Forecasting Related Time Series*

Ulrich K. Müller and Mark W. Watson
Department of Economics
Princeton University

This Draft: October 22, 2025

Abstract

A collection of time series are 'related' if they follow similar stochastic processes and/or they are statistically dependent. This paper proposes a Related Time Series (RTS) forecasting model that exploits these relationships. The model's foundation is a set of univariate autoregressions, one for each series, which are then augmented to incorporate stochastic volatility, heavy-tailed innovations, additive outliers, time varying parameters and common factors. The model is estimated and forecasts are computed using Bayes methods with hierarchical priors that pool information across series. Computationally efficient MCMC methods are proposed. The RTS model is applied to three datasets and yields encouraging pseudo-out-of-sample forecasting results.

Keywords: Bayes Forecasting, Factor Models, Hierarchical Priors
JEL: C11, C32, C53

*This paper was presented at the 13th Annual International Association of Applied Econometrics Conference in June 2025 as the *Journal of Applied Econometrics* Invited Paper. We thank the organizers for inviting us to write this paper and our discussants James Mitchell and Jonathan Wright for insightful comments.

1 Introduction

This paper considers "related" economic time series, where the series may be related in two distinct ways: They may be related because they follow similar stochastic processes and/or because they are statistically dependent. Series generated by autoregressive models with similar autoregressive (AR) coefficients are related in the first way, while variables that interact in a vector autoregression are related in the second way. In either case, information from one series may help forecast the value of another related series, where in the first instance this information involves the values of parameters used for forecasting, and in the second it exploits potential lead/lag relationships between the variables. This paper develops a forecasting model for related time series that utilizes both of these sources of information.

An empirical example used in the sections below helps set the stage. It uses a dataset comprised of monthly employment growth rates in each of the fifty U.S. states and the District of Columbia. While each state is different, they share many common features such as similar serial correlation patterns and variability. Moreover, macroeconomic forces in the U.S. affect all of the states, where these effects are stronger and/or materialize more quickly in some states than others, leading to dynamic cross-correlations among the series. Interest lies in forecasting the states' employment growth rates over the next several months.

This paper has three related goals. The first is to develop a flexible but tractable model that captures common features and comovement in the series and exploits these relationships for short-term forecasting. For lack of a better name, we call this the "Related Time Series" (RTS) model. The paper's second goal is to develop a set of numerical algorithms to reliably and efficiently carry out estimation and prediction for the RTS model. The third goal is to evaluate the forecasting performance of the RTS model using the state employment dataset and two unrelated datasets.

We construct the RTS model from familiar ingredients. At its foundation is a collection of univariate autoregressions, one for each series. It starts there because univariate AR models provide competitive point forecasts for a wide range of economic variables. Moving beyond point forecasts requires modeling higher-order moments, and the RTS model does this via stochastic volatility, heavy-tailed innovations, and additive outliers. Dependence across series is introduced through commonalities in stochastic volatility realizations and through a set of common factors that capture low-frequency, higher-frequency, and occasional large and

irregular comovements in the series. Parameters describing these features are allowed to slowly evolve through time, so that effectively, the model puts more weight on more recent data when constructing the forecast distribution. The RTS model is estimated using Bayesian methods, where hierarchical priors are used to pool information across series, and we develop a MCMC sampler to approximate posterior distributions.

Our analysis borrows from a vast literature. While it is impossible to do justice to all of the important contributions, we mention a few here. Perhaps the best place to start are the practical univariate methods described in Box and Jenkins (1970), and the incorporation of innovation and additive outliers in Fox (1972) and Abraham and Box (1979). The stochastic volatility in the RTS model has its genesis in Engle (1982) and the formulation we use builds on Kim, Shephard, and Chib (1998) and Omori, Chib, Shephard, and Nakajima (2007). The RTS model incorporates many features from unobserved component models with classic contributions in Nerlove, Grether, and Carvalho (1979) and Harvey (1989), and the work on factor model forecasting surveyed in Stock and Watson (2016b), related large-n forecasting methods in Banbura, Giannone, and Reichlin (2010), D'Agostino and Giannone (2012), Carriero, Clark, and Marcellino (2015), and multi-country VAR models of Chudik and Pesaran (2016), Bai, Carriero, Clark, and Marcellino (2022) and elsewhere. Several researchers have included many of the ingredients present in the RTS model: notable examples include Cogley and Sargent (2005), del Negro and Otrok (2008), Stock and Watson (2016a), Almuzara and Sbordone (2022), Antolin-Diaz, Drechsel, and Petrella (2024), and Carriero, Clark, Marcellino, and Mertens (2024). Gelman, Carlin, Stern, and Rubin (2004) provides a textbook overview of the Bayesian methods that we employ, and these are augmented with insights from Doan, Litterman, and Sims (1984), Meng and Wong (1996), Durbin and Koopman (2002), Geweke (2004), and Chan and Jeliazkov (2009).

The paper is organized as follows. Section 2 introduces the state employment dataset, and documents features of these data (similar autoregressive dynamics, stochastic volatility, outliers, comovement, etc.) that are incorporated in the RTS model. This section also outlines a pseudo-out-of-sample forecasting experiment and the loss functions we use to evaluate the RTS model's point, quantile and interval forecasts.

Section 3 develops the RTS model and is the heart of the paper. The RTS model includes autoregressive dynamics, heavy-tailed innovations and additive outliers (both modeled as Student-t random variables), low-frequency stochastic volatility with a component that

is common across series, time-varying level, autoregressive, and other parameters (modeled as random walks), and dynamic common factors. The model is estimated and predictions are formed using Bayesian methods, where hierarchical priors are used to pool information across the series. While each of these ingredients is straightforward, taken together they yield a notationally complex RTS model. With this in mind, Section 3 develops the RTS model sequentially: It begins with simple univariate AR models for each series, and then adds ingredients one at a time. This results in seven models, each more complex than its predecessor, and where the final model is the full RTS model.

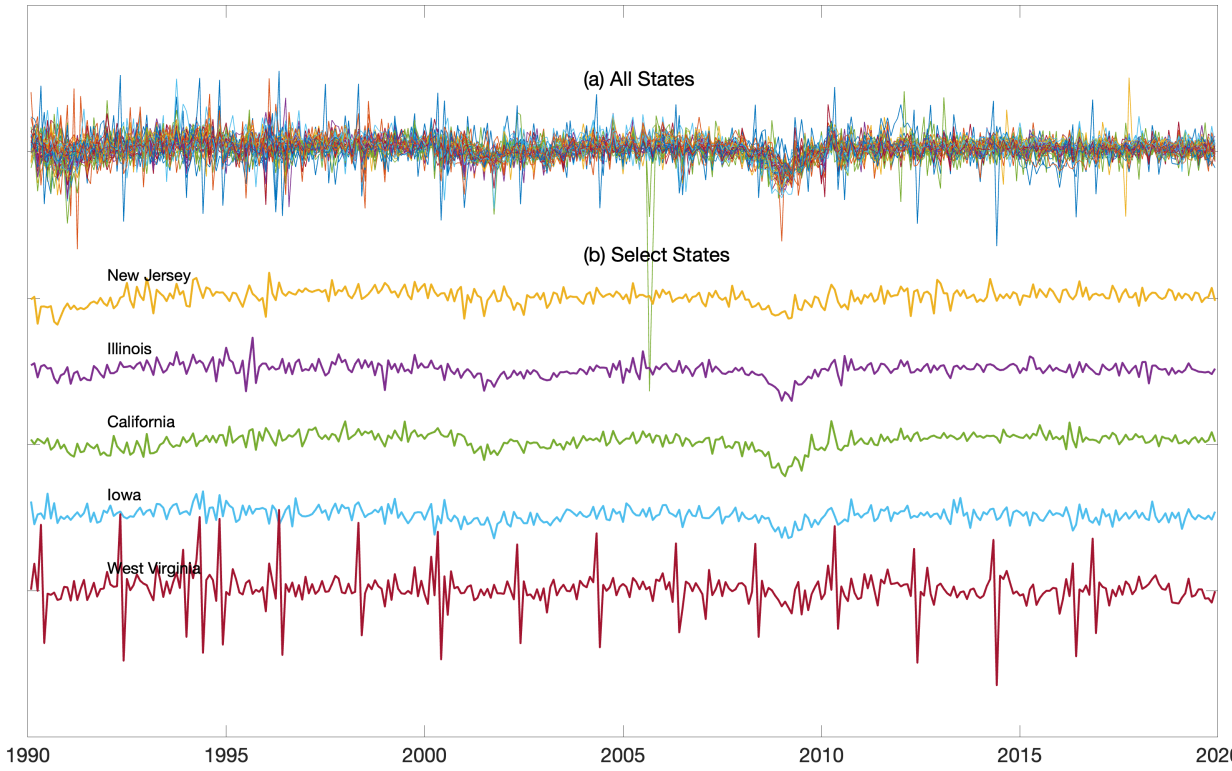
Section 3 also evaluates the fit and forecasting performance of the model using the state employment data. The results are encouraging. Over the pseudo-out-of-sample period from 2000-2019, the root mean square forecasting error from the RTS model is roughly 10 percent lower than a benchmark AR(12) model,[1] these gains are widespread across the states, and similar gains are obtained for quantile and interval forecasts.

Section 4 summarizes additional empirical exercises that serve as external validity checks for the usefulness of the RTS model. The structure of these exercises is simple. We introduce two new datasets—growth rates of industrial production from 16 Euro-area countries and inflation from 17 sectors making up personal consumption expenditures in the United States—and use the RTS model developed in Section 3, without any modification in the model or prior, in pseudo-out-of-sample forecasting experiments in these datasets. We also examine the forecasting performance of the model during the aftermath of the COVID recession. In each of these exercises, the RTS model performs well.

Section 5 provides concluding remarks, but much of the important work in the paper comes after these remarks: Appendix B provides a detailed description of the algorithms used to estimate the model. This discussion parallels the sequential development of the RTS model in Section 3 and highlights, in modular fashion, the features of the algorithm needed to accommodate the various ingredients in the RTS model. Interested users can mix and match these modules when building models that incorporate these ingredients.

[1] A 10% decrease of root-MSFE corresponds to having access to a variable whose correlation with the baseline forecast error is equal to 0.44 ($= \sqrt{1 - 0.9^2}$).

Figure 1: Monthly Growth Rates for U.S. States, 1990-2019



Notes: Panel (a) plots the monthly growth rates for employment in each of the 50 U.S. states and the District of Columbia. Panel (b) shows the growth rates for five selected states.

2 An initial look at the U.S. States Employment data and a benchmark model

2.1 Some Descriptive Statistics

Figure 1 plots seasonally adjusted monthly rates of growth for employment in $n = 51$ U.S. states and the District of Columbia (hereafter "states"). The sample starts in February 1990 and is plotted through the end of 2019, where for much of our analysis we truncate the sample in 2019m12 to avoid the COVID-19 pandemic and its aftermath. (Section 4.2 presents selected results for the post 2019 period.) Panel (a) plots all 51 series and panel (b) plots the growth rates for a handful of states.[2]

The figure highlights five features of the data that play a role in the RTS forecasting

[2]Appendix A includes sources and descriptions for the data used in this paper.

Table 1: Selected Statistics for U.S. States Employment Growth Rates

	Quantile				
	0.10	0.25	0.50	0.75	0.90
(a) AR(12) Model					
Sum of AR Coefficients	0.49	0.60	0.69	0.73	0.81
Innovation Standard Deviation	2.18	2.42	2.89	3.42	4.44
(b) GARCH(1,1) Student-t Model					
Sum of GARCH Coefficient	0.25	0.47	0.73	0.93	0.98
Student-t degrees of freedom	4.2	5.4	6.3	7.7	13.8
(c) p-values for Nyblom test statistics					
Level	< 0.01	0.04	0.30	0.51	0.67
Sum of AR Coefficients	0.02	0.23	0.56	0.78	0.90

Notes: The columns show the quantiles of the distribution of row-statistics across the 51 states.

model. We discuss these in turn.

First, the growth rates are correlated across states. For example, the figure shows that the 2007-2009 recession led to synchronized declines in employment across the U.S. states, and the 1990-1991 and 2001 recessions also led to widespread, but less severe, declines. The average pairwise correlation of year-over-year growth rates was 0.33 over the 1990-2019 sample period.

Second, the series have similar second moment properties. For example, consider the following state-specific univariate AR(12) models:

$$\phi_i(L)(y_{i,t} - \mu_i) = \varepsilon_{i,t}. \tag{1}$$

Panel (a) of Table 1 shows the distribution (across states) of the sum of the OLS-estimated AR coefficients, $\hat{\phi}_i(1)$, and the residual standard error, $\hat{\sigma}_{\varepsilon_i}$. They indicate similar patterns of long-run persistence and variability for many of the states. An examination of the scale-normalized spectra $\hat{S}_i(\omega) = |\hat{\phi}_i(e^{i\omega})|^2$ also indicates a similar pattern of persistence across all frequencies (where these spectra are not shown to conserve space).

Third, there are obvious outliers in the employment data. The largest outlier in Figure 1 is associated with Hurricane Katrina in August 2005 which caused large employment losses in Louisiana. Several outliers are also evident in the data for West Virginia plotted in panel (b), where many of these employment changes are associated with mining employment.

Fourth, time-varying volatility in growth rates is also evident. One simple way to gauge the

prevalence of volatility shifts is by examining estimated parameters from GARCH(1,1) models for the innovations in the AR(12) models in (1). The distribution (across states) for the sum of these GARCH coefficients is shown in panel (b) of Table 1, and indicates persistent changes in volatility for most states. The estimated GARCH models include Student-t innovations, and the table also shows the distribution of the estimated degrees of freedom; these are small, capturing the outliers and kurtosis in the innovations.

Fifth, but more difficult to see in the figure, is time variation in the levels of the growth rates—that is, time variation of μ_i in (1). Also of interest is potential time variation in the AR coefficients in $\phi_i(L)$. Panel (c) of Table 1 shows the distribution of the p-value of Nyblom (1989) tests for the null of constancy versus the alternative of random walk variation in these coefficients. The p-values suggest time variation in the level parameter for many of the states, but provide little evidence of time variation in the AR coefficients.

The RTS forecasting model, developed in Section 3, is designed to capture these five features of the data. Before presenting the model, we outline the pseudo-out-of-sample (POOS) forecasting experiment used to evaluate its performance.

2.2 Recursive pseudo-out-of-sample forecasting and criteria for forecast evaluation

2.2.1 Description of experiment

The AR(12) model in (1) will serve as a convenient benchmark for the forecasts, and we use that model to describe the recursive pseudo-out-of-sample (POOS) forecasting experiment. The experiment proceeds in the usual way. The full sample ranges from $t = 1, ..., T$ (where in this application $t = 1$ is 1990m2 and T is 2019m12). Let $\text{Emp}_{i,t}$ denote the level of employment in state i in month t, and let $y_{i,t} = 1200 \times \ln(\text{Emp}_{i,t} \,/\, \text{Emp}_{i,t-1})$ denote the monthly growth rates plotted in Figure 1, measured in percentage points at annual rate. Let $y^h_{i,t+h} = (1200/h) \times \ln(\text{Emp}_{i,t+h} \,/\, \text{Emp}_{i,t})$ denote the growth rate from t to $t+h$, again measured in annual percentage points. The goal is to forecast the values of $y^h_{i,t+h}$ for $h = 1, 3, 6$ months ahead. Forecasts are computed recursively: Using data from $t = 1, ..., T^*$ with $T^* = T_1$, the model parameters are estimated and forecasts (point, quantile and interval) are computed for y^h_{i,T^*+h}. This process is repeated for $T^* = T_1 + 1, T_1 + 2, ..., T - h$. The forecasting experiment is carried out using the data plotted in Figure 1, which are the currently available historical

data and not "real time" data; state level employment data undergo significant revisions and our analysis abstracts from these revisions. In the application the POOS experiment begins in T_1 =1999m12, so that the first POOS forecast is based on ten years of in-sample data.

