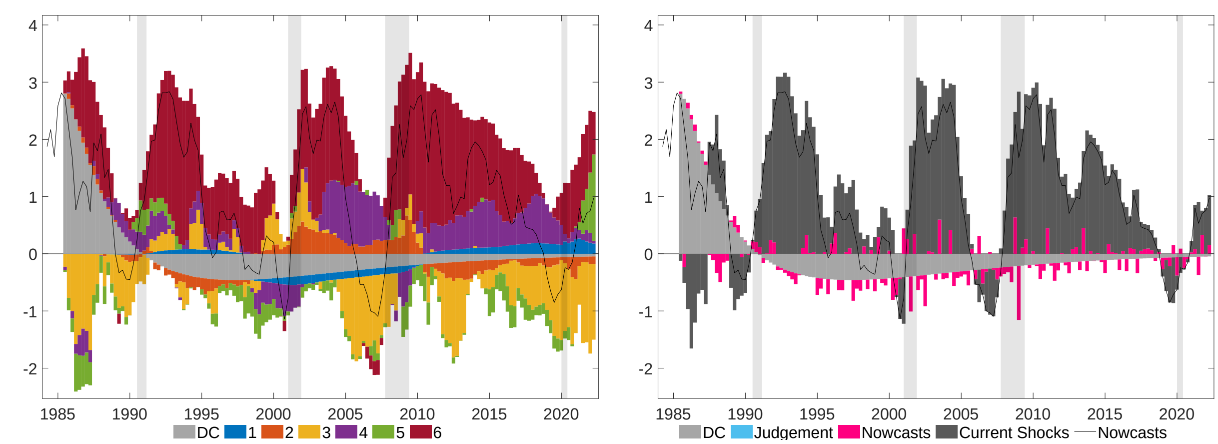
Figure F.7: Historical shock and judgement decompositions of nowcasts



(a) Real GDP, q-o-q growth rate



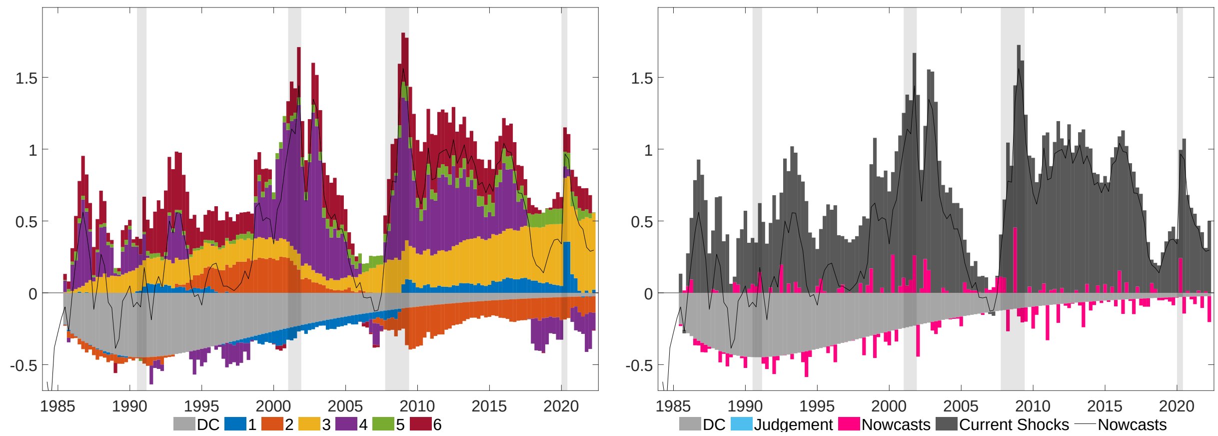
(b) Real Investment, q-o-q growth rate



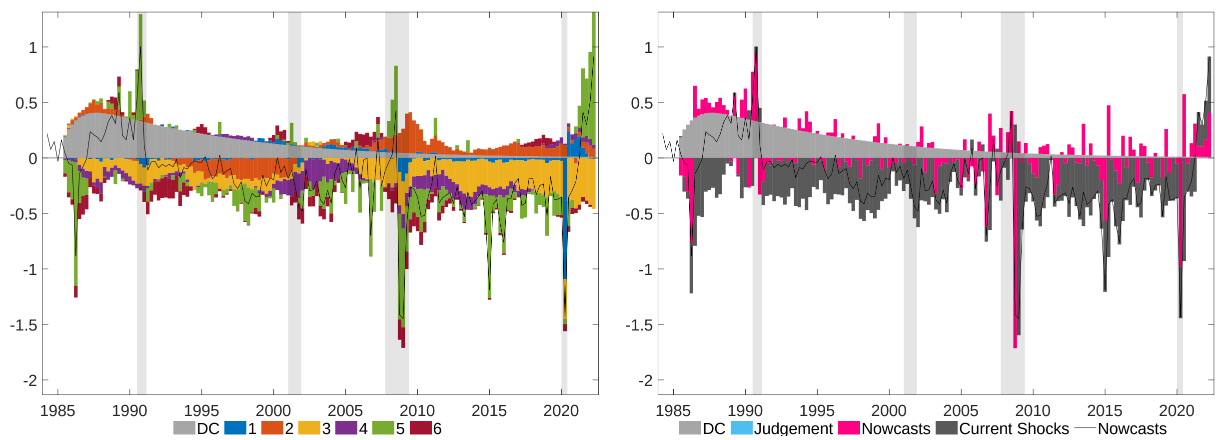
(c) Term spread

Note: The left panels show the historical shock decomposition for nowcasts in deviation from their long-run mean. The right panels show a decomposition of the nowcasts into deterministic conditions, current shocks, judgement about nowcasts and judgement about other horizons. We use the posterior mean of the historical decomposition as our point estimate. The shaded areas represent NBER recession periods.

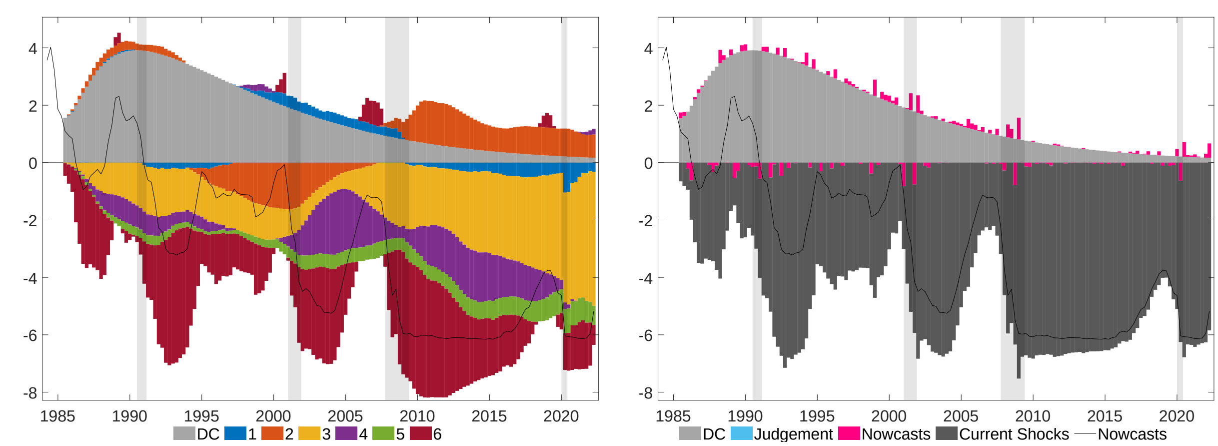
Figure F.7: Historical shock and judgement decompositions of nowcasts - continued



