ISSN 1936-5330

kcFED Research Working Papers

A New Approach to Integrating Expectations into VAR Models

Taeyoung Doh and A. Lee Smith

December 2018; revised February 2022
RWP 18-13
https://dx.doi.org/10.18651/RWP2018-13

Federal Reserve Bank of Kansas City

A New Approach to Integrating Expectations into VAR Models*

Taeyoung Doh† A. Lee Smith‡

First Draft: October 2020
This Draft: February 2022

Abstract

Expectations play a central role in macroeconomics. Expectations are empirically measured from surveys or financial markets and are frequently analyzed in Vector autoregression (VAR) models alongside realized data of the same variable. However, this leads to two different expectations for the same variable: the VAR-based forecast and the external forecast. This paper proposes a Bayesian prior over the VAR parameters which allows for varying degrees of consistency between these two forecasts. As we demonstrate in two applications, our approach can sharpen structural VAR identification of forward guidance shocks and enhances VAR forecasts of inflation tail risks.

Keywords: Bayesian Vector Autoregression (VAR), Sign Restrictions, Information Frictions, Monetary Policy, Forward Guidance, Inflation Expectations
JEL classification codes: C11, C32, E52, E31

*We thank Yuriy Gorodnichenko and an anonymous referee for incredibly constructive and helpful comments. We also thank participants at various conferences, including: the 2018 FRB-Chicago & NBER Midyear Meeting of the DSGE Models and Applications Group and the 2019 System Econometrics Group Meetings. We especially appreciate thoughtful comments from our formal discussants, Daniel Lewis and Paul Ho. We also benefited from conversations with Brent Bundick, Laura Gati, and John Keating. Jason Bae and Logan Hotz provided outstanding research support. We are grateful for computational and data support from CADRE. This paper supersedes an earlier working paper titled, "Reconciling VAR-based Forecasts with Survey Forecasts." The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System. The online appendix is available at `https://doi.org/10.18651/RWP2018-13`.

†Federal Reserve Bank of Kansas City. Email: Taeyoung.Doh@kc.frb.org
‡Federal Reserve Bank of Kansas City. Email: andrew.smith@kc.frb.org

1 Introduction

The conduct of monetary policy has shifted from primarily managing reserves and overnight interest rates to placing a greater emphasis on managing expectations, often through central bank communication. Therefore, analyzing expectations is necessary to study monetary policies which, for instance, aim to steer interest rate expectations through forward guidance or anchor inflation expectations. While vector autoregression (VAR) models remain one of the most flexible time-series tools for conducting such empirical macroeconomic research, there is no agreed upon way to integrate expectations into VAR models. At one extreme, including survey or financial market-implied expectations in a VAR model in a "model consistent" manner, in the sense that the survey or market forecast and the VAR-implied forecast align at every horizon, seems to contradict mounting empirical evidence against full information rational expectations (Coibion et al., 2018). Alternatively, at the other extreme, including expectations without imposing any connection between the VAR forecast and the survey or market forecast fails to fully leverage the rich information that expectations contain for VAR estimation, identification, and forecasting (see, for example, Cogley, 2005).

In this paper, we propose a Bayesian approach to integrating expectations, as measured by forecasts from surveys, into a VAR model that includes realized values of the same or a closely related variable. We leverage the fact that VARs themselves have the ability to generate a forecast for any variable included in the model. Then, using this VAR-implied forecast together with the survey forecast, we argue for constructing a non-degenerate prior over VAR coefficients that places greater mass on areas of the parameter space where the two forecasts align. We call this the *forecast consistent prior.* Our prior allows for a dynamic correlation between the realized data and its survey or market forecast, unlike the popular Minnesota prior, without dogmatically restricting the discrepancy between the two forecasts. From an implementation standpoint, our method relies on a computationally efficient importance sampling technique which simply re-weights the posterior draws. Intuitively, under the forecast consistent prior, a draw which is closer to satisfying forecast consistency receives a greater weight. These weights are informed by a hyperparameter which governs the precision of our prior and can be varied to allow for varying degrees of forecast consistency.

A novel aspect of our prior is its applicability to structural VAR models to aid in the identification of structural shocks. For some intuition, note that impulse responses are simply conditional VAR forecasts. When the forecast consistent prior is placed on unconditional forecasts then our prior shapes the reduced-form VAR coefficients. However, when the forecast consistent prior is placed on impulse responses — or forecasts that are conditional on the realization of a structural shock — then our prior can also inform the structural VAR coefficients. One natural application which we explore in this paper is the addition of our prior to a structural VAR identified using sign-restrictions. Structural VAR (SVAR) models identified by sign-restrictions on impulse responses only identify the model parameters up to a set, as discussed by Moon and Schorfheide (2012) and Uhlig (2017). Within this set, any two alternative SVAR models are equally probable. Forecast consistency restrictions on impulse responses provide probabilistic restrictions motivated by economic theory that can break this equality and distinguish alternative SVAR models.

We illustrate the usefulness of our approach in two applications to shed light on important issues for monetary policy, including: the effects of forward guidance shocks on output and the role that inflation expectations played in shaping inflation tail-risks after the Great Recession. The forward guidance application is particularly illustrative. In this application we add Blue Chip forecasts of one-year ahead short-term interest rates to the Uhlig (2005) VAR model. We apply similar sign restrictions that Uhlig (2005) proposes and find that forward guidance shocks which reduce survey expectations of future interest rates have an ambiguous effect on output. However, we also show that these sign-restrictions admit a wide range of VAR-implied forecasts of the federal funds rate, many of which deviate significantly from the path of rates predicted by forecasters following the forward guidance shock. We then layer the forecast consistent prior on top of the sign-restrictions to identify forward guidance shocks which better align these two forecasts. We find that such shocks lead output to rise and, moreover, the estimated output effects increase with the degree of forecast consistency. Monte Carlo simulations from a New-Keynesian model in which survey forecasts exhibit information rigidity suggest this result may arise from the ability of the forecast consistent prior to better recover the true forward guidance shocks as opposed to masquerading shocks (as in Wolf, 2020) which satisfy the sign restrictions but imply less forecast consistency.

2 The Forecast Consistent Prior

In this section we introduce our VAR notation, next we use a simple example with a bivariate VAR to introduce the forecast consistent prior, and then we show how to apply our prior in a more general VAR setting.

2.1 VAR Preliminaries

A reduced-form VAR(l) is given by:

$$Y_t = A_D + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_l Y_{t-l} + u_t, \tag{1}$$

where Y_t is an $m \times 1$ vector of data at date $t = 1 - l, \ldots, T$, A_i are coefficient matrices of size $m \times m$, and u_t is the one-step ahead prediction error, or reduced-form residuals, which are assumed to be distributed normally with mean 0 and covariance matrix Σ. A_D encapsulates constants and other deterministic components of the VAR model. We assume that this reduced-form VAR is derived from an underlying structural VAR model:

$$BY_t = B_D + B_1 Y_{t-1} + B_2 Y_{t-2} + \ldots + B_l Y_{t-l} + \varepsilon_t, \tag{2}$$

in which B is an $m \times m$ non-singular coefficient matrix which governs the contemporaneous interactions between the variables Y_t, B_i are coefficient matrices of size $m \times m$, and ε_t are the structural shocks which are independent of one another, mean zero, and have a standard deviation of one.

Combining equations (1) and (2) reveals that the reduced-form residuals u_t are related to the structural shocks by the mapping $u_t = B^{-1}\varepsilon_t$. Therefore, knowledge of B allows one to uncover the structural VAR model from the estimated reduced form VAR model. We parameterize B as follows:

$$B^{-1} = CQ, \tag{3}$$

where C is the lower-triangular Cholesky factor of Σ and Q is a square $m \times m$ orthogonal rotation matrix such that $Q'Q = QQ' = I_m$. This parameterization is able to encompass multiple identification strategies including sign-restrictions which consider a set of alternative rotation matrices Q and recursive short-run restrictions which achieve point identification by assuming that $Q = I_m$.

We take a Bayesian approach to estimation and inference throughout the paper. We describe the prior in terms of the stacked version of the reduced-form VAR in equation (1):

$$Y = XA + U, \tag{4}$$

where $X_t = [Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-l}]'$, $Y = [Y_1, \ldots, Y_T]'$, $X = [D, X_1, \ldots, X_T]'$, $U = [u_1, \ldots, u_T]'$ and $A = [A_D, A_1, \ldots, A_l]'$. The OLS estimates of A and Σ are given by:

$$\hat{A} = (X'X)^{-1}X'Y \tag{5}$$

$$\widehat{\Sigma} = \frac{1}{T-1}(Y - X\widehat{A})'(Y - X\widehat{A}) \tag{6}$$

We assume the following Normal and Inverse-Wishart conjugate priors for $\alpha = \text{vec}(A)$ and Σ, denoted by $p(\alpha, \Sigma) = p(\alpha|\Sigma)p(\Sigma)$, and parameterized by $\nu_0, V_0, \alpha_0 = vec(A_0),$ and S_0:

$$\Sigma \sim IW(S_0, \nu_0),$$
$$\alpha|\Sigma \sim \mathcal{N}(\alpha_0, \Sigma \otimes V_0).$$

Given these priors, the posterior distributions of α and Σ are:

$$\Sigma|Y \sim IW(S_T, \nu_T), \tag{7}$$

$$\alpha|\Sigma, Y \sim \mathcal{N}(\alpha_T, \Sigma \otimes V_T), \tag{8}$$

where:

$$\nu_T = \nu_0 + T,$$
$$V_T = [V_0^{-1} + X'X]^{-1},$$
$$A_T = V_T[V_0^{-1}A_0 + X'X\widehat{A}], \text{ and}$$
$$S_T = (Y - X\widehat{A})'(Y - X\widehat{A}) + S_0 + \widehat{A}'(X'X)\widehat{A} + A_0'V_0^{-1}A_0 - A_T'(V_0^{-1} + X'X)A_T,$$

with $\alpha_T = vec(A_T)$.[1] We assume non-informative priors throughout the empirical applications in this paper by setting V_0 to the zero matrix and $\nu_0 = 0$.

2.2 Introducing the Forecast Consistent Prior

We now introduce the forecast consistent prior. To generalize notation, let $\mathbf{A}(\alpha) = \mathbf{A}$ denote the companion form of the VAR(l) lag coefficients, $A_1, ..., A_l$, where the notation $\mathbf{A}(\alpha)$ makes explicit the fact that $\mathbf{A}$ is comprised of the elements of α. However, for illustrative purposes, assume for now that we are interested in a bi-variate VAR(1) with no deterministic components, so that $A_1 = \mathbf{A} = A$:

$$\begin{bmatrix} \pi_t \\ E_t^S(\pi_{t+1}) \end{bmatrix} = \mathbf{A} \begin{bmatrix} \pi_{t-1} \\ E_{t-1}^S(\pi_t) \end{bmatrix} + u_t = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ E_{t-1}^S(\pi_t) \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \tag{9}$$

where π_t is a variable of interest such as inflation and $E_t^S(\pi_{t+1})$ is the one-period ahead forecast of π_t obtained from survey data.[2] When left unconstrained, this model implies two forecasts for the same variable: the VAR-based forecast and the survey forecast. From the first equation of the VAR, we can generate the time t VAR-based forecast for π_{t+1}:

$$E_t^{\text{VAR}}(\pi_{t+1}) = a_{11}\pi_t + a_{12}E_t^S(\pi_{t+1}). \tag{10}$$

The only way for the VAR and survey forecasts to always be consistent with one another, so that $E_t^{\text{VAR}}(\pi_{t+1}) = E_t^S(\pi_{t+1})$ for all possible realizations of $u_t = [u_{1,t}, u_{2,t}]'$, is to impose the following restrictions on the VAR coefficients:

$$g(\mathbf{A}) = v_1'\mathbf{A} - v_2' I_{ml} = [0, 0] \;\Leftrightarrow\; [a_{11}, a_{12} - 1] = [0, 0], \tag{11}$$

where v_i is the companion-VAR variable selection vector defined as the i'th column of the I_{ml} identity matrix. However, imposing strict consistency at all times assumes that the two forecasts are conceptually and practically identical, assumptions which may be difficult to defend given issues ranging from the presence of information rigidities to measurement error.