2.2.2 Forecast Loss Function

We are interested in point estimates, predictive quantiles and interval forecasts. Let y denote the random variable of interest, $\hat{y}$ denote the point forecast, q_α denote the αth quantile of the predictive distribution of y with $\hat{q}_\alpha$ its estimate, $(q_{\alpha/2}, q_{(1-\alpha/2)})$ denote the equal-tailed $1-\alpha$ prediction interval and $(\hat{q}_{\alpha/2}, \hat{q}_{(1-\alpha/2)})$ denote its estimate. We evaluate the forecasts using standard loss functions:

- Squared error loss for point forecasts: $\ell(y, \hat{y}) = (y - \hat{y})^2$ with resulting mean squared forecast error (MSFE) as the expected loss. We will report root-MSFEs.
- Quantile loss: $\ell(y, \hat{q}_\alpha) = \begin{cases} \alpha(y - \hat{q}_\alpha) \text{ for } y \geq \hat{q}_\alpha \\ (1-\alpha)(\hat{q}_\alpha - y) \text{ for } y < \hat{q}_\alpha \end{cases}$.
- Interval loss:

$$\ell(y, \hat{q}_{\alpha/2}, \hat{q}_{(1-\alpha/2)}) = (\hat{q}_{(1-\alpha/2)} - \hat{q}_{\alpha/2}) + \frac{1}{\alpha}\left((\hat{q}_{\alpha/2} - y) \times \mathbf{1}[y < \hat{q}_{\alpha/2}] + (y - \hat{q}_{(1-\alpha/2)}) \times \mathbf{1}[y > \hat{q}_{(1-\alpha/2)}]\right).$$

As is well-known, the risks associated with these losses are minimized using the mean of the predictive distribution for $\hat{y}$, $\hat{q}_\alpha = q_\alpha$ and $(\hat{q}_{\alpha/2}, \hat{q}_{(1-\alpha/2)}) = (q_{\alpha/2}, q_{(1-\alpha/2)})$.

We compute sample average values of these losses (i.e., sample risk) for each of the forecasting models and forecast horizons.

3 Seven models

This section is the heart of the paper. Here we develop a tractable forecasting model for "related time series" that exhibit the features—similar dynamics, time-varying volatility, comovement, etc.—that are evident in the state employment data. The model that accommodates all of these features turns out to be rather complex and to understand its various ingredients and the associated computational methods for constructing forecasts, we find it useful to begin with the benchmark AR(12) model in (1) and sequentially add the ingredients.

This results in a sequence of seven increasingly rich forecasting models that culminate in the RTS model. Appendix B provides a detailed description of computational algorithms for estimating these models, where the increasing complexity of the algorithms allows us to highlight, in a modular way, the necessary modifications required for each model's new features.

Recall that the AR(12) model (1), estimated by OLS and with Gaussian i.i.d. errors, serves as the benchmark in the POOS forecast comparisons. The first of the seven models we consider is this univariate AR(12) model, but estimated by Bayesian methods using standard "Minnesota" priors. In this model, each of the series is modeled in isolation and any relationships between the series are ignored. The second model is similar to the first, but recognizes that the series are potentially related by having similar parameter values; it exploits this similarity through the use of hierarchical priors. The third model builds on the second, but replaces the Gaussian assumption for the innovations with Student-t distributions with series-specific degrees of freedom, where again hierarchical priors are used to exploit potential similarity of the degrees-of-freedom parameter. This model allows for heavy-tailed innovation outliers that induce dynamic effects on the series through the model's AR dynamics. The fourth model adds additive outliers, also modeled as Student-t distributed random variables, that induce one-off outliers in the series. The fifth model incorporates stochastic volatility, using a low-frequency formulation, that captures both idiosyncratic and common volatility shifts. The sixth model allows for time variation in the level and autoregressive parameters. Each of these models is a version of the univariate AR model so that, conditional on the parameter values, the conditional first moments of the series evolve independently; in these models the series' conditional first moments are "related" only because they (potentially) have similar parameter values, and this relationship is incorporated through the use of hierarchical priors. The final model is the RTS model that additionally includes comovement in the series by incorporating common latent factors; in this model, employment growth in some states might lead other states and improve the forecasts.

We use the following notation to describe the models. We write $x_t \sim RW(x_1, \gamma^2)$ to describe a Gaussian random walk with initial value x_1 and innovation variance γ^2. With x_t a scalar, the vector $(x_t, x_{t+1}, ..., x_{t+k})'$ is denoted by $x_{t:t+k}$. For a random variable x, we denote its mean and variance by m_x and v_x. Hierarchical priors play an important role in the models, and throughout we will use a Gaussian hierarchical distribution of the following form: Let

$\{\theta_j\}_{j=1}^n$ denote n random variables with

$$\theta_j|(m_\theta, v_\theta) \sim iid\mathcal{N}(m_\theta, v_\theta)$$

where m_θ and v_θ are independent with

$$m_\theta \sim \mathcal{N}(m_{m_\theta}, v_{m_\theta}), \qquad \ln(v_\theta) \sim \mathcal{N}(m_{\ln(v_\theta)}, v_{\ln(v_\theta)}).$$

We write this as

$$\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, v_{m_\theta}, m_{\ln(v_\theta)}, v_{\ln(v_\theta)}). \tag{2}$$

The six precursor models and the final RTS model are presented in the following seven sub-sections. These sections also evaluate the forecasting performance of the models using the POOS forecasting experiment and the state employment dataset. The change in forecast accuracy associated with each model provides a measure of the importance of that model's new feature, albeit in a way that depends on the order in which the features are added to the benchmark model. Ultimately, the forecasting performance of the final RTS model, which includes all of the features, is the object of interest. Section 3.8 complements these forecast comparisons with Bayes factors to evaluate the marginal importance of various features incorporated in the RTS model.

To facilitate the discussion of each model's forecasting performance, Tables 2 and 3 and Figure 2 summarize the results of the various POOS forecasting experiments. As described above, in these experiments, data from 1990m2 ($t = 1$) through $t = T^*$ is used to construct forecasts for time periods $T^* + h$, where in this exercise, T^* ranges from 1999m12 through 2019m6 and $h = 1, 3$, and 6. Table 2 shows the root mean square forecast error (root-MSFE) for each of the methods, where forecast errors results are pooled across the 51 states. The first panel of this table shows the root-MSFE for the benchmark AR(12) model and the values in panel (b) show the values of the root-MSFE for each of the methods relative to the benchmark model.[3] Figure 2 uses Box plots to summarize the distribution of these relative

[3]We also computed a version of Table 2 that attenuated the effect of outliers evident in Figure 1. Specifically, when computing the forecast errors used for the root mean square forecast error we replaced realizations of $y_{i,t}$ that deviated from the sample median by more than four times the interquartile range with the local median of y_{i,t^*} for $t^* \in (t-3, t+3)$. The resulting relative root-MSFEs for Models I-VII were similar to, but somewhat smaller than the values shown in Table 2. For example, the values for the RTS model (Model VII) were (0.90, 0.88, 0.89) for $h = 1, 3, 6$.

root-MSFEs across the 51 states. Table 3 shows the sample risk values for the predictive quantiles and the 80 percent (90-10) equal-tail prediction intervals, where the values are relative to the benchmark AR(12) model and are pooled across states. We discuss these results in the following subsections, but for now we simply note that the results show progressively more accurate point, quantile and interval forecasts as features are added to the model. In particular, the final RTS model yields a root-MSFE that is 8-11% lower than the benchmark model and shows similar gains for the quantile and interval forecasts.

Table 2: Root Mean Square Forecast Errors (Pooled Across States)

Forecasting Model	**Forecast of Employment Growth Rate from *T* to *T+h***		
	h = 1	**h = 3**	**h = 6**
	(a) Root MSFE (percentage points at an annual rate)		
Benchmark AR(12)	3.3	2.1	1.8
	(a) Relative Root MSFE		
Benchmark AR(12)	1.00	1.00	1.00
I: Bayes Shrinkage	0.98	1.00	1.02
II: Hier. Priors	0.96	0.95	0.96
III: Innov. Outliers	0.96	0.95	0.96
IV: Add. Outliers	0.96	0.95	0.96
V: Stoch. Vol.	0.96	0.95	0.95
VI: TVP	0.95	0.94	0.96
VII: RTS Model	0.92	0.89	0.89

Notes: Pseudo-out-of-sample forecasting root mean square forecast errors for the benchmark AR(12) and seven forecasting models discussed in Sections 3.1-3.7.

3.1 Model I: Bayesian shrinkage

The first model is the univariate AR(12) that utilizes independent (and standard) priors for the model's parameters. We present the model using different notation than we used in (1) as this facilitates the presentation of Models II-VII. In particular, we write Model I as

$$y_{j,t} = \mu_j + \omega u_{j,t} \tag{3}$$

$$u_{j,t} = \sum_{l=1}^{p} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t} \tag{4}$$

$$\epsilon_{j,t} = \sigma_j \varepsilon_{j,t} \tag{5}$$

$$\varepsilon_{j,t} \sim iid\mathcal{N}(0,1) \tag{6}$$

Table 3: Relative Values of Sample Quantile and Interval Risk (Pooled Across States)

Model	Quantile							90-10
	0.05	0.10	0.25	0.75	0.90	0.95		Interval
(a) h = 1								
Benchmark	1.00	1.00	1.00	1.00	1.00	1.00		1.00
I: Bayes Shrinkage	1.00	1.00	0.99	0.99	1.00	0.99		1.00
II: Hier. Priors	0.98	0.98	0.97	0.97	0.98	0.98		0.98
III: Innov. Outliers	0.97	0.96	0.95	0.93	0.94	0.96		0.95
IV: Add. Outliers	0.97	0.96	0.95	0.93	0.94	0.96		0.95
V: Stoch. Vol.	0.96	0.96	0.95	0.92	0.91	0.92		0.93
VI: TVP	0.95	0.95	0.94	0.92	0.91	0.91		0.93
VII: RTS Model	0.92	0.92	0.92	0.89	0.89	0.89		0.90
(b) h = 3								
Benchmark	1.00	1.00	1.00	1.00	1.00	1.00		1.00
I: Bayes Shrinkage	1.04	1.04	1.03	1.00	1.00	0.99		1.02
II: Hier. Priors	0.94	0.95	0.96	0.97	0.97	0.97		0.96
III: Innov. Outliers	0.92	0.94	0.95	0.95	0.96	0.97		0.95
IV: Add. Outliers	0.92	0.94	0.95	0.95	0.96	0.97		0.95
V: Stoch. Vol.	0.92	0.93	0.94	0.93	0.91	0.91		0.92
VI: TVP	0.88	0.9	0.94	0.95	0.93	0.93		0.92
VII: RTS Model	0.83	0.85	0.88	0.91	0.89	0.87		0.87
(c) h = 6								
Benchmark	1.00	1.00	1.00	1.00	1.00	1.00		1.00
I: Bayes Shrinkage	1.08	1.08	1.06	0.97	0.96	0.94		1.03
II: Hier. Priors	0.90	0.94	0.96	0.97	0.96	0.94		0.95
III: Innov. Outliers	0.88	0.92	0.95	0.97	0.97	0.96		0.94
IV: Add. Outliers	0.88	0.92	0.95	0.97	0.97	0.96		0.94
V: Stoch. Vol.	0.90	0.92	0.94	0.93	0.9	0.88		0.91
VI: TVP	0.83	0.88	0.94	0.99	0.99	0.98		0.92
VII: RTS Model	0.79	0.83	0.88	0.92	0.89	0.87		0.86

Notes: Pseudo-out-of-sample sample average losses relative to the benchmark AR(12) for the seven forecasting models discussed in Sections 3.1-3.7.

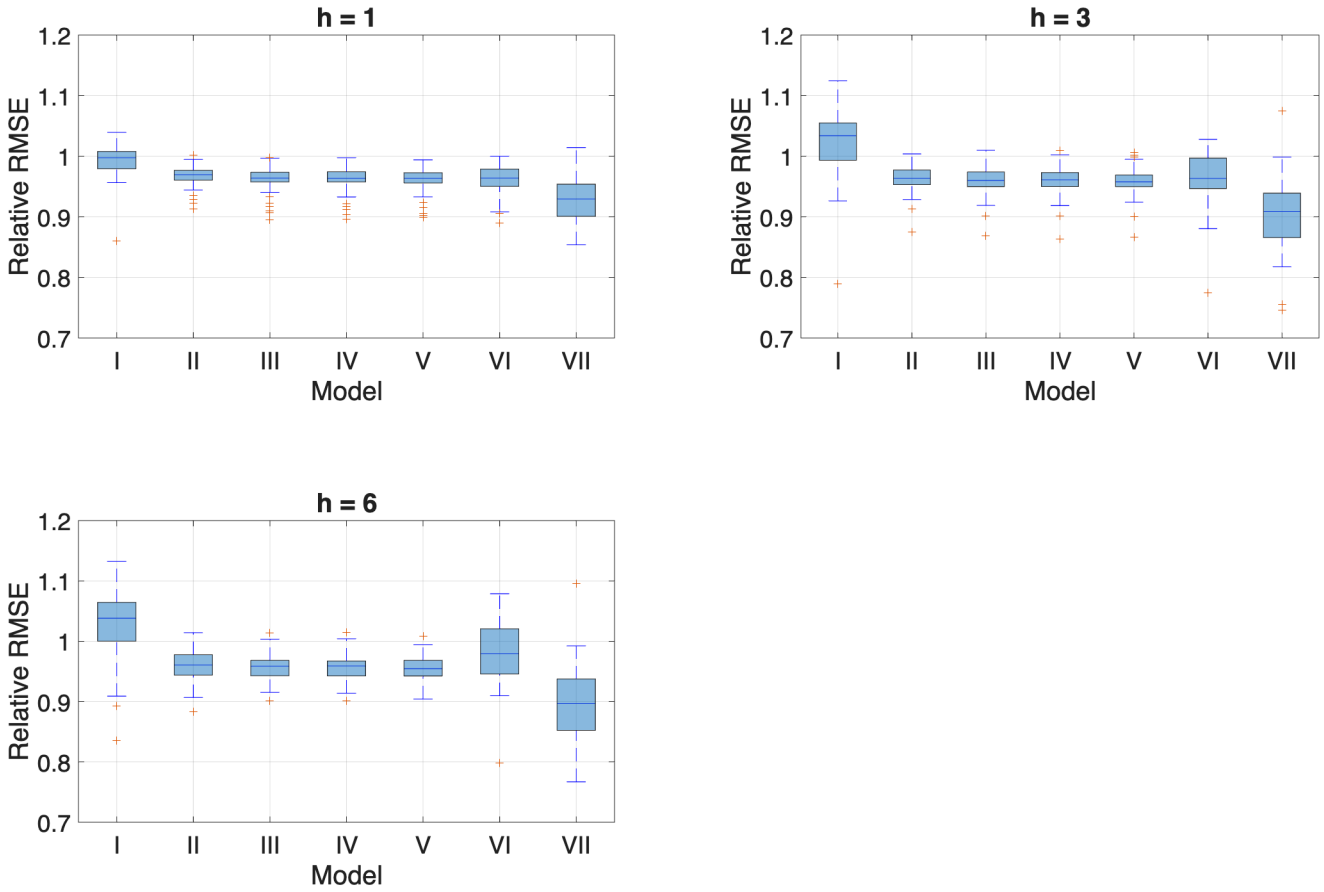
where the number of lags in the AR model (4) is $p = 12$. Note that the model introduces a common scale parameter ω, so the parameters σ_j capture the relative volatilities of the series. The level of each series is given by μ_j.

The model is estimated using the following priors:

- To induce scale and location equivariance, the priors for ω and $\{\mu_j\}_{j=1}^n$ are diffuse:

$$\ln(\omega^2) \sim \mathcal{N}(0, \infty) \tag{7}$$

Figure 2: Relative Root Mean Forecast ErrorDistribution Across States



Notes: Box plots for the relative root-MSFEs for the 51 states. The models are discussed in Sections 3.1-3.7, where Model VII is the RTS model.

and

$$\mu_j \sim \mathcal{N}(0, \infty). \tag{8}$$

- The model requires values for the initial values $\{u_{j,t}\}_{t=-11}^{0}$, and we assume that these are draws from a stationary AR model. Specifically, with $\phi_j = \{\phi_{j,l}\}_{l=1}^{12}$, the prior for $\{u_{j,t}\}_{t=-11}^{0}$ is

$$u_{j,-11:0}|(\sigma_j, \phi_j) \sim \mathcal{N}(0, \sigma_j^2 \Sigma(\phi_j)) \text{ with } \Sigma(\phi) = \Sigma_{AR}(c\phi) \tag{9}$$

where $\Sigma_{AR}(\phi)$ is the $p \times p$ covariance matrix of the stationary AR(p) model with AR

coefficient vector ϕ and unit innovation variance, and where the constant $c \leq 1$ in (9) is chosen so that the largest root of the companion matrix is no larger than 0.98. This prior allows the distribution of the initial conditions to depend on the AR parameters (σ_j, ϕ_j), but in a way that bounds the variance when the AR parameters yield explosive dynamics.