(d) AAA-10y spread



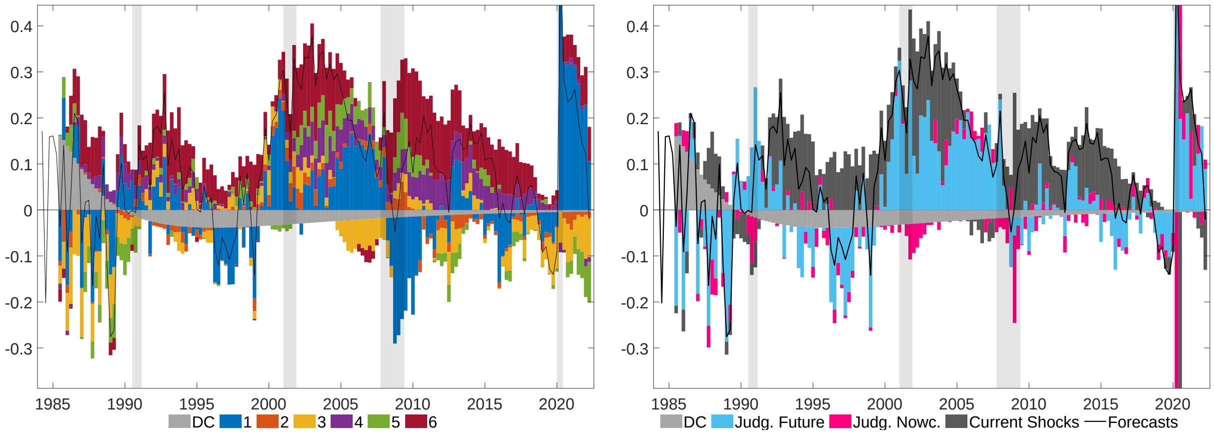
(e) CPI inflation



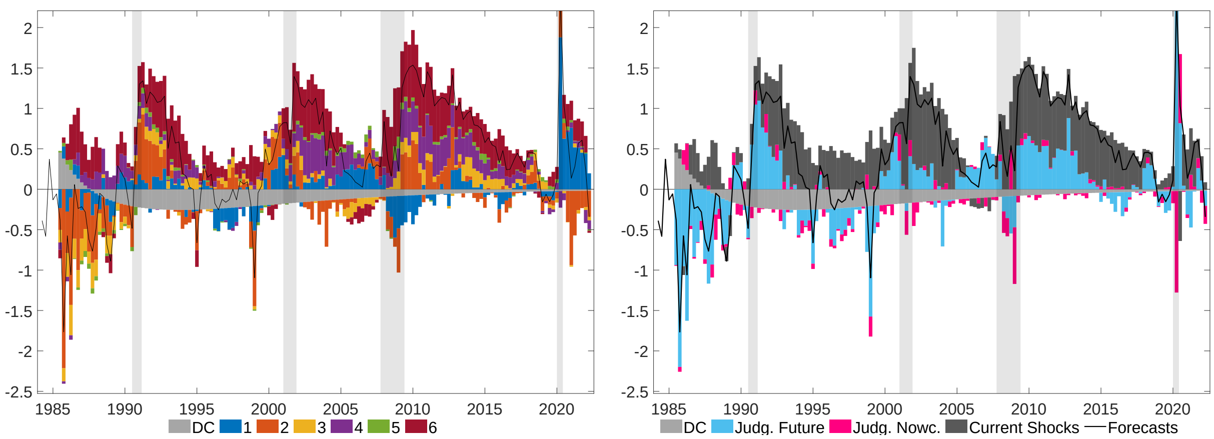
(f) T-bill

See note for Figure F.7 above.

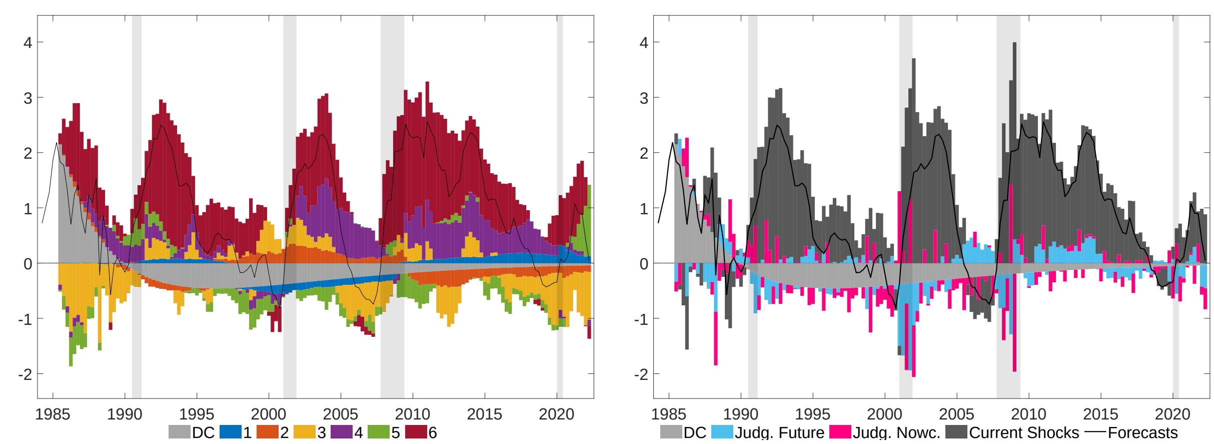
Figure F.8: Historical shock and judgement decompositions of one-year-ahead forecasts



