SUPPLEMENTARY APPENDIX

A Time Series Model of Interest Rates With the Effective Lower Bound*

Benjamin K. Johannsen†
Federal Reserve Board

Elmar Mertens
Bank for International Settlements

April 16, 2018

Abstract

This appendix contains supplementary results as well as further descriptions of computational procedures for our paper. Section I, describes the MCMC sampler used in estimating our model. Section II describes the computation of predictive densities. Section III reports additional estimates of trends and stochastic volatilities well as posterior moments of parameter estimates from our baseline model. Section IV reports estimates from an alternative version of our model, where the CBO unemployment rate gap is used as business cycle measure instead of the CBO output gap. Section V reports trend estimates derived from different variable orderings in the gap VAR of our model. Section VI compares the forecasting performance of our model to the performance of the no-change forecast from the random-walk model over a period that begins in 1985 and ends in 2017:Q2. Sections VII and VIII describe the particle filtering methods used for the computation of marginal data densities as well as the impulse responses.

*The views in this paper do not necessarily represent the views of the Bank for International Settlements, the Federal Reserve Board, any other person in the Federal Reserve System or the Federal Open Market Committee. Any errors or omissions should be regarded as solely those of the authors.

†For correspondence: Benjamin K. Johannsen, Board of Governors of the Federal Reserve System, Washington D.C. 20551. email benjamin.k.johannsen@frb.gov. Tel.: +(202) 530 6221.

Contents

List of Tables

List of Figures

I Priors and Posterior Sampling

MCMC estimates of the model are obtained from a Gibbs sampler. The sampler is run multiple times with different starting values and convergence is assessed with the scale reduction test of Gelman et al. (2003).[1] For each run, $10,000$ draws are stored after a burn-in period of $100,000$ draws; the post-burnin draws from each run are then merged.

Our models features two layers of latent states as well as various parameters; the first layer of latent states is given by:

$$\boldsymbol{\xi}_t = \begin{bmatrix} \bar{\boldsymbol{X}}_t' & \tilde{\boldsymbol{X}}_t' & \tilde{\boldsymbol{X}}_{t-1}' & \ldots & \tilde{\boldsymbol{X}}_{t-p+1}' \end{bmatrix}' \tag{A.1}$$

$$\bar{\boldsymbol{X}}_t = \begin{bmatrix} \bar{\pi}_t & \bar{r}_t & \bar{r}_t^2 & \bar{r}_t^5 & \bar{r}_t^{10} \end{bmatrix}' \tag{A.2}$$

$$\tilde{\boldsymbol{X}}_t = \begin{bmatrix} \tilde{\pi}_t & \tilde{c}_t & \tilde{s}_t & \tilde{y}_t^2 & \tilde{y}_t^5 & \tilde{y}_t^{10} \end{bmatrix}' \tag{A.3}$$

with $r_t^i = r_t + p^i \; \forall \, i = 2, 5, 10$, and where p is the lag length of the gap VAR in equation (8).

The second layer of latent variables consists of the vector of stochastic log-variances of the gap shocks, $\tilde{\boldsymbol{h}}_t \equiv \log(\boldsymbol{\sigma}_t^2)$ in equation (9) together with the stochastic log-variance process of shocks to the inflation trend, $\bar{h}_t \equiv \log(\sigma_{\bar{\pi},t}^2)$ in equation (4).[2] All told, our model consists of equations (1), (2), (3), (4), (5), (6), (7), (8) and (9); out of these, equations (2), (3), (5), (6), (7), (8) are represented as a conditionally linear state-space given by equations (13), (12), as well as the truncated measurement equation (11) for the nominal policy rate.[3]

In our baseline case — with constant-variance shocks in the trend real rate — details of the state space system given by equations (13) and (12), which are also reproduced below, are given

[1] Specifically, for every model, 4 independent runs for the Gibbs sampler were evaluated; each run initialized with different starting values drawn from the model's prior distribution. Convergence is deemed satisfactory when the scale reduction statistics for every parameter and latent variable are below 1.2; (values close to 1 indicate good convergence).

[2] Stochastic volatility in shocks to the real rate trend — if part of the model specification — can also be wrapped into $\bar{h}_t$.

[3] As noted already in the main text, in principle, a separate version of equation (1) applies for nominal interest rates of different maturities. However, in estimating our model, the distinction between longer-term yields and their respective shadow rates would be moot because the ELB never binds for the other yields in our data. Nevertheless, when simulating predictive densities for longer-term yields, the truncation implied by (1) is, of course, applied.

as follows:

$$\boldsymbol{\xi}_t = \boldsymbol{\mathcal{A}}\boldsymbol{\xi}_{t-1} + \boldsymbol{\mathcal{B}}\,\boldsymbol{\Sigma}_t^{1/2}\,\boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{I}) \tag{13}$$

$$\boldsymbol{X}_t = \boldsymbol{\mathcal{C}}\,\boldsymbol{\xi}_t \tag{12}$$

$$\boldsymbol{\mathcal{C}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \mathbf{0} \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \mathbf{0} \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \mathbf{0} \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{0} \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & \mathbf{0} \end{bmatrix} \tag{A.4}$$

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \end{bmatrix} \tag{A.5}$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are the lag coefficients matrices of the gap VAR in equation (8) and boldface symbols denote (sub-)matrices. Furthermore, we have

$$\boldsymbol{\mathcal{B}}_t = \boldsymbol{\mathcal{B}}\,\boldsymbol{\Sigma}_t^{1/2} \tag{A.6}$$

$$\boldsymbol{\mathcal{B}} = \begin{bmatrix} 1 & 0 & \mathbf{0} \\ 0 & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{A.7}$$

where $\boldsymbol{B}$ is a unit-lower-triangular matrix, and the stochastic volatilities — as well as the constant-variance parameter $\sigma_{\bar{r}}$ — are stacked into $\boldsymbol{\Sigma}_t^{1/2}$ as follows:

$$\boldsymbol{\Sigma}_t^{1/2} = \begin{bmatrix} \sigma_{\bar{\pi},t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\bar{r}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\tilde{\pi},t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\tilde{c},t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\tilde{s},t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\tilde{y}^2,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\tilde{y}^5,t} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\tilde{y}^{10},t} \end{bmatrix}. \tag{A.8}$$

The parameter vector of the model comprises the means, persistence and variance-covariances of shocks to $\bar{h}_t$ and $\tilde{\boldsymbol{h}}_t$, see equations (4) and (9). as well as the variance of shocks to the trend real rate, denoted $\sigma^2_{\bar{r}}$, the transition coefficients of the gap VAR in equation (8), stacked in a vector $\boldsymbol{a}$, and the lower diagonal elements of gap shock loadings $\boldsymbol{B}$ in equation (8) that can be stacked in a vector denoted $\boldsymbol{b}$. For ease of reference, all parameters are collected in the vector $\boldsymbol{\theta}$. Furthermore, since MCMC estimates of $\bar{h}^T$ and $\tilde{\boldsymbol{h}}^T$ will be obtained from the multi-move filter of Kim et al. (1998), the use of a set of discrete indicator variables, $\boldsymbol{s^T}$, is required to to approximate $\log \boldsymbol{\eta}^2_{\bar{\pi},t}$ and $\log \tilde{\boldsymbol{\eta}}^2_t$ in equations (4) and (9), respectively, with a mixture of normals. For ease of exposition, we stack the log of the stochastic variances of trend and gap shocks into the vector $\boldsymbol{h}_t$. Combining equations (4) and (9) yields the following vector system:

$$\boldsymbol{h}_t = (\boldsymbol{I} - \boldsymbol{\rho})\,\boldsymbol{\mu} + \boldsymbol{\rho}\boldsymbol{h}_{t-1} + \boldsymbol{\Phi}\boldsymbol{\eta}_t \qquad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{I}) \tag{A.9}$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_{\bar{\pi}} & \tilde{\boldsymbol{\mu}}' \end{bmatrix}$, $\boldsymbol{\rho}$ is a diagonal matrix with $\rho_{\bar{\pi}}$ and the diagonal values of $\tilde{\boldsymbol{\rho}}$ on its diagonal and $\boldsymbol{\Phi}$ is a block-diagonal matrix consisting of $\phi^2_{\bar{\pi}}$ and $\tilde{\boldsymbol{\Phi}}$.[4]

Conditional on draws for the various parameters, ($\boldsymbol{\theta}$), and log-volatilities $\boldsymbol{h}^T$, we can construct

[4] Since trend stochastic volatilities in equation (4) are independent of the gap stochastic volatilities in equation (9) — $\boldsymbol{\rho}$ is diagonal and $\boldsymbol{\Phi}$ is block-diagonal — both stochastic volatility blocks can be estimated in separate Gibbs steps.

matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$. and $\{\boldsymbol{\Sigma}_t^{1/2}\}_{t=1}^T$ and obtain the linear, Gaussian state space system described by equation (13) in the main paper.

For the initial values of the latent states, the following priors were used:

$$\boldsymbol{\xi_0} \sim N\left(E(\boldsymbol{\xi}_0), \boldsymbol{\Omega}\right) \qquad \text{with} \quad E(\boldsymbol{\xi}_0) = \begin{bmatrix} \bar{\boldsymbol{\xi}} \\ \mathbf{0} \end{bmatrix} \qquad \text{and} \quad \boldsymbol{\Omega} = \begin{bmatrix} \bar{\boldsymbol{\Omega}} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Omega}} \end{bmatrix} \tag{A.10}$$

An uninformative prior for the initial gap levels is obtained by setting $\tilde{\boldsymbol{\Omega}}$ equal to the ergodic variance-covariance matrix of the gaps implied by the VAR in equation (8), evaluated at the time zero draws for the stochastic volatilities, encoded in $\boldsymbol{\Sigma}_0$, for every MCMC draw.[5] The prior for the initial trend levels are set to be consistent with

$$\begin{bmatrix} \bar{\pi}_0 \\ \bar{r}_0 \\ \bar{r}_0^2 \\ \bar{r}_0^5 \\ \bar{r}_0^{10} \end{bmatrix} \sim N\left(\begin{bmatrix} 2.0 \\ 2.0 \\ 2.5 \\ 3.0 \\ 3.5 \end{bmatrix}, 100 \cdot \boldsymbol{I} \right) \tag{A.11}$$

which implies generally vague prior levels for the various trend components.

The prior for the average levels of the log-variances is normal, with a mean value of $\log(0.1)^2$ and variance corresponding to the ergodic distribution implied by draws for the shock variance and AR(1) lag coefficients associated with the corresponding log-variance process. For each AR(1) lag coefficient, the prior is $N(0.8, 0.2^2)$, as in Clark and Ravazzolo (2015). For the variance of shocks to the volatility of the inflation trend, the prior is a univariate inverse-Wishart distribution with 6 degrees of freedom and centered around a mean of 0.2^2, which coincides with the fixed coefficient-value of 0.2 used by Stock and Watson (2007) in their univariate model for inflation.[6]

[5] In the case of a VAR(1), the ergodic variance-covariance matrix solves $\tilde{\boldsymbol{\Omega}} = \boldsymbol{A}\tilde{\boldsymbol{\Omega}}\boldsymbol{A}' + \boldsymbol{B}\,\boldsymbol{\Sigma}_0\,\boldsymbol{B}'$ for given values of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\Sigma}_0$.

[6] The univariate inverse-Wishart distribution corresponds to an inverse-gamma distribution, whose shape parameter is typically expressed in units that correspond to half the degrees of freedom of the inverse-Wishart.

For the vector of shocks to the gap volatilities vector, $\boldsymbol{\eta}_t$ in equation (9), the prior is inverse Wishart, centered around a mean of $0.2^2 \cdot \boldsymbol{I}$ and $N + 11$ degrees of freedom where N is the number of gap variables ($N = 6$ in our baseline model).

The parameter governing the variability of real-rate trend shocks, $\sigma^2_{\bar{r}}$, has a univariate inverse-Wishart distribution with 3 degrees of freedom and is centered around a prior mean of 0.2^2. The prior mean for $\sigma^2_{\bar{r}}$ is thus similar to the estimated value for the corresponding parameter reported by Holston et al. (2017) in the context of their model; see the value of $\sigma_{r^*} = 0.194$ in their Table 1. With three degrees of freedom, our prior is only vaguely informative while also embedding some belief that trend shocks explain only a small share of variations in real rates. In addition, this prior also helps to avoid the pile-up problem — known, for example, from Stock and Watson (1998) and considered in the context of estimating $\sigma^2_{\bar{r}}$ also by Laubach and Williams (2003) as well as Clark and Kozicki (2005) and Holston et al. (2017) — by keeping posterior parameter draws from zero; see our estimates shown in Figure A.4 further below.

A Minnesota-style prior (centered around a mean of zero) is used for the VAR coefficients $\boldsymbol{a}$, with hyperparameters $\lambda_1 = 0.5$ (own lags) and $\lambda_2 = 0.2$ (cross lags). The prior $\boldsymbol{b}$ is multivariate normal, $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{I})$.

The Gibbs sampler is initialized with values drawn from the prior for $\boldsymbol{h}^T$, and $\boldsymbol{\theta}$ and then generates draws from the joint posterior distribution

$$p\left(\boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \sigma^2_{\bar{r}}, \boldsymbol{s}^T \mid \boldsymbol{Z^T}\right)$$

by iterating over draws from the following conditional distributions:[7]

1. Draw from $p\left(\boldsymbol{\xi}^T \mid \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \sigma^2_{\bar{r}}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right)$ with the disturbance smoothing sampler of Durbin and Koopman (2002) and rejection sampling for the shadow rate when the observed nominal short-term rate is at the ELB as described in Appendix A.

2. Draw from $p\left(\boldsymbol{a} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \sigma^2_{\bar{r}}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right) = p\left(\boldsymbol{a} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{b}\right)$, a normal conjugate

[7] For ease of notation, $\bar{h}^T$ and $\tilde{\boldsymbol{h}}^T$ are stacked into $\boldsymbol{h}^T$ unless when the distinction becomes material.

posterior for a VAR with known heteroscedasticity, with rejection sampling to ensure a stationary VAR (Cogley and Sargent, 2005; Clark, 2011)

3. Draw from $p\left(\boldsymbol{b} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \sigma^2_{\bar{r}}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right)$ via recursive Bayesian regressions with known heteroscedasticity to orthogonalize the gap shocks of the VAR in (8).

4. Draw from the univariate inverse-Wishart conjugate posteriors for $\sigma^2_{\bar{r}}$:

$$p\left(\sigma^2_{\bar{r}} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\Phi}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right) = p\left(\sigma^2_{\bar{r}} \mid \boldsymbol{\xi}^T\right)$$

5. Draw from the normal conjugate posterior for $\boldsymbol{\rho}$:

$$\begin{aligned} &p\left(\boldsymbol{\rho} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\mu}, \boldsymbol{\Phi}, , \sigma^2_{\bar{r}} \boldsymbol{s}^T, \boldsymbol{Z}^T\right) \\ &\qquad = p\left(\boldsymbol{\rho} \mid \boldsymbol{h}^T, \boldsymbol{\mu}, \boldsymbol{\Phi}\right) = p\left(\rho_{\bar{\pi}} \mid h^T_{\bar{\pi}}, \mu_{\bar{\pi}}, \phi^2_{\bar{\pi}}\right) \cdot p\left(\tilde{\boldsymbol{\rho}} \mid \tilde{\boldsymbol{h}}^T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Phi}}\right) \end{aligned}$$

Due to the assumed independence between the stochastic volatilities affecting shocks to trend inflation the residuals of the gap VAR, which are specified in equation (9), this step can be broken out into two separate Bayesian regression steps. The presence of correlated shocks in equation (9) necessitates a SUR regression to construct the posterior for $\tilde{\boldsymbol{\rho}}$. Rejection sampling is applied to ensure that all elements of $\boldsymbol{\rho}$ are inside the unit circle.