- We use Minnesota-like priors for AR coefficients:

$$\phi_{j,l} \sim \mathcal{N}(0, (0.2/l)^2). \tag{10}$$

  The value of 0.2 means the values are strongly shrunk toward zero, increasingly so for longer lags.

- The prior for σ_j is

$$\ln(\sigma_j^2) \sim \mathcal{N}(0, 0.3^2). \tag{11}$$

  Recall that ω captures the overall scale of the data, with σ_j parameterizing the relative variability for the jth state. This prior yields a median of 1.0 for σ_j and places approximately 90 percent of its mass on values of σ_j that are between 0.8 and 1.3.

In summary, Model I is given by equations (3)-(6), with priors given by (7)-(11).

Appendix B contains a detailed description of the algorithm used to obtain draws from the posterior distribution of the model's parameters. The algorithm contains familiar steps, although these are modified in important ways for the inclusion of the common scale ω and the initial conditions $\{u_{j,t}\}_{t=-11}^{0}$. Draws from the predictive distribution use these draws, augmented with draws of $\varepsilon_{j,t}$ for $t = T^* + 1, ..., T^* + h$ using (6), to obtain draws of y_{j,T^*+h}^h.

The pseudo-out-of-sample forecasting results reported in Tables 2 and 3 and Figure 2 suggest this model's forecast accuracy is roughly on par with the benchmark. Perhaps unsurprisingly, Bayesian forecasts using these priors do not improve on simple OLS-based forecasts since this model ignores any relationship between the time series.

3.2 Model II: Hierarchical priors

Model II replaces Model I's priors for the AR coefficients and innovation variances with hierarchical priors. These priors exploit similarities in parameter values across the series,

potentially sharpening the accuracy of the parameter estimates and forecasts. Specifically, Model II uses the following prior for ϕ:

$$\{\phi_{j,l}\}_{j=1}^{n} \sim \mathcal{HN}(0, 0.5^2(0.2/l)^2, \ln((0.2/l)^2), 0.5^2). \tag{12}$$

Unpacking the prior using the notation in (2): (12) says that $\phi_{j,l}|(m_{\phi_l}, v_{\phi_l}) \sim iid\mathcal{N}(m_{\phi_l}, v_{\phi_l})$, but with $m_{\phi_l} \sim \mathcal{N}(0, 0.5^2(0.2/l)^2)$ and $\ln(v_{\phi_l}) \sim \mathcal{N}(\ln((0.2/l)^2), 0.5^2)$. The resulting estimates of $\phi_{j,l}$ are shrunk towards m_{ϕ_l} with a strength governed by v_{ϕ_l}, where the data help inform appropriate values of m_{ϕ_l} and v_{ϕ_l}.

A hierarchical prior is also used for σ_j:

$$\{\ln(\sigma_j^2)\}_{j=1}^{n} \sim \mathcal{HN}(0, 0.5^2, \ln(0.3^2), 0.5^2) \tag{13}$$

so again, information from all of the series is used to inform the posterior for σ_j.

The estimation algorithm presented in Appendix B provides a computationally efficient method for incorporating these hierarchical priors into the analysis.

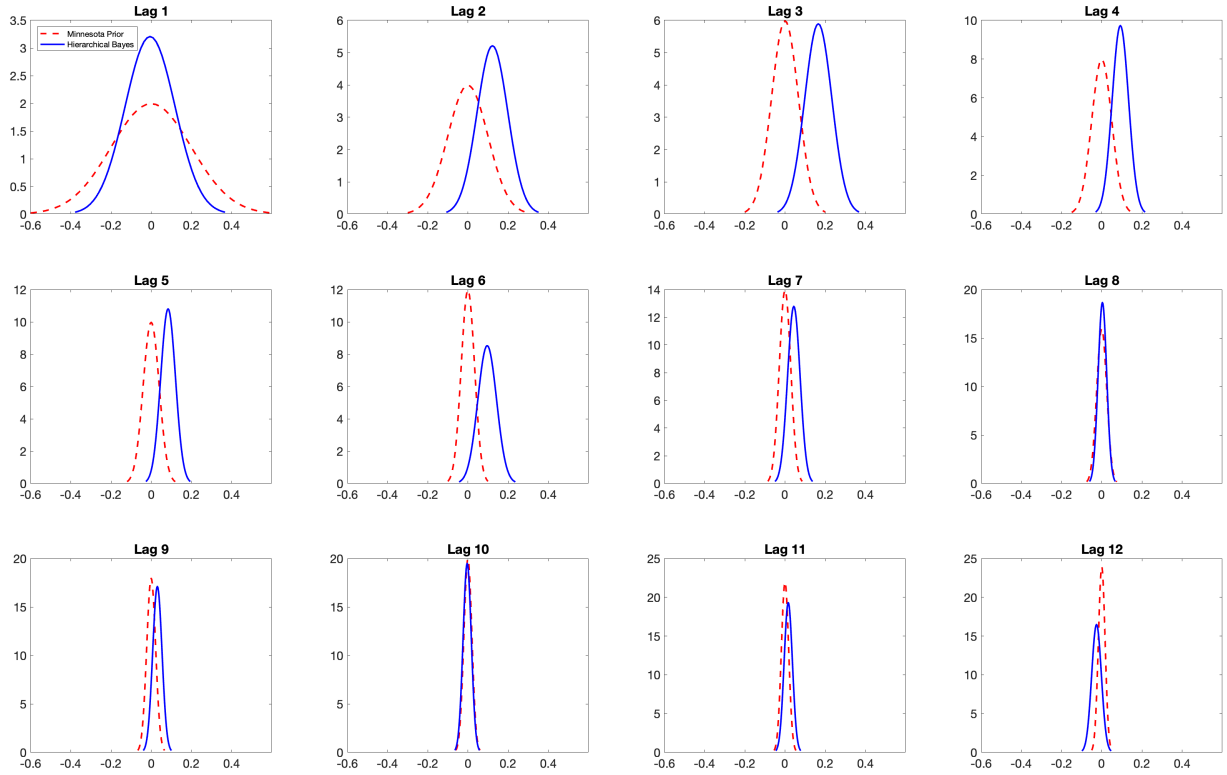
Tables 2 and 3 and Figure 2 indicate that Model II provides markedly more accurate forecasts than the benchmark AR(12) and its Bayesian implementation in Model I. The root-MSFEs are 4-6 percent smaller than the benchmark across the different horizons (Table 2), the improvement is widespread across the states (Figure 2), and quantile and interval forecasts are improved (Table 3).

Figure 3 provides some insight into this gain. It plots the prior for the AR coefficients from Model I together with the "prior" in Model II, $\phi_{j,l}|(m_{\phi_l}, v_{\phi_l}) \sim iid\mathcal{N}(m_{\phi_l}, v_{\phi_l})$, evaluated at the posterior mean of m_{ϕ_l} and $\ln v_{\phi_l}$ using data through 2019m12. While the prior for Model I shrinks the AR coefficients towards zero (the prior mean), Model II's hierarchical prior shrinks the coefficients toward non-zero values, particularly for lags two through six. This data-dependent shrinkage results in substantially better forecasts for state employment.

3.3 Model III: Student-t innovations

In Models I and II, the innovations $\varepsilon_{j,t}$ are normally distributed, which is at odds with the outliers evident in Figure 1. In Models III and IV, Student-t errors are introduced to describe

Figure 3: Model I's Prior and Model II's Estimated 'Prior'



Notes: For Model I these are the priors in (10). For Model II, this is the normal prior $\phi_{j,l}|(m_{\phi_l}, v_{\phi_l}) \sim iid\mathcal{N}(m_{\phi_l}, v_{\phi_l})$, evaluated at the posterior mean of m_{ϕ_l} and $\ln v_{\phi_l}$ from 2019m12.

these outliers. In particular, in Model III, the Gaussian assumption (6) is replaced with

$$\varepsilon_{j,t} \sim \mathcal{T}(\nu_j) \tag{14}$$

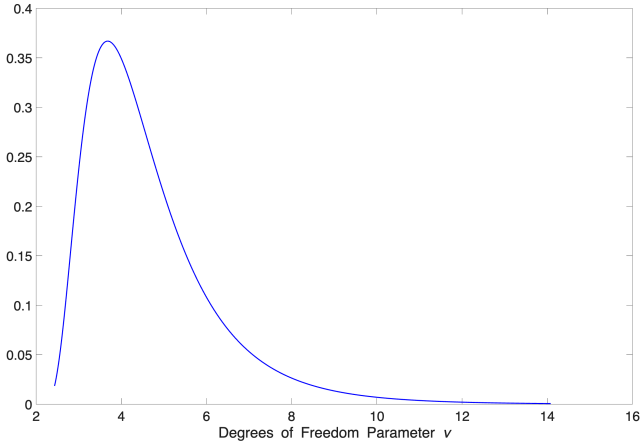
where $\mathrm{T}(\nu_j)$ denotes the Student-t distribution with ν_j degrees of freedom.

A hierarchical prior is used for the series-specific values of ν_j:

$$\{\ln(\nu_j - 2)\}_{j=1}^{n} \sim \mathcal{HN}(\ln(12-2), 0.5^2, \ln(0.5^2), 0.5^2). \tag{15}$$

This prior restricts ν_j to be greater than 2, it has a median of 12 and roughly 95 percent of its mass between 4 and 30. Figure 4 plots the prior evaluated at the (2019m12) posterior mean of the hyperparameters; it implies heavy-tailed innovations with degrees of freedom in the range of 3-7, which are broadly consistent with the values shown previously in Table 1.

Figure 4: Estimated 'Prior' ν in Model III



Note: This is the density of ν from $\ln(\nu_j - 2)|(m_{\ln(\nu-1)}, v_{\ln(\nu-2)}) \sim iid\mathcal{N}(m_{\ln(\nu-1)}, v_{\ln(\nu-2)})$ evaluated at the posterior mean of $m_{\ln(\nu-2)}$ and $v_{\ln(\nu-2)}$ from 2019m12.

The Bayes factors that will be presented in Section 3.8 indicate that these heavy-tailed innovations are important for describing the state employment data. In principle, by more accurately modeling the tails of the innovations, they should yield more accurate estimates of the mean parameters, μ and ϕ, and therefore produce more accurate point forecasts. The root-MSFE results in Table 2 and Figure 2 suggest that these forecasting gains are small. However, by more accurately modeling the tails, they do improve the quantile and interval forecasts (Table 3).

3.4 Model IV: Additive outliers

The heavy-tailed errors in Model III were in the innovations, and therefore produced outliers that propagated forward using the model's autoregressive dynamics. Model IV introduces new errors that are additive and have one-off effects on the variables. These additive outliers are denoted $o_{j,t}$ and enter the model with equation (3) replaced by

$$y_{j,t} = \mu_j + \omega(u_{j,t} + o_{j,t}) \tag{16}$$

where

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(\nu_j^o). \tag{17}$$

The additive outliers are governed by two parameters: κ_j determines their scale, and the degree-of-freedom parameter ν_j^o determines their kurtosis. Hierarchical priors are used for these parameters:

$$\{\ln(\kappa_j^2)\}_{j=1}^n \sim \mathcal{HN}(\ln(0.1^2), 0.5^2, \ln(0.3^2), 0.5^2) \tag{18}$$

$$\{\ln(\nu_j^o - 2)\}_{j=1}^n \sim \mathcal{HN}(\ln(4-2), 0.5^2, \ln(0.5^2), 0.5^2), \tag{19}$$

where the prior medians of κ_j and ν_j^o are 0.1 and 4, respectively.

As we will see from the Bayes factors reported in Section 3.8, these outliers are important for describing the distribution of the data—an unsurprising result given the handful of large outliers seen in Figure 1. That said, Tables 2 and 3 suggest that they have little effect on the forecasts.

3.5 Model V: Time varying volatility

Model V extends Model IV by incorporating stochastic volatility. Specifically, it replaces (5) with

$$\epsilon_{j,t} = \sigma_{j,t}\varepsilon_{j,t} \tag{20}$$

where $\ln(\sigma_{j,t})$ evolves as a random walk.

A standard way of including random walk stochastic volatility relies on methods like those pioneered in Kim, Shephard, and Chib (1998). We have found computational gains from using an alternative method that builds on a low-frequency approximation to the random walk; these gains are particularly important when modeling comovement in the volatility paths across the series. To explain the method, it is useful to take a short digression to describe the spectral representation of a random walk.

Consider a generic random walk $x_t \sim RW(x_1, \gamma^2)$ for $t = 1, ..., T$. The typical representation of the vector $x_{1:T}$ involves the initial value x_1 and the $T-1$ random variables $\Delta x_t \sim iid\mathcal{N}(0, \gamma^2)$ for $t = 2, ..., T$. This representation, along with a numerical approximation for the distribution of $\ln(\varepsilon_{j,t}^2)$, is used in the Kim, Shephard, and Chib (1998) method. We use an alternative representation for $x_{1:T}$ and an alternative approximation. The details are as follows. Let $\overline{x}$ denote the sample mean of x and $\tilde{x}_t = x_t - \overline{x}$ denote its demeaned value. Write the covariance matrix of $\tilde{x}_{1:T}$ as $\gamma^2\bar{\Sigma}_{RW}$, where $\bar{\Sigma}_{RW}$ is the $T \times T$ covariance

matrix of a demeaned random walk with unit innovation variance. Write the spectral decomposition of $\bar{\Sigma}_{RW}$ as $\bar{\Sigma}_{RW} = \sum_{l=1}^{T-1} \varphi_l \varphi_l'$ where $\varphi_l = \lambda_l^{1/2} e_l$ with λ_l the lth eigenvalue of $\bar{\Sigma}_{RW}$ and e_l the corresponding unit-length eigenvector. The vector $\tilde{x}_{1:T}$ can then be represented as $\tilde{x}_{1:T} = \sum_{l=1}^{T-1} \varphi_l \xi_l$ where $\xi_l \sim iid\mathcal{N}(0, \gamma^2)$. Examination of the matrix $\bar{\Sigma}_{RW}$ shows that $e_{l,t} \propto \cos\left(\frac{t-1/2}{T} l\pi\right)$ and, approximately, $\lambda_l^{1/2} \propto l^{-1}$, which implies that most of the variance of $\tilde{x}_t$ is associated with the first few eigenvalues and their associated eigenvectors describe the low-frequency variation in $\tilde{x}_t$. This motivates the approximation $x_t = \overline{x} + \tilde{x}_t \approx \overline{x} + \sum_{l=1}^{q} \varphi_{l,t} \xi_l$, where $\varphi_{l,t}$ is the tth element of φ_l and the approximation truncates the sum using only $q << T-1$ terms. This approximation models the low-frequency variation in the random walk, but ignores high-frequency variation; in particular, it captures periodicities longer than $2T/q$.