[1] In principle, the forecast consistent prior can be paired with any prior over the VAR parameters.

[2] Throughout the paper we analyze survey-based expectations. However, the same methodology applies to expectations obtained from financial markets or other sources.

While acknowledging these issues, completely ignoring the relevance of the survey forecasts for the VAR forecasts neglects the potentially useful information the econometrician possesses about the a-priori relationships between the variables in the VAR.

By adopting a Bayesian approach, we can vary the tightness of these cross-equation restrictions to allow for the possibility that survey forecasts and VAR-based forecasts are related to one another, albeit imperfectly. We can consider the following three cases which vary based on the degree of forecast consistency imposed:

1. **Strict Forecast Consistency**: $g(\mathbf{A}) = [0, 0]$. Therefore, $a_{11} = 0$ and $a_{12} = 1$ is dogmatically imposed.

2. **Forecast Consistent Prior**: $g(\mathbf{A}) \sim \mathcal{N}(0, (\lambda W)^{-1})$. The forecast consistent prior is centered at zero so that, on average, the VAR and survey forecasts are consistent with one another. However, the two forecasts may deviate from each other from time to time with the size of the potential forecast deviations governed by the weighting matrix W and the tuning parameter λ.

3. **No Forecast Consistency**: $g(\mathbf{A})$ is left unrestricted which is equivalent to a diffuse prior over $g(\mathbf{A})$.

Strict forecast consistency and no forecast consistency can be regarded as limiting cases when λ approaches ∞ and 0, respectively. In terms of A, the prior density function of $g(\mathbf{A})$ can be treated as the likelihood function for A using observations satisfying these restrictions.[3]

We can calculate the posterior density of α under the forecast consistent prior at low computational cost by importance sampling. First, define the posterior density of α obtained without imposing forecasting consistency by $p(\alpha|Y, \Sigma)$ and define the forecast consistent prior density kernel for α by $h(\alpha) = p(g(\mathbf{A}))$. Second, obtain the forecast consistent posterior density of α:

$$p(\alpha|Y, \Sigma, g(\mathbf{A})) = \frac{h(\alpha)p(\alpha|Y, \Sigma)}{\int h(\alpha)p(\alpha|Y, \Sigma)d\alpha}. \tag{12}$$

[3] We specify the forecast consistent prior as a normal distribution which is the maximum entropy prior for $g(\mathbf{A})$ under the constraint that the first moment of $g(\mathbf{A})$ is zero and the second moment of $g(\mathbf{A})$ is $(\lambda W)^{-1}$ which, in the information-theoretic sense, minimizes the amount of prior information on non-targeted moments of $g(\mathbf{A})$ (Robert, 2007; Cover and Thomas, 2012).

Using the above definition, we can simulate posterior draws of α from $p(\alpha|Y, \Sigma, g(\mathbf{A}))$ simply by re-weighting the posterior draws from $p(\alpha|Y, \Sigma)$. For $\alpha(m)$, a random draw from $p(\alpha|Y, \Sigma)$, we define the following importance weight:

$$w(\alpha(m)) = \frac{h(\alpha(m))}{\sum_{j=1}^{M} h(\alpha(j))} \text{ , } \alpha(j) \sim p(\alpha|Y, \Sigma). \tag{13}$$

We then re-sample these draws according to the weights $[w(\alpha(1)), \ldots, w(\alpha(M))]$ to simulate the posterior density $p(\alpha|Y, \Sigma, g(\mathbf{A}))$.

2.3 The Forecast Consistent Prior in a Structural VAR Model

The previous section motivates our forecast consistent prior in a reduced-form VAR. In this setting, the forecast consistent prior offers a theoretically grounded approach to parameter shrinkage. However, the full conceptual appeal of this prior is best illustrated in the context of a structural VAR model. In a structural VAR, the forecast consistent prior informs the structural VAR parameters for which, even with infinite observations, the data is indeterminate. Therefore, the benefits of the forecast consistent prior can extend beyond shrinkage when the econometrician is attempting to identify structural innovations.

As discussed above, the mapping between the reduced form and structural VAR models is defined by the coefficient matrix B which we parameterize by $B^{-1} = CQ$ where C is the lower-triangular Cholesky factor of Σ and Q is a square $m \times m$ orthogonal rotation matrix. For illustrative purposes, consider for a moment the VAR(1), which generalizes to a VAR(l) by writing the VAR in companion form:

$$Y_t = A_1 Y_{t-1} + u_t = \mathbf{A} Y_{t-1} + \mathbf{C} Q \varepsilon_t, \tag{14}$$

where $\mathbf{C} = [C; 0_{m(l-1)\times m}]$ is the companion form of C and ε_t are the structural innovations of interest. Suppose that we are interested in the effects of one particular structural shock which, without loss of generality, we label $\varepsilon_{1,t}$. Then the initial or impact-effect of a 1 unit realization of $\varepsilon_{1,t}$ can be recovered directly from the first column of $\mathbf{C}Q$ which is simply the forecast of Y_t conditional on $\varepsilon_{1,t} = 1$ in period t and 0 in all other periods:

$$E_t^{VAR}(Y_t|\varepsilon_{1,t} = 1) = \mathbf{C}Qe_1, \tag{15}$$

where e_i is a selection vector defined as the i'th column of the I_m identity matrix. The $h-$step ahead impulse response to this structural shock can be written as:

$$IRF(\mathbf{A}, \mathbf{C}, Q, h|\varepsilon_{1,t} = 1) = E_t^{VAR}(Y_{t+h}|\varepsilon_{1,t} = 1) = \mathbf{A}^h\mathbf{C}Qe_1, \tag{16}$$

for $h = 0, \ldots, H$. The elements $\mathbf{A}$ and $\mathbf{C}$ of the h-step ahead impulse response can be informed by the observed data contained in Y. However, the matrix Q must be identified from prior economic reasoning since the likelihood function of the VAR is invariant to Q.

When the VAR contains both realized and survey data, forecast consistency provides theoretically grounded restrictions on the matrix Q which, together with Σ, provides the mapping between reduced form VAR residuals and structural shocks. For example, if we suppose once again that $y_t = [\pi_t, E_t^S(\pi_{t+1})]'$, where $E_t^S(\pi_{t+1})$ is the 1-step ahead forecast of π_t obtained from survey data, then forecast consistency suggests the restrictions:

$$[v_1'\mathbf{A} - v_2' I_{ml}]\mathbf{A}^{h-1}\mathbf{C}Qe_1 = 0, \tag{17}$$

should hold for $h = 1, \ldots, H$. More formally, the forecast consistent prior can be expressed as $g(\mathbf{A}, \mathbf{C}, Q|H) \sim \mathcal{N}(0_H, (\lambda W)^{-1})$ where W is a $H \times H$ weighting matrix and λ calibrates the overall tightness of the forecast consistency restrictions.

Equation (17) draws an important connection between unconditional and conditional forecast consistency. In the limiting case of strict unconditional forecast consistency, $v_1'\mathbf{A} = v_2' I_{ml}$, and the forecast consistent prior places no restrictions on Q. This extreme case might arise if, for example, the underlying data-generating process is a full information rational expectations model. However, in our forward guidance application, we find that unconditional forecast consistency doesn't hold and therefore the forecast consistent prior does in fact shape the posterior distribution of Q. Monte Carlo simulations from a rational expectations model therefore are not able to replicate our empirical findings suggesting that, in reality, the existence of information frictions may make this extreme case unlikely.

Structural VAR identification in the presence of survey data can benefit from the use of the forecast consistent prior by weighting posterior draws of α, Σ, and Q by the degree to which VAR-based and survey forecasts align. While this is true regardless of the identification strategy pursued, a growing VAR literature aims to identify structural shocks of interest by restricting the shape of the impulse responses to identify parameters in Q. VAR models identified by sign restrictions typically find a large set of Q matrices that are compatible with these restrictions (Moon and Schorfheide, 2012; Uhlig, 2005, 2017). In this setting, forecast consistent priors may be especially of interest. In particular, the sign-restricted VAR literature has been criticized on the grounds that sign-restrictions alone do not rule out structural VAR models which are inconsistent with narrative evidence or plausible empirical specifications of equilibrium models (Arias et al., 2018; Antolín-Díaz and Rubio-Ramírez, 2018; Wolf, 2020). While addressing the same shortcoming, our approach instead provides further restrictions to narrow the set of potential structural VAR models based on the dynamic linkages between two different forecasts of the same variable, which may be preferred in applications of sign-restrictions beyond identifying conventional monetary policy shocks.

We briefly illustrate the potential for our conditional forecast consistent prior to sharpen the identification of structural VAR parameters in the context of sign restrictions. First, define $\mathcal{Q}$ to be the identified set of rotation matrices Q that satisfy the sign restrictions:

$$\mathcal{Q} = \{Q | B^{-1} = CQ, v_i' \cdot IRF(\widehat{\mathbf{A}}, \widehat{\mathbf{C}}, Q, h | \varepsilon_{1,t} = 1)_r \geq (\leq) 0 \,, \forall\, r = 1, \cdots, R\}, \tag{18}$$

where, we assume for a moment that, $\mathbf{A}$ and $\mathbf{C}$ are fixed at their OLS estimates, and $v_i' \cdot IRF(\widehat{\mathbf{A}}, \widehat{\mathbf{C}}, Q, h | \varepsilon_{1,t} = 1)_r \geq (\leq) 0$ denotes the r-th restriction on the impulse response of VAR variables with R representing the total number of restrictions on impulse responses. Since the data do not provide additional information to distinguish different Q matrices in the set of $\mathcal{Q}$, and since all matrices in the set $\mathcal{Q}$ satisfy the sign restrictions, there is no basis for preferring one element of $\mathcal{Q}$ over another. In contrast, our forecast consistent prior explicitly breaks this symmetry by penalizing structural coefficients that generate greater divergence between the survey and VAR-based forecasts. Hence, our forecast consistent prior induces a conceptually meaningful prior over the set $\mathcal{Q}$ and hence over alternative structural VAR models.[4]

To make matters concrete, consider the simple VAR(1) model:

$$\begin{bmatrix} \pi_t \\ E_t^S(\pi_{t+1}) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ E_{t-1}^s(\pi_t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \tag{19}$$

where $E_t^S(\pi_{t+1})$ is the 1-step ahead forecast of π_t obtained from survey data. Suppose that we aim to identify a "news" shock for which survey forecasts increase. In addition to this sign normalization, we impose the sign restriction that the impact effect of this news shock on the realized data in π_t is positive. Identifying this shock requires requires identifying a column of Q. Without loss of generality, we label the news shock $\varepsilon_{1,t}$ and therefore aim to identify the elements of the first column of Q, $[q_{11}, q_{21}]'$. Without identifying restrictions, the identified set of q_{11} and q_{21} is the entire unit circle. We impose the following restrictions to further narrow the identified set of q_{11} and q_{21}.

1. **Normalization Restriction: The "news" shock increases survey forecasts in period t.**
$$R1(q_{11}, q_{21}) = \left\{ (q_{11}, q_{21}) \middle| \frac{\partial E_t^S(\pi_{t+1})}{\partial \varepsilon_{1,t}} = 0.5 q_{11} + q_{21} > 0 \right\}.$$

2. **Sign Restriction: Realized data increases alongside survey forecasts in period t.**
$$R2(q_{11}, q_{21}) = \left\{ (q_{11}, q_{21}) \middle| \frac{\partial \pi_t}{\partial \varepsilon_{1,t}} = q_{11} > 0 \right\}.$$

3. **Forecast Consistency Restriction: The VAR forecast in period $t+1$ is consistent with the increase in survey forecasts in period t.**

$$R3(q_{11}, q_{21}) = \left\{ (q_{11}, q_{21}) \middle| \frac{\partial E_t^S(\pi_{t+1})}{\partial \varepsilon_{1,t}} - \frac{\partial E_t^{VAR}(\pi_{t+1})}{\partial \varepsilon_{1,t}} = q_{21} - 0.5 q_{11} = 0 \right\}.$$

Under the forecast consistent prior, the forecast consistency restriction in $R3$ is loosely imposed by assuming that $q_{21} - 0.5q_{11}$ follows a Normal distribution centered around 0. We denote this prior by $g(Q|\mathbf{A}, \mathbf{C}, h = 1) = q_{21} - 0.5q_{11} \sim \mathcal{N}(0, \lambda^{-1})$.