6. Draw from the inverse-Wishart conjugate posterior for $\boldsymbol{\Phi}$, which can again be broken out into two independent steps:

$$\begin{aligned} &p\left(\boldsymbol{\Phi} \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\mu}, \sigma^2_{\bar{r}}, \boldsymbol{\phi}^{\mathbf{2}}_{\bar{\boldsymbol{\pi}}}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right) \\ &\qquad = p\left(\boldsymbol{\Phi} \mid \boldsymbol{h}^T, \boldsymbol{\mu}, \boldsymbol{\rho}\right) = p\left(\phi^2_{\bar{\pi}} \mid h^T_{\bar{\pi}}, \mu_{\bar{\pi}}, \rho_{\bar{\pi}}, \phi^2_{\bar{\pi}}\right) \cdot p\left(\tilde{\boldsymbol{\Phi}} \mid \tilde{\boldsymbol{h}}^T, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\rho}}\right) \end{aligned}$$

7. Draw the mixture indicators $\boldsymbol{s}^T$ from $p\left(\boldsymbol{s}^T \mid \boldsymbol{\xi}^T, \boldsymbol{h}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \boldsymbol{Z}^T\right)$.

8. Draw from $p\left(\boldsymbol{h}^T, \boldsymbol{\mu} \mid s^T, \boldsymbol{\xi}^T, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\Phi}, \sigma^2_{\bar{r}}, \boldsymbol{s}^T, \boldsymbol{Z}^T\right)$ by applying the disturbance smoothing

sampler of Durbin and Koopman (2002) to a linear state space for $\left(\boldsymbol{h}^T, \boldsymbol{\mu}\right)$ as in Kim et al. (1998).[8]

Strictly speaking, this is not a simple Gibbs sampler consisting of steps 1 – 8 , but rather a Gibbs-within-Gibbs sampler with the outer Gibbs sampler iterating between

$$p\left(s^T, \boldsymbol{\xi}^T, \boldsymbol{\theta} \mid \boldsymbol{h}^T, \boldsymbol{\mu}, \boldsymbol{Z}^T\right) \qquad \text{(thus, a block consisting of steps 1 through 7)}$$
$$\text{and} \quad p\left(\boldsymbol{h}^T, \boldsymbol{\mu} \mid s^T, \boldsymbol{\xi}^T, \boldsymbol{\theta}, \boldsymbol{Z}^T\right) \qquad \text{(step 8),}$$

similar to the discussion in Del Negro and Primiceri (2015).

II Computation of Predictive Densities

In order to derive interest-rate forecasts that conform to the ELB (and other data in $\boldsymbol{Z}_t$), we first proceed by characterizing the predictive density for the shadow rate. Forecasts for actual rates can then be computed by integrating over the censored shadow-rate density. The shadow rate is included in the non-censored vector of variables $\boldsymbol{X}_t$ described in Appendix A. Apart from handling the truncation issues related to the ELB, our approach is fairly standard, building, for example, on the work by Geweke and Amisano (2010), Christoffel et al. (2010), and Warne et al. (2015). Given the truncation issues for interest rates and the fat tails introduced into the predictive density by the stochastic volatility specification, we have chosen to compute the predictive density based on the mixture of normals that is implied by the draws from our MCMC sampler, instead of approximating the predictive density solely based on its first two moments, treating the predictive density as a normal distribution, as has been done, for example, by Adolfson et al. (2007) in the case of linearized, constant-parameter DSGE models.

In order to compute the predictive density for $\boldsymbol{Z}_{t+h}$ jumping off data at time t, we first employ the MCMC sampler described in Appendix I to re-estimate all model parameters and latent vari-

[8] The constant $\boldsymbol{\mu}$ is embedded in the state space as a unit root without shocks, which improves the efficiency of the Gibbs sampler by jointly sampling $\boldsymbol{h}^T$ and $\boldsymbol{\mu}$.

ables ($\boldsymbol{\theta}$, $\boldsymbol{\xi}^t$ and $\boldsymbol{h}^t$) conditional on data available through time t. Draws from this MCMC sampler will henceforth be indexed by k.

Conditional on draws $(\boldsymbol{\xi}_t^k, \boldsymbol{h}_t^k, \boldsymbol{\theta}^k)$, it is straightforward to compute the predictive mean for uncensored variables:

$$E\left(\boldsymbol{X}_{t+h} \mid \boldsymbol{\xi}_t^k, \boldsymbol{h}_t^k, \boldsymbol{\theta}^k\right) = \boldsymbol{\mathcal{C}}^k \left(\boldsymbol{\mathcal{A}}^k\right)^h \boldsymbol{\xi}_t^k \tag{A.12}$$

and the predictive mean, conditional solely on data through t, can then be approximated by averaging over the means derived from each MCMC draw:

$$E\left(\boldsymbol{X}_{t+h} | \boldsymbol{Z}^t\right) \approx \sum_k E\left(\boldsymbol{X}_{t+h} \mid \boldsymbol{\xi}_t^k, \boldsymbol{h}_t^k, \boldsymbol{\theta}^k\right) \tag{A.13}$$

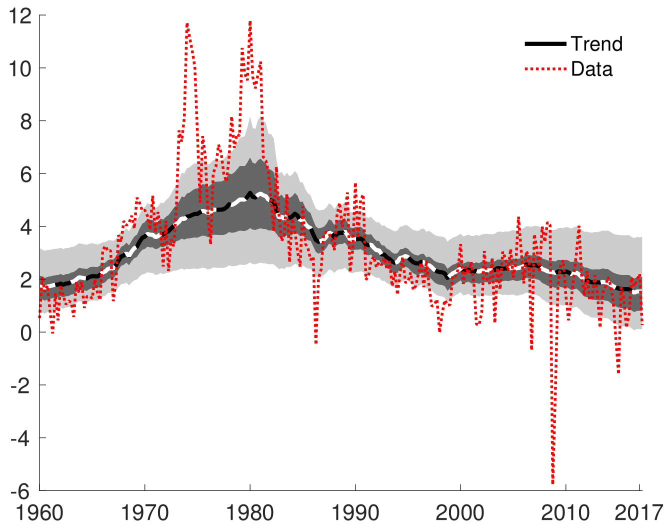
However, in order to characterize the entire predictive density for uncensored variables or even the predictive density for interest rates, which are subject to censoring due to the ELB constraint, we need to account for non-linearities in the distribution for future $\boldsymbol{\xi}_{t:t+h}$ arising from the stochastic volatility shocks in our model. We simulate $J = 100$ trajectories, each indexed by j, of $\boldsymbol{h}_{t:t+h}^{k,j}$ as well as shocks to $\boldsymbol{\xi}_{t:t+h}$ for each draw k from the MCMC sampler. From each draw $\boldsymbol{\xi}_{t:t+h}^{k,j}$ we construct $\boldsymbol{X}_{t+h}^{j,k}$; the ensemble of these draws across k and j approximates the predictive density for $\boldsymbol{X}_{t+h}$.

III Additional Estimates

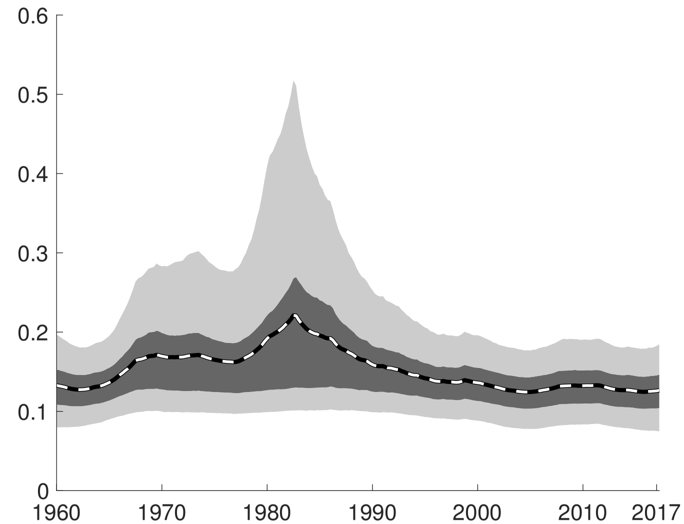
III.1 Latent Variables

This section reports estimates of additional latent variables, not shown in the main paper. Specifically, Figure A.1 displays estimates of the level of trend inflation as well as stochastic volatility affecting shocks to trend inflation. Figure A.2 reports the stochastic volatilities affecting residuals of the gap VAR in equation (8).

Figure A.1: Inflation Trend



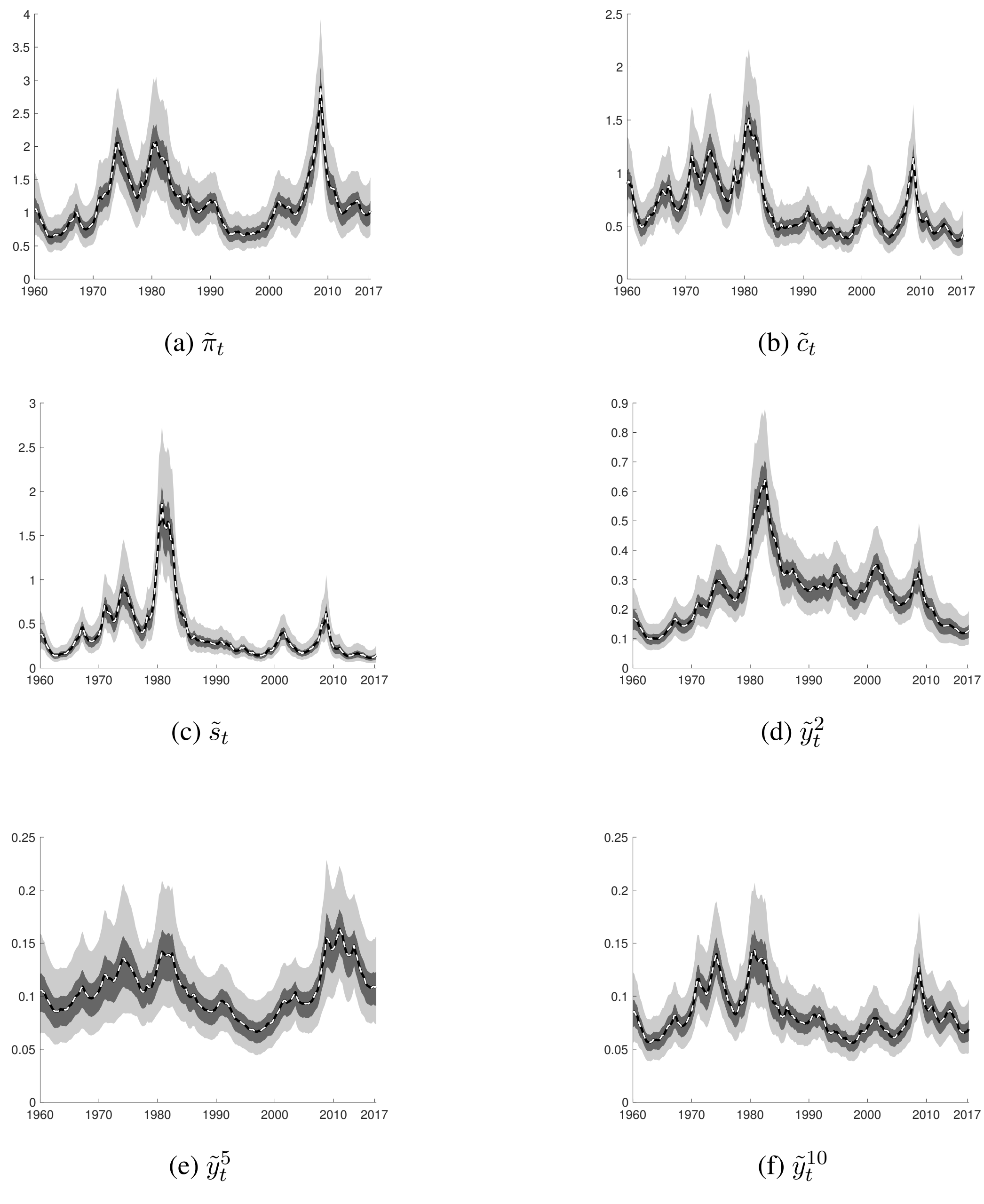
(a) Level of $\bar{\pi}_t$



(b) SV of shocks to $\bar{\pi}_t$

Note: Smoothed estimates using all available observations from 1960:Q1 through 2017:Q2. Thick dashed lines are posterior medians, shaded areas depict 90% and 50% uncertainty bands that reflect the joint uncertainty about model parameters and states.

Figure A.2: Stochastic Volatilities for Gap Variables



(a) $\tilde{\pi}_t$ (b) $\tilde{c}_t$

(c) $\tilde{s}_t$ (d) $\tilde{y}_t^2$

(e) $\tilde{y}_t^5$ (f) $\tilde{y}_t^{10}$

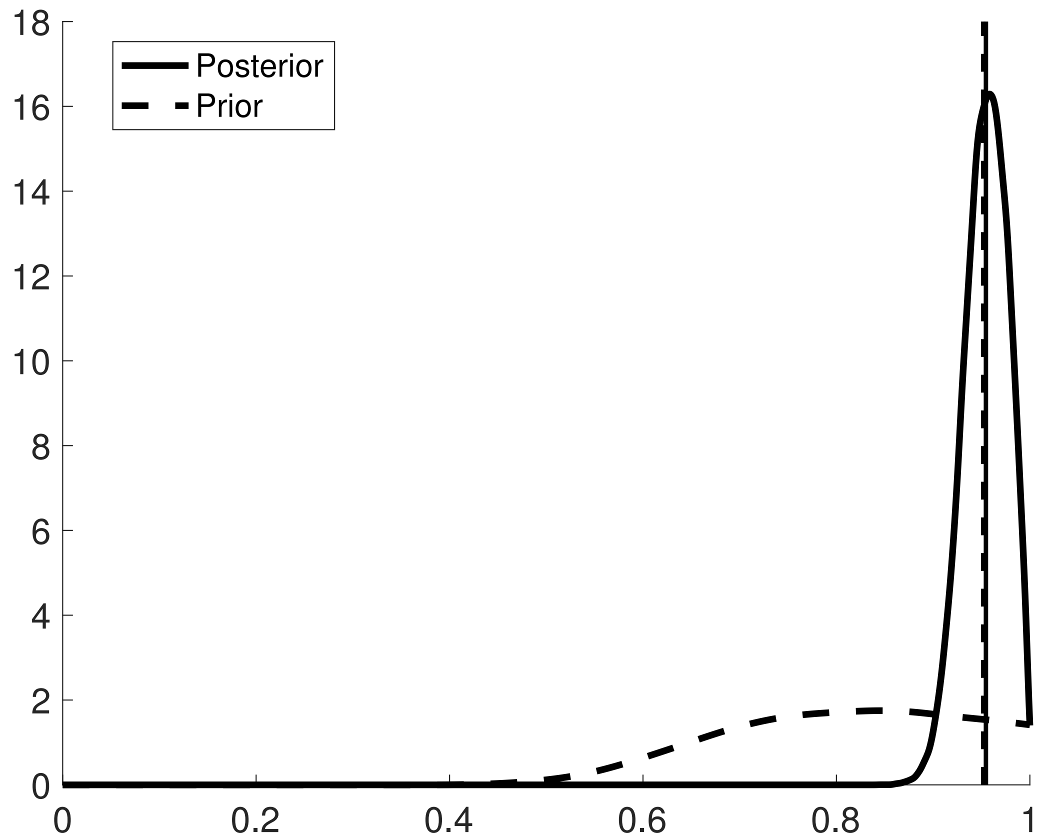
Note: Stochastic volatilities affecting the residuals in the gap VAR in equation (8). Specifically, these volatilities affect the Choleski residuals of the VAR and thus correspond to the diagonal elements of $\tilde{\boldsymbol{\Sigma}}_t^{1/2}$ in equation (8). Smoothed estimates using all available observations from 1960:Q1 through 2017:Q2. Thick dashed lines are posterior medians, shaded areas depict 90% and 50% uncertainty bands that reflect the joint uncertainty about model parameters and states.

III.2 Parameter Estimates

This appendix reports prior and posterior moments of various model parameters. Figure A.3 reports estimates of the persistence in the gap VAR (8) as measured by the largest, absolute eigenvalue of the associated companion-form transition matrix. Figure A.4 reports prior and posterior densities for $\sigma^2_{\bar{r}}$, the variance of shocks to the trend real rate. Figure A.5 depicts prior and posterior densities for average term premia between longer-term yields and the short-term shadow rate implied by the model estimates.