With this background, we model the evolution of $\ln(\sigma_{j,t}^2)$ using the low-frequency random walk:

$$\ln(\sigma_{j,t}^2) = \ln(\sigma_j^2) + \sum_{l=1}^{q} \varphi_l \left(\frac{t-1/2}{T}\right) \xi_{j,l} \tag{21}$$

where $q = \lfloor T/36 \rfloor$, and priors

$$\{\xi_{j,l}\}_{j=1}^{n} | (m_{\xi_l}, v_\xi) \sim iid\mathcal{N}(m_{\xi_l}, v_\xi) \tag{22}$$

with

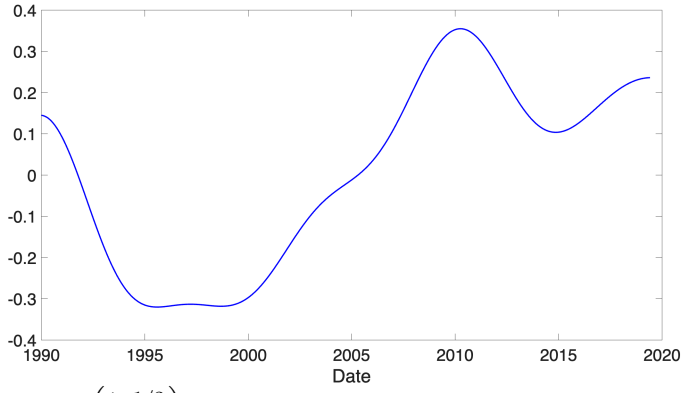
$$m_{\xi_l} \sim \mathcal{N}(0, 0.01^2) \tag{23}$$

and

$$\ln(v_\xi) \sim \mathcal{N}(\ln(0.01^2), 0.5^2). \tag{24}$$

We highlight four features of this specification. First, with $m_{\xi_l} = 0$ and $v_\xi = \gamma^2$, (21) is the representation for a random walk that was developed above, truncated after q terms. Second, the choice of q means that (21) captures variation in $\ln(\sigma_{j,t}^2)$ for periods longer than 72 months. Higher-frequency stochastic volatility is not modeled through variation in $\sigma_{j,t}$ (although it is captured in part through the Student-t errors, which can be represented as scale mixtures of normal random variables). Third, the hierarchical prior allows m_{ξ_l} to vary across l in response to the data, capturing comovement in the low-frequency volatility patterns in the series. Figure 5 gives a sense of this common evolution of volatility across the 51 states

Figure 5: Estimated Prior Mean for the log-volatility



Notes: The figure plots $\sum_{l=1}^{q} \varphi_l \left(\frac{t-1/2}{T}\right) \hat{m}_{\xi_l}$, where $\hat{m}_{\xi_l}$ is the posterior mean of m_{ξ_l} from 2019m12.

by plotting $\sum_{l=1}^{q} \varphi_l \left(\frac{t-1/2}{T}\right) \hat{m}_{\xi_l}$, where $\hat{m}_{\xi_l}$ is the posterior mean of m_{ξ_l} from 2019m12; it shows a period of relative calm in the late 1990s followed by a marked increase in volatility during the 2007-2009 recession and its aftermath. The fourth feature concerns the role of v_ξ in (21). Notice that, while the mean, m_{ξ_l}, depends on l, the variance, v_ξ, does not. Thus v_ξ serves as an overall scale for the variation in $\ln(\sigma_{j,t})$, and the posterior for v_ξ summarizes the information in the sample about this scale.

The time variation in $\ln(\sigma_{j,t})$ introduces an additional step for constructing the predictive distribution of y^h_{j,T^*+h}, as these are affected by the uncertainty in out-of-sample values of $\sigma_{j,t}$. Consistent with the random walk model, we set $\ln(\sigma_{j,T^*+t}) \sim RW(0, v_\xi)$ for independent random walks across j.

The accuracy of the resulting forecasts is summarized in Tables 2 and 3 and Figure 2. Unsurprisingly, the incorporation of stochastic volatility has little effect on the accuracy of point forecasts (see Table 2 and Figure 2), but significantly improves the accuracy of the predictive quantiles and intervals (Table 3).

3.6 Model VI: Time varying conditional mean parameters

The descriptive statistics reported in Section 2 suggest that the levels of the growth rates, μ_j in (16), varied over the sample period in some of the states. There was less evidence of time variation in the autoregressive coefficients, $\phi_{j,l}$. In Model VI we allow these coefficients to evolve as random walks, where the amount of time variation is estimated using hierarchical

priors.

In particular, Model VI replaces (16) and (4) with

$$y_{j,t} = \mu_{j,t} + \omega(u_{j,t} + o_{j,t}) \tag{25}$$

and

$$u_{j,t} = \sum_{l=1}^{p} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t} \tag{26}$$

where

$$\mu_{j,t} \sim RW(\mu_j, \omega^2 \gamma^2_{\mu_j}) \tag{27}$$

$$\phi_{j,l,t} \sim RW(\phi_{j,l}, \gamma^2_{\phi_{j,l}}). \tag{28}$$

The key parameters governing the evolution of these coefficients are the standard deviations γ_{μ_j} and $\gamma_{\phi_{j,l}}$, which are estimated by pooling information across series using hierarchical priors:

$$\{\ln(\gamma^2_{\mu_j})\}_{j=1}^{n} \sim \mathcal{HN}(\ln(0.005^2), 2^2, \ln(0.3^2), 0.5^2) \tag{29}$$

and

$$\{\ln(\gamma^2_{\phi_{j,l}})\}_{j=1}^{n} \sim \mathcal{HN}(\ln((0.005/l)^2), 0.5^2, \ln(0.3^2), 0.5^2). \tag{30}$$

The priors for the initial values of the coefficients, μ_j and $\phi_{j,l}$, are unchanged from the earlier models and are given by (8) and (12). The scaling by ω^2 in (27) ensures overall location and scale equivariance under the flat priors (8) and (7) on $\mu_{j,t}$ and ω.

Interestingly, while the Bayes factors reported below document the importance of this feature for describing the data, exploiting this time variation for forecasting is difficult. Table 2 indicates that the accuracy of the point estimates, pooled across states, are largely unchanged from Model 5, but Figure 2 shows significantly greater variability in the relative accuracy across states. The relative accuracy of quantile forecasts is also variable showing improvement for some quantiles at some horizons, but deterioration at others.

3.7 Common factors and the RTS model

In the models discussed thus far, the series are related through the values of their parameters and, as described in Section 3.5, through dependencies in their stochastic volatility processes.

However, conditional on their parameter values, the series are uncorrelated. The RTS model includes common factors to capture covariability in the series.

The model replaces (25) with

$$y_{j,t} = \mu_{j,t} + \mu_{n+1,t} + \omega(u_{j,t} + o_{j,t} + o_{n+1,t} + c_{j,t}) \tag{31}$$

where the new ingredients are the factors $\mu_{n+1,t}$, $o_{n+1,t}$ and $c_{j,t}$. There are no observations for $j = n+1$, so these factors are latent. The variable $\mu_{n+1,t}$ is a common time-varying level that evolves in the same way as the series-specific level from Model VI, that is as the random walk (27) with $j = n+1$. Similarly, the variable $o_{n+1,t}$ is a common additive outlier that evolves in the same way as the series-specific outliers in Model IV, that is as (17) with $j = n+1$.

The variable $c_{j,t}$ is more complicated. It captures potential lead/lag relationships between the series. To explain its evolution, let $u_{n+1,t}$ evolve the same way as the series-specific values of u, that is as the AR(12) model (26) for $j = n+1$, with the autoregressive coefficients allowed to vary in time as in (28). The variable $c_{j,t}$ is a series-specific moving average of $u_{n+1,t}$:

$$c_{j,t} = \sum_{l=0}^{5} \lambda_{j,l,t} u_{n+1,t-l} \tag{32}$$

where the moving average weights, $\lambda_{j,l}$ are allowed to vary through time as

$$\lambda_{j,l,t} \sim RW(\lambda_{j,l}, \gamma^2_{\lambda_{j,l}}). \tag{33}$$

In this model, series that load on the first few lags of $u_{n+1,t}$ serve as leading indicators for series that load on more distant lags, improving the model's forecasts for the lagging series. Moreover, even if all series load only on the contemporaneous value of $u_{n+1,t}$, forecasts are potentially improved because the different series provide independent information about $u_{n+1,t}$, which in turn follows its own dynamics; the forecasts for the different series are thus effectively shrunk towards a common non-zero value.

The new priors for this model are

$$\ln(\sigma^2_{n+1}) \sim \mathcal{N}(\ln(0.2^2), 0.5^2) \tag{34}$$

$$\mu_{n+1} = 0 \text{ (a normalization)}, \quad \ln(\gamma^2_{\mu_{n+1}}) \sim \mathcal{N}(\ln(0.005^2), 0.5^2) \tag{35}$$

$$\ln(\kappa_{n+1}^2) \sim \mathcal{N}(\ln(0.1^2), 0.5^2). \tag{36}$$

Notice that these priors are not tied to the hierarchical priors for the series-specific values of these parameters, so the parameters characterizing the common factors are unrelated to those governing the series-specific factors.

In contrast, the parameters characterizing the factor loadings, $\lambda_{j,l,t}$, are potentially similar across series, so that hierarchical priors are used:

$$\{\lambda_{j,0}\}_{j=1}^n \sim \mathcal{HN}(1, 0, \ln(0.2^2), 0.5^2), \tag{37}$$

$$\{\lambda_{j,l}\}_{j=1}^n \sim \mathcal{HN}(0, 0.5^2(0.05/l)^2, \ln((0.05/l)^2), 0.5^2) \text{ for } l > 0 \tag{38}$$

and

$$\{\ln(\gamma_{\lambda(j,l)}^2)\}_{j=1}^n \sim \mathcal{HN}(\ln((0.0005/(l+1))^2), 0.5^2, \ln(0.3^2), 0.5^2). \tag{39}$$

As we mentioned at the outset of this section, the resulting complete model incorporates autoregressive dynamics, stochastic volatility, innovation and additive outliers, common factors and time-varying parameters. It uses hierarchical priors to pool information across series. The complete RTS model is given by equations (31), (27), (20), (14), (21), (17), (26)-(28), (32) and (33), with priors (7)-(9), (12), (13), (15), (18), (19), (22)-(24), (29), (30), and (34)-(39).

Tables 2 and 3 and Figure 2 show the forecasting gains from the complete RTS model: the relative root-MSFEs show a 10 percent gain over the benchmark AR(12) model (Table 2), these gains are evident in nearly every state (Figure 2) and similar gains are evident in the quantile and interval forecasts (Table 3). Interestingly, additional calculations indicate that nearly all of these gains can be realized in a model in which the c-factors load only on contemporaneous values of $u_{n+1,t}$ in (32).[4]

[4]We have also compared the forecasting performance of the RTS model to a dynamic version of the common correlated effects model (Pesaran (2006)) that augments the benchmark AR(12) model with 6 lags of the mean employment growth rates across the 51 states. The resulting model performed marginally better than the AR(12) model at some horizons—for example, its pooled relative root-MSFEs are (0.99, 0.98 and 0.99) for $h = (1, 3, 6)$—but these were markedly larger than the relative root-MSFE for the RTS model (which are (0.92, 0.89, and 0.89) for $h = (1, 3, 6)$).

3.8 Bayes factors

The POOS forecast comparisons summarized in Tables 2-3 and Figure 2 show the marginal importance of the various features that appear in the RTS model, but these comparisons are lacking in two respects. First, they are based on the somewhat arbitrary sequencing of how the features appear in the models. For example, the relative importance of Student-t innovations (Model III) might well change if they had been added before, rather than after, stochastic volatility (Model V). Second, the forecasting results focus on the particular aspects of accuracy associated with the loss functions from Section 2.2.2, and while these are important, they do not fully capture the overall fit of the model.

In this subsection we investigate the importance of the model's features using an alternative approach: We begin with the RTS model with all features present, and compare the fit of that model to models where selected features are down-weighted. We use Bayes factors to gauge the fit of the RTS model relative to eight alternative models.

The first alternative model investigates the importance of the hierarchical priors that appear throughout the RTS model. Recall that the hierarchical normal prior assumes that a set of parameters, $\{\theta_j\}_{j=1}^n$, are characterized by the conditional normal prior $\theta_j|(m_\theta, v_\theta) \sim iid\mathcal{N}(m_\theta, v_\theta)$, where m_θ and v_θ are hyperparameters with their own prior distributions, $m_\theta \sim \mathcal{N}(m_{m_\theta}, v_{m_\theta})$ and $\ln(v_\theta) \sim \mathcal{N}(m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$; we have denoted the resulting hierarchical prior as $\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, v_{m_\theta}, m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$. When m_θ and v_θ have degenerate distributions, that is when $v_{m_\theta} = v_{\ln(v_\theta)} = 0$, the prior for θ_j collapses to the non-hierarchical normal prior $\theta_j \sim iid\mathcal{N}(\mu_\theta, \sigma_\theta^2)$ with $\mu_\theta = m_{m_\theta}$ and $\ln(\sigma_\theta^2) = m_{\ln(v_\theta)}$. (Versions of this non-hierarchical prior were used in Model I.) To gauge the importance of the hierarchical priors, it is instructive to compare the fit of the RTS model to an alternative model that uses non-hierarchical priors—that is to a model in which the hierarchical variance parameters v_{m_θ} and $v_{\ln(v_\theta)}$ are set to zero across all hierarchical priors. This can be achieved by computing the Bayes factor for the RTS model, with $\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, v_{m_\theta}, m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$ relative to the model using the alternative priors $\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, 0, m_{\ln(v_\theta)}, 0)$.

Unfortunately, computing Bayes factors is difficult, and we were unsuccessful in computing an accurate estimate of the Bayes factor for the RTS model versus the model with $\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, 0, m_{\ln(v_\theta)}, 0)$. However, we did succeed at a less ambitious task: We accurately estimated the Bayes factor for a less extreme alternative that sets the variance

Table 4: Log-Bayes Factors: RTS Model versus Models with Alternative Priors

Num.	RTS Model Prior	Alternative Prior	Log Bayes-Factor
(1)	$\{\theta_j\}_{j=1}^n \sim$ $\mathcal{HN}(m_{m_\theta}, v_{m_\theta}, m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$	$\mathcal{HN}(m_{m_\theta}, 0.5 \times v_{m_\theta}, m_{\ln(v_\theta)}, 0.5 \times v_{\ln(v_\theta)})$	86.1
(2)	$\{\ln(\nu_j - 2)\}_{j=1}^n \sim$ $\mathcal{HN}(\ln(10), 0.5^2, \ln(0.5^2), 0.5^2)$	$\mathcal{HN}(\ln(10) + 1.5, 0.5^2, \ln(0.5^2), 0.5^2)$	13.8
(3)	$\{\ln(\kappa_j^2)\}_{j=1}^n \sim$ $\mathcal{HN}(\ln(0.1^2), 0.5^2, \ln(0.3^2), 0.5^2)$	$\mathcal{HN}(\ln(0.1^2) - 1.5, 0.5^2, \ln(0.3^2), 0.5^2)$	10.9
(4)	$m_{\xi_l} \sim \mathcal{N}(0, 0.01^2)$ $\ln(v_\xi) \sim \mathcal{N}(\ln(0.01^2), 0.5^2)$	$\mathcal{N}(0, 0.5 \times 0.01^2)$ $\mathcal{N}(\ln(0.01^2) - 1.5, 0.5^2)$	30.6
(5)	$\{\ln(\gamma_{\mu_j}^2)\}_{j=1}^n \sim$ $\mathcal{HN}(\ln(0.005^2), 2^2, \ln(0.3^2), 0.5^2)$	$\mathcal{HN}(\ln(0.005^2) - 6, 2^2, \ln(0.3^2), 0.5^2)$	17.8
(6)	$\{\ln(\gamma_{\phi_{j,l}}^2\}_{j=1}^n \sim$ $\mathcal{HN}(\ln((0.005/l)^2), 0.5^2, \ln(0.3^2), 0.5^2)$	$\mathcal{HN}(\ln((0.005/l)^2) - 1.5, 0.5^2, \ln(0.3^2), 0.5^2)$	5.8
(7)	$\{\ln(\gamma_{\lambda(j,l)}^2)\}_{j=1}^n \sim$ $\mathcal{HN}(\ln((0.0005/(l+1))^2), 0.5^2, \ln(0.3^2), 0.5^2)$	$\mathcal{HN}(\ln((0.0005/(l+1))^2) - 1.5, 0.5^2, \ln(0.3^2), 0.5^2)$	17.9
(8)	$\ln(\sigma_{n+1}^2) \sim$ $\mathcal{N}(\ln(0.2^2), 0.5^2)$	$\mathcal{N}(\ln(0.2^2) - 1.5, 0.5^2)$	3.2

Notes: The table shows the prior used in the baseline RTS model and the prior used in the alternative model. The final column shows the log-Bayes factor of the baseline RTS model relative to the alternative. Bayes factors are computed using the 1990m2-2019m12 sample.

of the hyperparameters equal to one-half of their values in the RTS model. (These Bayes factors were computed using the bridge sampling approach of Meng and Wong (1996) that is described in Appendix B.3.) This less-extreme alternative model attenuates, but does not eliminate, the pooling of information across series for estimating the parameter values. The resulting Bayes factor is shown in the first row of Table 4. The log-Bayes factor exceeds 80 in favor of the RTS model relative to the alternative model. Evidently, the hierarchical priors used in the RTS model more accurately describe the data than the priors in the alternative model.