[4] In practice, the Haar prior implicitly induces an informative, but arbitrary distribution over $\mathcal{Q}$ (Baumeister and Hamilton, 2015; Wolf, 2020).

We can graphically illustrate how the forecast consistent prior shapes the posterior set of the structural VAR parameters. Panel (a) of Figure 1 highlights the identified set of (q_{11}, q_{21}) when we impose only the normalization and sign restrictions ($R1$ and $R2$). The blue line on the unit circle outlines the identified set under these two restrictions. We cannot discriminate different locations in the blue line without additional identifying restrictions. When we augment $R1$ and $R2$ with $g(Q|\mathbf{A}, \mathbf{C}, h = 1)$, the forecast consistent prior, we can discriminate different points on the blue line in panel (a) in a probabilistic way. Panel (b) of Figure 1 illustrates the forecast consistent prior's impact on the posterior distribution of (q_{11}, q_{21}) by shading with darker colors the region of the parameter space that has a higher probability mass under our forecast consistent prior. The forecast consistent prior adds curvature to the posterior distribution of structural parameters which enables the econometrician to discriminate between alternative structural VAR models that are otherwise equally likely according to the sign-restrictions. To sharpen this argument further, in the online appendix, we show that it is possible to go beyond set identification and instead achieve point identification by selecting the elements of the appropriate column of Q to maximize the degree of forecast consistency within the set of rotation matrices which satisfy the sign restrictions.

2.4 Selecting The Degree of Forecast Consistency

Before turning to our applications, we discuss a critical issue in implementing the forecast consistent prior: selecting the value of the hyperparameter λ. The choice between alternative values of λ amounts to selecting among alternative VAR models. Therefore, in keeping with our Bayesian perspective, we can compare competing models indexed by alternative values of λ by way of the marginal likelihood. In the context of unconditional forecast consistency, using the marginal likelihood to select λ follows much of the existing Bayesian VAR literature which selects hyperparameters on priors over the reduced-form VAR parameters (α and Σ) by maximizing the marginal likelihood.

However, we can also use the marginal likelihood to select λ in structural VAR applications since the degree of conditional forecast consistency is jointly determined by both the reduced-form VAR parameters and Q. Therefore, although the Q matrix does not directly appear in the likelihood function, borrowing the insight of Poirier (1998), the marginal posterior of Q is updated from the marginal prior indirectly through its correlation with the posterior distribution of the reduced-form parameters — α and Σ — which are directly incorporated into the likelihood function. Intuitively, the degree of conditional forecast consistency embedded in impulse responses depends not only on Q but also on the reduced-form VAR parameters, α and Σ. Selecting λ to maximize the marginal likelihood provides a generally applicable criterion which we implement to select λ in the following applications. However, we also show how information outside of the VAR model can be incorporated to select λ in one variant of our forward guidance application, to which we now turn.

3 Identifying Forward Guidance Shocks

The textbook view in macroeconomics is that what matters for consumption and investment decisions is the entire path of expected future interest rates, not just interest rates today (Woodford, 2003). This notion underpins the Federal Reserve's regular use of communication to manage expectations of future interest rates through forward guidance. However, there is little agreement regarding the effectiveness of this increasingly used policy tool, with recent research identifying a number of frictions that could dampen or even reverse the intended effects of forward guidance on real activity (e.g. Nakamura and Steinsson, 2018).

We contribute to the forward guidance literature by empirically estimating the output effects of forward guidance using a combination of sign and forecast consistency restrictions. Forward guidance shocks are inherently shocks to expectations of future short-term interest rates; therefore we augment a standard monetary VAR model with Blue Chip survey forecasts of future interest rates. While, as Ramey (2016) stresses, it is crucial to include forward-looking variables in the VAR to identify policy news shocks, embedding survey forecasts of future interest rates alongside the federal funds rate results in two simultaneous forecasts for future interest rates: the VAR forecast and the survey forecast.

Our results indicate that the estimated output effects of forward guidance depend importantly on the degree of forecast consistency restrictions. A pure sign restrictions approach which leaves the consistency between the two interest rate forecasts unconstrained suggests that forward guidance has an ambiguous effect on output. However, when we add the forecast consistent prior to the sign restrictions, we find expansionary effects on output for large enough values of λ which governs the discrepancy across interest rate forecasts. Using the marginal likelihood criterion to calibrate λ, we estimate that output has modestly increased following FOMC communication which signaled a lower path of future interest rates.

Monte Carlo simulations from a New-Keyensian style DSGE model reveal that our empirical results may arise from a problem of misidentification, similar to the issue Wolf (2020) identifies in the original Uhlig (2005) model. Estimating structural VAR models on DSGE-generated data reveals that sign restrictions alone are generally insufficient to recover the true forward guidance shocks. However, when sign restrictions are combined with forecast consistency restrictions, the SVAR model does a much better job of recovering the true forward guidance shocks as opposed to masquerading forward guidance shocks which sign restrictions alone can not rule out.

3.1 Data and VAR Model

To estimate the effects of forward guidance shocks on output, we augment the monetary VAR model specified in Uhlig (2005). In particular, we use monthly real GDP as produced by Macroeconomic Advisers, the consumer price index, an index of commodity prices, the effective federal funds rate, non-borrowed reserves, total reserves, and 1 year ahead Blue Chip consensus economic forecasts for the 3-month Treasury bill rate. We take 100 times the natural log of all variables except for the federal funds rate and Blue Chip forecast for the 3-month Treasury bill rate. We model these time series as a VAR(3) based on the Akaike information criteria over the sample January 1994 to December 2007.[5]

3.2 Sign Restrictions & The Forecast Consistent Prior

We identify a forward guidance shock by restricting the sign of the impulse response of commodity prices, the consumer price index, and forecasts of future interest rates for the first 6 months after a forward guidance shock.[6] In particular, we restrict prices and future

[5] We find similar results in VAR models with longer lag lengths.

[6] Baumeister and Hamilton (2015) propose an alternative approach of restricting the signs of the parameters in the matrix B which governs the contemporaneous relationships between variables in the VAR, including the monetary policy rule. However, Kilian and Lütkepohl (2017) caution that this approach as it can induce unintentionally informative priors on B^{-1}. Uhlig (2017) suggests that whether one imposes

interest rates to move persistently in opposing directions following a forward guidance shock. These restrictions follow Uhlig (2005) who uses similar restrictions, though without expected interest rates, to distinguish conventional monetary policy shocks from other demand and supply disturbances. The notion that forward guidance shocks cause expected nominal interest rates and prices to move in opposite directions is supported by standard sticky-price models (Eggertsson and Woodford, 2003), models which attribute a large role to a "Fed information effect" (Nakamura and Steinsson, 2018), and models which dampen the output effects of forward guidance through limited information or other frictions (e.g., Del Negro et al., 2015; Kiley, 2016; McKay et al., 2016; Angeletos and Lian, 2018; Farhi and Werning, 2019; García-Schmidt and Woodford, 2019; Gabaix, 2020).

In addition to sign restrictions, we also employ the forecast consistent prior to identify forward guidance shocks. The forecast consistent prior imposes probabilistic restrictions over the impulse responses of the consensus Blue Chip survey forecast of the 3-month Treasury bill rate and the VAR forecast for short-term interest rates implied by the impulse responses of the federal funds rate. Consistency between these two forecasts requires:

$$g(Q, \mathbf{A}, \mathbf{C}|h) = \frac{\partial}{\partial \varepsilon_{fg,t}} E^{BC}_{t+h} R^{T-Bill}_{t+h+12} - \frac{1}{3}\sum_{j=0}^{2}\left[\frac{\partial}{\partial \varepsilon_{fg,t}} FF_{t+h+12-1+j}\right], \tag{20}$$

where $\frac{\partial y_{t+h}}{\partial \varepsilon_{fg,t}}$ denotes the period h impulse response of variable y_t to a forward guidance shock that occurred in period t, FF_t denotes the period t federal funds rate, and $E^{BC}_t R^{T-Bill}_{t+12}$ denotes the Blue Chip consensus economic forecast for the 3-month Treasury bill 4-quarters from time t. In the monthly Blue Chip survey, forecasters report what they expect the 3-month T-bill to average over the three months ending 4 quarters ahead. So, in November, forecasters report what they expect the yield on the 3-month Treasury bill to average in the three months of October, November, and December of the following year.[7] Historically, the yield on the 3-month Treasury bill rate has closely tracked the federal funds rate, therefore this forecast should be linked with the average federal funds rate over the three months ending in December of the following year. According to the VAR forecast, the response of

restrictions on B or B^{-1} likely depends on the application at hand. Given the difficulty in specifying a policy *rule* for forward guidance shocks, we directly impose restrictions on B^{-1}.

[7] Depending on whether the month is at the end or beginning of the quarter, the horizon of the forecast varies between 10-12 and 12-14 months. However, we must fix the horizon when implementing forecast consistency. This suggest that even under full information rational expectations, forecast consistency might only weakly hold due to measurement error.

the average federal funds rate over the three months ending 4-quarters from now, conditional on a forward guidance shock in period t, is given by $\frac{1}{3}\sum_{j=0}^{2}\left[\frac{\partial}{\partial \varepsilon_{fg,t}} FF_{t+12-1+j}\right]$. We can stack the forecast consistency restriction in equation (20) for $h = 0, 1, \ldots, H-13$ to form the forecast consistent prior $g(Q|\mathbf{A}, \mathbf{C}, H) \sim \mathcal{N}(0_{H-12}, I_{H-12}\lambda^{-1})$ where λ tunes the precision over the forecast consistent prior.[8]

3.3 Implementation of Sign Restrictions & The Forecast Consistent Prior

We specify our reduced form VAR as:

$$y_t = A_D + A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + u_t, u_t \sim \mathcal{N}(0, \Sigma),\ A = [A_D, A_1, A_2, A_3]', \tag{21}$$

where A_D includes a constant term but no time trend. Using our earlier parameterization of Σ from Section 2.1, $\Sigma = CQQ'C'$, where where C is the lower-triangular Cholesky factor of Σ and Q is a square $m \times m$ orthogonal rotation matrix. The mapping between the reduced form VAR residuals u_t and structural VAR shocks ε_t, is governed by the linear mapping $u_t = CQ\varepsilon_t$.

To focus on the issue of identification, rather than inference, we initially keep $\alpha = vec(\widehat{\mathbf{A}})$ and $\Sigma = (Y - X\widehat{A})'(Y - X\widehat{A})/(T-1)$ and consider only random draws of Q. However, in a robustness check, we jointly draw from the posterior of α and Σ as well as Q. We draw random orthogonal rotation matrices Q by drawing a 6×6 random square matrix denoted by χ, with each element of χ independently drawn from a standard normal distribution, and then we take the QR decomposition of χ using MATLAB's `[Q,R]=qr(`χ`)` function. Each draw of Q represents a candidate structural VAR model. Structural VAR models are kept if they satisfy the sign restrictions and are otherwise discarded.

After accumulating $M = 5,000$ orthogonal rotation matrices Q which satisfy the sign restrictions we then calculate the posterior weight of any given draw m under the forecast

[8]In practice, we set $H = 60$ to encompass the 48 period impulse responses we plot.

consistent prior according to:

$$w(Q(m)) = \frac{\exp(-0.5\lambda g(Q(m)|\mathbf{A}, \mathbf{C}, H)' g(Q(m)|\mathbf{A}, \mathbf{C}, H))}{\sum_{k=1}^{M} \exp(-0.5\lambda g(Q(k)|\mathbf{A}, \mathbf{C}, H)' g(Q(k)|\mathbf{A}, \mathbf{C}, H))}. \tag{22}$$

These weights form the importance sampling weights we use to simulate the posterior distribution of the SVAR. [9]

3.4 Impulse Response Functions

Figure 2 shows the median and 68 percent error bands of impulse response functions across the draws that satisfy the forward guidance sign restrictions. In this figure we set $\lambda = 0$. Therefore, we caution that the distribution of impulse responses is purely an artifact of the Haar prior (Baumeister and Hamilton, 2015). Nevertheless, this specification provides a close analogue to Uhlig (2005). Per the imposed restrictions, forecasted interest rates decline and commodity prices along with the overall price level rise for the first 6 months. In the months that follow the restricted periods, survey forecasts of interest rates remain low and commodity prices remain elevated. The price level continues to gradually climb throughout the impulse response horizon. Recall that our sign restrictions leave non-borrowed reserves, total reserves, and the path of the actual federal funds rate unconstrained. However, 12 months after the forward guidance shock, measures of bank reserves increase which precipitates a decline in the federal funds rate. After reaching a trough around the 12-month horizon, the federal funds rate begins rising and eventually overshoots its pre-shock path.