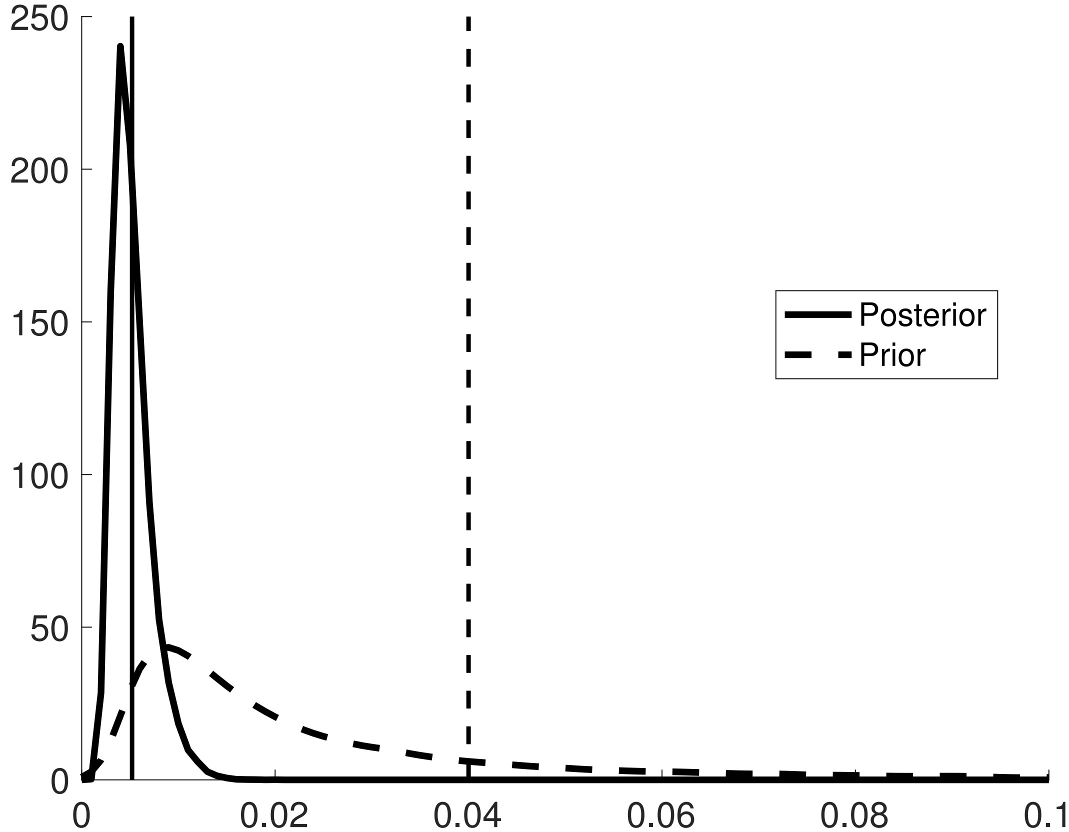
Tables A.1, A.2 and A.3 report prior and posterior moments of the various SV processes as well as the Choleski factorization of shocks to the gap VAR in equation (8).

Figure A.3: Persistence of Gap VAR



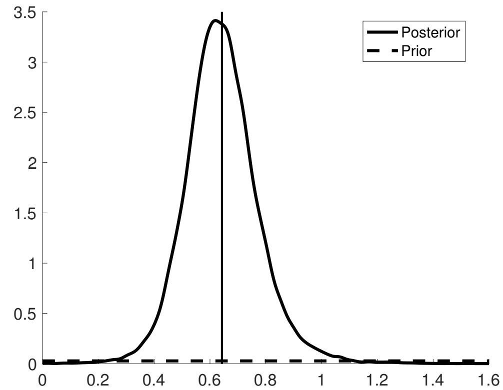
Note: Prior and posterior distribution of maximum, absolute eigenvalue of the companion-form transition matrix associated with the gap VAR in equation (8); prior and posterior means are depicted with horizontal lines. Full-sample estimates using all available observations from 1960:Q1 through 2017:Q2. A Minnesota-style prior (centered around a mean of zero) has been used for the VAR coefficients $\boldsymbol{a}$, with hyperparameters $\lambda_1 = 0.5$ (own lags) and $\lambda_2 = 0.2$ (cross lags).

Figure A.4: Variance of Shocks to the Trend Real Rate

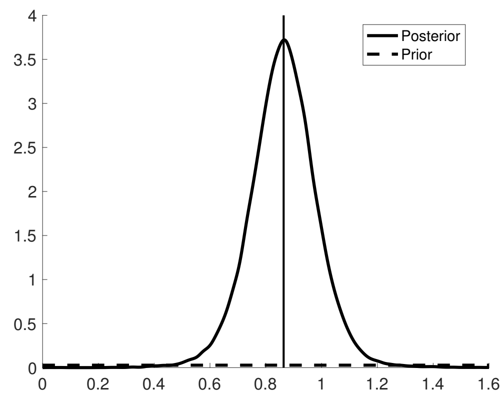


Note: Prior and posterior distribution of $\sigma_{\bar{r}}^2$, the variance of shocks to the trend real rate, see equation (6); prior and posterior means are depicted with horizontal lines. Full-sample estimates using all available observations from 1960:Q1 through 2017:Q2. The prior for $\sigma_{\bar{r}}^2$ is a univariate inverse Wishart distribution with mean of $0.2^2 = 0.04$ and three degrees of freedom. Our prior mean is thus similar to the estimated value for the corresponding parameter reported by Holston et al. (2017) in the context of their model; see the value of $\sigma_{r^*} = 0.194$ in their Table 1.

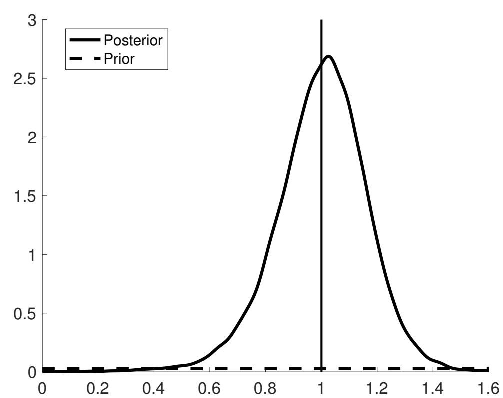
Figure A.5: Average Term Premia



(a) 2-year rate



(b) 5-year rate



(c) 10-year rate

Note: Reflecting the assumed cointegration between interest rates of different maturities, average term premia, $p^i = \bar{r}_t^i - \bar{r}_t$, are the constant offsets between real rate trends of longer-term yields, $\bar{r}_t^i$, relative to the trend of the short-term (shadow) real rate, $\bar{r}_t$. The prior, implied by equation (A.11), is essentially flat. Posterior means are depicted with horizontal lines. Full-sample estimates using all available observations from 1960:Q1 through 2017:Q2.

Table A.1: Shock Loadings $\boldsymbol{B}$ in Gap VAR

	Choleski Residuals of Shocks to …					
Variables	$\tilde{\pi}_t$	$\tilde{c}_t$	$\tilde{s}_t$	$\tilde{y}^2_y$	$\tilde{y}^5_t$	$\tilde{y}^{10}_y$
$\tilde{\pi}_t$	**1.0**					
$\tilde{c}_t$	**0.003**	**1.0**				
	[-0.060 — 0.065]					
$\tilde{s}_t$	**0.031**	**0.134**	**1.0**			
	[-0.010 — 0.069]	[0.066 — 0.202]				
$\tilde{y}^2_y$	**0.042**	**0.145**	**0.817**	**1.0**		
	[0.001 — 0.081]	[0.076 — 0.213]	[0.734 — 0.900]			
$\tilde{y}^5_t$	**0.055**	**0.102**	**0.573**	**1.063**	**1.0**	
	[0.018 — 0.090]	[0.042 — 0.162]	[0.479 — 0.665]	[1.008 — 1.118]		
$\tilde{y}^{10}_y$	**0.064**	**0.068**	**0.401**	**0.895**	**1.228**	**1.0**
	[0.030 — 0.097]	[0.013 — 0.123]	[0.306 — 0.492]	[0.813 — 0.976]	[1.111 — 1.344]	

Note: Posterior moments for coefficients of the unit-lower-triangular matrix $\boldsymbol{B}$ that maps shocks to the gap variables in the VAR of equation (8) to its Choleski residuals (scaled by stochastic volatility): $\boldsymbol{A}(L)\tilde{\boldsymbol{X}}_t = \boldsymbol{B}\tilde{\boldsymbol{\Sigma}}_t^{1/2}\tilde{\boldsymbol{\varepsilon}}_t$, where $\tilde{\boldsymbol{\Sigma}}_t^{1/2}$ is a diagonal matrix of stochastic volatilities. By definition, diagonal elements of $\boldsymbol{B}$ have unit value, and upper-diagonal elements are zero (indicated by blank entries above). Bold, lower-diagonal entries report posterior means; below the means, numbers in square brackets report 5% and 95% quantiles. Each coefficient has a standard normal prior distribution; this with mean zero, 5% (95%) quantile equal to -1.65 (1.65). Full-sample estimates using all available observations from 1960:Q1 through 2017:Q2. Estimated trajectories of the stochastic volatilities on the diagonal of $\tilde{\boldsymbol{\Sigma}}_t^{1/2}$ are depicted in Figure A.2.

Table A.2: Coefficients of AR(1)-SV Processes

PANEL A: Trend Inflation			
SV in ...	$\exp\left(\mu_{\bar{\pi}}/2\right)$	$\rho_{\bar{\pi}}$	$\phi_{\bar{\pi}}$
$\bar{\pi}_t$	**0.143**	**0.815**	**0.199**
	[0.096 — 0.201]	[0.499 — 0.992]	[0.140 — 0.279]
PANEL B: Gap Variables			
SV in ...	$\exp\left(\tilde{\mu}_i/2\right)$	$\tilde{\rho}_i$	$\sqrt{\tilde{\Phi}_{ii}}$
$\tilde{\pi}_t$	**1.038**	**0.932**	**0.291**
	[0.708 — 1.365]	[0.859 — 0.984]	[0.199 — 0.405]
$\tilde{c}_t$	**0.598**	**0.903**	**0.334**
	[0.469 — 0.735]	[0.829 — 0.960]	[0.229 — 0.455]
$\tilde{s}_t$	**0.272**	**0.943**	**0.452**
	[0.152 — 0.409]	[0.901 — 0.979]	[0.322 — 0.604]
$\tilde{y}_y^2$	**0.203**	**0.977**	**0.253**
	[0.063 — 0.361]	[0.946 — 0.997]	[0.183 — 0.339]
$\tilde{y}_t^5$	**0.102**	**0.929**	**0.214**
	[0.075 — 0.131]	[0.829 — 0.988]	[0.153 — 0.296]
$\tilde{y}_y^{10}$	**0.078**	**0.909**	**0.242**
	[0.062 — 0.095]	[0.804 — 0.975]	[0.172 — 0.330]

Note: Posterior moments of coefficients of the AR(1) stochastic volatility (SV) processes affecting different variables. Bold numbers are posterior means with 5% and 95% quantiles reported in square brackets underneath. All estimates reflect data on all available observations from 1960:Q1 through 2017:Q2; estimated trajectories of the stochastic volatilities on the diagonal of $\tilde{\boldsymbol{\Sigma}}_t^{1/2}$ are depicted in Figure A.2. Panel A reports the coefficients of the SV process affecting shocks to trend inflation, see equation (4):

$$\log\left(\sigma^2_{\bar{\pi},t}\right) = (1-\rho_{\bar{\pi}})\mu_{\bar{\pi}} + \rho_{\bar{\pi}}\log\left(\sigma^2_{\bar{\pi},t-1}\right) + \phi_{\bar{\pi}}\eta_{\bar{\pi},t} \qquad \eta_{\bar{\pi},t} \sim N(0,1) \tag{4}$$

Prior distributions are $\rho_{\bar{\pi}} \sim N(0.8, 0.2^2)$, $\phi_{\bar{\pi}}$ is univariate inverse Wishart with mean 0.2^2 and 12 degrees of freedom. The (conditional) prior distribution of $\mu_{\bar{\pi}}$ is normal with mean equal to $\log\left(0.1^2\right)$ and variance corresponding to the ergodic mean of the SV process implied by given values of $\rho_{\bar{\pi}}$ and $\phi_{\bar{\pi}}$.
Panel B reports coefficients of the SV processes affecting the orthogonalized shocks in the gap VAR of equation (8) given by the SUR system (9):

$$\log\left(\tilde{\boldsymbol{\sigma}}^2_t\right) = \left(\boldsymbol{I} - \tilde{\boldsymbol{\rho}}\right)\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\rho}}\log\left(\tilde{\boldsymbol{\sigma}}^2_{t-1}\right) + \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\eta}}_t \qquad \tilde{\boldsymbol{\eta}}_t \sim N(\boldsymbol{0}, \boldsymbol{I}) \tag{9}$$

In Panel B, $\tilde{\mu}_i$ and $\tilde{\rho}_i$ refer to the ith (diagonal) element of $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\rho}}$, respectively. $\tilde{\boldsymbol{\Phi}}$ is the variance-covariance matrix of shocks to the gap SV processes, and $\sqrt{\tilde{\Phi}_{ii}}$ is the volatility of shocks to the ith gap SV process. Correlation coefficients implied by the off-diagonal elements of $\tilde{\boldsymbol{\Phi}}$ are reported in Table A.3. Prior distributions are $\tilde{\rho}_i \sim N(0.8, 0.2^2)$, $\tilde{\mu}_i \sim N(\log\left(0.1^2\right), 5^2)$, and $\tilde{\boldsymbol{\Phi}}$ is inverse Wishart with mean $0.2^2 \cdot \boldsymbol{I}$ and $N_y + 11 = 17$ degrees of freedom.