Table 4 shows results for several other alternative models, where in each case these alternatives involve a change in the prior that downweights a particular feature incorporated in the RTS model.

The second row of the table considers an alternative prior that increases the degrees-of-freedom in the Student-t distribution for the innovations; that is, the prior suggests innovation distributions that are closer to the normal. The degrees of freedom parameter is denoted by

ν, and as indicated in Table 4, the alternative increases the mean of the hierarchical prior for ν. (In the RTS model, the prior median for ν is 12 (see (15)), while in the alternative the prior median increases to 47.) The Bayes factor reported in the table strongly favors the lower degrees of freedom in the RTS model.

The third row of the table compares the RTS model to an alternative in which the additive outliers are less important. This is achieved by using a prior with more mass on small values of κ_j, the scale associated with the outliers (see (17)-(18)). The Bayes factor implies that these additive outliers are quite important for describing the state employment data.

The fourth row of the table considers a model with reduced stochastic volatility. This is achieved by changing the prior to reduce the variability of the ξ parameters (see (21)-(24)); which in turn is achieved by reducing the variance of m_ξ and decreasing the mean of $\ln(v_\xi)$. Again, the RTS model is strongly favored over this alternative.

The next three rows focus on time variation in the model parameters. Recall that time variation is modeled using random walk processes, and the magnitude of this time variation is governed by the variance of the associated first differences. This parameter is denoted by γ^2 in the various equations (see (27)-(28) and (33)). We consider three alternative models, all of which are characterized by priors that lower the mean of γ^2. Row five of the table considers the level parameters μ, row six considers the autoregressive coefficients ϕ, and row seven considers the factor loadings λ. In all cases, Bayes factors suggest that time variation is important for describing the data, and this is despite the negligible impact of time variation for μ and ϕ for forecasting performance discussed above for Model VI.

The final row of the table considers an alternative in which the common factors $c_{j,t}$ are downweighted. Recall that the c-factors are series-specific moving averages of a common autoregressive factor, $u_{n+1,t}$. These factors can be down-weighted by reducing the variance of u_{n+1}, that is by reducing the value of σ^2_{n+1} (see (34)). The final row of the table considers this alternative and here too the Bayes factor favors the prior from the RTS model, but by an amount ($e^{3.2} \approx 25$) that is smaller than the other alternatives. One interpretation of this result is that common movements in the series remain well explained by the level and outlier common factors, μ_{n+1} and o_{n+1}.

4 Additional empirical results

The previous sections have used the state employment dataset in the pre-COVID sample period to motivate the structure of the RTS model and to evaluate its relative forecast accuracy. In this section we investigate how the RTS model performs in other applications that involve two new datasets of related series—the growth rates of industrial production (IP) in 16 Euro-area countries and the inflation rates for 17 sectors making up personal consumption expenditures (PCE) in the United States.

4.1 Two new datasets

Figure 6 plots the new datasets. The top panel shows the monthly growth rates of industrial production in 16 Euro-area countries beginning in 1975m8 through 2019m12. The bottom panel shows the monthly growth rate of prices for 17 PCE sectors from 1959m2 to 2025m4. As in the earlier application using state employment data, the Euro-area IP dataset is truncated to only include pre-COVID observations; we consider the COVID recovery period below for the employment and IP datasets. The dynamics of the inflation data over the COVID period are generally consistent with its behavior over the earlier sample period, so we analyze these data over the full sample period.

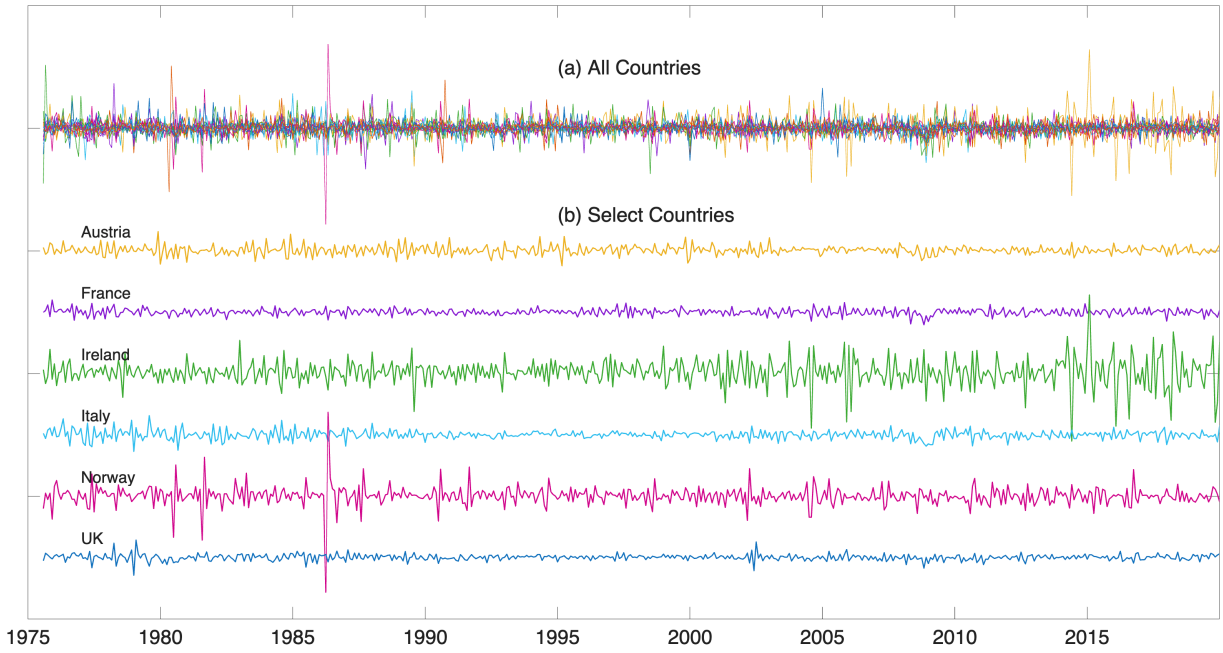
While the details differ, these datasets exhibit many of the same features as state employment. In particular they exhibit comovement, outliers, time-varying volatility, and level shifts. The series appear to follow similar stochastic processes (although inflation in one of the PCE sectors—energy—is much more volatile than the other sectors). A notable difference is that these datasets have many fewer series than the $n = 51$ states in the employment dataset.

In Section 3 we developed the RTS model in seven steps and showed the forecasting performance of each of the models for the state employment data. This was a pedagogical device that allowed us to sequentially describe the various features of the model and (in the appendix) the computational algorithms to accommodate these features. In this section, we avoid these preliminary models and move directly to the RTS model. We conduct the same POOS forecasting experiment used in Section 3, where the POOS forecasts are computed over 1985m6-2019m6 for the IP data and over 1984m12-2024m10 for the inflation data. The results of the experiment are summarized in Figure 7 and Table 5.

Figure 7 shows the distribution of relative root-MSFEs across the entities (states for the

Figure 6: Two Datasets

Industrial Production Growth Rates in 16 Euro Area Countries: 1975m8-2019m12



U.S. PCE Inflation in 17 Sectors: 1959m2-2025m4

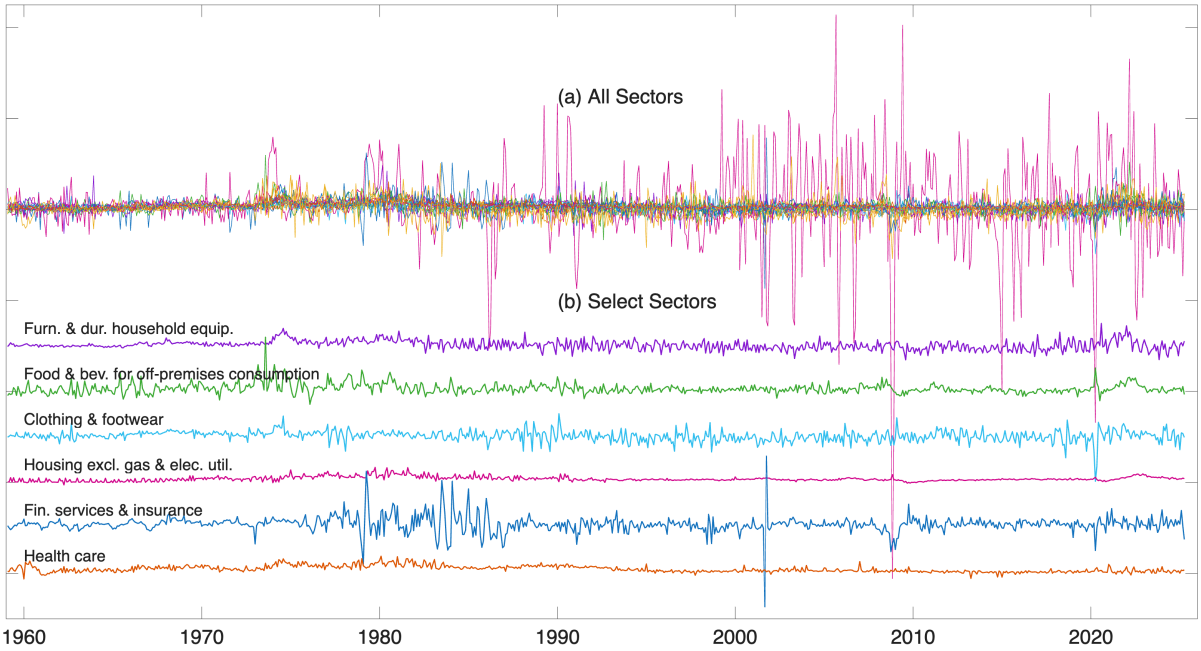
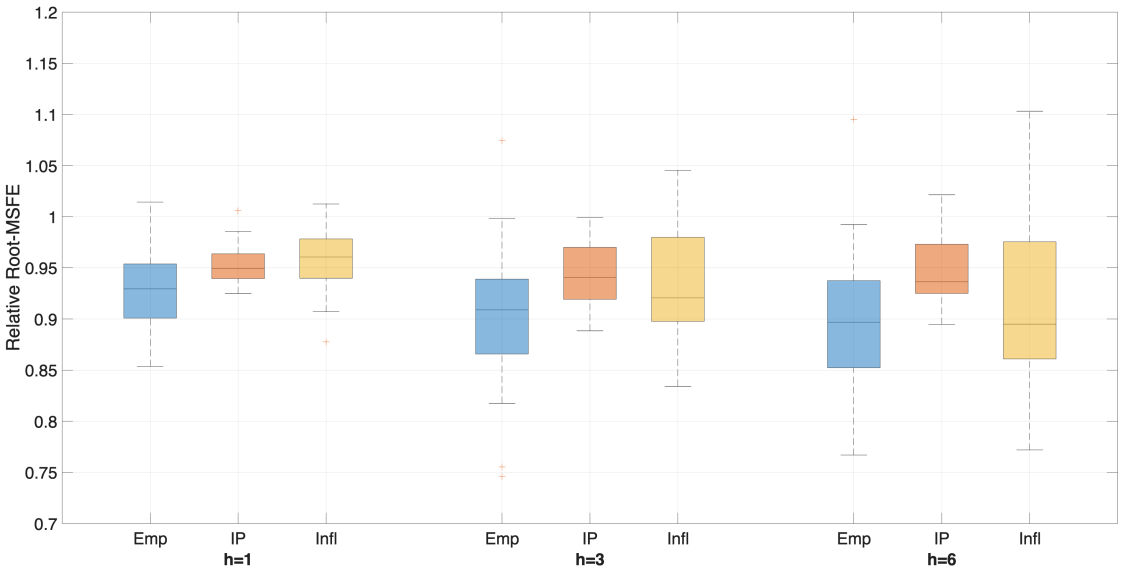


Figure 7: Relative root-MSFE for 3 Datasets: Distribution Across Entities



Notes: Box plots for the relative root-MSFEs across entities the state employment (Emp), industrial production in Euro-area countries (IP), inflation in sectoral PCE (Infl) datasets for the RTS model relative to the benchmark AR(12) model.

employment, countries for Euro-area IP, and consumption sectors for PCE inflation) for the three datasets, and where the results for state employment were shown earlier in Figure 2. Panel (a) of Table 5 shows the relative sample average values of predictive quantile and interval losses for $h = 3$ step ahead forecasts; the values of $h = 1$ and $h = 6$ (not shown) are similar. Examination of these results shows that the RTS model provides forecasting gains for these new datasets that are somewhat smaller than, but broadly consistent with, the results for the state employment data.

Panel (b) of Table 5 compares the log-Bayes factors for the alternative models listed in Table 4 for the three datasets, each computed using the data through 2019m12. Here too, these Bayes factors are similar to those obtained for the state employment dataset, but with two exceptions: additive outliers are less important for the two new datasets, and the c-factor is less important for inflation. Regarding the latter, recall that the model includes three sets of common factors, level shifts (μ), outliers (o) and the serially correlated c-factors. The Bayes factor for alternative model 8 only concerns the importance of the c-factor; one interpretation of its value is that the common inflation trends in the sectors that are visible in Figure 6 are well described by the common level shift factor $\mu_{n+1,t}$.

Table 5: Selected Results for Three Datasets

(a) Relative value of sample quantile and interval loss: $h = 3$

Dataset	Quantile							90-10 Interval
	0.05	0.10	0.25	0.75	0.90	0.95		
State employment	0.83	0.85	0.88	0.91	0.89	0.87		0.87
Euro area Ind. Prod.	0.93	0.93	0.95	0.94	0.92	0.92		0.92
U.S. PCE Inflation	0.89	0.95	0.96	0.91	0.88	0.86		0.91

(b) Log-Bayes Factors

Dataset	Alternative Model							
	1	2	3	4	5	6	7	8
State employment	86.1	13.8	10.9	30.6	17.8	5.8	17.9	3.2
Euro area Ind. Prod.	43.5	10.8	0.0	23.4	8.2	5.9	3.7	5.7
U.S. PCE Inflation	76.8	9.8	-0.4	33.9	11.7	28.8	0.9	-1.3

Notes: Panel (a): Relative values of pooled sample quantile and interval risk. Panel (b): log-Bayes factors for the RTS model versus the 8 alternative models listed in Table 4.

The forecasting gains reported in Figure 7 and Table 5 are relative to the same univariate AR(12) benchmark used in Section 3. We have also compared the forecasting performance of the RTS model for inflation to a more sophisticated model, the Federal Reserve Bank of New York's multivariate core trend inflation (MCT) model described in Almuzara and Sbordone (2022). The MCT model is a monthly extension of the quarterly model developed in Stock and Watson (2016a) for the same 17 sector decomposition of PCE inflation, and, like the RTS model, incorporates stochastic volatility, common factors and outliers. As its name suggests, the MCT model produces a monthly estimate of the "trend" in core-inflation, which can serve as a forecast of future aggregate inflation. The RTS model can also be used to forecast aggregate inflation using a share-weighted average of the sectoral inflation forecasts. Table 6 compares the relative root-MSFE of the RTS and MCT models over the 1984m12 through 2024m10 POOS sample period.[5] For forecasting the all-items aggregate inflation, the MCT model is comparable to the benchmark AR(12); however, it provides substantial gains for core inflation—that is, for aggregate inflation excluding the volatile food and energy sectors. The RTS model provides improvements for both all-items and core-inflation relative to both the benchmark AR(12) model and the MCT forecasts.[6]

[5]We thank Martin Almuzara for sharing the MCT code, and the FRBNY for allowing us to use it for this purpose.