(a) Real GDP, q-o-q growth rate



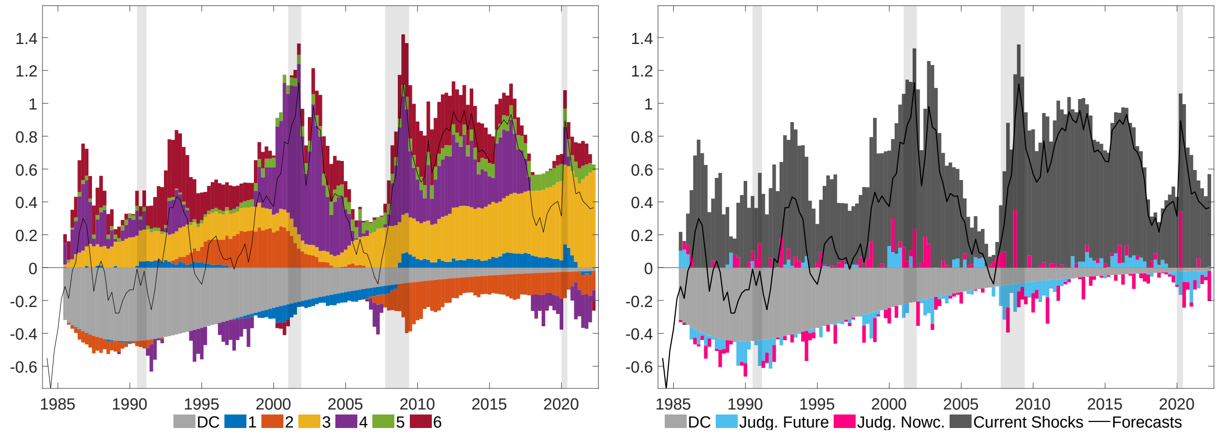
(b) Real Investment, q-o-q growth rate



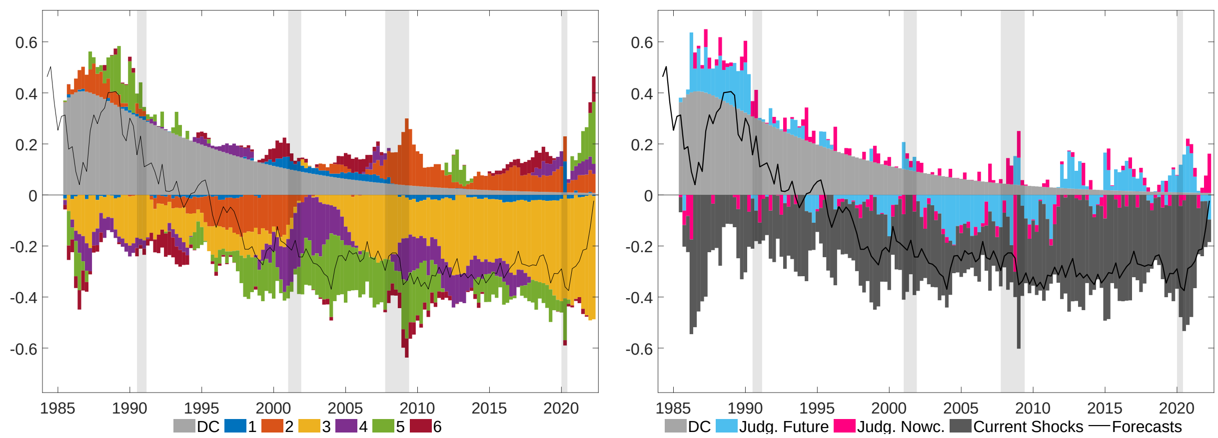
(c) Term spread

Note: The left panels show the historical shock decomposition for the one-year-ahead forecasts in deviation from their long-run mean. The right panels show a decomposition of the forecasts into deterministic conditions, current shocks, judgement about nowcasts and judgement about other horizons. We use the posterior mean of the historical decomposition as our point estimate. The shaded areas represent NBER recession periods.

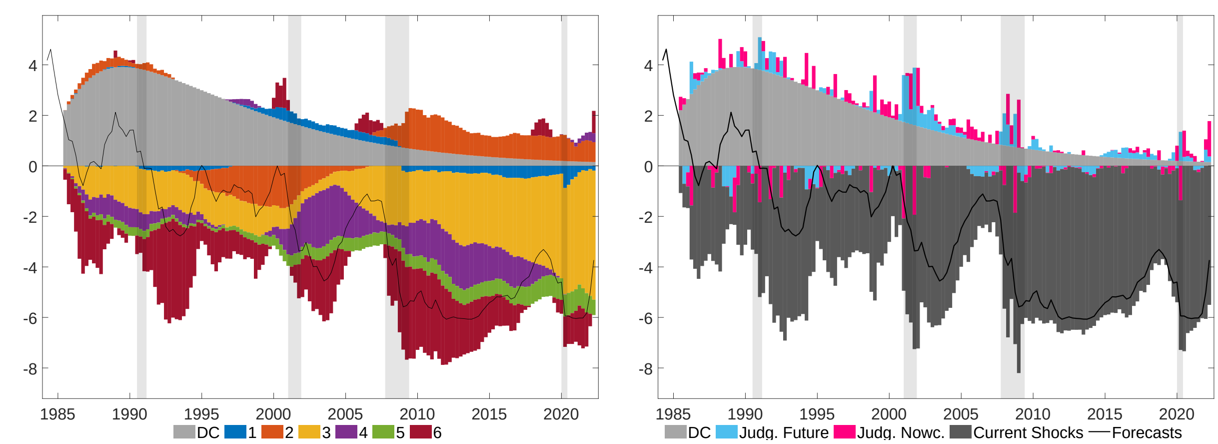
Figure F.8: Historical shock and judgement decompositions of one-year-ahead forecasts - continued



(d) AAA-10y spread



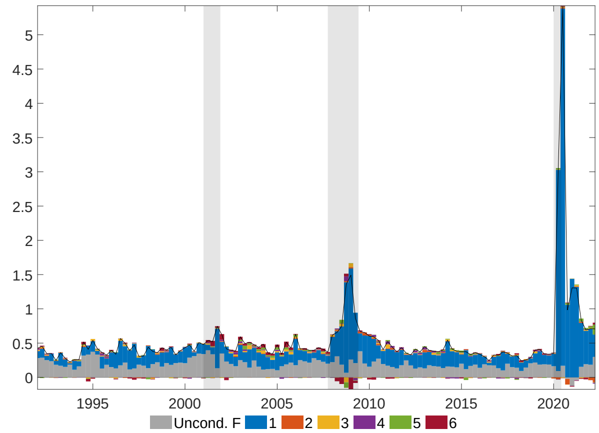
(e) CPI inflation



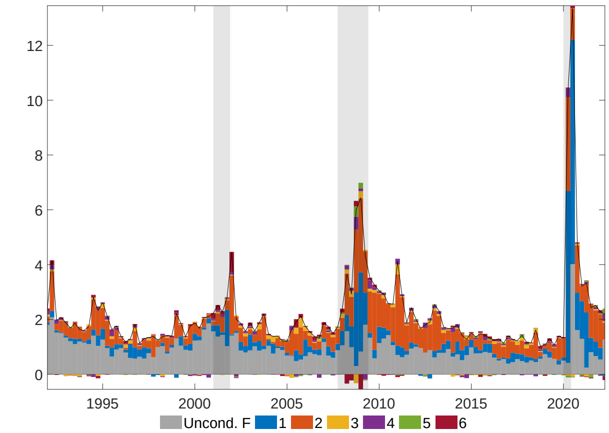
(f) T-bill

See note for Figure F.8 above.

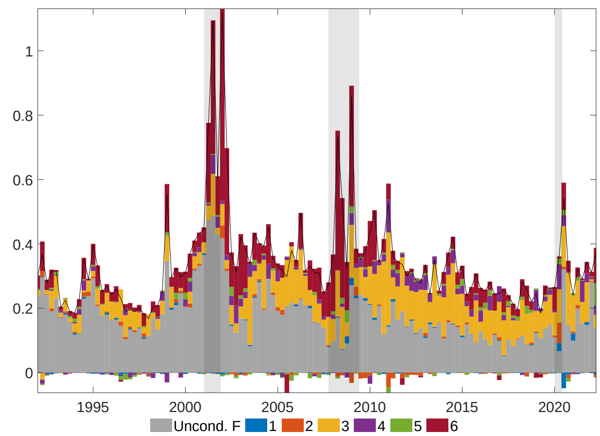
Figure F.9: Historical decomposition of one-year-ahead disagreement from interpolated data