Table A.3: Correlation between Gap SV Shocks

	$\tilde{\pi}_t$	$\tilde{c}_t$	$\tilde{s}_t$	$\tilde{y}_y^2$	$\tilde{y}_t^5$	$\tilde{y}_y^{10}$
$\tilde{\pi}_t$	**1.0**	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
		[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]
$\tilde{c}_t$	**0.549**	**1.0**	*0.000*	*0.000*	*0.000*	*0.000*
	[0.224 — 0.792]		[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]
$\tilde{s}_t$	**0.594**	**0.760**	**1.0**	*0.000*	*0.000*	*0.000*
	[0.264 — 0.829]	[0.558 — 0.894]		[*-0.457 — 0.457*]	[*-0.457 — 0.457*]	[*-0.457 — 0.457*]
$\tilde{y}_y^2$	**0.407**	**0.552**	**0.679**	**1.0**	*0.000*	*0.000*
	[0.015 — 0.709]	[0.244 — 0.778]	[0.432 — 0.852]		[*-0.457 — 0.457*]	[*-0.457 — 0.457*]
$\tilde{y}_t^5$	**0.326**	**0.332**	**0.315**	**0.204**	**1.0**	*0.000*
	[-0.184 — 0.701]	[-0.187 — 0.695]	[-0.238 — 0.706]	[-0.299 — 0.611]		[*-0.457 — 0.457*]
$\tilde{y}_y^{10}$	**0.520**	**0.571**	**0.619**	**0.473**	**0.362**	**1.0**
	[0.173 — 0.778]	[0.262 — 0.799]	[0.307 — 0.837]	[0.123 — 0.735]	[-0.083 — 0.691]	

Note: Prior and posterior moments of correlation coefficients implied by the variance-covariance matrix of shocks to the SUR system of SV processes in equation (9) affecting residuals in the gap VAR equation (8):

$$\log\left(\tilde{\boldsymbol{\sigma}}_t^2\right) = \left(\boldsymbol{I} - \tilde{\boldsymbol{\rho}}\right)\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\rho}}\log\left(\tilde{\boldsymbol{\sigma}}_{t-1}^2\right) + \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\eta}}_t \qquad \tilde{\boldsymbol{\eta}}_t \sim N(\boldsymbol{0}, \boldsymbol{I}) \tag{9}$$

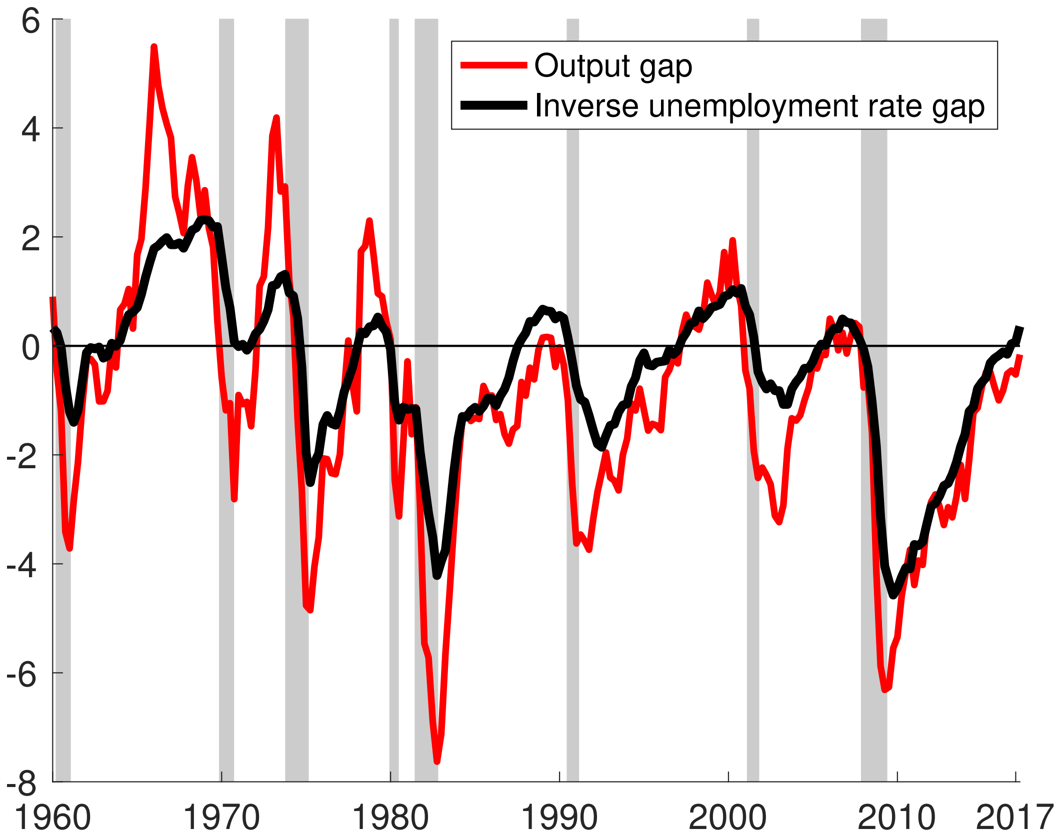
Upper-diagonal entires report prior moments, lower-diagonal entries posterior moments. For each correlation coefficients, the table reports the mean as well as 5% and 95% quantiles (in square brackets). The prior distribution of $\tilde{\boldsymbol{\Phi}}$ is inverse Wishart with mean $0.2^2 \cdot \boldsymbol{I}$ and $N_y + 11 = 17$ degrees of freedom.

IV Results using the Unemployment Rate Gap

This section reports estimates generated from an alternative specification where data for the CBO output gap is replaced by data measuring the CBO unemployment rate gap. As can be seen on Figure A.6, both measures convey a broadly similar picture of the business cycle in the U.S.; correspondingly, results from our model obtained using either measure are very similar. The unemployment rate gap is computed as the difference between the CBO's measure of the natural long-term rate of unemployment for a given quarter and the quarterly average rate of unemployment.[9] All other data series are identical to those used in the main paper.

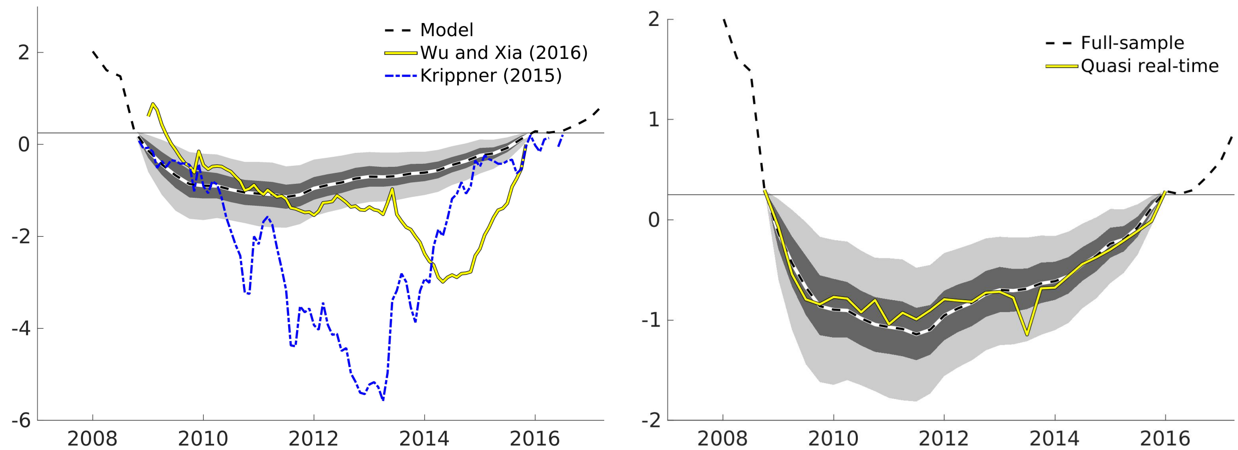
[9]Data for the unemployment rate and the CBO's estimate of the natural rate of unemployment in the long run are obtained from the FRED database, available at `https://fred.stlouisfed.org`, where they are labeled `UNRATE` and `NROU`, respectively.

Figure A.6: CBO Measures for Output Gap and Unemployment Rate Gap



Note: Shaded areas indicate NBER recessions. The unemployment rate gap is computed as the log-difference between the CBO's measure of the natural long-term rate of unemployment for a given quarter and the quarterly average rate of unemployment; shown in the figure is the *inverse* of the unemployment rate gap (corresponding to the log-difference between the natural and the actual rate of unemployment). The output gap is computed as the log difference between real GDP and the CBO's measure of potential real GDP for a given quarter. (In order to be comparable to annualized growth rates, for our computations the log difference between actual and potential GDP is been scaled by a factor of 400 when computing the output gap.) All computations are based on the vintage of FRED data available that has been available at the end of October 2017.

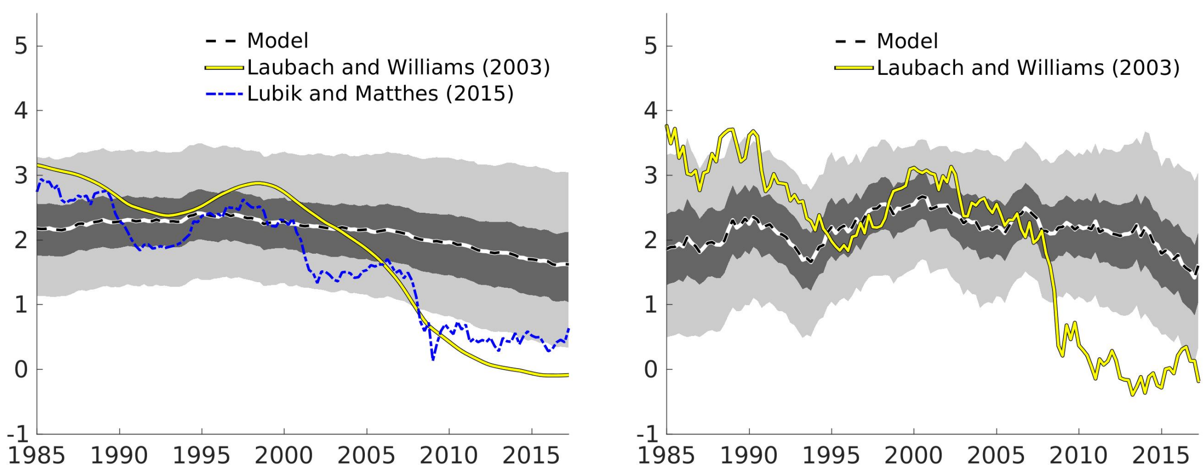
Figure A.7: Shadow Rate Estimates (w/Unemployment Rate Gap)



(a) Full-Sample Estimates and Other Estimates (b) Full-Sample and Quasi Real-Time Estimate

Note: Shaded areas indicate 50 and 90 percent uncertainty bands, dashed lines are posterior means. Results shown in Panel (a) reflect "smoothed" estimates using all available observations from 1960:Q1 through 2017:Q2. The red line in Panel (b) reflects mean of the endpoints of sequentially re-estimating the entire model over growing samples of quarterly observations starting in 1960:Q1, thus reflecting "filtered" estimates of the model's latent variables. Uncertainty bands reflect the joint uncertainty about model parameters and states.

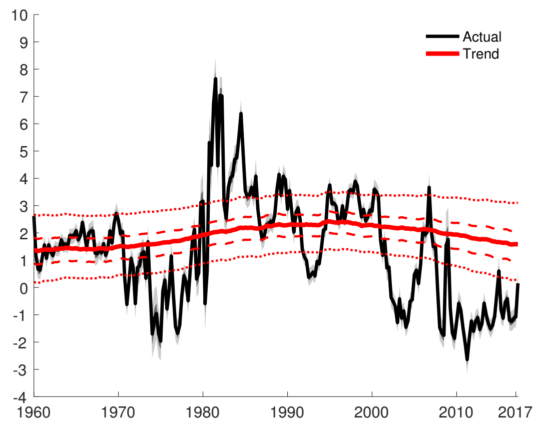
Figure A.8: The Real Rate in the Long Run (w/Unemployment Rate Gap)



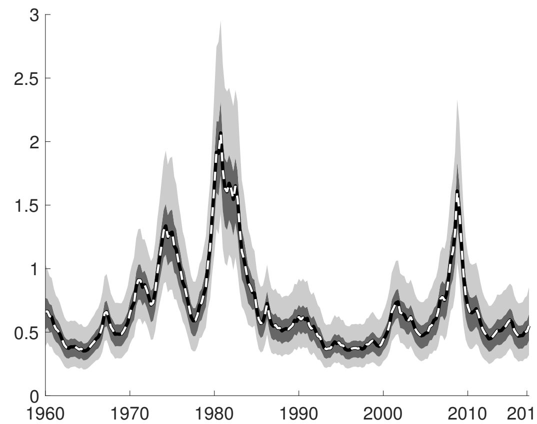
(a) Smoothed Estimates (b) Quasi Real-Time Estimate

Note: Shaded areas indicate 50 and 90 percent uncertainty bands, dashed lines are posterior means. Results shown in Panel (a) reflect "smoothed" estimates using all available observations from 1960:Q1 through 2017:Q2. Results shown in Panel (b) reflect the endpoints of sequentially re-estimating the entire model over growing samples of quarterly observations starting in 1960:Q1, thus reflecting "filtered" estimates of the model's latent variables. Uncertainty bands reflect the joint uncertainty about model parameters and states.

Figure A.9: Real-Rate Estimates: Trend Level and Gap Volatility (w/Unemployment Rate Gap)



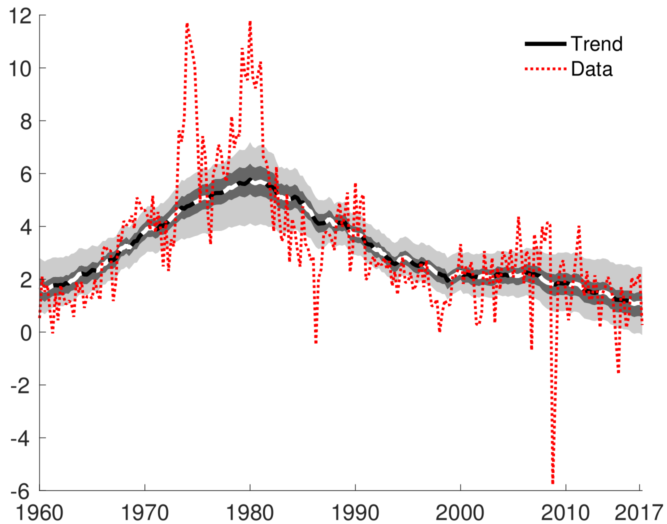
(a) $\bar{r}_t$



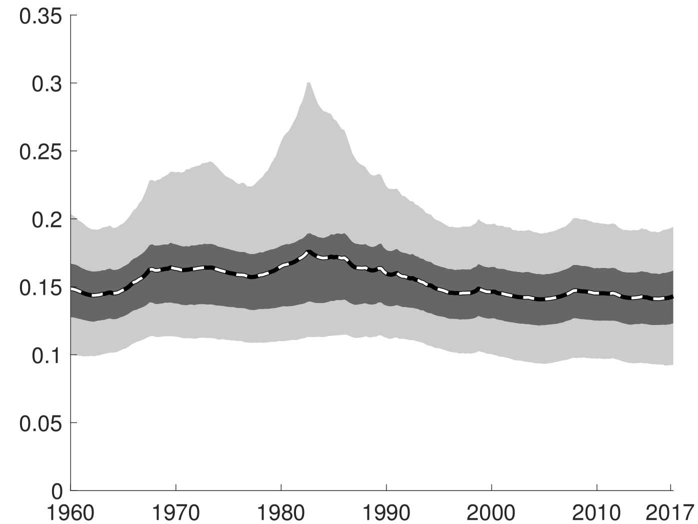
(b) $\text{Vol}_{t-1}(\tilde{r}_t)$

Note: Panel (a) depicts posterior means as well as 50% and 90% uncertainty bands for the trend real rate as well as for model estimates of the actual real rate, $r_t^* = i_t - E_t\pi_{t+1}$. Panel (b) reports estimates of the conditional volatility of shocks to the (shadow) real-rate gap, $\text{Vol}_{t-1}(\tilde{r}_t)$ where $\tilde{r}_t = \tilde{s}_t - E_t\tilde{\pi}_{t+1}$. Both panels reflect smoothed estimates computed using all available observations from 1960:Q1 through 2017:Q2.

Figure A.10: Inflation Trend (w/Unemployment Rate Gap)



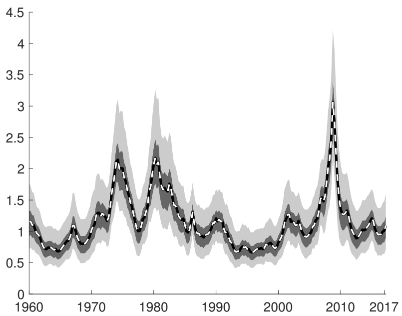
(a) Level of $\bar{\pi}_t$



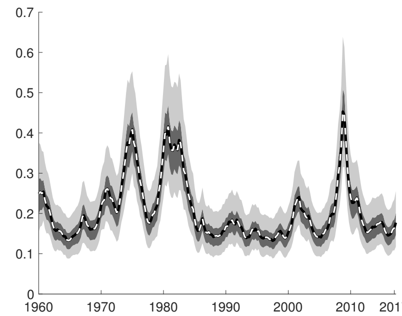
(b) SV of shocks to $\bar{\pi}_t$

Note: Smoothed estimates using all available observations from 1960:Q1 through 2017:Q2. Thick dashed lines are posterior means, shaded areas depict 90% and 50% uncertainty bands that reflect the joint uncertainty about model parameters and states.