[6]We have also compared the forecasting performance of the RTS model to a version of the 'random-walk' model used in Atkeson and Ohanian (2001); that paper considered quarterly data on inflation and used average

Table 6: Relative root-MSFE for Aggregate Inflation

Forecasting Model	All items				Core (excl. Food and Energy)		
	$h=1$	$h=3$	$h=6$		$h=1$	$h=3$	$h=6$
RTS Model	0.98	0.95	0.92		0.93	0.88	0.83
FRBNY-MCT	1.03	0.99	1.00		0.96	0.90	0.86

Notes: The table reports the root-MSFE relative to the sector-specific benchmark AR(12) model, where the forecast of aggregate inflation is the share-weighted average of the sectoral forecasts over the 1984m12-2024m10 POOS sample period.

4.2 Forecasting during the COVID recovery

The COVID-19 pandemic led to dramatic declines in employment and economic activity in the first half of 2020 and the subsequent recovery involved dynamics that differed from the pre-COVID period.[7] Both of these features are evident in Figure 8, which plots the state employment growth rates over 2017m1-2025m4 in panel (a) and the growth rate of Euro-area industrial production (IP) over 2017m1-2023m10 in panel (b). (The IP data is only available for this truncated sample.) Additionally, panel (b) shows one country—Ireland—with a marked increase in volatility and/or outliers during the post-COVID period. Needless to say, forecasting during the COVID recession and its aftermath was challenging.

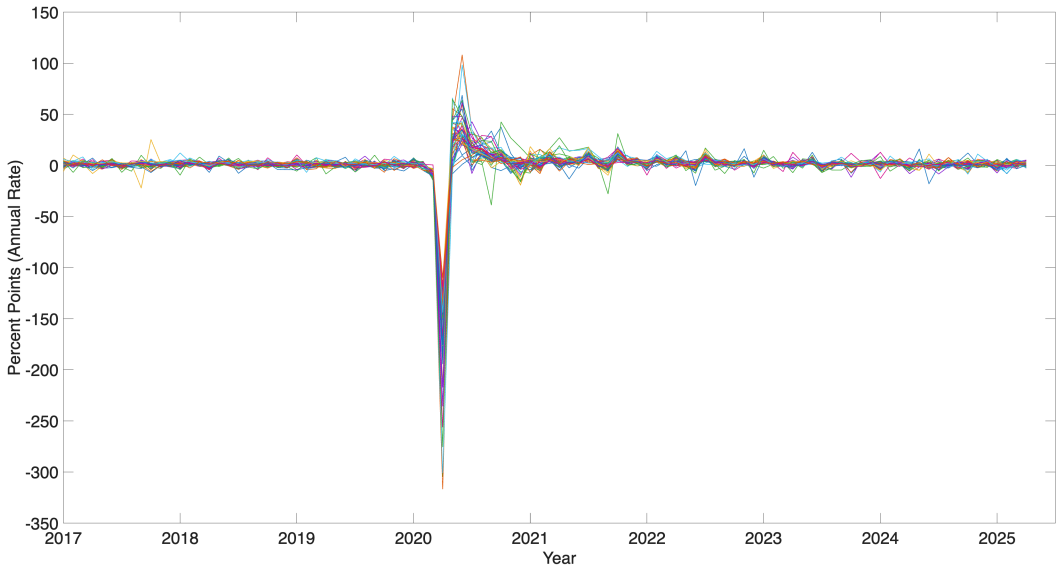
Time series (and other) models were ill-equipped to forecast the dramatic changes in employment that occurred during the first half of 2020, and we will continue to omit this period from our POOS forecasting experiment. A more interesting question is how well models performed in the aftermath of the COVID recession. For example, because AR(p) models rely on lags, the outliers in early 2020 will continue to affect the AR model forecasts p periods after the COVID outliers. The RTS model, with its allowance for outliers, stochastic volatility and time-varying coefficients, can more easily downweight these outliers and adapt to changes in the economy's dynamics. Thus, one expects that the RTS model will perform better than the AR(12) model following COVID. But how much better, and how quickly will these relative gains dissipate as the economy returns to its pre-COVID dynamics?

inflation over the past four quarters to predict the average value of inflation over the next four quarters. Here we use average inflation over the last twelve months to predict average inflation over the next h months, with $h = 1, 3, 6$. Pooled root-MSFE for the RTS forecasts relative to the AO forecasts for $h = 1, 3, 6$ are (0.90, 0.88, 0.80). The corresponding values for all-items aggregate inflation are (0.90, 0.90, 0.87), and for core aggregate inflation they are (0.98, 0.95, 0.93).

[7] See Stock and Watson (2025) for an empirical analysis of the dynamics of real activity in the U.S. in the recovery from the COVID recession.

Figure 8: State Employment and Euro-area IP Growth Rates During COVID

(a) State Employment Growth Rates, 2017m1-2025m4



(b) Euro-Area IP Growth Rates, 2017m1-2023m10

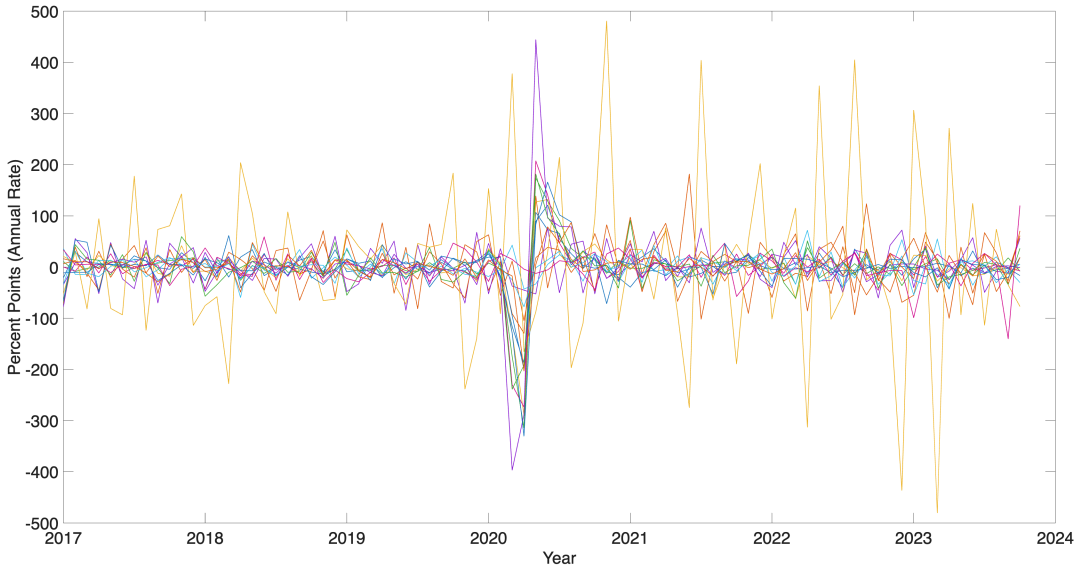


Table 7 reports results from the POOS experiment to answer these questions. In particular, it shows the root-MSFE pooled across the 51 states (panel (a)) and the 16 Euro-area countries for the benchmark AR(12) model over three POOS periods: pre-COVID and two post-COVID periods, the first beginning in 2020m6 and the second in 2021m6. It also shows the relative root-MSFE of the RTS model over these sample periods and relative sample loss for the 90-10 prediction interval.

Looking first at the results for the state employment in panel (a), the pre-COVID period results were shown previously: for example the benchmark AR(12) model has a root-MSFE of 2.1 percentage points for $h = 3$; the RTS model's root-MSFE is 11 percent lower. For the forecasts beginning in 2020m6, the root-MSFE for the AR(12) increases sharply to 16.5 percentage points; the RTS model's increases too, but only to 3.5 percentage points. For forecasts beginning in 2021m6, the AR(12) model is not directly affected by the COVID outliers and its forecasting performance returns to levels similar to the pre-COVID period. The RTS model performance also improves during this period, so much so that its relative root-MSFE is below its pre-COVID value.

The results for Euro-area industrial production are less dramatic. The fall and recovery of Euro-area industrial production during COVID was much smaller than for U.S. employment. The benchmark AR(12) root-MSFE for the period beginning in 2020m6 is roughly twice as large as its pre-COVID value, versus nearly eight times larger for the state employment data, and the relative gains from using the RTS, while greater in the immediate aftermath of COVID, are not as large as those for the U.S. employment data.

Overall, we are reassured by these external validity tests: the RTS performs reasonably well for two additional data sets and it adapts well to the dramatic changes in the economy associated with the COVID pandemic.

5 Concluding remarks

This paper has developed a model for forecasting "related" time series. The resulting RTS model exploits similarity in the stochastic processes describing the individual series as well as covariation between the series. Both are useful for forecasting. The paper also developed MCMC methods to efficiently obtain draws from the posterior distribution of the model's parameters and predictive distributions. Modeling and computation go hand-in-hand in fore-

Table 7: Root-MSFE and Forecast 90-10 Interval Risk in Pre- and Post-COVID Sample Periods

POOS forecast period	$h = 1$			$h = 3$			$h = 6$		
	AR root-MSFE	RTS relative MSFE	RTS relative 90-10 Interval	AR root-MSFE	RTS relative MSFE	RTS relative 90-10 Interval	AR root-MSFE	RTS relative MSFE	RTS relative 90-10 Interval
	(a) State Employment								
1999m12-2019m6	3.3	0.92	0.90	2.1	0.89	0.87	1.8	0.89	0.86
2020m6-2024m10	26.4	0.21	0.20	16.5	0.21	0.18	7.8	0.29	0.27
2021m6-2024m10	4.2	0.88	0.50	2.4	0.74	0.43	1.8	0.69	0.47
	(b) Euro-Area Industrial Production								
1985m6-2019m6	31.7	0.96	0.92	13.2	0.96	0.92	8.4	0.95	0.94
2020m6-2023m4	62.8	0.97	0.82	27.6	0.81	0.63	14.9	0.88	0.73
2021m6-2023m4	54.0	1.0	0.88	20.6	0.97	0.80	12.3	1.00	0.85

Notes: The entries are absolute root-MSFE of the AR(12) benchmark model, as well as the relative root-MSFE and relative interval risk of the RTS model over three sample periods.

casting applications such as this—a model is useful only to the extent that it can be implemented.

The RTS model was developed to capture features in macro datasets such as the state employment, Euro-area industrial production and sectoral inflation data analyzed in the paper. The variables in these datasets are, to first order, well described by simple univariate autoregressions with similar coefficient values, but additionally exhibited stochastic volatility, heavy-tailed innovations, occasional large outliers, common sources of variability, and slowly drifting parameters. The RTS model incorporates all of these.

In other applications, researchers may want to include all or only a subset of these features. The modular design of the MCMC algorithms described in the appendix makes them well-suited for such applications.

References

Abraham, B., and G. E. Box (1979): "Bayesian Analysis of Some Outlier Problems in Time Series," *Biometriika*, 66(2), 229–236.

Almuzara, M., and A. Sbordone (2022): "Inflation Peristence: How Much is There and Where Is It Coming From?," *Federal Reserve Bank of New York Liberty Street Economics*, https://libertystreeteconomics.newyorkfed.org/2022/04/inflation-persistence-how-much-is-there-and-where-is-it-coming-from/.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2024): "Advances in Nowcasting Economic Activity: The Role of Heterogeneous Dynamics and Fat Tails," *Journal of Econometrics*, 238, Article 105634.

Atkeson, A., and L. E. Ohanian (2001): "Are Phillips Curves Useful for Forecasting Inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), 2–11.

Bai, Y., A. Carriero, T. Clark, and M. Marcellino (2022): "Macroeconomic Forecasting in a Multi-Country Context," *Journal of Applied Econometrics*, 37(6), 1230–1255.

Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics*, 25(1), 71–92.

Box, G. E. P., and G. M. Jenkins (1970): *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Carriero, A., T. E. Clark, and M. Marcellino (2015): "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, 30(1), 46–73.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2024): "Addressing Covid-19 Outliers in BVARS with Stochastic Volatility," *Review of Economics and Statistics*, 106(5), 1403–1417.

Chan, J., and I. Jeliazkov (2009): "Efficient Simulation and Integrated Likelihood Estimation in State Space Models," *International Journal of Mathematical Modeling and Numerical Optimization*, 1, 101–120.

Chudik, A., and M. H. Pesaran (2016): "Theory and Practice of GVAR Modeling," *Journal of Economic Surveys*, 30(1), 165–197.

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8, 262–302.

D'Agostino, A., and D. Giannone (2012): "Camparing Alternative Predictors Based on Large-Panel Factor Models," *Oxford Bulletin of Economics and Statistics*, 74(2), 306–326.

del Negro, M., and C. Otrok (2008): "Dynamic Factor Models with Time-Varying Parameters: Changes in International Business Cycles," *Federal Reserve Bank of New York, Staff Report 326*.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Durbin, J., and S. J. Koopman (2002): "A Simple and Effcient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89(3), 603–615.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Ecomometrica*, 4, 987–1007.

Fox, A. (1972): "Outliers in Time Series," *Journal of the Royal Statistical Society, Series B*, 34(3), 350–363.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004): *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, Florida, 2nd edn.

Geweke, J. (2004): "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, 99, 799–804.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatillity: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393.

Meng, X. L., and W. H. Wong (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity: a Theoretical Exploration," *Stastistica Sinica*, 6, 831–860.

Müller, U. K., and M. W. Watson (2020): "Low-Frequency Analysis of Economic Time Series," *Chapter prepared for the Handbook of Econometrics*.

Nerlove, M., D. Grether, and J. Carvalho (1979): *Analysis of Economic Time Series*. Academic Press.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stocahstic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425–449.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogenous Panels with Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Stock, J. H., and M. W. Watson (2016a): "Core Inflation and Trend Inflation," *Review of Economics and Statistics*, 98(4), 770–784.

——— (2016b): "Factor Models and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2A, pp. 415–526. North Holland.

——— (2025): "Recovering From Covid," *Brookings Papers on Economic Activity*, Spring, 2025.

Appendix

A Data

The state employment data are from the U.S. Bureau of Labor Statistics and are "Total Nonfarm Employees" in each state. The data were downloaded from the FRB-St.Louis FRED database. As examples, data for Alaska and Wyoming are the FRED series AKNA and WYNA.

The Euro-area industrial production data include data from the 16 countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden and the UK. These data are from OECD and were downloaded from FRED. As examples, data for Austria and the UK are the FRED series AUTPROINDMISMEI and GBRPROINDMISMEI.

The 17 sectoral price series for personal consumption expenditures in the U.S. are from the U.S. Bureau of Economic Analysis and are taken from the monthly NIPA tables 2.3.4 and 2.3.5. These contain a 16-sector decomposition of the PCE. Following Stock and Watson (2016a), we decomposed the "housing and utilities" sector into the two sectors "housing-energy" and "housing excluding energy."

B Computation

B.1 General comments

The posteriors were obtained by standard MCMC with numerous Gibbs steps. A detailed description of the steps for each model are in the following subsections. Here we provide some generic comments on the numerical implementation.

B.1.1 Posterior Draw from Linear Gaussian State Space System (SSS)

Consider a generic state space system

$$\begin{aligned} y_t &= h_t' s_t + w_t,\ w_t \sim \mathcal{N}(0, R_t) \\ s_t &= F_t s_{t-1} + u_t,\ u_t \sim \mathcal{N}(0, Q_t),\ t = 2, \ldots, T \\ s_1 &\sim \mathcal{N}(s_{1|0}, P_{1|0}) \end{aligned}$$

where y_t is a scalar. The evaluation of the log-likelihood $\sum_{t=1}^{T} \ell_t$ of $\{y_t\}_{t=1}^{T}$ after integrating out $\{s_t\}_{t=1}^{T}$ can be obtained by the following Kalman iterations for $t = 1, \ldots, T$:

1. $\ell_t = -\frac{1}{2} e_t^2/\omega_t^2 - \frac{1}{2} \log \omega_t^2$ with $e_t = y_t - h_t' s_{t|t-1}$, $\omega_t^2 = h_t' v_t + R_t$ and $v_t = P_{t|t-1} h_t$

2. $s_{t|t} = s_{t|t-1} + v_t e_t/\omega_t^2$

3. $P_{t|t} = P_{t|t-1} - v_t v_t'/\omega_t^2$

4. $s_{t+1|t} = F_{t+1} s_{t|t}$

5. $P_{t+1|t} = F_{t+1} P_{t|t} F_{t+1}' + Q_{t+1}$.