How does the VAR-implied path of the federal funds rate compare to forecasters' expectations of short-term interest rates? The red-dashed line in the bottom-left panel shows the VAR-implied forecast for the one-year ahead short-term interest rate based on the impulse response of the federal funds rate. One year after the forward guidance shock, the VAR-forecast of short-term rates falls by an amount similar to what forecasters anticipated. However, in subsequent months, the VAR path of short-term interest rates exceeds the path anticipated by forecasters and remains above the survey forecast for interest rate for several years. The estimated response of output using only sign restrictions suggests that forward

[9] We scale the size of each draw to have the same effect on interest rate forecasts before calculating the importance sampling weights. This prevents larger shocks from being penalized simply due to the size of the forward guidance shock. This is similar to the approach taken in Uhlig (2005, pg. 413).

guidance may not be effective in stimulating economic activity as output initially declines after the forward guidance shock and then gradually increases towards its pre-shock path.

We next add forecast consistency restrictions to these sign restrictions. That is to say that we now consider $\lambda > 0$ whereas in Figure 2 we set $\lambda = 0$. We first build some intuition for how alternative values of λ will influence the posterior distribution of the impulse responses. In particular, consider two particular candidate SVAR models among the 5000 draws that satisfy the sign restriction: the SVAR draw that comes the closest to satisfying the forecast consistency restriction in equation (20) and the SVAR draw that is the furthest from satisfying the forecast consistency restriction in equation (20). In other words, these are the two SVAR draws that register the highest and lowest values in our forecast consistent prior density. We respectively refer to these as the "best" and "worst" draws.

Figure 3 plots the impulse responses for the SVAR models associated with these best and worst draws. The green-dashed-dotted line represents the best draw and the red-dashed line represents the worst draw. To visually understand what is behind these rankings, the top-right panel of Figure 3 shows the cumulative forecast deviation. By construction, the period 48 cumulative deviation is closer to zero for the best draw than for the worst draw. In other words, the best draw results in a VAR-implied forecast of the federal funds rate which more closely mirrors the path of rates that professional forecasters expected. For the worst draw, the VAR-implied forecast of the federal funds rate meaningfully diverges from the path expected by forecasters. Although the response of output played no role in our selection, the best draw implies a persistent expansion in output while the worst draw suggests that output persistently declines following a reduction in expectations of future interest rates.

The impulse response of output is heavily influenced by the choice of λ. A tighter forecast consistent prior, achieved by selecting higher values of λ, will place greater weight on draws like the "best" one under which output expands and lower weight on draws like the "worst" one under which output contracts. Figure 4 specifically focuses on the effect λ has on the response of output 18 months following the forward guidance shock. For smaller values of λ output fails to meaningfully rise — and may actually decline — following what ought to be an expansionary forward guidance shock. However, as λ increases so too does the response of output at 18 months. Figure 4 therefore illustrates that the degree of forecast consistency plays a significant role in shaping the estimated output effects from forward guidance.

Which of these responses best characterizes the U.S. experience with forward guidance? To answer this, we select a particular value of λ by maximizing the marginal likelihood function, $p_\lambda(Y)$:

$$p_\lambda(Y) \propto \int p(Y|\alpha,\Sigma)p_{\text{sign restriction}}(\alpha,\Sigma,Q)p(g(\mathbf{A}(\alpha),\mathbf{C}(\Sigma),Q)|\lambda)d(\alpha,\Sigma,Q), \qquad (23)$$

where $p_{\text{sign restriction}}(\alpha,\Sigma,Q)$ is the probability of the triplet (α,Σ,Q) satisfying the sign restrictions. The dependency of $p_\lambda(Y)$ on λ is made clear by the fact that by varying λ, the forecast consistent prior will alter the weights given to various draws of α, Σ, and Q. By reweighting posterior draws of (α,Σ,Q) based on the forecast consistent prior conditional on a particular value of λ, we generate posterior draws of (α,Σ,Q) from the posterior kernel of $p_\lambda(Y)$ given λ. We approximate the marginal likelihood by computing the harmonic mean from these posterior draws across a grid of values for λ and select the λ which maximizes the marginal likelihood on this grid. [10]

Figure 5 shows the median and 68 percent error bands of impulse response functions across the draws that satisfy the sign restrictions when reweighted with $\lambda > 0$. The top-left panel of Figure 5 shows that output rises in a gradual but persistent manner following an expansionary forward guidance shock for the value of λ which maximizes the marginal likelihood. The bottom left panel of Figure 5 illustrates that this setting of the forecast consistent prior reduces the deviation between the VAR-implied path of the federal funds rate and the path of rates anticipated by professional forecasters following a forward guidance shock. More precisely, the cumulative deviation between the VAR-based forecast and the Blue Chip forecast falls from 119 basis points when $\lambda = 0$ to 76 basis points under our calibration of λ, a 43 basis point reduction.

3.5 Impulse Response Functions for Alternative Specifications

The finding that identified forward guidance shocks which better align interest rate expectations, as calibrated through larger values of λ, lead to more expansionary estimated output effects is shared by several alternative VAR specifications. Figure 6 shows impulse responses from three variants of our baseline VAR model: one which extends the estimation sample

[10] This procedure, detailed in the online appendix, selects $\lambda = 6.36 \times 10^8$.

to include the 2008-2015 zero lower bound period, one which calibrates λ with the use of a high-frequency measure of forward guidance shocks, and one which sequentially draws from the posterior of α and Σ as well as Q. We briefly discuss each in turn.[11]

The first column of Figure 6 shows that over an extended sample that runs through 2015, we continue to find expansionary effects for the value of λ which maximizes the marginal likelihood and contractionary effects when $\lambda = 0$. The second column of Figure 6 shows that we continue to find evidence that output expands following a downward revision to interest rate forecasts if, instead of using the marginal likelihood criterion, we calibrate λ to maximize the correlation between our SVAR forward guidance shocks and high-frequency financial market measures of forward guidance shocks constructed by Swanson (2021). Finally, the third column of Figure 6 shows impulse responses when we extend our forecast consistent prior to shape the posterior of the full set of VAR parameters, including the VAR lag coefficients α and the covariance matrix Σ. The range of impulse responses drawn from the posterior of the full set of VAR parameters is understandably wider than in our baseline results. However, the ranking of responses remains with larger values of λ implying more expansionary effects from forward guidance.

3.6 Interpreting Forecast-Consistent Forward Guidance Shocks

Our principal empirical finding is that the estimated output effects of forward guidance shocks are largely shaped by the extent to which interest rate forecasts align. We now offer an interpretation for this result based on Monte Carlo evidence from a New-Keynesian DSGE model in which survey forecasts exhibit information rigidity. We show that in this setting better aligned interest rate forecasts across "survey" data and the estimated VAR model — implemented by a tighter prior over forecast consistency — improves the ability of the SVAR to recover the true underlying forward guidance shocks.

[11] We include the full impulse responses for these alternative specifications in the online appendix.

The core of our model can be defined by three familiar equations:

$$x_t = \mathbb{E}_t x_{t+1} - \left(i_t - \mathbb{E}_t \pi_{t+1} - (a_t - \mathbb{E}_t(a_{t+1}))\right), \tag{24}$$

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t \pi_{t+1} + \mu_t, \tag{25}$$

$$i_t = \phi_i i_{t-1} + (1-\phi_i)\left(\phi_x x_t + \phi_{\Delta x}(x_t - x_{t-1}) + \phi_\pi \pi_t\right) + \sigma_{fg}\epsilon^{fg}_{\tau,t-\tau} + \sum_{j=0}^{\tau-1}\sigma_j \epsilon^j_{j,t-j}, \tag{26}$$

where $\mathbb{E}_t$ is the rational expectations operator, x_t is output, π_t is the per-period inflation rate, and i_t is the one period nominal interest rate, each expressed as log-deviations from their non-stochastic steady state. The model follows the textbook treatment in Woodford (2003) and Galí (2008), in which one can find detailed derivations of these equations. In short, equation (24) is the dynamic IS curve, equation (25) is the New-Keynesian Phillips curve, and equation (26) is the monetary policy rule. Four underlying structural shocks drive the model: a serially correlated aggregate demand disturbance $a_t = \rho_a a_{t-1} + \sigma_a \epsilon^a_t$, a serially correlated aggregate supply disturbance $\mu_t = \rho_\mu \mu_{t-1} + \sigma_\mu \epsilon^\mu_t$, a forward guidance shock $\epsilon^{fg}_{\tau,t-\tau}$, and a noise shock to measured survey forecasts which will be discussed momentarily. The details of the model calibration are provided in the online appendix.

The innovation $\epsilon^{fg}_{\tau,t-\tau}$ is a monetary policy news shock which agents learn about in period $t-\tau$ but isn't realized until period t. Therefore, this represents a forward guidance shock. The additional policy innovations $\sum_{j=0}^{\tau-1}\sigma_j\epsilon^j_{j,t-j}$ are included so that an expansionary forward guidance announcement does not incite an immediate interest rate increase due to the endogenous policy response to higher inflation and output. Instead, we assume that a τ-period ahead forward guidance shock announced in period t is accompanied by an announcement that the policy rate will be held fixed at its pre-shock level until period $t+\tau$, similar to Campbell et al. (2017, pg. 329). For our baseline specification we set $\tau = 1$ and we choose $\sigma_0\epsilon^0_{0,t}$ so that the current policy rate is unchanged following a forward guidance announcement.[12]

[12] In the online appendix, we show results when we relax both the assumption that $\tau = 1$ and the assumption that forward guidance does not provoke an immediate policy response (i.e. we set set $\sum_{j=0}^{\tau-1}\sigma_j\epsilon^j_{j,t-j} = 0$). The results from these robustness exercises reassuringly remain consistent with improved identification of forward guidance shocks from the use of the forecast consistent prior, though the magnitude of improvement is somewhat diminished.

We assume the 1-period ahead survey forecast for the policy rate, $q_t = \mathbb{E}_t^S i_{t+1}$, evolves according to:

$$q_t = (1-\rho)\mathbb{E}_t(i_{t+1}) + \rho q_{t-1} + \sigma_n \varepsilon_t^n, \tag{27}$$

where $\mathbb{E}_t(i_{t+1})$ is the full-information rational expectation for 1-period ahead interest rates and ε_t^n is a mean zero iid noise shock. The parameter ρ governs the weight forecasters give to old information in forming their forecast. Therefore, when $\rho > 0$, measured survey forecasts feature information rigidities along the lines documented by Coibion and Gorodnichenko (2012) in that they only gradually adopt new information and are noisy measures of the true rational expectations. Most relevant for our purpose, when $\rho > 0$ and $\sigma_n > 0$, unconditional forecast consistency between survey and VAR forecasts will not hold, leaving room for the forecast consistent prior to shape structural shock identification. In our calibration, we set $\rho = 0.86$, so that equation (27) aligns with the information rigidities present in professional forecasts documented by Coibion and Gorodnichenko (2012, pg. 143).