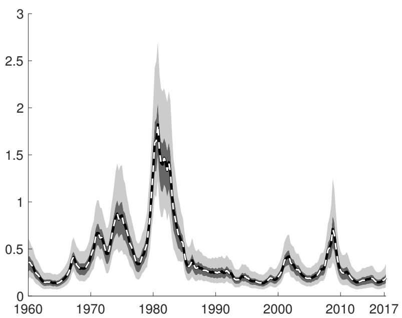
Figure A.11: Stochastic Volatilities for Gap Variables (w/Unemployment Rate Gap)



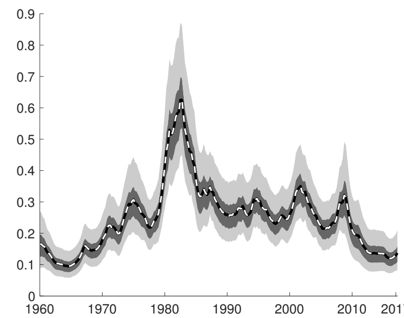
(a) $\tilde{\pi}_t$



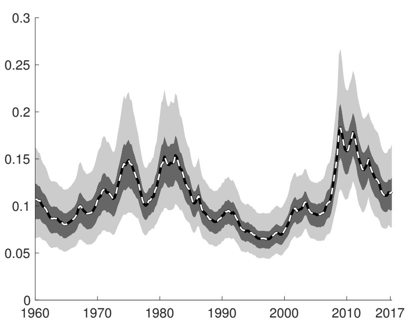
(b) $\tilde{c}_t$



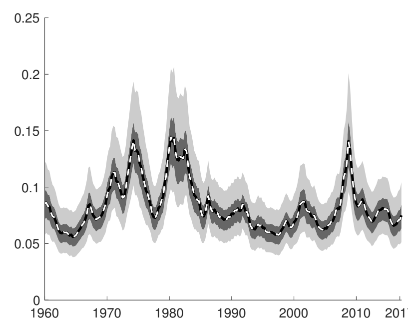
(c) $\tilde{s}_t$



(d) $\tilde{y}_t^2$



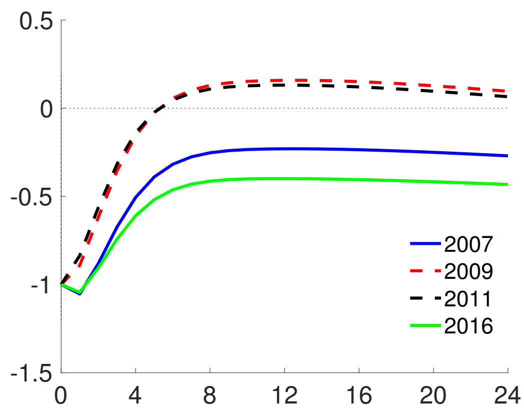
(e) $\tilde{y}_t^5$



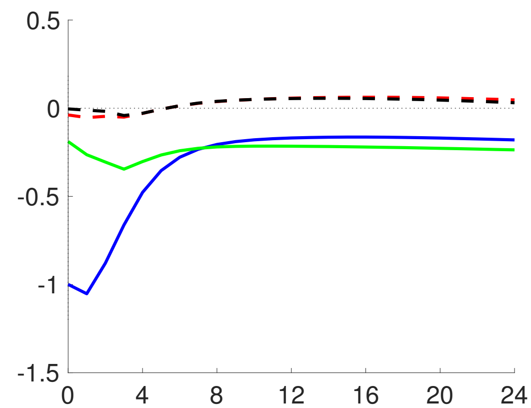
(f) $\tilde{y}_t^{10}$

Note: Stochastic volatilities affecting the residuals in the gap VAR (8). Specifically, these volatilities affect the Choleski residuals of the VAR and thus correspond to the diagonal elements of $\tilde{\mathbf{\Sigma}}_t^{1/2}$ in (8). Smoothed estimates using all available observations from 1960:Q1 through 2017:Q2. Thick dashed lines are posterior means, shaded areas depict 90% and 50% uncertainty bands that reflect the joint uncertainty about model parameters and states. The business cycle measure $\tilde{c}_t$ is given by the inverse unemployment rate gap.

Figure A.12: Short-Term Interest Rate Responses to Monetary Policy Shock (w/Unemployment Rate Gap)



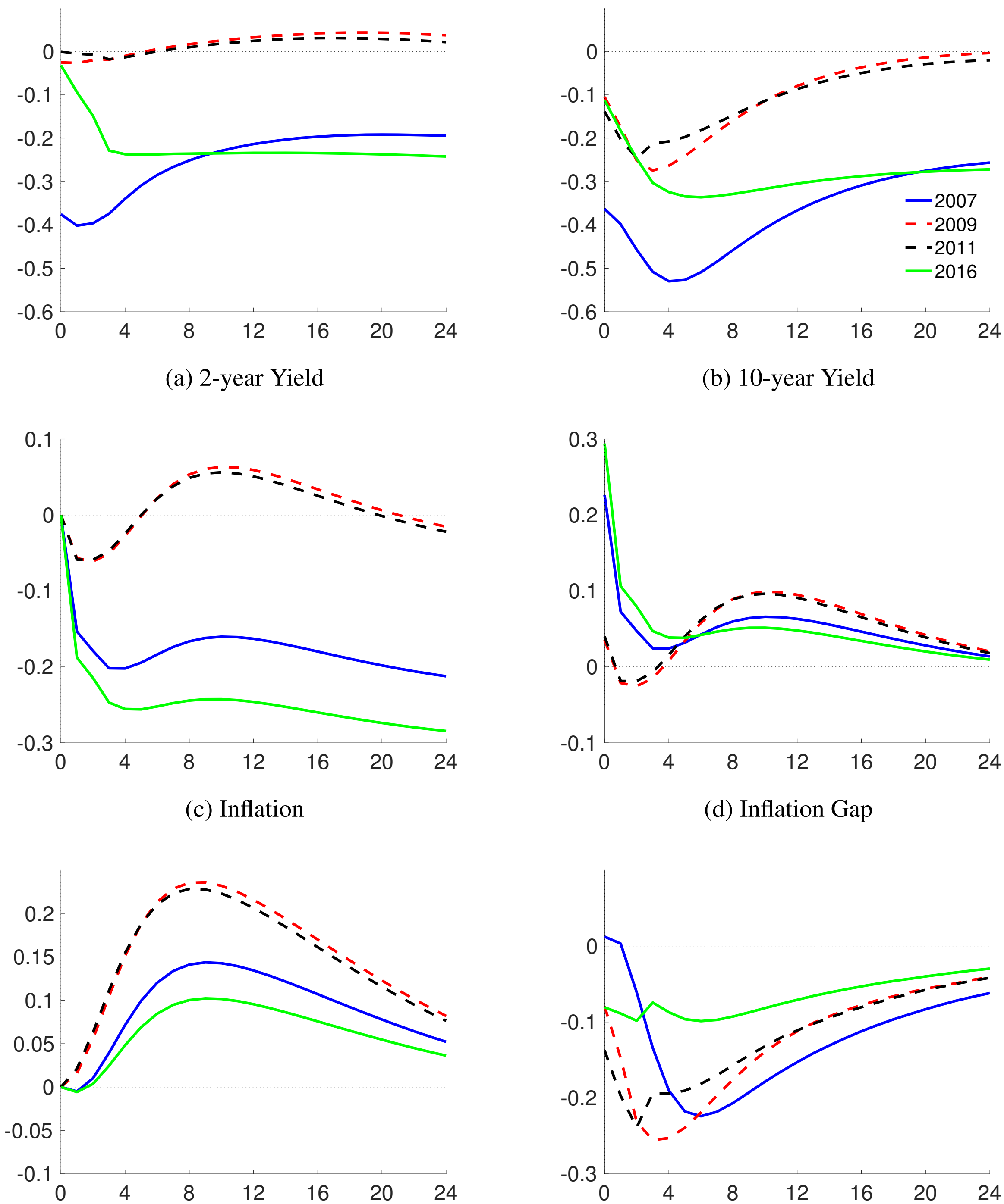
(a) Shadow Rate



(b) Policy Rate

Note: Responses to monetary policy shocks estimated for 2007:Q4, 2009:Q4, 2011:Q4 as well as 2016:Q4; dashed lines indicate responses at times when the ELB was binding for actual data. Shocks are scaled to generate a 1 percentage point drop in the shadow rate on impact. Vertical axis units are in percentage points. Horizontal axis units are quarters after impact of the monetary policy shock, which occurs at quarter zero.

Figure A.13: Responses to Monetary Policy Shock (w/Unemployment Rate Gap)



(a) 2-year Yield

(b) 10-year Yield

(c) Inflation

(d) Inflation Gap

(e) Inverse Unemployment Rate Gap

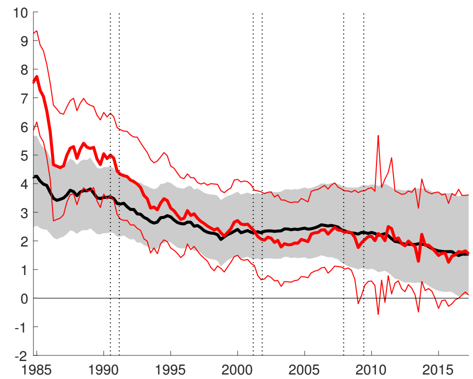
(f) Spread: 10- over 2-year Yield

Note: Responses to monetary policy shocks estimated for 2007:Q4, 2009:Q4, 2011:Q4 as well as 2016:Q4; dashed lines indicate responses at times when the ELB was binding for actual data. Shocks are scaled to generate a 1 percentage point drop in the shadow rate on impact. Vertical axis units are in percentage points. Horizontal axis units are quarters after impact of the monetary policy shock, which occurs at quarter zero.

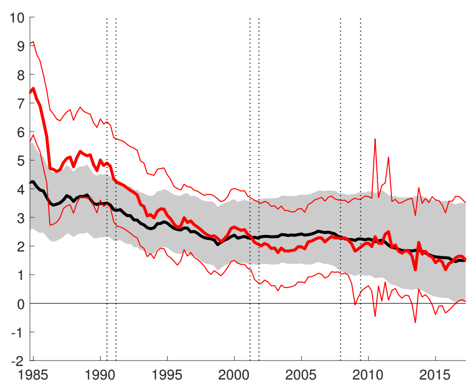
V Results with Alternative Orderings of Gap Variables

Our model embeds a VAR for the gap components with stochastic-volatility in its orthogonalized shocks, see equation (8). In a constant-variance case, estimation of the VAR would be invariant of the ordering of shocks in the Choleski decomposition implied by the unit-lower-triangular structure of $\boldsymbol{B}$ in equation (8) However, due to the stochastic volatilities, estimation of the VAR coefficients is not invariant to the ordering of variables in the stochastic-volatility case (Primiceri, 2005). This appendix documents the robustness of trend and gap estimates to different variable orderings.

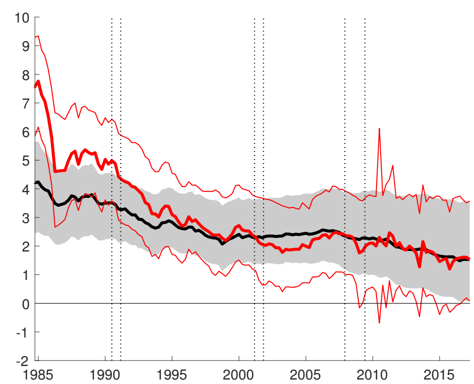
Figure A.14: Inflation Trend Estimates with re-ordered Gap VAR



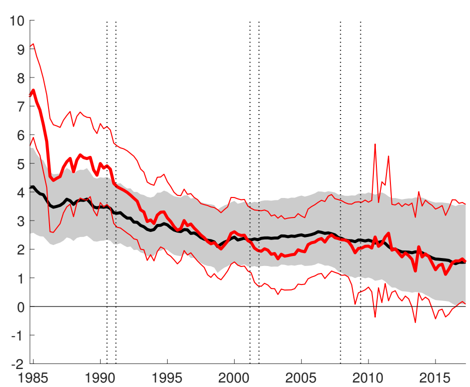
(a) π,c,s,y (baseline)



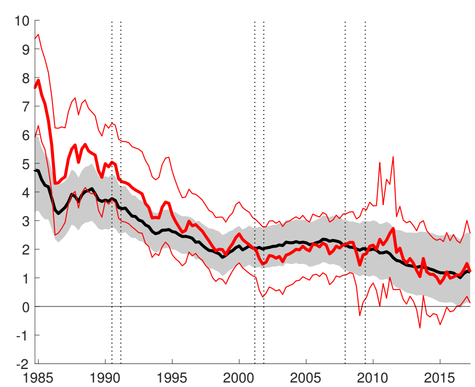
(b) π,s,c,y



(c) c,π,s,y



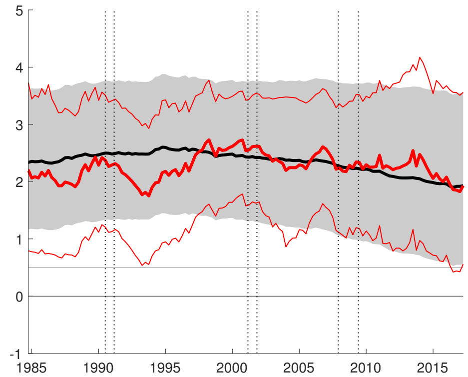
(d) s,c,π,y



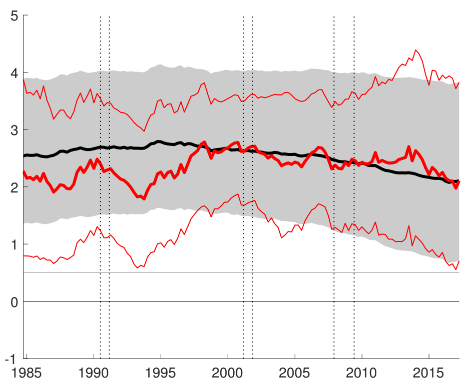
(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.14a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

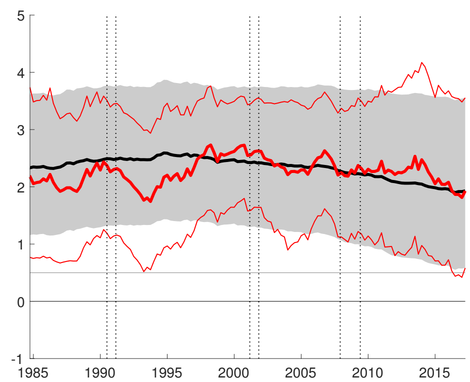
Figure A.15: Real Rate Trend Estimates with re-ordered Gap VAR



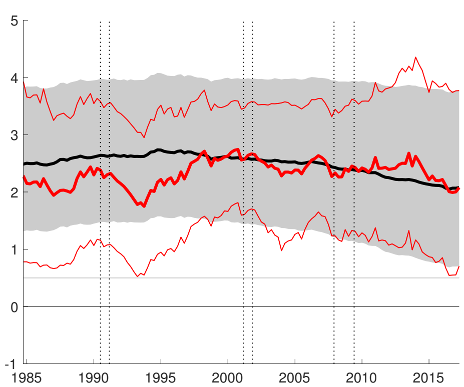
(a) π,c,s,y (baseline)



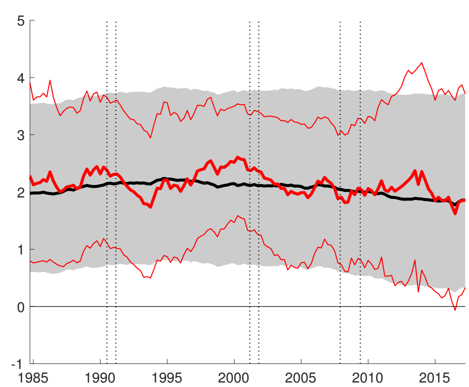
(b) π,s,c,y



(c) c,π,s,y



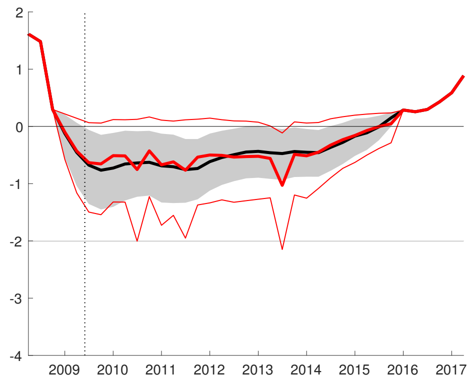
(d) s,c,π,y



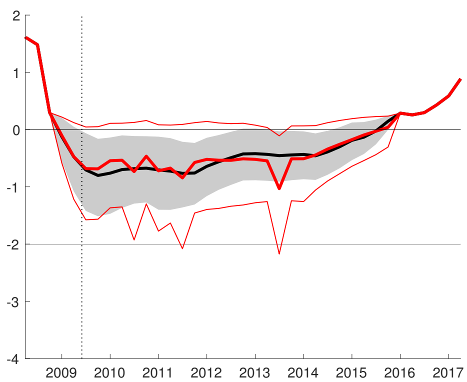
(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.15a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