To generate a random draw $\{s_t^{**}\}_{t=1}^T$ from the posterior $\{s_t\}_{t=1}^T|\{y_t\}_{t=1}^T$, we employ the algorithm developed in Durbin and Koopman (2002). The following expressions are simple rearrangements of the formulas given there, optimized for a scalar measurement and computational efficiency: Set $s_{1|0}^* = 0$ and $s_1^* \sim \mathcal{N}(0, P_{1|0})$ and iterate for $t = 1, \ldots, T$

1. $b_t = v_t/\omega_t^2$ where $\omega_t^2 = h_t' v_t + R_t$ and $v_t = P_{t|t-1} h_t$

2. $a_t = (y_t - h_t' s_{t|t-1})/\omega_t^2$ and $a_t^* \sim (h_t'(s_t^* - s_{t|t-1}^*) + \varepsilon_t^*)/\omega_t^2$ with $\varepsilon_t^* \sim \mathcal{N}(0, R_t)$

3. $s_{t|t} = s_{t|t-1} + v_t a_t$, $s_{t|t}^* = s_{t|t-1}^* + v_t a_t^*$

4. $P_{t|t} = P_{t|t-1} - b_t v_t'$

5. $s_{t+1}^* \sim \mathcal{N}(F_{t+1} s_t^*, Q_{t+1})$

6. $s_{t+1|t} = F_{t+1} s_{t|t}$, $s_{t+1|t}^* = F_{t+1} s_{t|t}^*$

7. $P_{t+1|t} = F_{t+1} P_{t|t} F_{t+1}' + Q_{t+1}$

followed by an iteration $t = T, T-1, \ldots, 1$ with initial values $r_T = r_T^* = 0$

1. $r_{t-1} = F_t' r_t + h_t(a_t - b_t' F_t' r_t)$, $r_{t-1}^* = F_t' r_t^* + h_t(a_t^* - b_t' F_t' r_t^*)$

2. $s_t^{**} = s_{t|t-1} + P_{t|t-1} r_{t-1} + s_t^* - s_{t|t-1}^* - P_{t|t-1} r_{t-1}^*$

so the scalars $\{a_t\}_{t=1}^T, \{a_t^*\}_{t=1}^T$, vectors $\{b_t\}_{t=1}^T$, $\{s_{t|t-1}\}_{t=1}^T$, $\{s_t^*\}_{t=1}^T$, and matrices $\{P_{t|t-1}\}_{t=1}^T$ must be saved during the first set of iterations to generate the draws $\{s_t^{**}\}_{t=1}^T$.

If the posterior means $s_{t|T}$ and covariance matrices $P_{t|T}$ of $s_t|\{y_t\}_{t=1}^T$ are also needed, then they can be computed by adding to the above iteration $t = T, T-1, \ldots, 1$ the following three steps with initial value $G_T = 0$

3. $G_{t-1} = A_t - h_t x_t' - x_t h_t' + h_t h_t'(b_t' x_t + 1/\omega_t^2)$ with $A_t = F_t' G_t F_t$ and $x_t = A_t b_t$

4. $P_{t|T} = P_{t|t-1} - P_{t|t-1} G_{t-1} P_{t|t-1}$

5. $s_{t|T} = s_{t|t-1} + P_{t|t-1} r_{t-1}$

In applications of this algorithm, the vector h_t and the matrices $P_{1|0}$, Q_t and F_t are often sparse, which we exploit to further gain computational efficiency.

B.1.2 Posterior Draw from Gaussian Hierarchical Model with Conjugate Likelihood

Suppose $\theta_0 \sim \mathcal{N}(\mu_0, \Omega_0)$, $\theta_j|\theta_0 \sim \mathcal{N}(\theta_0, \Omega_j)$ and the log-likelihood of the observations is of the form $C - \frac{1}{2}\sum_{j=1}^n (Y_j - \theta_j)'\Sigma_j^{-1}(Y_j - \theta_j) + \log\det\Sigma_j$ for some conformable Y_j (as would be the case for independent observations $Y_j \sim \mathcal{N}(\theta_j, \Sigma_j)$). Elementary calculations show that the posterior of θ_0 is Gaussian $\mathcal{N}(V_0(\Omega_0^{-1}\mu_0 + \sum_{j=1}^n (\Sigma_j + \Omega_j)^{-1}Y_j), V_0)$ with $V_0^{-1} = \Omega_0^{-1} + \sum_{j=1}^n (\Sigma_j + \Omega_j)^{-1}$. Furthermore, conditional on θ_0, the posteriors for $\{\theta_j\}_{j=1}^n$ are independent Gaussian $\mathcal{N}(V_j\Sigma_j^{-1}(Y_j - \theta_0) + \theta_0, V_j)$ with $V_j^{-1} = \Omega_j^{-1} + \Sigma_j^{-1}$.

B.1.3 Draws from Posterior with Hierarchical Normal Prior

Suppose the prior for n scalar parameters is $\{\theta_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_\theta}, v_{m_\theta}, m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$, and the likelihood factors in θ_j. Then the standard way of generating draws from θ_j is to (i) condition on m_θ and v_θ and update θ_j from the likelihood information and the prior $\theta_j \sim \mathcal{N}(m_\theta, v_\theta)$ separately for $j = 1, \ldots, n$; (ii) condition on $\{\theta_j\}_{j=1}^n$ and update m_θ and v_θ.

This standard approach will not lead to a well mixing chain when n is large, however. To see why, consider the extreme case where the likelihood is uninformative about θ_j. Then in the first step, $\theta_j \sim iid\mathcal{N}(m_\theta, v_\theta)$. By the law of large numbers, $n^{-1}\sum_{j=1}^n \theta_j \approx m_\theta$ and $n^{-1}\sum_{j=1}^n (\theta_j - m_\theta)^2 \approx v_\theta$, so in the second step, we recover nearly the same parameters (m_θ, v_θ) that we started with.

A better approach, especially if the likelihood is not very informative, is to condition in the second step on the current "z-scores" $\{z_j\}_{j=1}^n$ with $z_j = \frac{\theta_j - m_\theta}{\sqrt{v_\theta}}$. Given the current value (m_θ^c, v_θ^c), a random walk Metropolis proposal $(m_\theta^p, \ln v_\theta^p)' \sim \mathcal{N}((m_\theta^c, \ln v_\theta^c)', \Lambda)$ then induces the values $\theta_j^p = m_\theta^p + \sqrt{v_\theta^p / v_\theta^c}(\theta_j^c - m_\theta^c)$, $j = 1, \ldots, n$. The acceptance probability for the proposal involves the probability of $(m_\theta^p, \ln v_\theta^p)$ relative to $(m_\theta^c, \ln v_\theta^c)$ (in the hierarchical normal model, computed from $m_\theta \sim \mathcal{N}(m_{m_\theta}, v_{m_\theta})$ and $\ln(v_\theta) \sim \mathcal{N}(m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$), and the likelihood of $\{\theta_j^p\}_{j=1}^n$ relative to $\{\theta_j^c\}_{j=1}^n$, but it does not involve the prior $\theta_j \sim \mathcal{N}(m_\theta, v_\theta)$, as the z-scores are, by construction, equally likely for all values of (m_θ, v_θ). Of course, if the evaluation of the likelihood is computationally expensive, then performing this step is much slower than conditioning on $\{\theta_j\}_{j=1}^n$. But note that in the flat likelihood example, this alternative approach has excellent mixing properties, as it mixes as well as a random walk Metropolis chain that explores (m_θ, v_θ) under the only information that $m_\theta \sim \mathcal{N}(m_{m_\theta}, v_{m_\theta})$ and $\ln(v_\theta) \sim \mathcal{N}(m_{\ln(v_\theta)}, v_{\ln(v_\theta)})$.

B.1.4 Geweke (2004) Test

It is notoriously easy to make coding mistakes in posterior samplers. We tested (each component of) our code with the Geweke (2004) test, in the implementation described in Müller and Watson (2020).

B.1.5 Random Walk Metropolis Step Size

The algorithms involve random walk Metropolis draws. Let $\theta = (\theta_1, \ldots, \theta_k)$ be a vector valued parameter that is subject to a Metropolis step. If the current value is θ_c, the proposed value is drawn from $\theta_p \sim \mathcal{N}(\theta_c, \kappa_0^2 \tau \operatorname{diag}(\kappa_1^2 v_1, \ldots, \kappa_k^2 v_k))$, where v_j are the prior variances, $\tau \in \{n^{-1}, T^{-1}, (nT)^{-1}\}$ depending on the order of accumulation of information about θ, and the κ_j are positive constants. Note that in the presence of hierarchical priors, the prior vector $(v_1, \ldots, v_k)$ might itself be different from draw to draw. The constants κ_j are determined in the burn-in phase to approximately yield a 50% acceptance rate. In particular, we keep track of the $k+1$ acceptance rates $\hat{a}_i$, $i = 0, \ldots, k$ in the last 200 draws. Here $\hat{a}_0$ corresponds to the rate under the proposal above, and $\hat{a}_i$ for $i > 0$ is the rate for the proposal that moves only one element, $\mathcal{N}(\theta_c, \operatorname{diag}(0, \ldots, 0, \kappa_0^2 \kappa_i^2 \tau v_i, 0, \ldots, 0))$. We then update $\{\kappa_i\}_{i=0}^k$ via

$$\kappa_i \leftarrow \left(\left(\left(\frac{\hat{a}_i}{1 - \hat{a}_i} \right)^{0.4} \wedge 3 \right) \vee 1/3 \right) \kappa_i$$

that is, the more $\hat{a}_i$ deviates from 50%, the larger the adjustment, but only up to a maximal increase or decrease by factor of 3 or 1/3, respectively. This process is repeated every 200 draws for a phase in the burn-in period. The initial baseline values of κ_i where also determined in this fashion, and then hard-coded and held constant across variations of the sample size (n, T), data sets and variations of the model (which more plausibly yields a reasonable acceptance rate due to the presence of the deterministic τ).

The process to update the κ_i is computationally costly, as it requires $2(k+1)$ evaluations of the likelihood, rather than just 2. To reduce this burden, we treat vectors of a parameter that corresponds to the $p = 12$ AR coefficients $\{\phi_{j,t,l}\}_{l=1}^p$ as one block with a single corresponding κ. For the common value of the AR coefficients $\{m_{\phi_l}\}_{l=1}^p$, the proposal variance is further multiplied by $\operatorname{diag}(1, 2, \ldots, p)$, with the idea that the likelihood has much more curvature for l small relative to the sharply decreasing prior variances. We similarly treat the q stochastic volatility components as one block and rescale the proposal variance by $\operatorname{diag}(1, 2, \ldots, q)$.

B.1.6 Burn in and Number of Draws

We initialize the sampler with $u_{n+1,t} = 0$, $\mu_{j,t} = o_{j,t} = 0$ for $t = 1, \ldots, T$ and $j = 1, \ldots, n+1$ and all parameters are equal to the prior mean. For a sampler that generates N total usable MCMC draws, we use a burn-in period that consists of three phases: First, we take 200 draws that do not update any priors or other common components such as $u_{n+1,t}$, $\mu_{n+1,t}$ or $o_{n+1,t}$. Second, for $l = 1, \ldots, N/3$ further draws that involves all steps, we set the Metropolis step size standard deviations equal to $5^{(1-3l/N)}$ of its baseline value (that is, κ_0 of the previous subsection is inflated by a factor $5^{(1-3l/N)}$). The idea is to allow the sampler to make larger movements initially to quickly approach values where the posterior is high. Finally, for another $N/3$ draws, we adjust the Metropolis step sizes as

described in the last subsection.

We set $N = 1000$ when $n < 20$ and $N = 2000$ for the employment application for models I-VI, and a twice as large N for the RTS model. It takes about one minute to generate the $N = 4000$ draws for the employment application in the RTS model, including the burn-in, in a Fortran implementation on a 24 core workstation.

B.2 Details on the Gibbs Steps for each Model

We condition on all variables if not stated otherwise. To ease notation, we do not explicitly mention updates to $\epsilon_{j,t} = \sigma_{j,t}\varepsilon_{j,t}$ that arise from changes in $\varepsilon_{j,t}$ and $\sigma_{j,t}$.

B.2.1 Model I: Bayesian shrinkage

1. $\{\omega, \{\sigma_j\}_{j=1}^n\}$: Draw $\ln \omega^2 | \{\ln \omega^2 + \ln \sigma_j^2\}_{j=1}^n$ from conjugate normal obtained from (11) and (7), then update $\{\sigma_j\}_{j=1}^n$ according to new ω. [The model depends on $\{\omega, \{\sigma_j\}_{j=1}^n\}$ only through the products $\{\omega\sigma_j\}_{j=1}^n$, so there is no additional contribution to the posterior.]

2. $\{\sigma_j, \mu_j, u_{j,-11:0}, \{\varepsilon_{j,t}\}_{t=1}^T\}$ looping over j:

   (a) Draw σ_j: RW Metropolis step with prior (11) and likelihood computed from Kalman filter with state $(\mu_j, u_{j,t-1}, u_{j,t-2}, \ldots, u_{j,t-12})$, measurement equation (3), state evolution (4), initial state drawn from $\mu_j \sim \mathcal{N}(0, \infty)$ (approximated by using large but finite variance) and (9).

   (b) Draw $\{\mu_j, u_{j,-11:0}, \{\varepsilon_{j,t}\}_{t=1}^T\} | \sigma_j$: Kalman smoother draw from same SSS as in Step 2a.

3. $\{\phi_j, \{\varepsilon_{j,t}\}_{t=1}^T\}$ looping over j: Metropolis-Hastings step with proposal generated from Kalman smoother draw from SSS with state $\phi_j = (\phi_{j,1}, \ldots, \phi_{j,12})$, measurement equation (4) and initial state (10). The proposal ϕ_j^p is accepted over the current value ϕ_j^c with probability $1 \wedge \frac{L_9(\phi_j^p)}{L_9(\phi_j^c)}$, where $L_9(\phi_j)$ is the likelihood of (9). (This Metropolis step adjusts the Gibbs step for the initial values.)

B.2.2 Model II: Hierarchical priors

1. $\{\omega, m_{\ln \sigma^2}\}$: Draw $\ln \omega^2 | \ln \omega^2 + m_{\ln \sigma^2}$ from conjugate normal obtained from (13) and (7), then update $m_{\ln \sigma^2}$ accordingly.

2. $\{m_{\ln \sigma^2}, v_{\ln \sigma^2}, \{\mu_j, u_{j,-11:0}, \sigma_j, \{\varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $\{m_{\ln \sigma^2}, v_{\ln \sigma^2}\} | \{(\ln \sigma_j^2 - m_{\ln \sigma^2})/\sqrt{v_{\ln \sigma^2}}\}_{j=1}^n$: Bivariate RW Metropolis step with prior (13) and likelihood computed from Kalman filter of Model I Step 2a applied to $j = 1, \ldots, n$. Update $\{\sigma_j^2\}_{j=1}^n$ if accepted.

(b) $\{\sigma_j, \mu_j, u_{j,-11:0}, \{\varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n | m_{\ln\sigma^2}, v_{\ln\sigma^2}$: Same as Model I Step 2, except that the prior $\ln\sigma_j^2 \sim iid\mathcal{N}(m_{\ln\sigma^2}, v_{\ln\sigma^2})$ is used in place of (11).

3. $\{\{m_{\phi_l}, v_{\phi_l}\}_{l=1}^{12}, \{\phi_j, \{\varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $\{m_{\phi_l} v_{\phi_l}\}_{l=1}^{12} | \{\{(\phi_{j,l} - m_{\phi_l})/\sqrt{v_{\phi_l}}\}_{l=1}^{12}\}_{j=1}^n$: 24-dimensional RW Metropolis step, with prior (12) and likelihood computed from the product over $j = 1, \ldots, n$ of (9) and Kalman filters with state $\phi_j = (\phi_{j,1}, \ldots, \phi_{j,12})$ and measurement equations (4). Update $\{\phi_j\}_{j=1}^n$ if accepted.