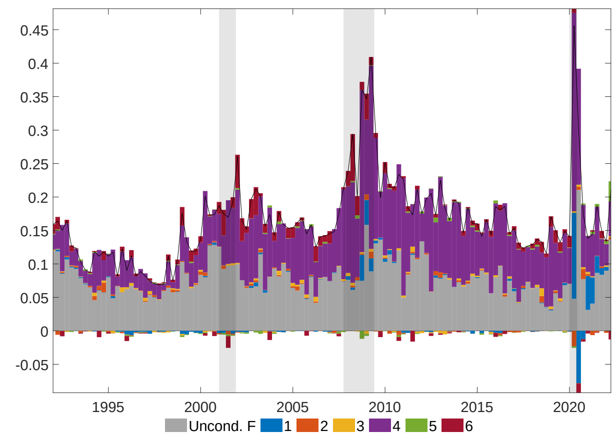
(a) Real GDP, y-o-y growth rate



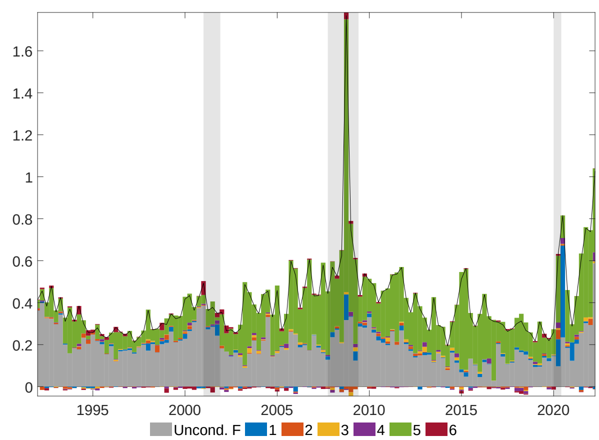
(b) Real Investment, y-o-y growth rate



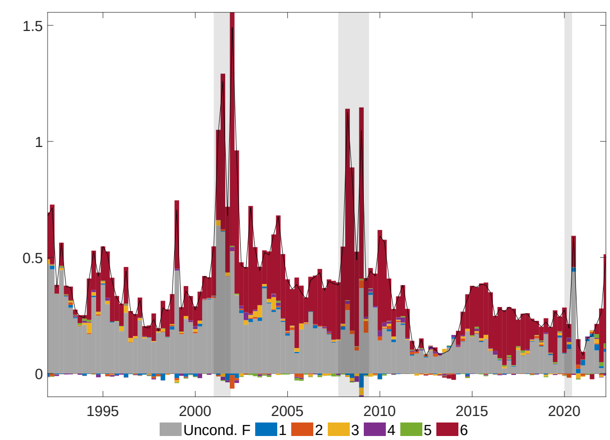
(c) Term spread



(d) AAA-10y spread



(e) CPI, y-o-y log-growth rate



(f) T-bill

Note: The figure shows the historical decomposition of the one-year-ahead disagreement from interpolated data. The shaded areas represent NBER recession periods. The numbers in the legend correspond to the shock numbering of Figure 2.

Table F.1: Relation of external shock proxies to estimated structural shocks

Mnemonic	Source	Type	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	Cand.	LRT
LPR20DMASS7	Lagerborg et al. (2022)	CONF							3	
LPR20MASSFAT7VEGAS	Lagerborg et al. (2022)	CONF							5	
LPR20MASSFATM	Lagerborg et al. (2022)	CONF							5	
LPR20MASSFATMMORE7	Lagerborg et al. (2022)	CONF				-			4	
LPR20MASSVIC7	Lagerborg et al. (2022)	CONF				-			4	
BH2019D	Baumeister and Hamilton (2019)	D	+++					++	1	*
BCDZ14	Bassett et al. (2014)	F				++			4	
BCDZ14b	Bassett et al. (2014)	F				+			4	
GZ12	Gilchrist and Zakrajšek (2012)	F				+++			4	
BZP17	Ramey (2016) and Zeev and Pappa (2017)	G	+					++	6	
FP10	Fisher and Peters (2010)	G							2	
R11	Ramey (2011)	G					++		5	
R11scaled	Ramey (2011) and Caldara and Kamps (2017)	G							1	
RZ18	Ramey and Zubairy (2017)	G					++		5	
RZ18scaled	Ramey and Zubairy (2017)	G							5	
BBE05	Bernanke et al. (2005)	M						+++	6	
BC13	Barakchian and Crowe (2013)	M		+	- - -	-			3	***
BRW19	Bu et al. (2020)	M		+++					2	
BRW19u	Bu et al. (2020)	M		+++					2	
CH19MCGCS	Caldara and Herbst (2019)	M							5	
CH19MHF	Caldara and Herbst (2019)	M							6	
CH19MRR	Caldara and Herbst (2019)	M							6	
CH19MRRCS	Caldara and Herbst (2019)	M							4	
CH19MRRCSOLD	Caldara and Herbst (2019)	M							4	
DJL23AP	Lewis (2019)	M							1	*
DJL23FF	Lewis (2019)	M						+++	6	*
DJL23FG	Lewis (2019)	M							3	
DJL23FI	Lewis (2019)	M						++	6	
GK15ED2TC	Gertler and Karadi (2015) and Ramey (2016)	M	+					+++	6	
GK15ED2VR	Gertler and Karadi (2015) and Ramey (2016)	M			- -			+++	6	
GK15ED2ramey	Gertler and Karadi (2015) and Ramey (2016)	M			- -				3	
GK15ED3TC	Gertler and Karadi (2015) and Ramey (2016)	M						+++	6	

Continued on next page

Table F.1 – continued

Mnemonic	Source	Type	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	Cand.	LRT
GK15ED4TC	Gertler and Karadi (2015) and Ramey (2016)	M						++	6	
GK15FF1VR	Gertler and Karadi (2015) and Ramey (2016)	M			-			+++	6	
GK15FF1ramey	Gertler and Karadi (2015) and Ramey (2016)	M	-					+	6	
GK15FF4TC	Gertler and Karadi (2015) and Ramey (2016)	M						+++	6	
GK15FF4VR	Gertler and Karadi (2015) and Ramey (2016)	M	+		- -			+++	6	**
GK15FF4ramey	Gertler and Karadi (2015) and Ramey (2016)	M							6	
GK15MP1TC	Gertler and Karadi (2015) and Ramey (2016)	M			-		-	+++	6	
GK15MP1VR	Gertler and Karadi (2015) and Ramey (2016)	M			- -			+++	6	
JK20CBImedian	Jarociński and Karadi (2020)	M			- -	- - -	++	+++	6	***
JK20CBImedianm	Jarociński and Karadi (2020)	M			- -	- - -	++	+++	6	***
JK20CBIpm	Jarociński and Karadi (2020)	M						+++	6	
JK20CBIpmm	Jarociński and Karadi (2020)	M						+++	6	
JK20MPmedian	Jarociński and Karadi (2020)	M		+	- -			+++	6	*
JK20MPmedianm	Jarociński and Karadi (2020)	M		+	- -			+++	6	*
JK20MPpm	Jarociński and Karadi (2020)	M			- - -			+++	6	***
JK20MPpmm	Jarociński and Karadi (2020)	M			- - -			+++	6	***
MAR16IV1	Miranda-Agrippino and Ricco (2021)	M						++	6	
MAR16IV5	Miranda-Agrippino and Ricco (2021)	M							6	
MAR2021CBINFO	Miranda-Agrippino and Ricco (2021)	M	+++	++			+++	+++	1	***
MAR2021FF4	Miranda-Agrippino and Ricco (2021)	M			- -			+	3	
MAR2021MPI	Miranda-Agrippino and Ricco (2021)	M						++	6	
MJ2023u1	Jarocinski (2021)	M			- - -			+++	6	
MJ2023u2	Jarocinski (2021)	M						-	6	
MJ2023u3	Jarocinski (2021)	M	+++	+	++	- -		- - -	1	***
MJ2023u4	Jarocinski (2021)	M				- -	++		5	
RR0483	Romer and Romer (2004) and Ramey (2016)	M	- - -					+++	6	*
RR0483b	Romer and Romer (2004) and Ramey (2016)	M	- - -					+++	6	*
RR04full	Romer and Romer (2004) and Wieland (2021)	M	- -					+++	6	*
RR04orig	Romer and Romer (2004) and Wieland (2021)	M	-			- -			4	**
RR04origreg	Romer and Romer (2004) and Wieland (2021)	M	- -					++	6	*
RR04ramey	Romer and Romer (2004) and Ramey (2016)	M	- -					+++	6	*
RRDUMMY23	Romer and Romer (2023)	M							4	
RRDUMMYORIG23	Romer and Romer (2023)	M							4	