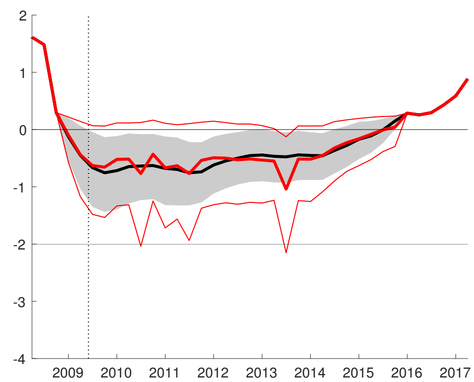
Figure A.16: Shadow Rate Estimates with re-ordered Gap VAR



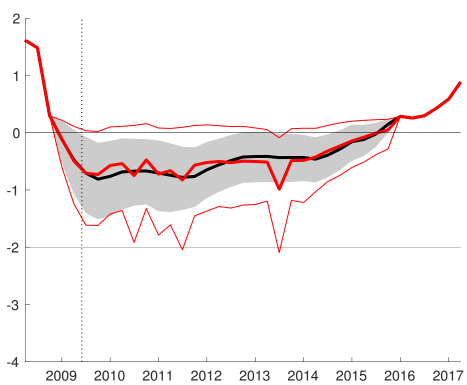
(a) π,c,s,y (baseline)



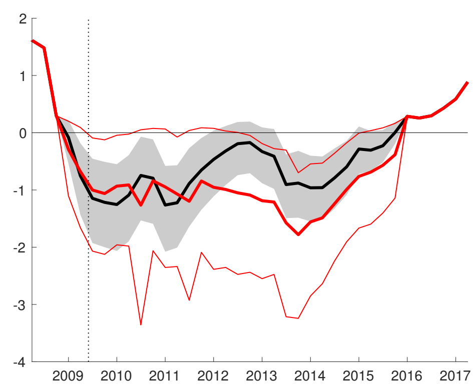
(b) π,s,c,y



(c) c,π,s,y



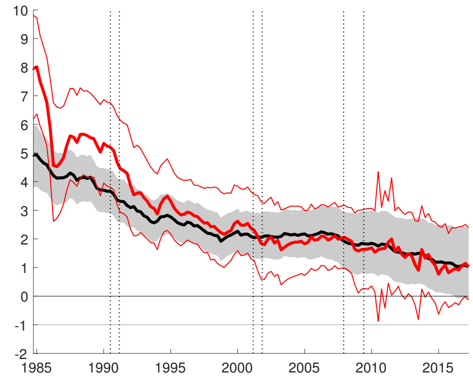
(d) s,c,π,y



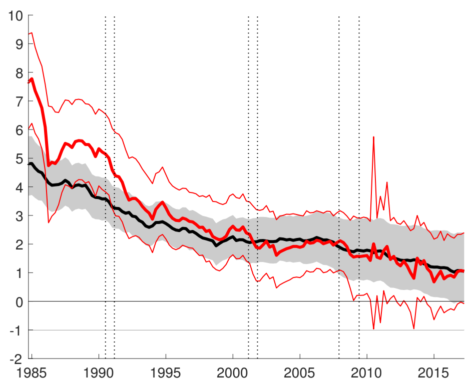
(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.16a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

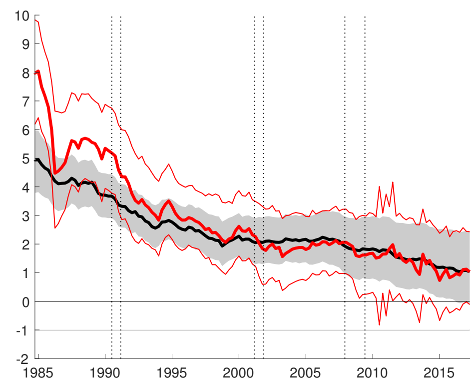
Figure A.17: Inflation Trend Estimates with re-ordered Gap VAR (w/Unemployment Rate Gap)



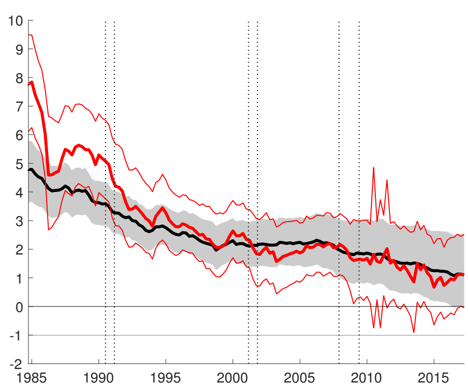
(a) π,c,s,y (baseline)



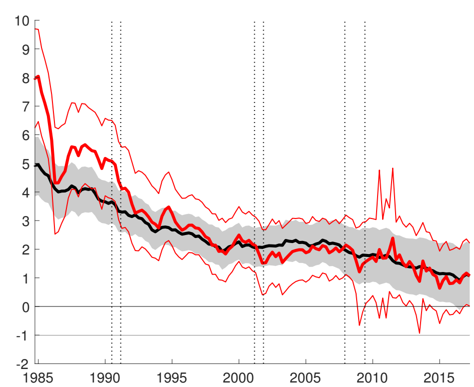
(b) π,s,c,y



(c) c,π,s,y



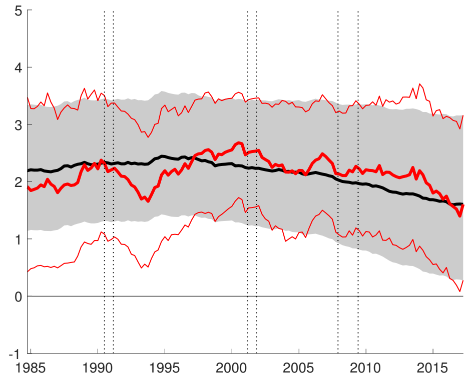
(d) s,c,π,y



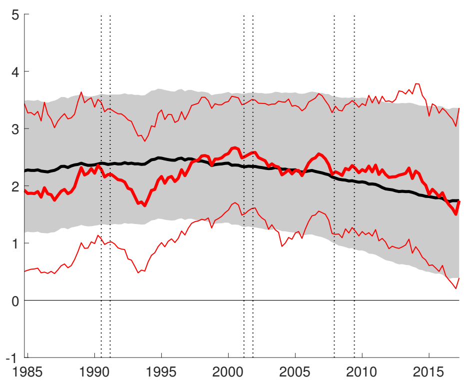
(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.17a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

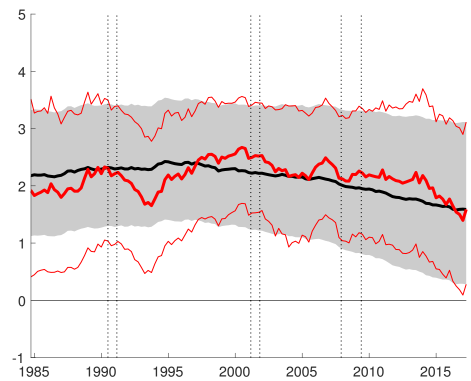
Figure A.18: Real Rate Trend Estimates with re-ordered Gap VAR (w/Unemployment Rate Gap)



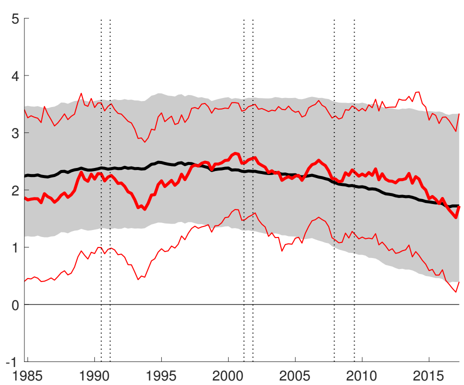
(a) π,c,s,y (baseline)



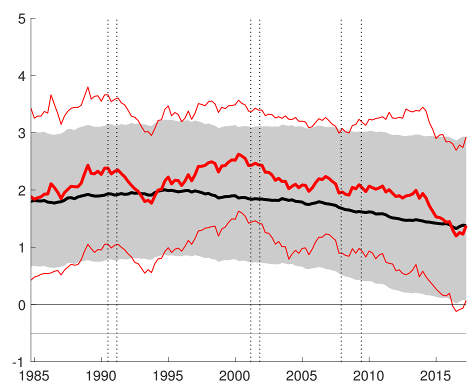
(b) π,s,c,y



(c) c,π,s,y



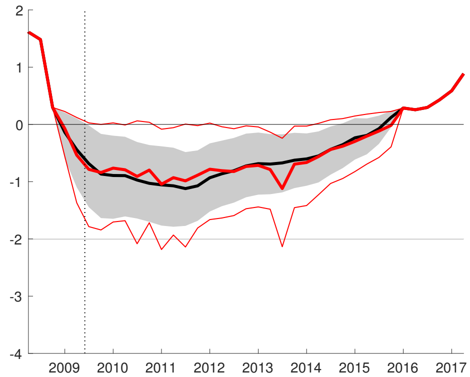
(d) s,c,π,y



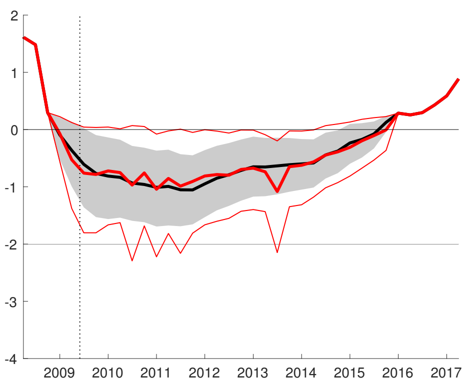
(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.18a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

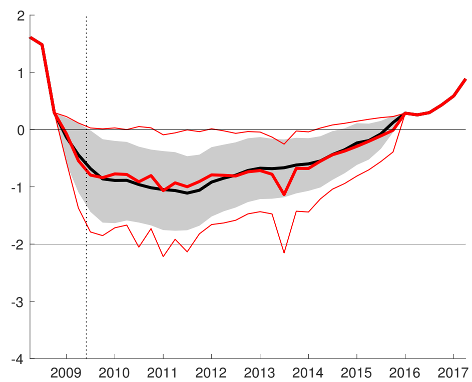
Figure A.19: Shadow Rate Estimates with re-ordered Gap VAR (w/Unemployment Rate Gap)



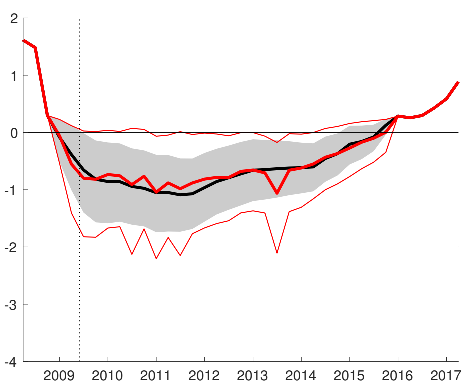
(a) π,c,s,y (baseline)



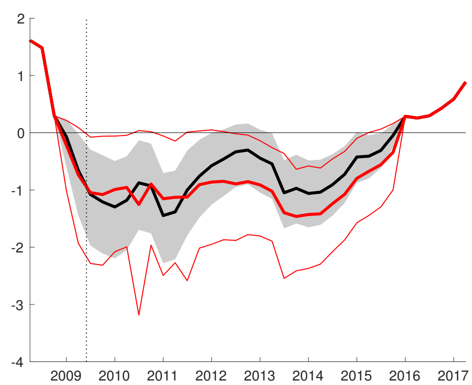
(b) π,s,c,y



(c) c,π,s,y



(d) s,c,π,y



(e) y,s,c,π

Note: Estimates generated from alternative orderings of gap variables in the VAR described in equation (8). Panel A.19a depicts baseline estimates; alternative orderings are as indicated above where "y" indicates the block of yield gaps y^2, y^5, y^{10}. Filtered estimates in red, smoothed estimates in black; both surrounded by 90% uncertainty bands (Filtered estimates reflect the endpoints of sequentially re-estimating the entire model over growing samples starting in 1960:Q1. Smoothed estimates use all available observations from 1960:Q1 through 2017:Q2.

VI Forecasting Performance vs. a Random Walk since 1985

Table A.4 compares the forecasts from our model with the no-change forecasts from a random-walk model for the period after 1985. We start in 1985 so that our model has several periods to use a training sample before producing out-of-sample forecasts. The statistics for our model are shown as calculated and the statistics for the random-walk model are shown on a relative basis to our model. We include results from versions of our model where the CBO's measure of the output gap is used as our business cycle measure and where the CBO's measure of the unemployment rate gap is used as our business cycle measure. For both the short- and long-term rate, our model performs about as well, on balance, as the random-walk model.

Table A.4: Comparison of Interest Rate Forecasts against Random Walk (starting 1985:Q1)

	Forecast horizon h					
	1	2	3	4	5	8
	Panel A: Short-term interest rate i_{t+h}					
	Model (output gap)					
MAD	0.22	0.45	0.67	0.88	1.09	1.59
RMSE	0.12	0.45	0.92	1.55	2.24	2.04
	RW rel. to model (output gap)					
rel. MAD	1.14*	1.06	1.01	0.99	0.96	0.93
rel. RMSE	1.12*	1.05	1.02	1.00	0.98	0.98
	Model (unemployment rate gap)					
MAD	0.23	0.46	0.68	0.88	1.08	1.58
RMSE	0.13	0.47	0.95	1.58	2.25	2.04
	RW rel. to model (unemployment rate gap)					
rel. MAD	1.10	1.03	0.98	0.98	0.97	0.94
rel. RMSE	1.10*	1.03	1.01	0.99	0.98	0.98
	Panel B: 10-year interest rate y^{10}_{t+h}					
	Model (output gap)					
MAD	0.31	0.49	0.62	0.73	0.80	0.87
RMSE	0.15	0.38	0.59	0.82	0.98	1.11
	RW rel. to model (output gap)					
rel. MAD	1.00	1.02	1.00	0.99	0.98	0.97
rel. RMSE	1.01	1.01	1.02	1.01	1.01	0.94
	Model (unemployment rate gap)					
MAD	0.31	0.50	0.62	0.74	0.81	0.90
RMSE	0.15	0.38	0.60	0.83	1.00	1.13
	RW rel. to model (unemployment rate gap)					
rel. MAD	0.99	1.01	0.99	0.97	0.96	0.93
rel. RMSE	1.00	1.01	1.01	1.01	1.00	0.92

Note: *RMSE* are root-mean-squared errors computed from using the medians of our model's and the no-change forecasts from a random-walk model; *MAD* are mean absolute deviations obtained from using the same forecasts. Relative RMSE and MAD are expressed as ratios relative to the corresponding statistics from the baseline model (values below unity denoting better performance than our model). Predictive densities are re-estimated over growing samples that start in 1985:Q1 for our model. Stars indicate significant differences, relative to baseline, in squared losses, absolute losses and density scores, respectively, as assessed by the test of Diebold and Mariano (1995); ***, ** and * denote significance at the 1%, 5% respectively 10% level.