We simulate a sample of 50,000 observations from the DSGE model and estimate a four variable VAR model with $Y_t = [x_t, \pi_t, i_t, \mathbb{E}_t^S i_{t+1}]'$. We identify forward guidance shocks using sign restrictions that call for $\mathbb{E}_t^S i_{t+1}$ and π_t to move in opposite directions for the first 6 periods following a forward guidance shock. Just as in our empirical application, we then layer forecast consistency restrictions on top of these sign restrictions. For each value of λ, which governs the tightness of the forecast-consistent prior, we report pseudo shock weights by calculating the pairwise correlations between the SVAR-identified forward guidance shock and each of the four DSGE model shocks, scaled by the norm of these four correlations.[13]

Figure 7 shows results from these Monte Carlo simulations. The solid blue line shows that as λ increases —and therefore the tightness of the forecast consistency restrictions increases —the SVAR-identified forward guidance shock places more weight on the true forward guidance shock. Moreover, as λ increases, the weight that the SVAR-identified shocks places on the DSGE noise and supply shocks is diminished. Intuitively, forecast consistency restrictions can deliver better alignment between the SVAR-identified and true forward guidance shocks because forward guidance shocks in the DSGE model imply a greater

[13] While not the same as the shock weights in Wolf (2020), these correlation-based weights are closely related to the coefficients from regressing the SVAR-identified forward guidance shock on the four (standardized) DSGE model shocks and therefore can be compared to give a sense of the relative relationship between our SVAR-identified forward guidance shock and each of the DSGE model shocks.

degree of forecast consistency than does the linear combination of noise and other shocks that could masquerade as forward guidance shocks based on sign-restrictions alone. For example, we can construct a linear combination of the four DSGE shocks which places a relatively low weight on the true forward guidance shocks but nevertheless satisfies the sign restrictions.[14] This masquerading shock produces a cumulative forecast discrepancy of about 19 basis points whereas the cumulative forecast discrepancy following the true forward guidance shock is just 3 basis points. Therefore, imposing forecast consistency in the SVAR helps to distinguish forward guidance shocks from linear combinations of other shocks that could masquerade as a forward guidance shock based on sign-restrictions alone.

4 Inflation Tail Risks

According to former Federal Reserve Chairman, Alan Greenspan, "the conduct of monetary policy in the United States has come to involve, at its core, crucial elements of risk management." One element of this policy strategy involves managing risks around the Federal Reserve's price stability mandate. While during much of the 1980's and 1990's the FOMC was primarily concerned with defending its inflation mandate from above, in more recent decades it has confronted the risk of too low inflation. Deflation concerns were especially elevated following the Global Financial Crisis, sparking broad interests in analyzing inflation tail risks over this period.

Several recent papers have measured inflation tail risks using financial market data and identify an elevated risk of deflation in the aftermath of the Global Financial Crisis. Fleckenstein et al. (2017) estimate a model of time-varying deflation risks identified from inflation swaps and options. Similarly, Anene and D'Amico (2017) and Hattori et al. (2016) use inflation derivatives to study the impact of the FOMC's unconventional policy actions on stemming the risk of deflation. We contribute to the literature by studying inflation tail risks implied by a time-varying parameter (TVP-)VAR model of inflation that includes both near-term and longer-term survey forecasts of inflation. Relative to this previous literature,

[14] This is the "median" identified SVAR forward guidance shock using solely sign restrictions with respective weights on the DSGE forward guidance, demand, supply, and noise shocks of (0.32, -0.50, -0.50, 0.63).

our approach has the appeal that, unlike financial market data, survey forecasts are not contaminated by time-varying risk premiums.[15]

This application illustrates some novel features of our forecast consistent prior that we have yet to demonstrate. First, in this application, the forecast consistent prior is applied in a setting with multiple survey measures at multiple horizons. Second, by simulating tails of the predictive density of inflation from a VAR model with survey forecasts, we demonstrate the appeal of augmenting VAR models with survey data to elicit moments of the forecast density that may not be provided by surveys.

Our results suggest that inflation expectations as measured from survey forecasts play an important role in shaping inflation tail risks. In particular, estimates from our forecast consistent TVP-VAR suggest that the risk of deflation during the Global Financial Crisis and its aftermath was generally lower than unconstrained VAR models and financial market estimates might suggest. The role of inflation expectations in influencing tail risks is especially vivid when survey forecasts for near-term inflation expectations hit all-time lows in 2009:Q1, which led deflation risks from the TVP-VAR model to sharply rise. However, as survey forecasts rebounded in subsequent quarters deflation probabilities quickly fell. In contrast, deflation risks from the TVP-VAR model without forecast consistency remain persistently elevated after 2009. The marginal likelihood criterion favors the TVP-VAR model with some degree of forecast consistency imposed, implying that the data prefers a model which more tightly links deflation risks to variation in survey forecasts for inflation.

4.1 Data and TVP-VAR Model

The specification of our VAR follows Clark and Davig (2011) closely by including near- and long-term survey forecasts of inflation alongside realized inflation, a measure of real economic activity, and a measure of the policy rate. We use both 1 year and 10-year ahead forecasts for Consumer Price Index (CPI) inflation from the Survey of Professional Forecasters (SPF)

[15] Our approach is comparable to Kozicki and Tinsley (2012) who estimate a time-varying parameter TVP-AR(p) model for inflation together with survey forecasts. However, they allow only constant terms to drift over time while lag coefficients in the AR(p) model remain time-invariant. Keeping lag coefficients time-invariant has some limitations in modeling changes in the relationship between inflation and inflation expectations, particularly if the persistence of inflation has varied over time.

as well as realized CPI inflation.[16] We include the Chicago Fed National Activity Index to broadly measure real activity and the federal funds rate to account for the stance of monetary policy. Our formal estimation sample is 1982-2015 which includes the zero lower bound period. Therefore, to better account for the full spectrum of the FOMC's policy actions from 2009-2015 we splice the effective federal funds rate together with Wu and Xia (2016) shadow federal funds rate.

We model these five variables as a TVP-VAR(4) with stochastic volatility:

$$y_t = A_{D,t} + \sum_{j=1}^{4} A_{j,t} y_{t-j} + u_t \,,\; u_t \sim \mathcal{N}(0, B^{-1}\Sigma_{u,t}B^{-1'}), \tag{28}$$

where $\Sigma_{u,t}$ is a diagonal matrix with positive, time-varying entries and B is a constant-parameter lower-diagonal matrix. We detail the construction of our priors and the algorithm used to estimate and simulate the TVP-VAR model in the online appendix.

4.2 The Forecast Consistent Prior

The TVP-VAR model contains both near-term and long-term survey forecasts of inflation as well as realized inflation. Therefore, we impose our forecast consistent prior over the survey and VAR-based forecasts at both forecast horizons. To define these forecast consistency restrictions, we express the TVP-VAR model in equation (28) in companion form:

$$\tilde{y}_t = \mathbf{A}_{D,t} + \mathbf{A}_t \tilde{y}_{t-1} + \tilde{u}_t. \tag{29}$$

Let π_t denote realized CPI quarterly annualized inflation and let $\pi_t^{e,L}$ and $\pi_t^{e,S}$ denote the respective long-term forward and short-term weighted averages of expected inflation over different horizons under the expectation operator E^e:

$$\pi_t^{e,L} = \frac{\sum_{j=5}^{40} E_t^e(\pi_{t+j})}{36} \qquad \text{and} \qquad \pi_t^{e,S} = \frac{\sum_{j=1}^{4} E_t^e(\pi_{t+j})}{4},$$

[16] The 10-year ahead forecasts for CPI inflation from SPF are available beginning in 1991. Prior to 1991, we use long-run inflation forecasts obtained from the public release of the Federal Reserve Board of Governors's FRB/US econometric model which is constructed using alternative surveys and econometric estimates. We use realized inflation and inflation nowcasts to construct our inflation expectations measures to prevent overlap between long-term survey forecasts, near-term survey forecasts, and realized inflation.

where E_t^e is the survey expectation when $e = S$ and the VAR expectation when $e = VAR$.[17]

Forecast consistency at both the long- and short-forecast horizons, denoted respectfully by $g(A_t) = [g(A_t)_L, g(A_t)_S]'$, requires the following restrictions on $\mathbf{A}_{D,t}$ and $\mathbf{A}_t$:

$$\begin{aligned} g(A_t)_L &= \left[v_\pi' \frac{[\sum_{h=5}^{40} \sum_{j=0}^{h-1} \mathbf{A}_t^j \mathbf{A}_{D,t}])}{36}, v_{\pi,L}' - v_\pi' \frac{[\sum_{h=5}^{40} \mathbf{A}_t^h]}{36} \right]' \\ g(A_t)_S &= \left[v_\pi' \frac{[\sum_{h=1}^{4} \sum_{j=0}^{h-1} \mathbf{A}_t^j \mathbf{A}_{D,t}]}{4}, v_{\pi,S}' - v_\pi' \frac{[\sum_{h=1}^{4} \mathbf{A}_t^h]}{4} \right]' \end{aligned} \tag{30}$$

In the above calculations of the VAR-implied forecasts, we assume no future parameter drift so that the VAR coefficients are fixed at their time t estimates.

As before, we calibrate the hyperparamter λ that controls the tightness of forecast consistency prior restrictions by selecting λ to maximize the marginal data density:

$$p_(y^T|\lambda) = \int p(y^T|A^T, \Sigma_u^T, B)p(A^T, \Sigma_u^T, B)p(g(A^T)|\lambda)d(A^T, \Sigma_u^T, B). \tag{31}$$

The log marginal likelihood is maximized at $\lambda = 1.42$.[18] Therefore, imposing a modest degree of forecast consistency improves the time series fit of the TVP-VAR model.

4.3 Time-Varying Inflation Tail Risks

While survey forecasts are available for mean inflation outcomes, our interest in this application lies in assessing potential tail outcomes for inflation for which survey forecasts are not consistently available. Therefore, the forecast consistent prior is particularly useful as it tilts the mean of the predictive distribution of inflation in the TVP-VAR towards the available survey forecasts and then relies on the VAR model to generate other moments of the predictive distribution of inflation. We simulate the predictive distribution of inflation outcomes using posterior draws of parameters and shocks up to time t from the TVP-VAR model. In practice, we achieve this by generating a full trajectory of inflation for the m-th posterior

[17]Since the end-point of the SPF 10-year inflation forecasts changes only in the first quarter of each year, the notation in the text is illustrative and applies only to the first quarter of the year.

[18]We calculate the marginal likelihood using the harmonic mean of the likelihood implied by posterior draws. Further details are provided in the online appendix.

draw of parameters and shocks, denoted by $\pi_{t+h|t}(m)$, for each draw $m = 1, \cdots, M$ from the posterior. We then compute the probability that inflation will be less than 0 percent on average over the next H-periods according to:

$$P\left(\sum_{h=1}^{H} \pi_{t+h|t} < 0\right) = \frac{\sum_{m=1}^{M} \mathbb{I}(\sum_{h=1}^{H} \pi_{t+h|t}(m) < 0)}{M}. \tag{32}$$

We can perform a similar calculation to assess the likelihood of high inflation which, following Fleckenstein et al. (2017), we define as inflation above 4 percent.

Table 1 compares estimated deflation probabilities from the TVP-VAR with those from Fleckenstein et al. (2017) and the cross-sectional distribution of individual expectations from the University of Michigan consumer survey over the period of 2009:Q4-2015:Q4.[19] Although the underlying source data are completely different, the mean and median probabilities of deflation are quite comparable given the magnitude of the standard deviation of each measure. However, the TVP-VAR model with forecast consistency restrictions implies uniformly lower deflation probabilities across all three quantile estimates (minimum, median, maximum) at both the 1- and 2-year horizons. This is also true for the estimated probability of high inflation (inflation greater than 4 percent) in Table 2.

The lower levels of inflation tail risks from our forecast consistent TVP-VAR model suggests that both financial market measures as well as unconstrained VAR models tend to overstate tail risks to inflation. Moreover, simply including mean survey forecasts of inflation in the VAR fails to fully capture the potential role of survey expectations in driving inflation tail risks. In contrast, the forecast consistency restrictions more tightly link inflation tail risks to survey forecasts of inflation. For example in 2009, shortly after oil prices fell by more than \$100 per barrel, one-year ahead survey forecasts for inflation hit all-time lows. The bottom row of Figure 8 shows that the risk of deflation sharply increased at this time.Then, as near-term inflation expectations rebounded, the risk of deflation subsided according to the forecast consistent TVP-VAR but remained elevated in the unrestricted TVP-VAR.