Continued on next page

Table F.1 – continued

Mnemonic	Source	Type	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	Cand.	LRT
ACD20ShockC	Angeletos et al. (2020)	MBC	+++						1	*
ACD20ShockDP	Angeletos et al. (2020)	MBC					+++	++	5	
ACD20ShockH	Angeletos et al. (2020)	MBC	+++	+++				++	1	***
ACD20ShockI	Angeletos et al. (2020)	MBC	+++	+++	++		-	+++	1	***
ACD20ShockR	Angeletos et al. (2020)	MBC	+	+++				+++	6	***
ACD20ShockSW	Angeletos et al. (2020)	MBC	- -					-	1	
ACD20ShockTFP	Angeletos et al. (2020)	MBC		- -					2	
ACD20ShockU	Angeletos et al. (2020)	MBC	- - -	- - -	- -		+	- -	1	***
ACD20ShockY	Angeletos et al. (2020)	MBC	+++	++	++		- - -	+++	1	***
ACD20ShockYSH	Angeletos et al. (2020)	MBC	+++				- - -		1	**
BH2022E	Baumeister (2023)	OIL					+++		5	
BH2022S	Baumeister (2023)	OIL		-			+++	++	5	***
HAM03a	Hamilton (2003) and Caldara and Kamps (2017)	OIL					++		5	
HAM03b	Hamilton (2003)	OIL		-			+++		5	
BH2019OILD	Baumeister and Hamilton (2019)	OILD		- -			++	++	5	**
CCI19oild	Caldara et al. (2019)	OILD			-			++	6	
BH2019OILI	Baumeister and Hamilton (2019)	OILI			+		++		5	
BH2019OILS	Baumeister and Hamilton (2019)	OILS				+	- - -		5	*
BH2019OILS2	Baumeister and Hamilton (2019)	OILS					- - -		5	
CCI19inst	Caldara et al. (2019)	OILS					- -		5	
CCI19oils	Caldara et al. (2019)	OILS					- - -		5	
DK21nw	Känzig (2021)	OILS		- - -			+++	++	5	***
DK21nwprecovid	Känzig (2021)	OILS		- -			+++	++	5	**
DK21s	Känzig (2021)	OILS					+++		5	
DK21sprecovid	Känzig (2021)	OILS					+++		5	
LK08a	Kilian (2008)	OILS				+			4	
LK08b	Kilian (2008)	OILS	+						1	
LK08o	Kilian (2008)	OILS	+						1	
LK09	Kilian (2009)	OILS						+	6	
LPW12	Leeper et al. (2013) and Caldara and Kamps (2017)	TAX							1	
LPW122	Leeper et al. (2013) and Ramey (2016)	TAX							3	
MR12news	Mertens and Ravn (2012)	TAX			- -	- - -	-		4	*
MR12unc	Mertens and Ravn (2013)	TAX	- - -		-				1	

Continued on next page

Table F.1 – continued

Mnemonic	Source	Type	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	Cand.	LRT
MR12unc2	Mertens and Ravn (2013)	TAX	- - -		-				1	
MR2013TCI	Mertens and Ravn (2019)	TAX	-						1	
MR2013TPI	Mertens and Ravn (2019)	TAX	- - -		-				1	
MR2013mCI	Mertens and Ravn (2019)	TAX	-						1	
MR2013mPI	Mertens and Ravn (2019)	TAX	- - -		- -				1	*
RR10endo	Romer and Romer (2010) and Ramey (2016)	TAX	++	+		- - -			4	***
RR10exo	Romer and Romer (2010) and Ramey (2016)	TAX	- - -		- -				1	
BFK06	Basu et al. (2006)	TECH	+++				-		1	
BFK06dtfp	Basu et al. (2006)	TECH	+++			- -	-	+++	1	***
BFK06dtfpC	Basu et al. (2006)	TECH	+++	+		- -	- -	+++	1	***
BFK06dtfpCutil	Basu et al. (2006)	TECH	+++				-	+	1	*
BFK06dtfpI	Basu et al. (2006)	TECH	+++			-		+++	1	***
BFK06dtfpIutil	Basu et al. (2006)	TECH		- - -					2	
BFK06dtfputil	Basu et al. (2006)	TECH		- -					2	
BP14	Beaudry and Portier (2014)	TECH	++	- -	-	- -			1	**
BP14tfpnewslr	Beaudry and Portier (2014) and Ramey (2016)	TECH	+		- - -	- - -			4	***
BP14tfpnewssr	Beaudry and Portier (2014) and Ramey (2016)	TECH			- - -	- - -			3	***
BS11	Barsky and Sims (2011)	TECH					- -		5	
BZK15ist	Ben Zeev and Khan (2015) and Ramey (2016)	TECH	+				+++		5	*
BZK15istnews	Ben Zeev and Khan (2015) and Ramey (2016)	TECH		+++			- - -		5	**
BZK15tfp	Ben Zeev and Khan (2015) and Ramey (2016)	TECH	+++				- -	++	1	*
DV22PI	Cascaldi-Garcia and Vukotić (2022)	TECH			+	++			4	
DV22PNS	Cascaldi-Garcia and Vukotić (2022)	TECH	++						1	
DV22PNSE	Cascaldi-Garcia and Vukotić (2022)	TECH	++						1	
FORD14	Francis et al. (2014)	TECH	+++	- - -					2	***
JPT11ist	Justiniano et al. (2011) and Ramey (2016)	TECH					+++		5	
JPT11mei	Justiniano et al. (2011) and Ramey (2016)	TECH	+++	+++	++		+	+++	2	***
JPT11tfp	Justiniano et al. (2011) and Ramey (2016)	TECH		-	++		- - -		5	
KO13news07l	Kurmann and Otrok (2017)	TECH		- - -	++		- - -		2	***
KO13news07s	Kurmann and Otrok (2017)	TECH		- - -	+++	- -	- -		3	***
KO13news16l	Kurmann and Otrok (2017)	TECH			+	-	- -		4	**
KO13news16s	Kurmann and Otrok (2017)	TECH	++	- -	+++		- -	-	2	***
KO13slope07l	Kurmann and Otrok (2017)	TECH		- - -	+++				3	***