VII Particle Filtering the Likelihood for MDD Computations

Computation of the marginal data densities (MDD) in Section 3.3 relies on particle filter estimates of our model's likelihood. We estimate the MDD using the harmonic mean estimator of Geweke (1999), as presented by Herbst and Schorfheide (2014), given by:

$$p(\boldsymbol{Z}) \approx \left[\frac{1}{N} \sum_{n=1}^{N} \frac{f(\boldsymbol{\theta}^n)}{p(\boldsymbol{Z}|\boldsymbol{\theta}^n)p(\boldsymbol{\theta}^n)} \right]^{-1}, \tag{15}$$

where N is the number of draws from the posterior distribution, $\boldsymbol{\theta}$ is a vector that collects all of the estimated parameters and $\boldsymbol{\theta}^n$ is a particular draw from the posterior distribution, f is a function of the parameter vector that integrates to one, $p(\boldsymbol{\theta}^n)$ is the prior density of $\boldsymbol{\theta}^n$, and $p(\boldsymbol{Z}|\boldsymbol{\theta}^n)$ is the likelihood of the data, stacked into the vector $\boldsymbol{Z}$. As in Geweke (1999), we use the following choice of f:

$$\begin{aligned} f(\boldsymbol{\theta}) \; = \; & \tau^{-1}(2\pi)^{-d/2}|\boldsymbol{V}_{\boldsymbol{\theta}}|^{-1/2} \exp\left[-0.5\,(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\boldsymbol{V}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})\right] \\ & \times \mathcal{I}\left\{(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\boldsymbol{V}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}) \leq F_{\chi_d^2}^{-1}(\tau)\right\} \end{aligned}$$

where $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{V}_\theta$ be the mean and variance of the posterior distribution of $\boldsymbol{\theta}$, d the length of $\boldsymbol{\theta}$, and $F_{\chi_d^2}$ is the cumulative distribution function of the χ^2 distribution with d degrees of freedom, and $\mathcal{I}$ is the indicator function. We set $\tau = 0.9$, and our results are robust to other choices of τ.

Since our model has two layers of latent variables, the likelihood $p(\boldsymbol{Z}|\boldsymbol{\theta}^n)$ in equation (15) cannot be computed analytically.[10]

Using notation introduced in Appendix I, the two layers of latent state variables are the trends and gaps stacked in $\boldsymbol{\xi}_t$ as well as the stochastic volatilities captured by $\boldsymbol{h}_t$. Effectively, our model

[10]The prior, described in Section I, can be broken down into products of normal and inverse-Wishart density functions. For some parameters, like the coefficients of the gap VAR, denoted $\boldsymbol{a}$ in Section I, or the AR(1) lag coefficients $\boldsymbol{\rho}$) in equation (A.9), the priors are *truncated* normals and the necessary rejection probabilities are straightforward to sample.

is a non-linear state-space model with the following composite state vector:[11]

$$\boldsymbol{S_t} = \begin{bmatrix} \boldsymbol{h}_t \\ \boldsymbol{\xi}_t \end{bmatrix} \tag{A.14}$$

For the likelihood computation, this composite state vector needs to be integrated out, which is not provided by the MCMC sampler.[12]

However, a particle filter can be used to approximate the likelihood of such a non-linear model. In fact, apart from the ELB constraint, estimation of an unobserved component model with stochastic volatility (UC-SV) like our is fairly straightforward with a Rao-Blackwellized particle filter (RB-PF) that exploits the conditionally linear structure of the model. RB-PFs are surveyed, for example by Creal (2012) and Lopes and Tsay (2011); see also the applications by Carvalho et al. (2017), Mertens and Nason (2017) and Mertens (2016). After describing a standard RB-PF that neglects any ELB issues, this section turns to our handling of the ELB constraint within a RB-PF.

VII.1 Rao-Blackwellized Particle Filter

For ease of exposition, let us first describe the RB-PF without considering issues arising from the ELB. For now, we thus treat shadow rates, and thus $\boldsymbol{X}_t$ as observable. The particle filter approximates the likelihood $p(\boldsymbol{X}^t|\boldsymbol{\theta})$ and filtered posterior density of the latent state vector $\boldsymbol{S_t} \mid \left(\boldsymbol{X}^t, \boldsymbol{\theta}\right)$ for given $\boldsymbol{X}^t$ and parameter values $\boldsymbol{\theta}$. The priors for the initial values $\boldsymbol{S_0}$ are as described in Appendix I.[13]

At each point in time, indexed by t, the filter tracks a swarm of M "particles", indexed by i, that consist of the stochastic volatilities $\boldsymbol{h}_t^{(i)}$ and Kalman filtered estimates of the linear state that

[11] Conditional on the stochastic volatilities, the dynamics of $\boldsymbol{\xi}_t$ are linear, and we will refer to $\boldsymbol{\xi}_t$ also as "linear state variables."

[12] Recall that the MCMC sampler iterates only between generating smoothed conditional draws of $\boldsymbol{\xi}^T \Big| \left(\boldsymbol{h}^T; \boldsymbol{\theta}\right)$ and $\boldsymbol{h}^T \Big| \left(\boldsymbol{\xi}^T; \boldsymbol{\theta}\right)$ — as well as generating appropriate draws of $\boldsymbol{\theta}$ — but does not *jointly filter* $\boldsymbol{\xi}_t$ and $\boldsymbol{h}_t$.

[13] As before, the initial variance of the gaps components in $\boldsymbol{\xi}_0$, depends on the parameter vector $\boldsymbol{\theta}$ as well as stochastic volatilities for the gap shocks, $\tilde{\boldsymbol{\Sigma}}_t$; the gap variance is recomputed for each particle draw of $\tilde{\boldsymbol{\Sigma}}_0$ and based on the specific parameter vector $\boldsymbol{\theta}^n$ used when evaluating the particle filter.

condition on the particles history of $\boldsymbol{h}_t^{(i)}$; the particle's Kalman filtering distribution of the linear states and conditional on data $\boldsymbol{X}^t$ is characterized by the mean vector $\boldsymbol{\xi}_{t|t}^{(i)}$ and variance-covariance matrix $\boldsymbol{\Psi}_{t|t}^{(i)}$.

We utilize an auxiliary particle filter (APF) as described by Lopes and Tsay (2011). First introduced by Pitt and Shephard (1999), the APF is a refinement of the bootstrap filter. While the bootstrap filter propagates particles from one period to the next based on their prior distribution, the APF seeks to adapt new particles based on the likelihood they imply for the data. Before turning to the unique steps of the auxiliary particle filtering steps, we describe the standard bootstrap algorithm.

The filter begins by generating M initial particles $\boldsymbol{h}_0^{(i)}$, $\boldsymbol{\xi}_{t|t}^{(i)}$ and $\boldsymbol{\Psi}_{t|t}^{(i)}$ from the their respective priors described in Appendix I. For $t = 1, \ldots, T$, the filter repeats the following steps:

1. For $i = 1, \ldots, M$ draw new particles $\boldsymbol{h}_t^{(i)}$ based on its prior conditional on $\boldsymbol{h}_{t-1}^{(i)}$ implied by (A.9) and construct the corresponding diagonal matrix of log-volatilities $\boldsymbol{\Sigma}_t^{1/2,(i)}$ and $\boldsymbol{\mathcal{B}}_t^{(i)} = \boldsymbol{\mathcal{B}}\,\boldsymbol{\Sigma}_t^{1/2,(i)}$.

2. For each particle i, engage the Kalman filter for equation (13) to compute:

$$\boldsymbol{\Psi}_{t|t-1}^{(i)} = \boldsymbol{\mathcal{A}}_t\ \boldsymbol{\Psi}_{t-1|t-1}^{(i)}\ \boldsymbol{\mathcal{A}}_t' + \boldsymbol{\mathcal{B}}_t^{(i)}\boldsymbol{\mathcal{B}}_t^{(i)\prime} \tag{A.15}$$

$$\boldsymbol{\Omega}_{t|t-1}^{(i)} = \boldsymbol{\mathcal{C}}_t\ \boldsymbol{\Psi}_{t|t-1}^{(i)}\ \boldsymbol{\mathcal{C}}_t{}' \tag{A.16}$$

$$\boldsymbol{e}_t^{(i)} = \boldsymbol{X}_t - \boldsymbol{\mathcal{C}}_t\ \boldsymbol{\mathcal{A}}_t\ \boldsymbol{\xi}_{t-1|t-1}^{(i)} \tag{A.17}$$

$$l_t^{(i)} = -\frac{1}{2}\left\{\log\left(2\cdot\pi\right)\cdot N_x^* + \log\left|\boldsymbol{\Omega}_{t|t-1}^{(i)}\right|_+ + \boldsymbol{e}_t^{(i)\prime}\left(\boldsymbol{\Omega}_{t|t-1}^{(i)}\right)^+ \boldsymbol{e}_{t|t-1}^{(i)}\right\} \tag{A.18}$$

$$\boldsymbol{K}_t^{(i)} = \boldsymbol{\Psi}_{t|t-1}^{(i)}\ \boldsymbol{\mathcal{C}}_t{}'\ \left(\boldsymbol{\Omega}_{t|t-1}^{(i)}\right)^+ \tag{A.19}$$

$$\boldsymbol{\xi}_{t|t}^{(i)} = \boldsymbol{\xi}_{t|t-1}^{(i)} + \boldsymbol{K}_t^{(i)}\boldsymbol{e}_t^{(i)} \tag{A.20}$$

$$\boldsymbol{\Psi}_{t|t}^{(i)} = \boldsymbol{\Psi}_{t|t-1}^{(i)} - \boldsymbol{\Psi}_{t|t-1}^{(i)}\ \boldsymbol{\mathcal{C}}_t{}'\ \left(\boldsymbol{\Omega}_{t|t-1}^{(i)}\right)^+\ \boldsymbol{\mathcal{C}}_t\ \boldsymbol{\Psi}_{t|t-1}^{(i)}{}' \tag{A.21}$$

In light of the possibility of missing data — encoded as elements of $\boldsymbol{X}_t$ fixed at zero and a rank deficient $\boldsymbol{\mathcal{C}}_t$ — note that N_x^* is the number of actual observations in $\boldsymbol{X}_t$, corresponding to the number of non-zero rows of $\mathcal{C}_t$, and $|\cdot|_+$ and $\cdot^+$ denote the pseudo-determinant and pseudo-inverse operators, respectively.

3. Compute the particle weights

$$w_t^{(i)} = \frac{\exp\left(l_t^{(i)}\right)}{\sum_i^M \exp\left(l_t^{(i)}\right)}.$$

The filtered distribution of $\boldsymbol{h}_t$ is approximated by the discrete distribution of particle draws $\boldsymbol{h}_t^{(i)}$ using the *pdf* described by $w_t^{(i)}$. The associated filtered distribution of $\boldsymbol{\xi_t}$ is approximated by a mixture of normals $N\left(\boldsymbol{\xi_{t|t}^{(i)}}, \boldsymbol{\Psi_{t|t}^{(i)}}\right)$ with weights $w_t^{(i)}$. (For this purpose, the values for $\boldsymbol{h}_t^{(i)}$, $\boldsymbol{\xi_{t|t}^{(i)}}$ and $\boldsymbol{\Psi_{t|t}^{(i)}}$ are stored before the resampling described in the next step.)

4. For $t < T$ prepare the next iteration by applying systematic resampling to the particles $\boldsymbol{h}_t^{(i)}$, $\boldsymbol{\xi_{t|t}^{(i)}}$ and $\boldsymbol{\Psi_{t|t}^{(i)}}$ based on the particle weights $w_t^{(i)} = \frac{\exp\left(l_t^{(i)}\right)}{\sum_i^M \exp\left(l_t^{(i)}\right)}$.

The likelihood of the date t observation, conditional on parameters and the previous history of observations is estimated by averaging over the likelihoods of each particle:[14]

$$p\left(\boldsymbol{X_t}|\boldsymbol{X^{t-1}}, \boldsymbol{\theta}^*\right) \propto \frac{1}{M}\sum_{i=1}^{M} \exp\left(l_t^{(i)}\right) \tag{A.22}$$

The log-likelihood, which corresponds also to the log-predictive score for given parameter values $\boldsymbol{\theta}^*$ (Geweke and Amisano, 2010; Creal et al., 2010), is then given by

$$\mathcal{L}\left(\boldsymbol{X^T}|\boldsymbol{\theta}^*\right) = \sum_{t=1}^{T} \log\left\{p\left(\boldsymbol{X_t}|\boldsymbol{X^{t-1}}, \boldsymbol{\theta}^*\right)\right\} . \tag{A.23}$$

The bootstrap filter described above generates new particles for item t merely by propagating stochastic volatilities based on their prior values, $\boldsymbol{h}_{t-1}^{(i)}$ and the law of motion (A.9) but without regard for the likelihood they will attract at t. The APF seeks to adapt new particle draws $\boldsymbol{h}_t^{(i)}$

[14]Since particles get reweighed at every step, the simple average is appropriate, see, for example, Creal (2012).

to data at t. To do so, we employ an algorithm described by Lopes and Tsay (2011) for Rao-Blackwellized particle filters that performs two resampling steps. First, before simulating new particles $\boldsymbol{h}_{t-1}^{(i)}$ from equation (A.9), time $t-1$ particles are resampled based on the particle weights implied at t when using the auxiliary particles $\boldsymbol{h}_t^{(i),*} = \boldsymbol{\mu} + \boldsymbol{\rho}\left(\boldsymbol{h}_{t-1}^{(i)} - \boldsymbol{\mu}\right)$.[15] Denoting the weights implied by the auxiliary particles $w_t^{*,(i)}$, weights are then given by $w_t^{(i)} = \tilde{w}_t^{(i)} / \left(\sum_i \tilde{w}_t^{(i)}\right)$ where $\tilde{w}_t^{(i)} = l_t^{(i)} / w_t^{*,(i)}$ and, as in equation (A.18), $l_t^{(i)}$ are the likelihood contributions but now generated by the particles obtained from propagating $\boldsymbol{h}_{t-1}^{(i)}$ after the auxiliary reweighting. Further details are provided by Lopes and Tsay (2011).

VII.2 Adjusting the Particle Filter to account for the ELB

Neglecting the ELB, the Rao-Blackwellized particle filter described above relies on the conditionally linear structure of the UC-SV model so that, for each particle, posterior distributions of the linear states $\boldsymbol{\xi}_t$ are normal and can be tracked by following the evolution of their first two moments, $\boldsymbol{\Psi}_{t|t}^{(i)}$ and $\boldsymbol{\Psi}_{t|t}^{(i)}$. Indeed, for data histories where the ELB has not been binding, we have $\boldsymbol{X}^t = \boldsymbol{Z}^t$ and can directly employ the RB-PF described above. However, once the ELB binds, the shadow rate becomes a latent variable, and thus $\boldsymbol{X}_t \neq \boldsymbol{Z}_t$. The shadow rate is a linear combination of $\boldsymbol{\xi}_t$. Thus, to ensure that shadow rates are below the ELB when the actual rate is at the ELB, the posterior for $\boldsymbol{\xi}_t$ must be characterized by a truncated distribution.

To account for the ELB, we adapt the RB-PF as follows: Since a conjugate prior-posterior characterization for ξ_t is not available at the ELB, we abandon the Rao-Blackwellization once the ELB binds in the data and proceed as follows. Instead, we add draws for $\boldsymbol{\xi}_t^{(i)}$ to the particle vector that satisfy the truncation constraint $s_t = \boldsymbol{c}_s \, \boldsymbol{\xi}_t^{(i)} \leq ELB$.

Consider the case where the ELB has not been binding for $\boldsymbol{Z}_{t-1}$ but binds for $\boldsymbol{Z}_t$. In this case, for every particle, we inherit normal priors for the linear states characterized by $\boldsymbol{\xi}_{t|t-1}^{(i)}$ and $\boldsymbol{\Psi}_{t|t-1}^{(i)}$. Analogously to the MCMC algorithm described in Appendix A, these priors for $\boldsymbol{\xi}_t$ can be updated to $\boldsymbol{\xi}_{t|t}^{NC,(i)}$ and $\boldsymbol{\Psi}_{t|t}^{NC,(i)}$ by conditioning only on the elements of $\boldsymbol{Z}_t$ for which the ELB did not bind.

[15] The auxiliary particles are thus generated at the means implied by the prior particles.

To account for the ELB binding at t, we then sample particles $\boldsymbol{\xi}_t^{(i)}$ from $N\left(\boldsymbol{\xi}_{t|t}^{NC,(i)}, \boldsymbol{\Psi}_{t|t}^{NC,(i)}\right)$ that are consistent with $s_t^{(i)} = \boldsymbol{c}_s \, \boldsymbol{\xi}_t^{(i)} \leq ELB$.