[19]For the University of Michigan survey, we calculate the percentage of respondents who anticipated prices would go down among all the respondents who provided answers on expected prices.

5 Conclusion

A growing literature has incorporated survey measures of expectations into VAR models to capture the importance of forward-looking behavior. In this paper, we have proposed the *forecast consistent prior* as a computationally efficient Bayesian approach for the estimation and inference of VAR models with survey forecasts and realized data of the same variable. We highlight possible applications of our framework in the context of both structural VAR shock identification as well as VAR forecasting. The applications shed light on the benefits of imposing forecast consistency to identify forward guidance shocks and the role that inflation expectations play in shaping inflation tail risks. However, many applications remain given the growing interest in identifying news shocks as well as the increased interest in macroeconomics of understanding the formation and evolution of expectations. Therefore, as these literatures advance, there appears to be a growing need for frameworks which can flexibly and efficiently incorporate expectations into VAR models.

References

Anene, D., D'Amico, S., 2017. A Tale of Four Tails: Inflation, the Policy Rate, Longer-Term Rates, and Stock Prices. Working Paper 2017-26. Federal Reserve Bank of Chicago.

Angeletos, G.M., Lian, C., 2018. Forward guidance without common knowledge. American Economic Review 108, 2477–2512.

Antolín-Díaz, J., Rubio-Ramírez, J.F., 2018. Narrative sign restrictions for svars. American Economic Review 108, 2802–29.

Arias, J.E., Caldara, D., Rubio-Ramirez, J.F., 2018. The systematic component of monetary policy in svars: an agnostic identification procedure. Journal of Monetary Economics Forthcoming.

Baumeister, C., Hamilton, J.D., 2015. Sign restrictions, structural vector autoregressions, and useful prior information. Econometrica 83, 1963–1999.

Campbell, J.R., Fisher, J.D.M., Justiniano, A., Melosi, L., 2017. Forward Guidance and Macroeconomic Outcomes since the Financial Crisis. NBER Macroeconomics Annual 31, 283–357.

Clark, T.E., Davig, T., 2011. Decomposing the declining volatility of long-term inflation expectations. Journal of Economic Dynamics and Control 35, 981–999.

Cogley, T., 2005. Changing beliefs and the term structure of interest rates: Cross-equation restrictions with drifting parameters. Review of Economic Dynamics 8, 420–451.

Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities? Journal of Political Economy 120, 116–159.

Coibion, O., Gorodnichenko, Y., Kamdar, R., 2018. The formation of expectations, inflation, and the phillips curve. Journal of Economic Literature 56, 1447–91.

Cover, T.M., Thomas, J.A., 2012. Elements of information theory. John Wiley & Sons.

Del Negro, M., Giannoni, M., Patterson, C., 2015. The Forward Guidance Puzzle Working Paper.

Eggertsson, G.B., Woodford, M., 2003. The Zero Lower Bound on Interest Rates and Optimal Monetary Policy. Brookings Papers on Economic Activity 1, 139–211.

Farhi, E., Werning, I., 2019. Monetary policy, bounded rationality, and incomplete markets. American Economic Review 109, 3887–3928.

Fleckenstein, M., Longstaff, F.A., Lustig, H., 2017. Deflation risk. The Review of Financial Studies 30, 2719–2760.

Gabaix, X., 2020. A behavioral new keynesian model. American Economic Review 110, 2271–2327.

Galí, J., 2008. Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework. Princeton University Press.

García-Schmidt, M., Woodford, M., 2019. Are low interest rates deflationary? a paradox of perfect-foresight analysis. American Economic Review 109, 86–120.

Hattori, M., Schrimpf, A., Sushko, V., 2016. The response of tail risk perceptions to unconventional monetary policy. American Economic Journal: Macroeconomics 8, 111–36.

Kiley, M.T., 2016. Policy paradoxes in the new keynesian model. Review of Economic Dynamics 21, 1–15.

Kilian, L., Lütkepohl, H., 2017. Structural vector autoregressive analysis. Cambridge University Press.

Kozicki, S., Tinsley, P.A., 2012. Effective use of survey information in estimating the evolution of expected inflation. Journal of Money, Credit and Banking 44, 145–169.

McKay, A., Nakamura, E., Steinsson, J., 2016. The Power of Forward Guidance Revisited. American Economic Review 106, 3133–3158.

Moon, H.R., Schorfheide, F., 2012. Bayesian and frequentist inference in partially identified models. Econometrica 80, 755–782.

Nakamura, E., Steinsson, J., 2018. High-frequency identification of monetary non-neutrality: the information effect. The Quarterly Journal of Economics 133, 1283–1330.

Poirier, D.J., 1998. Revising beliefs in nonidentified models. Econometric Theory 14, 483–509.

Ramey, V.A., 2016. Macroeconomic shocks and their propagation, in: Handbook of Macroeconomics. Elsevier. volume 2, pp. 71–162.

Robert, C., 2007. The Bayesian Choice. Springer Science & Business Media.

Swanson, E.T., 2021. Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. Journal of Monetary Economics 118, 32–53.

Uhlig, H., 2005. What are the effects of monetary policy on output? results from an agnostic identification procedure. Journal of Monetary Economics 52, 381–419.

Uhlig, H., 2017. Shocks, sign restrictions, and identification, in: Advances in Economics and Econometrics: Volume 2: Eleventh World Congress, Cambridge University Press. p. 95.

Wolf, C.K., 2020. Svar (mis)identification and the real effects of monetary policy shocks. American Economic Journal: Macroeconomics 12, 1–32.

Woodford, M., 2003. Interest and Prices. Princeton University Press.

Wu, J.C., Xia, F.D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. Journal of Money, Credit and Banking 48, 253–291.

Table 1: Summary Statistics for Deflation Probabilities: 2009:Q4 - 2015:Q4

Horizon	Source	Mean	Standard Deviation	Minimum	Median	Maximum
1 year	Fleckenstein et al. (2017)	18.76	10.16	2.00	15.74	47.76
1 year	University of Michigan Survey	15.91	6.12	6.06	14.14	32.65
1 year	TVP-VAR($\lambda = 0$)	13.74	9.62	2.72	10.54	45.98
1 year	TVP-VAR($\lambda = 1.4$)	9.03	8.22	1.66	7.32	42.78
2 years	Fleckenstein et al. (2017)	13.80	7.26	1.80	11.62	35.42
2 years	TVP-VAR ($\lambda = 0$)	16.34	6.27	8.38	15.14	32.06
2 years	TVP-VAR ($\lambda = 1.4$)	9.54	4.59	4.10	7.98	23.96

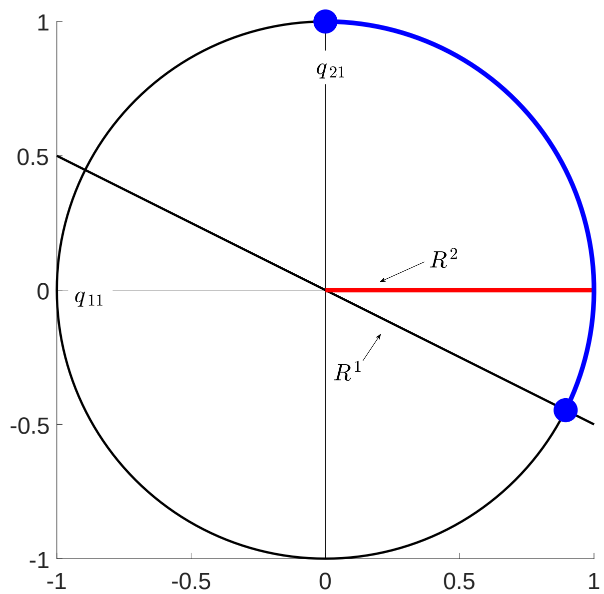
Notes: Deflation probabilities from Fleckenstein et al. (2017) are based on daily observations of inflation swaps and options from October 5, 2009 to October 28, 2015 while those from the University of Michigan survey are quarterly average values of monthly observations from October, 2009 to October, 2015.

Table 2: Summary Stastistics for High Inflation (>4 percent) Probabilities: 2009:Q4 - 2015:Q4

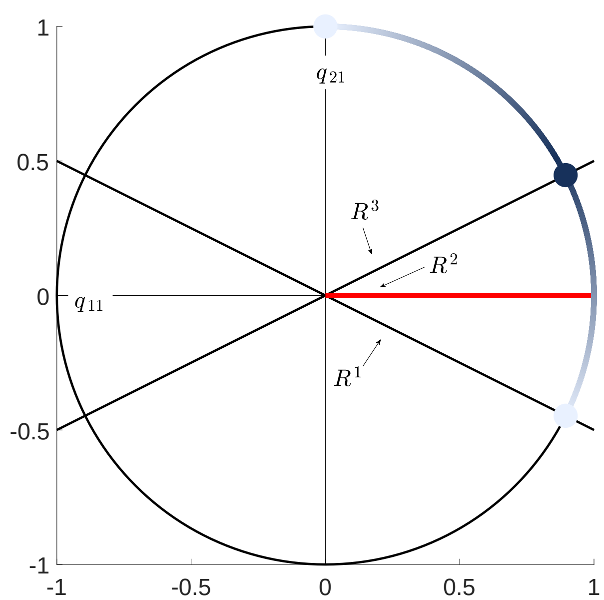
Horizon	Source	Mean	Standard Deviation	Minimum	Median	Maximum
1 year	Fleckenstein et al. (2017)	2.62	2.46	0.14	2.09	16.27
1 year	TVP-VAR ($\lambda = 0$)	1.34	2.30	0.04	0.40	10.92
1 year	TVP-VAR ($\lambda = 1.4$)	0.64	1.29	0.00	0.16	6.06
2 years	Fleckenstein et al. (2017)	3.09	2.39	0.30	2.70	15.32
2 years	TVP-VAR ($\lambda = 0$)	4.10	2.86	0.88	3.14	14.02
2 years	TVP-VAR ($\lambda = 1.4$)	2.30	2.17	0.32	1.32	9.64

Notes: High Inflation probabilities from Fleckenstein et al. (2017) are based on daily observations of inflation swaps and options from October 5, 2009 to October 28, 2015.