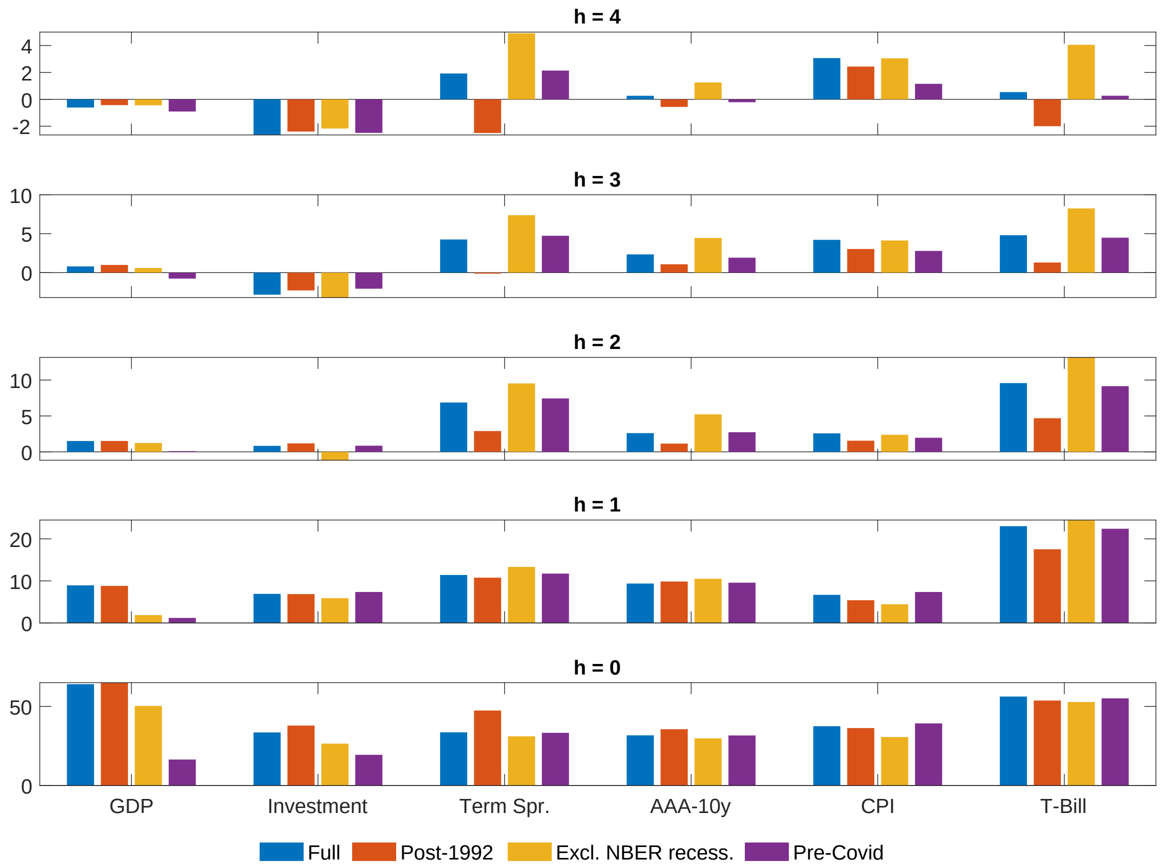
Continued on next page

Table F.1 – continued

Mnemonic	Source	Type	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	Cand.	LRT
KO13slope07s	Kurmann and Otrok (2017)	TECH		-	+++	- - -			3	***
KO13slope16l	Kurmann and Otrok (2017)	TECH		- - -	+++				3	***
KO13slope16s	Kurmann and Otrok (2017)	TECH		-	+++	- - -			3	***
KS21meanBS	Kurmann and Sims (2021)	TECH		- - -			- - -		2	**
KS21meanU	Kurmann and Sims (2021)	TECH		- - -			- - -		5	***
KS21medBS	Kurmann and Sims (2021)	TECH		- - -			- - -		2	**
KS21medU	Kurmann and Sims (2021)	TECH		- - -			- - -		5	***
MN15istp	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							1	
MN15istpn4	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							5	
MN15istpn8	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							2	
MN15ists	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH	++				+		1	*
MN15istsn4	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							4	
MN15istsn8	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH			- -				3	
MN15tfpp	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH						++	6	
MN15tfppn4	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							4	
MN15tfppn8	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH		- -			+	+	5	
MN15tfps	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH	+++		+		- -		1	*
MN15tfpsn4	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH							4	
MN15tfpsn8	Ramey (2016) and Miyamoto and Nguyen (2020)	TECH		- -			+++	+	5	
SW07prod	Smets and Wouters (2007)	TECH							3	
BBD16	Baker et al. (2016)	UNC	- -			+++	++	- -	4	**
NB09	Bloom (2009)	UNC			+	+++			4	
NB09FMT	Bloom (2009)	UNC				++			4	
NB09MMT	Bloom (2009)	UNC				+++			4	

Note: The table presents the shock series, sourced from the literature, that we use to label our structural shocks, as described in section 2.3.1. Each row depicts a different shock series sourced from a specific study (column "Source"), to which we assign a mnemonic for the sake of brevity (column "Mnemonic"). In some instances, different estimates in one study attempt to capture the same structural innovation but differ due to assumptions about sample size, underlying information etc. Column "Shock type" assigns each shock to a different class: "CONF" stands for confidence shocks; "F" - Financial; "G" - government spending; "M" - monetary; "MBC" - main business cycle à la (Angeletos et al., 2020); "OIL" - oil; "OILD" - oil demand; "OILI" - oil inventory; "OILS" - oil supply; "TAX" - tax policy; "TECH" - technology; "UNC" - uncertainty. The columns of ψ_i indicate the sign of the posterior mean of coefficients obtained by regressing shocks from the literature on our shock estimates. The number of signs denotes different levels of high probability density intervals that do not include the zero value (+++ (- - -)=99%, ++(- -)=95%, +(-)=90%). "Cand" is the shock with the highest absolute correlation and "LRT" is the significance level of the p-value from the likelihood ratio test.

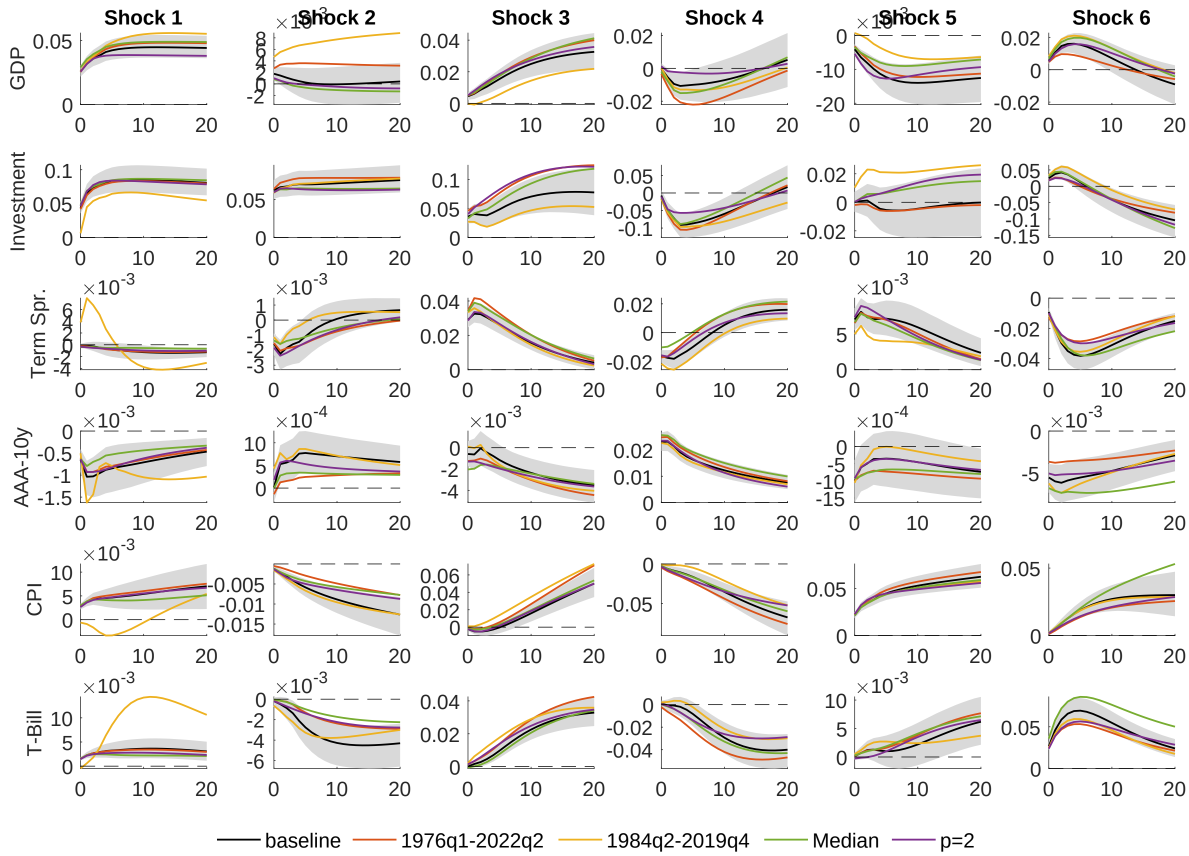
Figure F.10: Forecast performance gains of SPF versus unconditional forecasts



Note: The figure shows percentage gains in terms of root mean squared forecast errors (RMSFE) for the SPF forecasts compared to model-consistent unconditional forecasts: $100(1-RMSE_{SPF}/RMSE_{UC})$. The five panels represent different forecast horizons, while the coloured bars represent results using different sub-periods of forecast errors: "Full" is the full sample; "Post-1992" is the sample from 1992q1; "Excl. NBER recess." is the sample without NBER recessions; "Pre-Covid" is the sample until 2019q4.