The RB-PF algorithm described in the previous section above can then simply be amended by replacing the priors for every particle at $t+1$, originally denoted $\left(\boldsymbol{\xi}_{t+1|t}^{(i)}, \boldsymbol{\Psi}_{t+1|t}^{(i)}\right)$ by the draw $\boldsymbol{\xi}_t^{(i))}$ as well as a matrix of zeros before evaluating the time $t+1$ Kalman filtering steps described above. Effectively, the algorithm proceeds by using a degenerate normal posterior at t for $\boldsymbol{\xi}_t$ that has all point mass at $\boldsymbol{\xi}_t^{(i)}$. While the *Rao-Blackwellized* particle filter tracks at least part of the uncertainty about $\boldsymbol{\xi}_t$ via particle-specific conditional distributions, uncertainty about $\boldsymbol{\xi}_t$ becomes entirely captured by the range of values for $\boldsymbol{\xi}_t^{(i)}$ generated across particles when the ELB binds.

In addition to abandoning the Rao-Blackwellization, the likelihood contributions $l_t^{(i)}$ in (A.18) need to be augmented by adding the probability of the shadow rate being below the ELB conditional on all other elements of $\boldsymbol{Z}_t$, which is straightforward to compute based on the normal CDF for s_t implied by $\boldsymbol{\xi}_{t|t}^{NC,(i)}$, and $\boldsymbol{\Psi}_{t|t}^{NC,(i)}$. This add factor accounts for the information contained in observing $i_t = ELB$ by tilting particle weights towards particles that place higher likelihood on the ELB binding. Specifically, (A.18) is replaced by

$$l_t^{(i)} = -\frac{1}{2}\left\{\log\left(2\cdot\pi\right)\cdot N_x^* + \log\left|\boldsymbol{\Omega_{t|t-1}^{(i)}}\right|_+ + \boldsymbol{e_t^{(i)\prime}}\left(\boldsymbol{\Omega_{t|t-1}^{(i)}}\right)^+ \boldsymbol{e_{t|t-1}^{(i)}}\right\}$$
$$+ \log\left(\text{Prob}\left(s_t \leq ELB \middle| \boldsymbol{\xi}_{t|t}^{NC,(i)}, \boldsymbol{\Psi}_{t|t}^{NC,(i)}\right)\right). \quad \text{(A.24)}$$

VIII Constructing IRF with a Particle Filter

Neglecting issues related to the ELB, the impulse responses described in Section 5 of the main paper, are straightforward to compute from the Rao-Blackwellized particle filter (RB-PF) described in Section VII.1 above. In particular, computing responses to one-standard deviation shocks, simplifies the computation of particle weights for the impulse responses as described below. To account for the ELB, we employ a mean-variance approximation to the UC-SV model's innovations representation described in the remainder of this section.

VIII.1 Particle Weights used for Impulse Responses

Impulse responses reflect the difference between a baseline forecast and a revised forecast prompted by the hypothetical observation of an identified shock. In the context of the RB-PF, the arrival of new information generally prompts also a reweighting of particles reflecting changes in each particles' likelihood given the new data. To simplify the IRF computation, we compute responses to one-standard deviation shocks. As such, these shocks provide information about the direction of the impulse in the space of innovations $\boldsymbol{e}_t$ but do not lead to a reassessment of the likelihood of different values for the stochastic volatilities that are tracked by the different particles. As such, the arrival of such a standardized shock does not affect a change in particle weights relative to the baseline forecast. Particle weights for the baseline forecast made at time $t-1$, denoted $w_{t-1}^{(i)}$, are computed as described in Section VII above. When computing the revised forecast, these weight are simply propagated as $w_{t|t-1}^{(i)} = w_{t-1}^{(i)}$ and we can compute impulse responses as:

$$\text{IRF}_t^k = \sum_i w_{t|t-1}^{(i)} \cdot \left(\boldsymbol{d}_{k,t}^m\right)^{(i)} \qquad \text{with} \qquad \left(\boldsymbol{d}_{k,t}^m\right)^{(i)} = \boldsymbol{\mathcal{C}}\, \boldsymbol{\mathcal{A}}^k\, \boldsymbol{K}_t^{(i)}\, \boldsymbol{q}_t^{m,(i)} \,. \tag{21}$$

Of course, the analysis could be extended to consider also the effects from reweighting the different particles prompted by observing shocks of a particular *absolute* size. In this case, the impulse responses would reflect an additional term, $\left(w_{t-1}^{(i)} - w_t^{(i)}\right) \cdot E(\boldsymbol{X}_t|\boldsymbol{X}^{t-1})$, that captures the reevaluation of the weight attached to each particle's baseline forecast in light of the incoming shock.

VIII.2 Accounting for the ELB

The impulse responses are based on innovations in the shadow rate and rely on the innovations representation of the UC-SV model with linear measurements and Gaussian shocks. As indicated in Section 5.1.4 of the main paper, when the ELB binds, we use a mean-variance approximation of the innovations representation. This approximation differs from our handling of the ELB for the purpose of computing the model's likelihood as described in Section VII.2 for the MDD computation. There, we could simply abandon the Rao-Blackwellization at the ELB, albeit at the cost of removing any uncertainty about trend and gap levels *per particle*, since each particle conditions on a specific draw $\boldsymbol{\xi}_t^{(i)}$ before propagating when at the ELB. But, time-varying uncertainty about trend and gap levels is a key driver for time-variation in the VMA resulting from the model's innovations representation. Instead of replacing the Rao-Blackwellization with draws of $\boldsymbol{\xi}_t^{(i)}$ that obey $s_t^{(i)} = \boldsymbol{c}_s\,\boldsymbol{\xi}_t^{(i)} \leq ELB$ when the ELB binds, we have thus chosen to resort to the following mean-variance approximation for the construction of IRFs:

When the ELB binds at t, we generate $N\left(\boldsymbol{\xi}_{t|t}^{NC,(i)}, \boldsymbol{\Psi}_{t|t}^{NC,(i)}\right)$ as described in Section VII.2. We then approximate the truncated normal

$$\boldsymbol{\xi}_t^{(i)}\bigg|\left(\boldsymbol{c}_s\,\boldsymbol{\xi}_t^{(i)} \leq ELB\right) \sim TN\left(\boldsymbol{\xi}_{t|t}^{NC,(i)}, \boldsymbol{\Psi}_{t|t}^{NC,(i)}\bigg|\boldsymbol{c}_s\,\boldsymbol{\xi}_t^{(i)} \leq ELB\right) \tag{A.25}$$

$$\boldsymbol{\xi}_{t|t}^{*,(i)} \equiv E\left(\boldsymbol{\xi}_t^{(i)}\bigg|\boldsymbol{c}_s\,\boldsymbol{\xi}_t^{(i)} \leq ELB\right) \tag{A.26}$$

$$\boldsymbol{\Psi}_{t|t}^{*,(i)} \equiv \mathrm{Var}\left(\boldsymbol{\xi}_t^{(i)}\bigg|\boldsymbol{c}_s\,\boldsymbol{\xi}_t^{(i)} \leq ELB\right) \tag{A.27}$$

with a normal distribution $N\left(\boldsymbol{\xi}_{t|t}^{*,(i)}, \boldsymbol{\Psi}_{t|t}^{*,(i)}\right)$ that matches first and second moments of the true truncated distribution. Estimates of the model's latent variables obtained with this approximation are close to this obtained from the particle filter described in Section VII where the ELB is enforced exactly.

Mean and variance coefficients in (A.26) and (A.27) can be computed analytically as follows: First, we derive the conditional distribution of $\boldsymbol{\xi}$ for given values of the shadow rate. Second, we spell out the mean and variance of the shadow rate conditional on the shadow rate being below the ELB. Third, combining results from the previous two steps yields the desired moments of $\boldsymbol{\xi}$ conditional on the shadow rate being below the ELB. The details are described in the remainder of this section.

For ease of notation, we drop time and particles indices t and i and let $\boldsymbol{\mu} = \boldsymbol{\xi}_{t|t}^{NC,(i)}$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}_{t|t}^{NC,(i)}$ as well as $\mu_s = \boldsymbol{c}_s\boldsymbol{\mu}$ and $\psi_s = \boldsymbol{c}_s\boldsymbol{\Psi}\boldsymbol{c}_s'$ and obtain the following setup:

$$\boldsymbol{\xi} \sim N(\boldsymbol{\mu}, \boldsymbol{\Psi}) \qquad\qquad s = \boldsymbol{c}_s\boldsymbol{\xi} \sim N(\mu_s, \psi_s) \tag{A.28}$$

For a given realization of the shadow rate s, the conditional distribution of $\boldsymbol{\xi}$ is given by standard signal extraction formulas:

$$\boldsymbol{\xi}\big|s \sim N\left(\hat{\boldsymbol{\mu}}(s), \hat{\boldsymbol{\Psi}}\right) \tag{A.29}$$

$$\text{with } \hat{\boldsymbol{\mu}}(s) \equiv (\boldsymbol{I} - \boldsymbol{\beta}\boldsymbol{c}_s)\boldsymbol{\mu} + \boldsymbol{\beta}\, s, \qquad \boldsymbol{\beta} = \boldsymbol{\Psi}\boldsymbol{c}_s'\psi_s^{-1}, \qquad \text{and } \hat{\boldsymbol{\Psi}} \equiv \boldsymbol{\Psi} - \boldsymbol{\Psi}\boldsymbol{c}_s'\psi_s^{-1}\boldsymbol{c}_s\boldsymbol{\Psi}\ . \tag{A.30}$$

Importantly for what follows, we obtain the following, *observationally equivalent*, orthogonal decomposition for the linear states:

$$\boldsymbol{\xi} = \boldsymbol{\beta}s + \boldsymbol{\varepsilon} \qquad\qquad E(s\,\boldsymbol{\varepsilon}) = 0 \tag{A.31}$$

where $\mathrm{Var}\,(\boldsymbol{\varepsilon}) = \mathrm{Var}\,(\boldsymbol{\xi}|s) = \hat{\boldsymbol{\Psi}}$ does not depend on s.[16] Since s and $\boldsymbol{\varepsilon}$ are jointly normal, the orthogonality condition $E(s\,\boldsymbol{\varepsilon}) = 0$ establishes also that both components are independently distributed from each other.[17]

[16]Note that the setup in equation (A.28) implies $|\hat{\boldsymbol{\Psi}}| = 0$, which does, however, not impede any of the subsequent calculations.

[17]Of course, for a given realization of $\boldsymbol{\xi}$, the distributions of $s|\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}|\boldsymbol{\xi}$ are interdependent. However, in order to derive the moments of $\boldsymbol{\xi}$ conditional on $s < ELB$, it is sufficient to consider equation (A.31), which is observationally equivalent to equation (A.28), as primitive representation.

Conditional on $s \leq ELB$, the shadow rate has the following mean and variance:

$$\mu_s^* \equiv E(s|s < ELB) = \mu_s - \sqrt{\psi_s} \cdot \frac{\phi(b)}{\Phi(b)} \tag{A.32}$$

$$\psi_s^* \equiv \operatorname{Var}(s|s < ELB) = \psi_s \left(1 - b \cdot \frac{\phi(b)}{\Phi(b)} - \left(\frac{\phi(b)}{\Phi(b)}\right)^2\right) \tag{A.33}$$

$$\text{given } b \equiv \frac{ELB - \mu_s}{\sqrt{\psi_s}} \tag{A.34}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are pdf and cdf of the standard normal distribution, respectively.

Finally, putting everything together, we obtain:

$$E(\boldsymbol{\xi}|s < ELB) = E\left(E(\boldsymbol{\xi}|s) \;\middle|\; s < ELB\right) = \hat{\boldsymbol{\mu}}(\mu_s^*) \tag{A.35}$$

$$\operatorname{Var}(\boldsymbol{\xi}|s < ELB) = \boldsymbol{\beta} \operatorname{Var}(s|s < ELB)\, \boldsymbol{\beta}' + \operatorname{Var}(\boldsymbol{\varepsilon}|s < ELB) = \boldsymbol{\beta}\, \psi_s^*\, \boldsymbol{\beta}' + \hat{\boldsymbol{\Psi}} \tag{A.36}$$

In (A.35), notice that $\hat{\boldsymbol{\mu}}(\mu_s^*)$ is linear in μ_s^*. Furthermore, in (A.36), note that the independence between s and $\boldsymbol{\varepsilon}$ leads to the the absence of covariance terms as well as $\operatorname{Var}(\boldsymbol{\varepsilon}|s < ELB) = \operatorname{Var}(\boldsymbol{\varepsilon}) = \hat{\boldsymbol{\Psi}}$.

References

Adolfson, Malin, Jesper Linde, and Mattias Villani (2007) "Forecasting performance of an open economy DSGE model." *Econometric Reviews*, 26, 289–328.

Carvalho, Carlos, Stefano Eusepi, Emanuel Moench, and Bruce Preston (2017) "Anchored inflation expectations." *mimeo*, August.

Christoffel, Kai, Anders Warne, and Gnter Coenen (2010) "Forecasting with DSGE models." Working Paper Series 1185, European Central Bank.

Clark, Todd E. (2011) "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility." *Journal of Business & Economic Statistics*, 29, 327–341.

Clark, Todd E. and Sharon Kozicki (2005) "Estimating equilibrium real interest rates in real time." *The North American Journal of Economics and Finance*, 16, 395–413.

Clark, Todd E. and Francesco Ravazzolo (2015) "Macroeconomic forecasting performance under alternative specifications of time-varying volatility." *Journal of Applied Econometrics*, 30, 551–575.

Cogley, Timothy and Thomas J. Sargent (2005) "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics*, 8, 262–302.

Creal, Drew (2012) "A survey of sequential Monte Carlo methods for economics and finance." *Econometric Reviews*, 31, 245–296.

Creal, Drew, Siem Jan Koopman, and Eric Zivot (2010) "Extracting a robust us business cycle using a time-varying multivariate model-based bandpass filter." *Journal of Applied Econometrics*, 25, 695–719.

Del Negro, Marco and Giorgio E. Primiceri (2015) "Time varying structural vector autoregressions and monetary policy: A corrigendum." *Review of Economic Studies*, 82, 1342–1345.

Diebold, Francis X. and Roberto S Mariano (1995) "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, 13, 253–63.

Durbin, J. and S. J. Koopman (2002) "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89, 603–615.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003) *Bayesian Data Analysis*: Chapman and Hall, 2nd edition.

Geweke, John (1999) "Using simulation methods for bayesian econometric models: inference, development, and communication." *Econometric reviews*, 18, 1–73.

Geweke, John and Gianni Amisano (2010) "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26, 216 – 230, Special Issue: Bayesian Forecasting in Economics.

Herbst, Edward and Frank Schorfheide (2014) “Bayesian inference for DSGE models.” September, *mimeo* Board of Governors of the Federal Reserve System.

Holston, Kathryn, Thomas Laubach, and John C. Williams (2017) “Measuring the natural rate of interest: International trends and determinants.” *Journal of International Economics*, 108, 59–75.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998) “Stochastic volatility: Likelihood inference and comparison with arch models.” *The Review of Economic Studies*, 65, 361–393.

Laubach, Thomas and John C. Williams (2003) “Measuring the natural rate of interest.” *The Review of Economics and Statistics*, 85, 1063–1070.

Lopes, Hedibert F. and Ruey S. Tsay (2011) “Particle filters and Bayesian inference in financial econometrics.” *Journal of Forecasting*, 30, 168–209.

Mertens, Elmar (2016) “Measuring the level and uncertainty of trend inflation.” *The Review of Economics and Statistics*, 98, 950–967.

Mertens, Elmar and James M. Nason (2017) “Inflation and professional forecast dynamics: An evaluation of stickiness, persistence, and volatility.” CAMA Working Papers 2017-60, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Pitt, Michael K. and Neil Shephard (1999) “Filtering via simulation: Auxiliary particle filters.” *Journal of the American Statistical Association*, 94, 590–599.

Primiceri, Giorgio E. (2005) “Time varying structural vector autoregressions and monetary policy.” *The Review of Economic Studies*, 72, 821–852.

Stock, James H. and Mark W. Watson (1998) “Median unbiased estimation of coefficient variance in a time-varying parameter model.” *Journal of the American Statistical Association*, 93, 349–358.

——— (2007) “Why has U.S. inflation become harder to forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.

Warne, Anders, Gnter Coenen, and Kai Christoffel (2015) “Marginalized predictive likelihood comparisons of linear Gaussian state-space models with applications to DSGE, DSGE-VAR, and VAR models.” December, *mimeo* European Central Bank.