Figure 1: Refining Sign Restrictions with the Forecast Consistent Prior



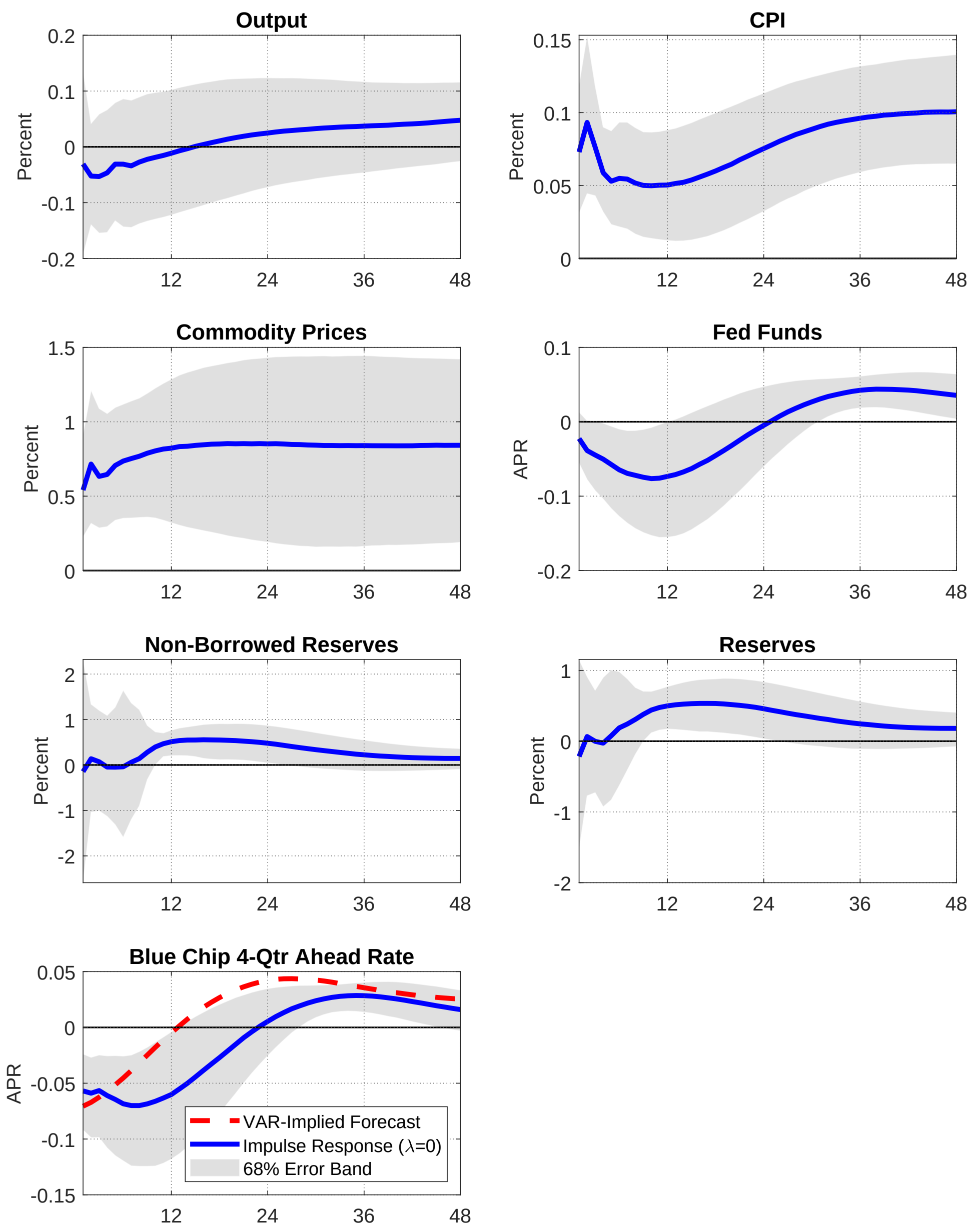
(a) Identified Set: Normalization and Sign Restriction



(b) Identified Set: Normalization, Sign Restriction, and Forecast Consistency

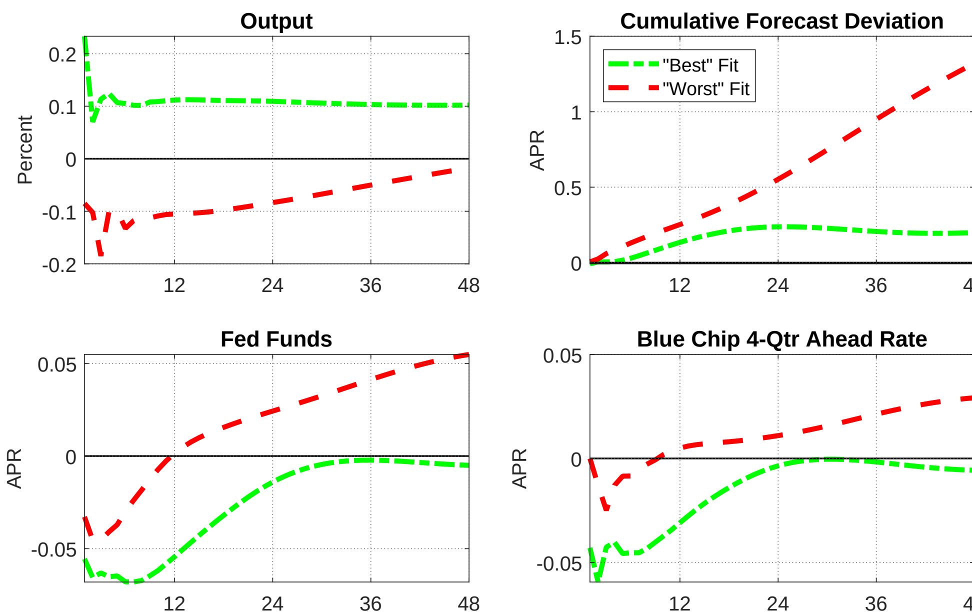
Notes: This figure illustrates how the forecast consistent prior, represented by R^3 in panel b, can shape the identified set from a sign-restricted structural VAR model. Darker shading around the black dot on the unit circle in panel b corresponds to higher values of the forecast consistent prior density.

Figure 2: Forward Guidance Shock: Sign Restrictions Only



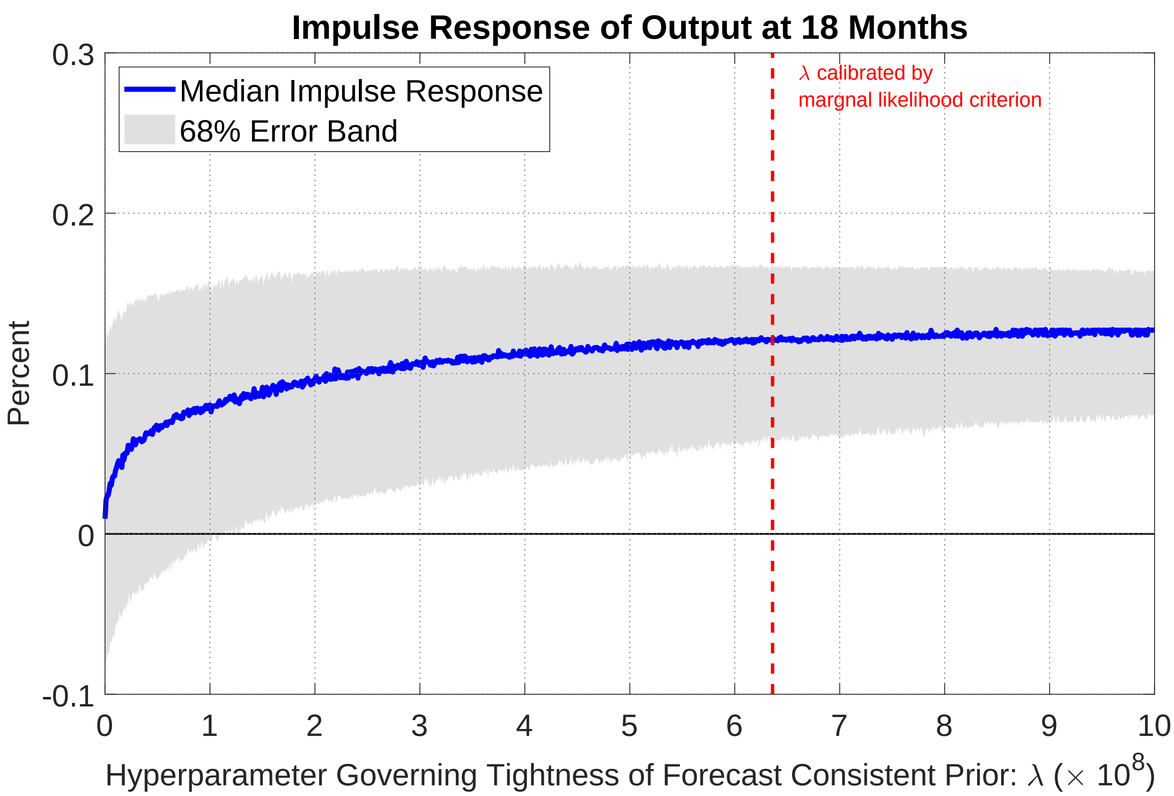
Notes: This figure shows the impulse responses to an identified forward guidance shock using only sign restrictions. The solid blue line is the median response and the shaded region is the 68% interval among structural VAR models. The red-dashed line shows the VAR-implied response of future short-term interest rates. The estimation sample period is 1994-2007.

Figure 3: Forward Guidance Shocks: "Best" and "Worst" Fitting Draws



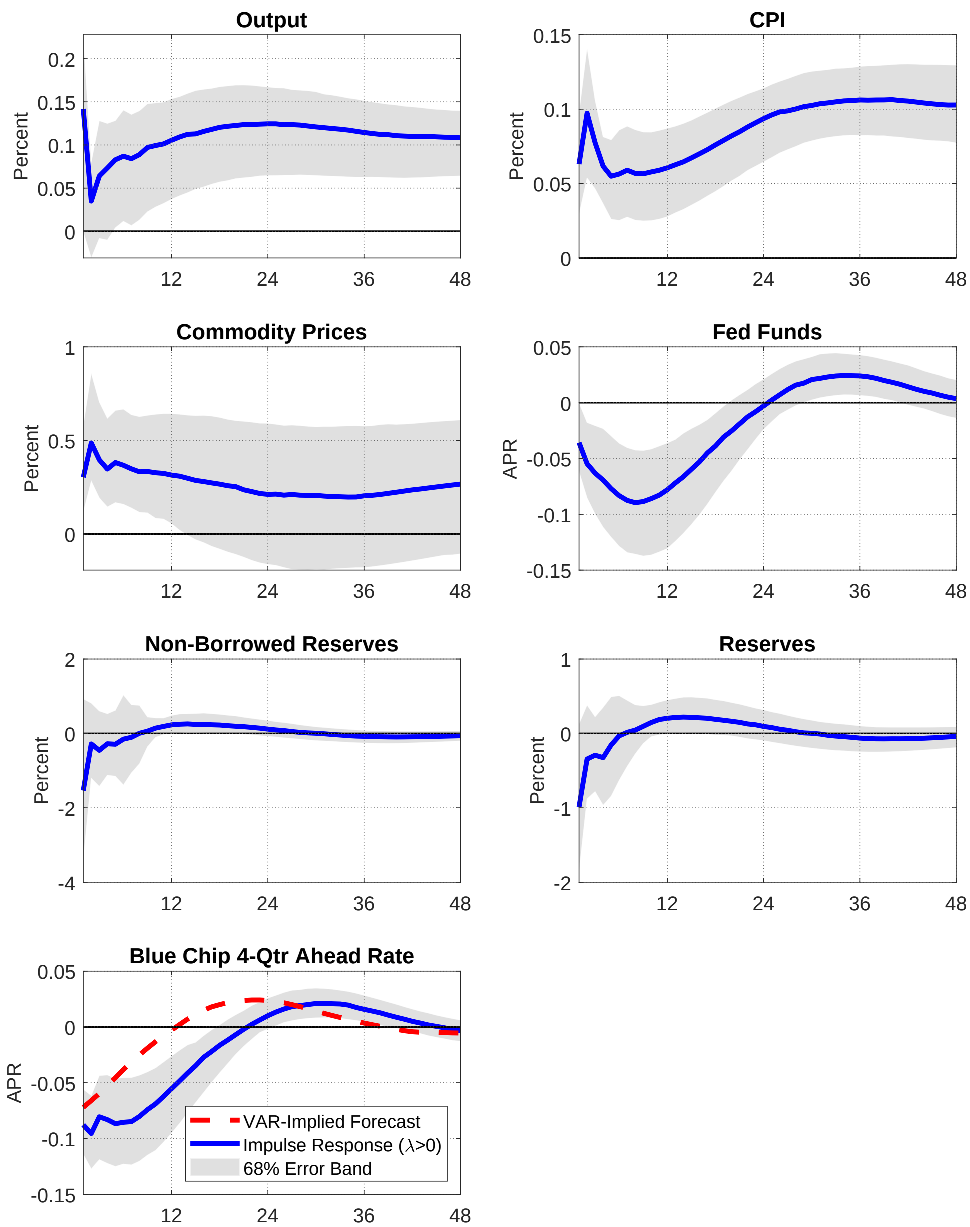
Notes: This figure shows the impulse responses to an identified forward guidance shock for the "best" and "worst" fitting models. The "best" fitting model is the draw that comes the closest to satisfying the forecast consistency restrictions and the "worst" fitting model is the draw that is the furthest from satisfying the forecast consistency restrictions. The estimation sample period is 1994-2007.