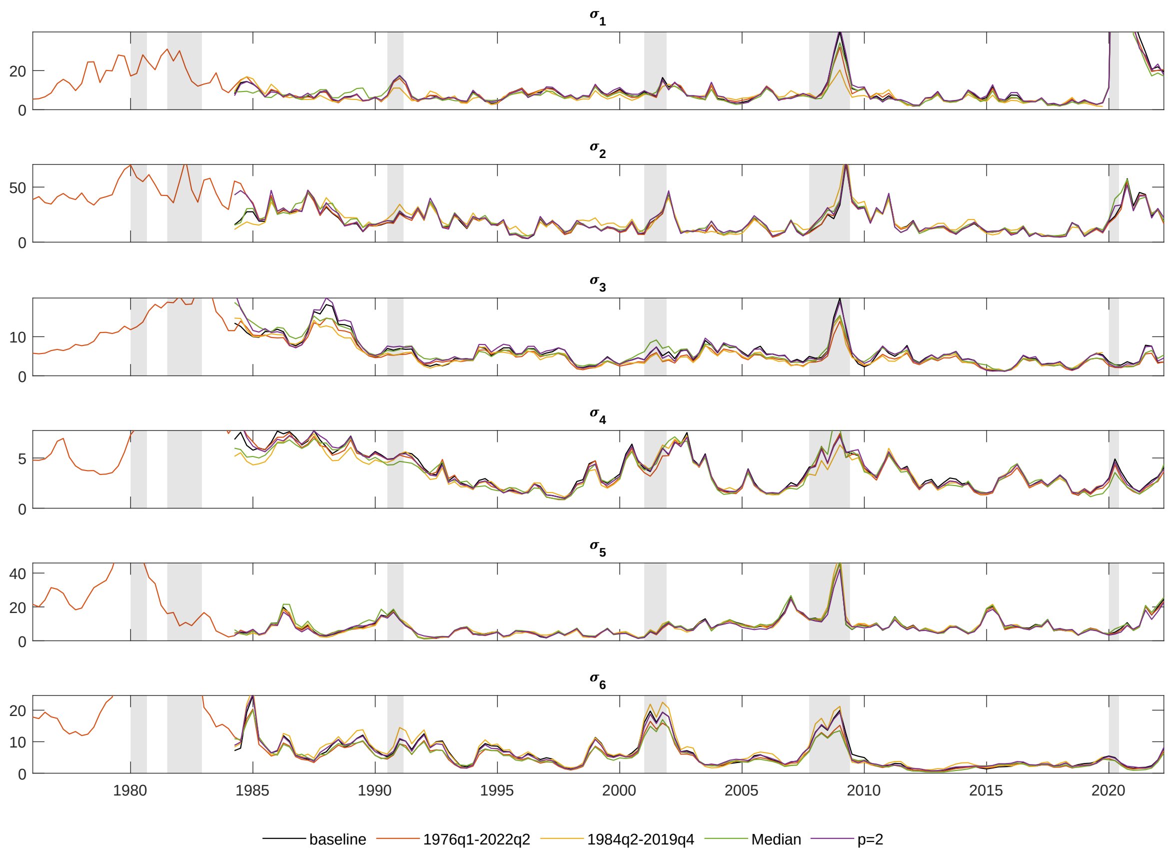
G Alternative specifications

Figure G.1: Impulse responses for alternative specifications



Note: The figure shows impulse response functions for the baseline model and alternative specifications: "baseline" stands for the baseline specification; "1976q1-2022q2" is the specification with the longer sample; "1984q2-2019q4" excludes the pandemic period; "Median" uses median SPF responses; "p=2" uses two lags. Each sub-panel shows the response of a variable (in the rows) to a shock (in the columns). The grey areas are the posterior 90% credible sets for the baseline specification, while the lines are the posterior means for the baseline and alternative specifications.

Figure G.2: Estimated stochastic volatility for alternative specifications



Note: The figure shows the estimated stochastic volatility for each structural shock. Lines in different colours represent alternative specifications: "baseline" stands for the baseline specification; "1976q1-2022q2" is the specification with the longer sample; "1984q2-2019q4" excludes the pandemic period; "Median" uses median SPF responses; "p=2" uses two lags.

References

Angeletos, G.-M., Collard, F., & Dellas, H. (2020). Business-cycle anatomy. *American Economic Review*, *110*(10), 3030–3070. https://doi.org/10.1257/aer.20181174

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Barakchian, S. M., & Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics*, *60*(8), 950–966. https://doi.org/10.1016/j.jmoneco.2013.09.006

Barsky, R. B., & Sims, E. R. (2011). News shocks and business cycles. *Journal of Monetary Economics*, *58*(3), 273–289. https://doi.org/10.1016/j.jmoneco.2011.03.001

Bassett, W. F., Chosak, M. B., Driscoll, J. C., & Zakrajšek, E. (2014). Changes in bank lending standards and the macroeconomy. *Journal of Monetary Economics*, *62*, 23–40. https://doi.org/10.1016/j.jmoneco.2013.12.005

Basu, S., Fernald, J. G., & Kimball, M. S. (2006). Are technology improvements contractionary? *American Economic Review*, *96*(5), 1418–1448. https://doi.org/10.1257/aer.96.5.1418

Baumeister, C. (2023, January 1). Measuring market expectations. In R. Bachmann, G. Topa, & W. van der Klaauw (Eds.), *Handbook of economic expectations* (pp. 413–441). Academic Press. https://doi.org/10.1016/B978-0-12-822927-9.00022-7

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, *109*(5), 1873–1910. https://doi.org/10.1257/aer.20151569

Beaudry, P., & Portier, F. (2014). News-driven business cycles: Insights and challenges. *Journal of Economic Literature*, *52*(4), 993–1074. https://doi.org/10.1257/jel.52.4.993

Ben Zeev, N., & Khan, H. (2015). Investment-specific news shocks and u.s. business cycles. *Journal of Money, Credit and Banking*, *47*(7), 1443–1464. https://doi.org/10.1111/jmcb.12250

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, *120*(1), 387–422. Retrieved April 2, 2023, from https://econpapers.repec.org/article/oupqjecon/v_3a120_3ay_3a2005_3ai_3a1_3ap_3a387-422..htm

Bertsche, D., & Braun, R. (2022). Identification of structural vector autoregressions by stochastic volatility. *Journal of Business & Economic Statistics*, *40*(1), 328–341. https://doi.org/10.1080/07350015.2020.1813588

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685. https://doi.org/10.3982/ECTA6248

Bu, C., Rogers, J., & Wu, W. (2020). A unified measure of fed monetary policy shocks. *Journal of Monetary Economics*. https://doi.org/10.1016/j.jmoneco.2020.11.002