Figure 4: Output Effects of Forward Guidance: The Role of the Forecast Consistent Prior



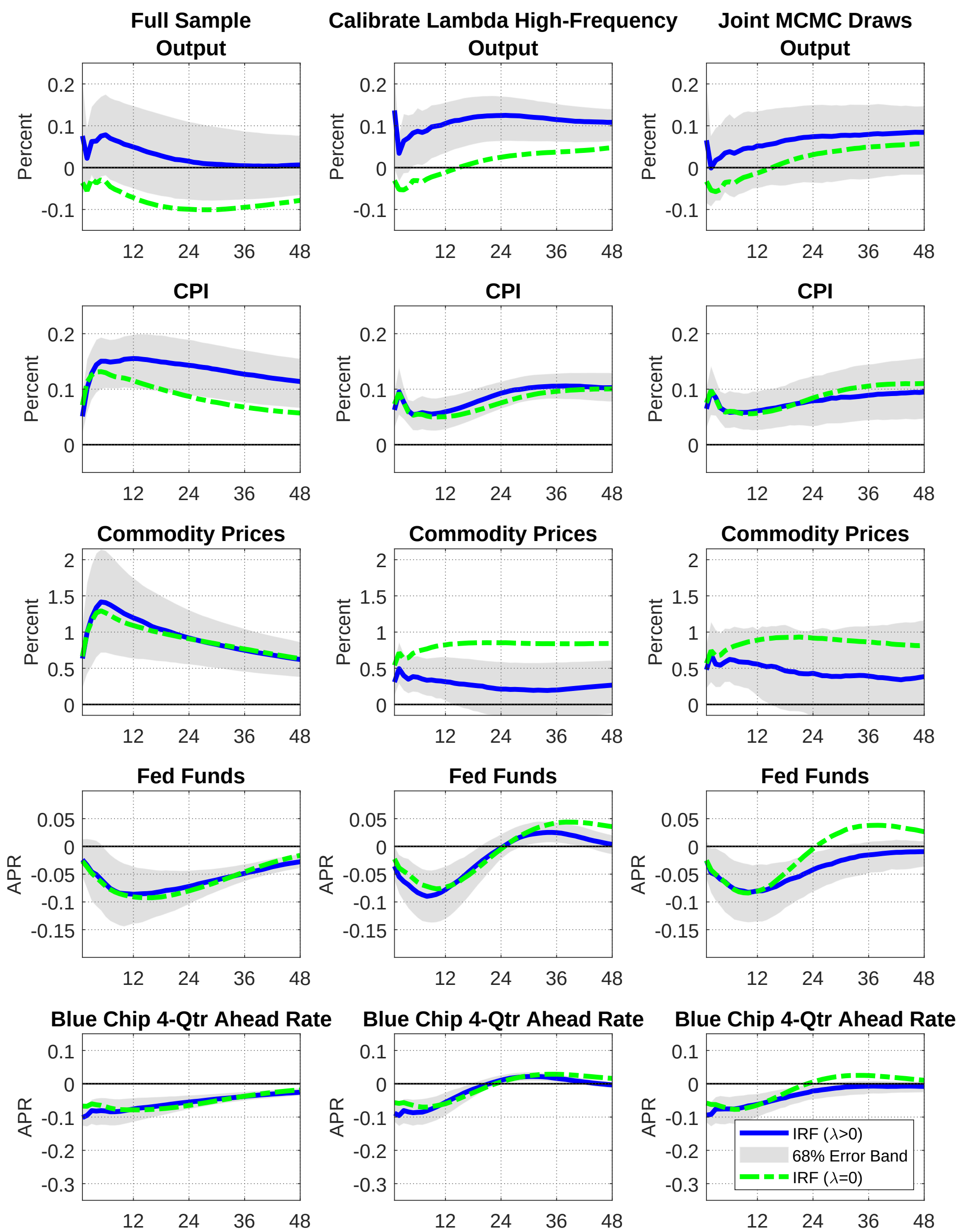
Notes: This figure shows the median output response and the corresponding 68% intervals after 18 months for alternative values of λ, which governs the tightness of the forecast consistent prior. The vertical red line denotes the value of λ selected to maximize the marginal likelihood criterion. The estimation sample period is 1994-2007.

Figure 5: Forward Guidance Shock: Sign Restrictions and the Forecast Consistent Prior



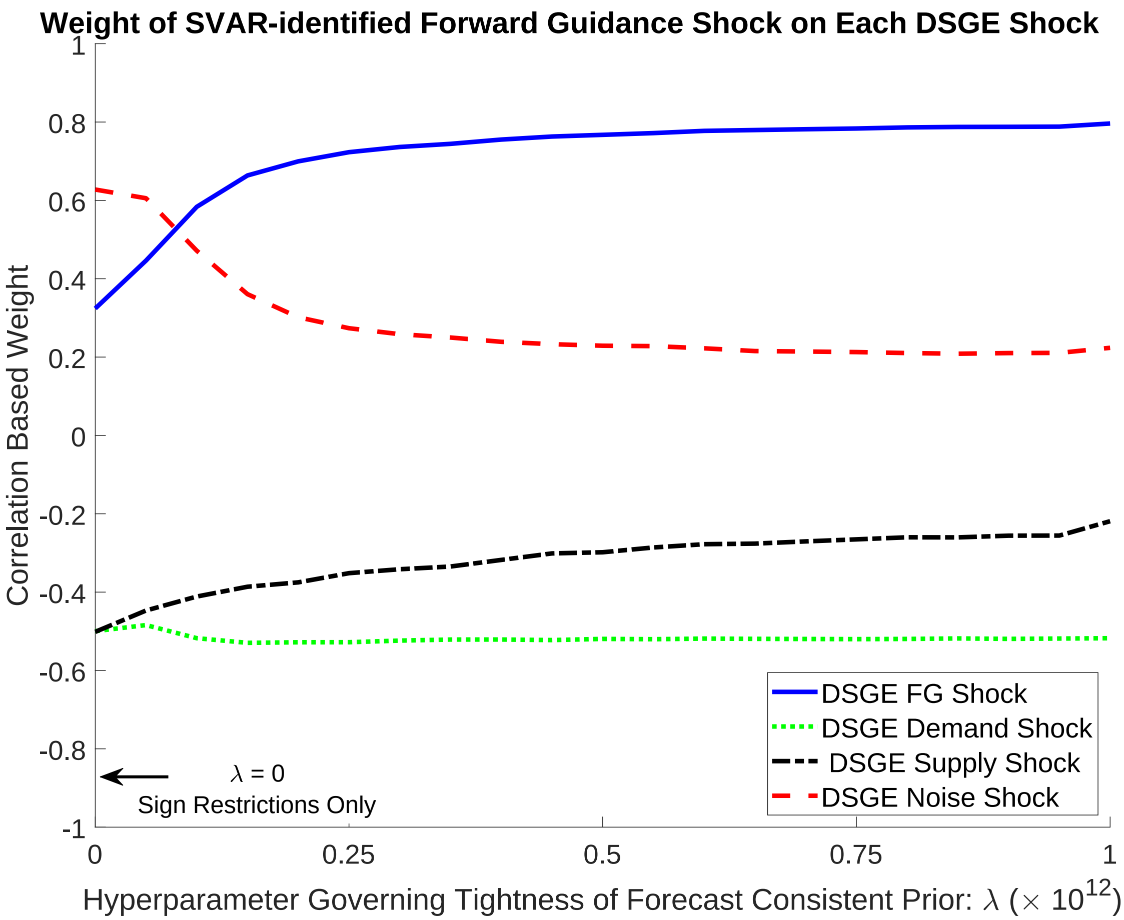
Notes: This figure shows the impulse responses to an identified forward guidance shock using sign restrictions as well as our forecast consistent prior. The solid blue line is the median response and the shaded region is the 68% error band. The red line shows the VAR-implied response of future short-term interest rates. The estimation sample period is 1994-2007.

Figure 6: Forward Guidance Shock: Alternative Specifications



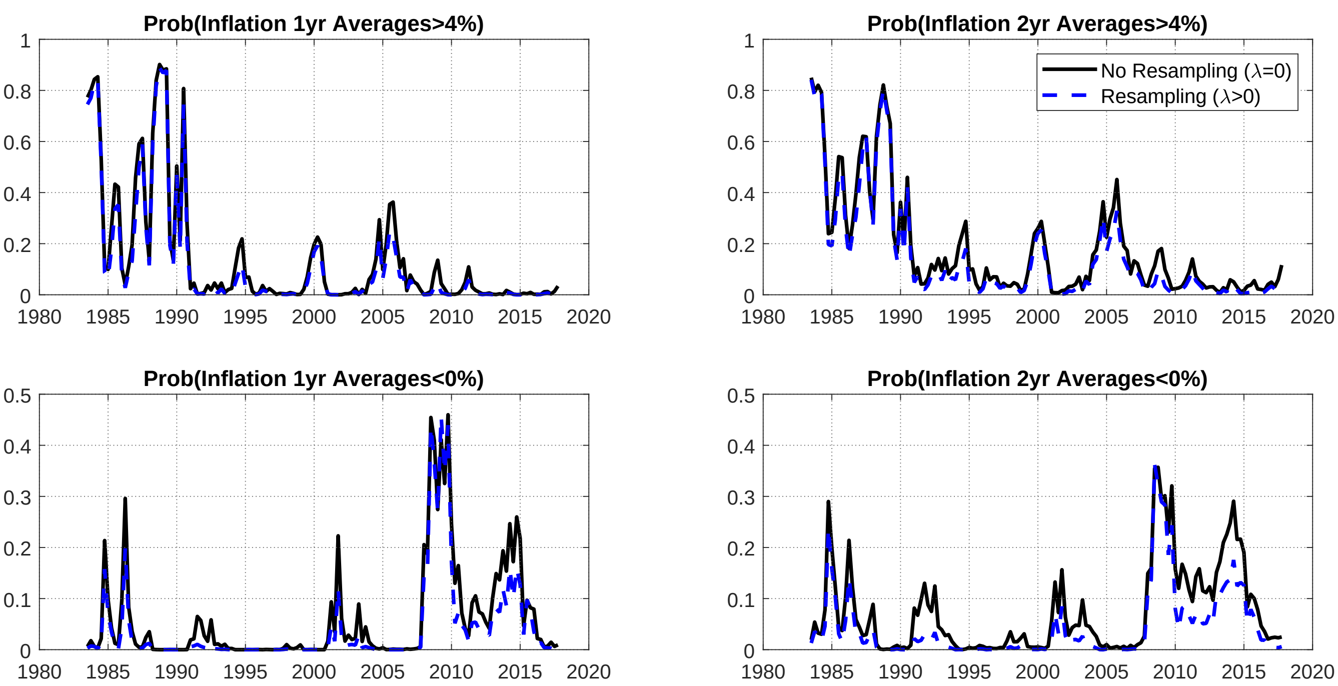
Notes: This figure shows the impulse responses to an identified forward guidance shock using sign restrictions as well as our forecast consistent prior. The solid blue line is the median response and the shaded region is the 68% error band. The green-dashed line shows the median impulse response to an identified forward guidance shock using only sign restrictions. Each column shows impulse responses from an alternative VAR model. For details, see Section 3.5.

Figure 7: Monte Carlo Simulation Results: Recovering Forward Guidance Shocks with Sign Restrictions and the Forecast Consistent Prior



Notes: This figure shows the correlation-based weights that forward guidance shocks identified from SVAR models place on various structural shocks from a New-Keynesian DSGE model with noisy survey forecasts which serves as the data-generating process. These weights are shown for alternative values of λ, which governs the tightness of the forecast consistent prior. When $\lambda = 0$, only sign restrictions are used to identify forward guidance shocks. The VAR is a 5-lag VAR estimated on one sample consisting of 50,000 DSGE model-generated observations with the number of selected lags based on the AIC.

Figure 8: INFLATION TAIL RISKS



Notes: This figure shows the time-varying probabilities of high inflation (inflation $>$ 4 percent) in the top row and the time-varying probability of deflation (inflation $<$ 0 percent) from our TVP-VAR model. The left column shows these probabilities over the 1 year horizon and the right column shows these probabilities over the 2 year horizon. The black solid lines show these probabilities from the TVP-VAR without the forecast consistent prior, labeled "No Resampling ($\lambda = 0$)" and the blue dashed lines show these probabilities from the TVP-VAR with the forecast consistent prior, labeled "Resampling ($\lambda > 0$)", which corresponds to $\lambda = 1.42$ as calibrated by the marginal likelihood criterion.