Caldara, D., Cavallo, M., & Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, *103*, 1–20. https://doi.org/10.1016/j.jmoneco.2018.08.004

Caldara, D., & Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy SVARs. *American Economic Journal: Macroeconomics*, *11*(1), 157–192. https://doi.org/10.1257/mac.20170294

Caldara, D., & Kamps, C. (2017). The analytics of SVARs: A unified framework to measure fiscal multipliers. *The Review of Economic Studies*, *84*(3), 1015–1040. Retrieved February 20, 2021, from https://www.jstor.org/stable/45106771

Carriero, A., Chan, J., Clark, T. E., & Marcellino, M. (2022). Corrigendum to "large bayesian vector autoregressions with stochastic volatility and non-conjugate priors" [j. econometrics 212 (1) (2019) 137–154]. *Journal of Econometrics*, *227*(2), 506–512. https://doi.org/10.1016/j.jeconom.2021.11.010

Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, *212*(1), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024

Cascaldi-Garcia, D., & Vukotić, M. (2022). Patent-based news shocks. *The Review of Economics and Statistics*, *104*(1), 51–66. https://doi.org/10.1162/rest_a_00943

Chan, J. C. C., Koop, G., & Yu, X. (2021). Large order-invariant bayesian VARs with stochastic volatility, 34.

Chan, J. C. C., Poon, A., & Zhu, D. (2023). High-dimensional conditionally gaussian state space models with missing data.

Chan, J. C., & Hsiao, C. Y. (2014). Estimation of stochastic volatility models with heavy tails and serial dependence. In *Bayesian inference in the social sciences* (pp. 155–176). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118771051.ch6

Chan, J. C., & Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*(1), 101. https://doi.org/10.1504/IJMMNO.2009.030090

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American Statistical Association*, *96*(453), 270–281. https://doi.org/10.1198/016214501750332848

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302. https://doi.org/10.1016/j.red.2004.10.009

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–615. Retrieved August 12, 2022, from http://www.jstor.org/stable/4140605

Fisher, J. D. M., Melosi, L., & Rast, S. (n.d.). Long-run inflation expectations.

Fisher, J. D., & Peters, R. (2010). Using stock returns to identify government spending shocks. *The Economic Journal*, *120*(544), 414–436. https://doi.org/10.1111/j.1468-0297.2010.02355.x

Francis, N., Owyang, M. T., Roush, J. E., & DiCecio, R. (2014). A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks. *The*

Review of Economics and Statistics, *96*(4), 638–647. Retrieved March 20, 2023, from https://www.jstor.org/stable/43554945

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76. https://doi.org/10.1257/mac.20130329

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics* (4th ed.). Clarendon Press.

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, *102*(4), 1692–1720. https://doi.org/10.1257/aer.102.4.1692

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*(2), 363–398. https://doi.org/10.1016/S0304-4076(02)00207-5

Herbst, E., & Winkler, F. (2021, July 1). *The factor structure of disagreement* (SSRN Scholarly Paper No. 3899646). Social Science Research Network. Rochester, NY. https://doi.org/10.17016/FEDS.2021.046

Jarocinski, M. (2021). Estimating the fed's unconventional policy shocks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3982819

Jarociński, M. (2015). A note on implementing the durbin and koopman simulation smoother. *Computational Statistics & Data Analysis*, *91*, 1–3. https://doi.org/10.1016/j.csda.2015.05.001

Jarociński, M., & Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, *12*(2), 1–43. https://doi.org/10.1257/mac.20180090

Justiniano, A., Primiceri, G. E., & Tambalotti, A. (2011). Investment shocks and the relative price of investment. *Review of Economic Dynamics*, *14*(1), 102–121. https://doi.org/10.1016/j.red.2010.08.004

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, *111*(4), 1092–1125. https://doi.org/10.1257/aer.20190964

Kilian, L. (2008). Exogenous oil supply shocks: How big are they and how much do they matter for the u.s. economy? *The Review of Economics and Statistics*, *90*(2), 216–240. Retrieved April 2, 2023, from https://www.jstor.org/stable/40043142

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, *99*(3), 1053–1069. https://doi.org/10.1257/aer.99.3.1053

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, *65*(3), 361–393. Retrieved March 7, 2023, from https://www.jstor.org/stable/2566931

Kurmann, A., & Otrok, C. (2017). News shocks and the slope of the term structure of interest rates: Reply. *American Economic Review*, *107*(10), 3250–3256. https://doi.org/10.1257/aer.20161946

Kurmann, A., & Sims, E. (2021). Revisions in utilization-adjusted TFP and robust identification of news shocks. *The Review of Economics and Statistics*, *103*(2), 216–235. https://doi.org/10.1162/rest_a_00896

Lagerborg, A., Pappa, E., & Ravn, M. O. (2022). Sentimental business cycles. *The Review of Economic Studies*, rdac053. https://doi.org/10.1093/restud/rdac053

Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica*, *81*(3), 1115–1145. https://doi.org/10.3982/ECTA8337

Lewis, D. J. (2019). Announcement-specific decompositions of unconventional monetary policy shocks and their macroeconomic effects. *Staff Reports*. Retrieved April 27, 2023, from https://ideas.repec.org//p/fip/fednsr/891.html

Lewis, D. J. (2021). Identifying shocks via time-varying volatility. *The Review of Economic Studies*, *88*(6), 3086–3124. https://doi.org/10.1093/restud/rdab009

Mertens, K., & Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks. *American Economic Journal: Economic Policy*, *4*(2), 145–181. https://doi.org/10.1257/pol.4.2.145

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, *103*(4), 1212–47. https://doi.org/10.1257/aer.103.4.1212

Mertens, K., & Ravn, M. O. (2019). The dynamic effects of personal and corporate income tax changes in the united states: Reply. *American Economic Review*, *109*(7), 2679–2691. https://doi.org/10.1257/aer.20180707

Miranda-Agrippino, S., & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics*, *13*(3), 74–107. https://doi.org/10.1257/mac.20180124

Miyamoto, W., & Nguyen, T. L. (2020). The expectational effects of news in business cycles: Evidence from forecast data. *Journal of Monetary Economics*, *116*, 184–200. https://doi.org/10.1016/j.jmoneco.2019.09.007

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, *126*(1), 1–50. https://doi.org/10.1093/qje/qjq008

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In *Handbook of macroeconomics* (pp. 71–162). Elsevier. https://doi.org/10.1016/bs.hesmac.2016.03.003

Ramey, V. A., & Zubairy, S. (2017). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, *126*(2), 850–901. https://doi.org/10.1086/696277

Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084. https://doi.org/10.1257/0002828042002651

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, *100*(3), 763–801. https://doi.org/10.1257/aer.100.3.763

Romer, C. D., & Romer, D. H. (2023). Presidential address: Does monetary policy matter? the narrative approach after 35 years. *American Economic Review*, *113*(6), 1395–1423. https://doi.org/10.1257/aer.113.6.1395

Schlaak, T., Rieth, M., & Podstawski, M. (2023). Monetary policy, external instruments, and heteroskedasticity. *Quantitative Economics*, *14*(1), 161–200. https://doi.org/10.3982/QE1511

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A bayesian DSGE approach. *American Economic Review*, *97*(3), 586–606. https://doi.org/10.1257/aer.97.3.586

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, *24*(4), 630–650. https://doi.org/10.1002/jae.1065

Wieland, J. (2021, March 23). Updated romer-romer monetary policy shocks. https://doi.org/10.3886/E135741V1

Zeev, N. B., & Pappa, E. (2017). Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks. *The Economic Journal*, *127*(603), 1568–1597. https://doi.org/https://doi.org/10.1111/ecoj.12349