   (b) $\{\phi_j, \{\varepsilon_{j,t}\}_{t=1}^T\} | \{m_{\phi_l}, v_{\phi_l}\}_{l=1}^{12}$ looping over j: Metropolis-Hastings step with proposal generated from Kalman smoother draw from SSS of Step 3a with initial state $\phi_j \sim \mathcal{N}((m_{\phi_1}, \ldots, m_{\phi_{12}}), \mathrm{diag}(v_{\phi_1}, \ldots, v_{\phi_{12}}))$. The proposal ϕ_j^p is accepted over the current value ϕ_j^c with probability $(1 \wedge \frac{L_9(\phi_j^p)}{L_9(\phi_j^c)})$, where $L_9(\phi_j)$ is the likelihood of (9).

4. $\{\{m_{\phi_l}\}_{l=1}^{12}, \{\phi_j, \{\varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$: Ignoring the likelihood contribution from (9), the log-likelihood is quadratic in ϕ_j. We can thus generate a Metropolis-Hastings proposal of $\{m_{\phi_l}\}_{l=1}^{12}$ and $\{\phi_j\}_{j=1}^n$ using the algorithm in Section B.1.2, and accept it with probability $1 \wedge \frac{\prod_{j=1}^n L_9(\phi_j^p)}{\prod_{j=1}^n L_9(\phi_j^c)}$ in obvious notation. [This step is not needed given Step 3, but it improves mixing.]

B.2.3 Model III: Student-t innovations

Let $S_{j,t}$ be independent draws of $\nu_j/\chi^2_{\nu_j}$. Then $\varepsilon_{j,t} \sim \mathcal{T}(\nu_j)$ from (14) can be represented as

$$\varepsilon_{j,t} = \sqrt{S_{j,t}} z_{j,t}, \quad z_{j,t} \sim iid\mathcal{N}(0,1).$$

The sampler for Model III (and those below) treats $\{\{S_{j,t}\}_{t=1}^T\}_{j=1}^n$ as an additional unobserved component (which we condition on if not explicitly part of a block), so that after conditioning, we recover a Gaussian model for $\varepsilon_{j,t}|S_{j,t}$ and thus $u_{j,t}|S_{j,t}, \phi_j, \sigma_j^2$.

1. $\{m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{\nu_j, \sigma_j^2, \{S_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) $\{m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}\} | \{(\ln(\nu_j - 2) - m_{\ln(\nu-2)})/\sqrt{v_{\ln(\nu-2)}}, (\ln\sigma_j^2 - m_{\ln\sigma^2})/\sqrt{v_{\ln\sigma^2}}\}_{j=1}^n$: Four dimensional RW Metropolis step with priors (13), (15) and likelihood computed from student-t density (14). Update $\{\nu_j, \sigma_j^2\}_{j=1}^n$ if accepted.

   (b) $\{\ln(\nu_j - 2), \sigma_j^2\}_{j=1}^n | m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}$: Looping over j, bivariate RW Metropolis step with priors (13), (15) and likelihood computed from (14).

   (c) $\{\{S_{j,t}\}_{t=1}^T\}_{j=1}^n | m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{\nu_j, \sigma_j^2\}_{j=1}^n$: Looping over j, draw $S_{j,t}$ independently from $(\nu_j + \varepsilon_{j,t}^2)/\chi^2_{\nu_j+1}$, $t = 1, \ldots, T$.

[Given that $m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{\sigma_j^2\}$ are also updated in Step 2 below, we could keep them fixed in this step, but it improves mixing to exploit the relatively weaker informativeness of the student-t likelihood to update the variance parameters in both steps.]

2. Perform Steps 1, 3 and 4 of Model II, except that in the SSS, the variance of the measurement equation is now given by $\sigma_j^2 S_{j,t}$.

B.2.4 Model IV: Additive outliers

In analogy to Model III, $\eta_{j,t} \sim \mathcal{T}(\nu_j^o)$ can be represented as

$$\eta_{j,t} = \sqrt{S_{j,t}^o} z_{j,t}^o, \qquad z_{j,t}^o \sim iid\mathcal{N}(0,1)$$

where $S_{j,t}^o$ are independent draws of $\nu_j^o/\chi_{\nu_j^o}^2$. The sampler for Model IV (and those below) treats $\{\{S_{j,t}^o\}_{t=1}^T\}_{j=1}^n$ as an additional unobserved component (which we condition on if not explicitly part of a block), so that after conditioning, we recover a Gaussian model for $\eta_{t,j}|S_{j,t}^o$. To ease notation, we do not explicitly mention updates to $o_{j,t} = \kappa_{j,t}\eta_{j,t}$ that arise from changes in $\eta_{j,t}$ and $\kappa_{j,t}$.

1. $\{\omega, m_{\ln\kappa^2}, m_{\ln\sigma^2}\}$: Draw $\ln\omega^2|(\ln\omega^2 + m_{\ln\sigma^2}, \ln\omega^2 + m_{\ln\kappa^2})$ from conjugate normal obtained from (13), (18) and (7), then update $(m_{\ln\sigma^2}, m_{\ln\kappa^2})$ accordingly.

2. $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{\mu_j, u_{j,-11:0}, \sigma_j, \kappa_j, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}\}|\{(\ln\sigma_j^2 - m_{\ln\sigma^2})/\sqrt{v_{\ln\sigma^2}}, (\ln\kappa_j^2 - m_{\ln\kappa^2})/\sqrt{v_{\ln\kappa^2}}\}_{j=1}^n$: 4-dimensional RW Metropolis step with priors (15), (18) and likelihood computed from Kalman filter with state $(\mu_j, u_{j,t-1}, u_{j,t-2}, \ldots, u_{j,t-12})$, measurement equation (16), state evolution (4), and initial state $\mu_j \sim \mathcal{N}(0,\infty)$ (approximated by using large but finite variance) and (9). Update $\{\sigma_j, \kappa_j\}_{j=1}^n$ if accepted.

   (b) $\{\ln\sigma_j^2, \ln\kappa_j^2\}_{j=1}^n|m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}$: Looping over j, bivariate RW Metropolis step with prior (15), (18) and likelihood computed from same Kalman filter as in Step 2a.

   (c) $\{\mu_j, u_{j,-11:0}, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\}|m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{\sigma_j, \kappa_j\}_{j=1}^n$: Looping over j, draw from Kalman smoother from same SSS as in Step 2a.

3. $\{m_{\ln(\nu^o-2)}, v_{\ln(\nu^o-2)}, m_{\ln\kappa^2}, v_{\ln\kappa^2}, \{\nu_j^o, \kappa_j, \{S_{j,t}^o\}_{t=1}^T\}_{j=1}^n\}$:

   (a) $\{m_{\ln(\nu^o-2)}, v_{\ln(\nu^o-2)}, m_{\ln\kappa^2}, v_{\ln\kappa^2}\}|\{(\ln(\nu_j^o - 2) - m_{\ln(\nu^o-2)})/\sqrt{v_{\ln(\nu^o-2)}}, (\ln\kappa_j^2 - m_{\ln\kappa^2})/\sqrt{v_{\ln\kappa^2}}\}_{j=1}^n$: Four dimensional RW Metropolis step with priors (18), (19) and likelihood computed from (17). Update $\{\nu_j^o, \kappa_j^2\}_{j=1}^n$ if accepted.

   (b) $\{\ln(\nu_j^o - 2), \kappa_j\}_{j=1}^n|m_{\ln(\nu^o-2)}, v_{\ln(\nu^o-2)}, m_{\ln\kappa^2}, v_{\ln\kappa^2}$: Looping over j, bivariate Metropolis step with priors (18), (19) and student-t likelihood (17).

(c) $\{\{S^o_{j,t}\}_{t=1}^T\}_{j=1}^n | m_{\ln(\nu^o-2)}, v_{\ln(\nu^o-2)}, m_{\ln\kappa^2}, v_{\ln\kappa^2}\{\nu^o_j, \kappa_j\}_{j=1}^n$: Looping over j, draw $S^o_{j,t}$ independently from $(\nu^o_j + \eta^2_{j,t})/\chi^2_{\nu^o_j+1}$, $t = 1, \ldots, T$.

4. Perform Step 1 of Model III and Steps 3-4 of Model II, except that in the SSS, the variance of the measurement equation is given by $\sigma^2_j S_{j,t}$.

B.2.5 Model V: Time varying volatility

1. $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, \{\mu_j, u_{j,-11:0}, \sigma_j, \{\xi_{j,l}\}_{l=1}^q, \kappa_j, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $(m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi) | \{(\ln\sigma^2_j - m_{\ln\sigma^2})/\sqrt{v_{\ln\sigma^2}}, (\ln\kappa^2_j - m_{\ln\kappa^2})/\sqrt{v_{\ln\kappa^2}}, \{(\xi_{j,l} - m_{\xi_l})/\sqrt{v_\xi}\}_{l=1}^q\}_{j=1}^n$: $(5+q)$-dimensional RW Metropolis step with prior (13), (18), (23)-(24) and likelihood computed from Kalman filter from same SSS as in Step 2 of Model IV (except that the measurement equation now has variance $\omega^2\sigma^2_{j,t}S_{j,t} + \omega^2\kappa^2_j S^o_{j,t}$). Update $\{\sigma_j, \kappa_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n$ if accepted.

   (b) $\{\ln\sigma^2_j, \ln\kappa^2_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n | m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi$: Looping over j, $q+2$-dimensional RW Metropolis step with prior (13), (18), (22) and likelihood computed from same Kalman filter as in Step 1a.

   (c) $\{\mu_j, u_{j,-11:0}, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\} | m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, \{\sigma_j, \kappa_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n$: Looping over j, Kalman smoother draw from same SSS as in Step 1a.

2. $\{m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, \{\nu_j, \sigma^2_j, \{\xi_{j,l}\}_{l=1}^q, \{S_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) $(m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi) | \{(\ln(\nu_j - 2) - m_{\ln(\nu-2)})/\sqrt{v_{\ln(\nu-2)}}, (\ln\sigma^2_j - m_{\ln\sigma^2})/\sqrt{v_{\ln\sigma^2}}, \{(\xi_{j,l} - m_{\xi_l})/\sqrt{v_\xi}\}_{l=1}^q\}_{j=1}^n$: Six dimensional RW Metropolis step with priors (13), (15), (23)-(24) and likelihood computed from student-t density (14). Update $\{\nu_j, \sigma^2_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n$ if accepted.

   (b) $\{\ln(\nu_j - 2), \sigma^2_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n | m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi$: Looping over j, three dimensional RW Metropolis step with priors (13), (15), (22) and likelihood computed from (14).

   (c) $\{\{S_{j,t}\}_{t=1}^T\}_{j=1}^n | m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, \{\nu_j, \sigma^2_j, \{\xi_{j,l}\}_{l=1}^q\}_{j=1}^n$: Looping over j, draw $S_{j,t}$ independently from $(\nu_j + \varepsilon^2_{j,t})/\chi^2_{\nu_j+1}$, $t = 1, \ldots, T$.

3. Perform Steps 1 and 3 of Model IV and Steps 3-4 of Model II, except that in the SSS, the variance of the measurement equation is given by $\sigma^2_j S_{j,t}$.

B.2.6 Model VI: Time varying conditional mean parameters

1. $\{\omega, m_{\ln\kappa^2}, m_{\ln\sigma^2}, m_{\ln\gamma^2_\mu}\}$: Draw $\ln\omega^2 | \ln\omega^2 + m_{\ln\sigma^2}, \ln\omega^2 + m_{\ln\kappa^2}, \ln\omega^2 + m_{\ln\gamma^2_\mu}$ from conjugate normal obtained from (13), (18), (29) and (7), then update $(m_{\ln\sigma^2}, m_{\ln\kappa^2}, m_{\ln\gamma^2_\mu})$ accordingly.

2. $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, m_{\ln\gamma_\mu^2}, v_{\ln\gamma_\mu^2}, \{\mu_j, u_{j,-11:0}, \sigma_j, \{\xi_{j,l}\}_{l=1}^q, \gamma^2_{\mu(j)}, \kappa_j, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, m_{\ln\gamma_\mu^2}, v_{\ln\gamma_\mu^2}\}|$ $\{(\ln\sigma_j^2 - m_{\ln\sigma^2})/\sqrt{v_{\ln\sigma^2}}, (\ln\kappa_j^2 - m_{\ln\kappa^2})/\sqrt{v_{\ln\kappa^2}}, \{(\xi_{j,l} - m_{\xi_l})/\sqrt{v_\xi}\}_{l=1}^q, (\ln\gamma^2_{\mu(j)} - m_{\ln\gamma_\mu^2})/\sqrt{v_{\ln\gamma_\mu^2}}\}_{j=1}^n$: $(7+q)$-dimensional RW Metropolis step with prior (15), (18), (23), (29) and likelihood computed from Kalman filter with state $(\mu_{j,t}, u_{j,t-1}, u_{j,t-2}, \ldots, u_{j,t-12})$, measurement equation (25), state evolution (26) and $\mu_{j,t}|\mu_{j,t-1} \sim \mathcal{N}(\mu_{j,t-1}, \gamma^2_{\mu(j)})$, and initial state $\mu_{j,0} \sim \mathcal{N}(0, \infty)$ (approximated by using large but finite variance) and (9). Update $\{\sigma_j, \kappa_j, \{\xi_{j,l}\}_{l=1}^q, \gamma^2_{\mu(j)}\}_{j=1}^n$ if accepted.

   (b) $\{\ln\sigma_j^2, \ln\kappa_j^2, \{\xi_{j,l}\}_{l=1}^q, \gamma^2_{\mu(j)}\}_{j=1}^n | m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, m_{\ln\gamma_\mu^2}, v_{\ln\gamma_\mu^2}$: Looping over j, $q+3$ dimensional RW Metropolis step with prior (15), (18), (22), (29) and likelihood computed from same Kalman filter as in Step 2a.

   (c) $\{\mu_j, u_{j,-11:0}, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^T\} | m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, m_{\ln\gamma_\mu}, v_{\ln\gamma_\mu}$, $\{\sigma_j, \kappa_j, \{\xi_{j,l}\}_{l=1}^q, \gamma^2_{\mu(j)}\}_{j=1}^n$: Looping over j, draw from Kalman smoother from same SSS as in Step 2a.

3. $\{\{m_{\phi_l}, v_{\phi_l}, m_{\ln(\gamma^2_{\phi(l)})}, v_{\ln(\gamma^2_{\phi(l)})}\}_{l=1}^{12}, \{\{\gamma^2_{\phi(j,l)}\}_{l=1}^{12}, \{\{\phi_{j,l,t}\}_{l=1}^{12}, \varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$:

   (a) Draw $\{m_{\phi_l}, v_{\phi_l}, m_{\ln(\gamma^2_{\phi(l)})}, v_{\ln(\gamma^2_{\phi(l)})}\}_{l=1}^{12} | \{\{(\phi_{j,l,1} - m_{\phi_l})/\sqrt{v_{\phi_l}}\}_{l=1}^{12}, \{(\ln(\gamma^2_{\phi(j,l)}) - m_{\ln(\gamma^2_{\phi(l)})})/\sqrt{v_{\ln(\gamma^2_{\phi(l)})}}\}_{l=1}^{12}\}_{j=1}^n$: 48-dimensional RW Metropolis step with prior (12), (30) and likelihood computed from (9) and Kalman filter with state evolution $\phi_{j,t,l}|\phi_{j,t-1,l} \sim \mathcal{N}(\phi_{j,t-1,l}, \gamma^2_{\phi(l)})$, measurement equation (26) and initial state $\phi_{j,1,l}$, $l = 1, \ldots, 12$. Update $\{\{\phi_{j,l,1}, \gamma^2_{\phi(j,l)}\}_{l=1}^{12}\}_{j=1}^n$ if accepted.

   (b) $\{\ln\gamma^2_{\phi(j,l)}\}_{l=1}^{12} | \{m_{\phi_l}, v_{\phi_l}, m_{\ln(\gamma^2_{\phi(l)})}, v_{\ln(\gamma^2_{\phi(l)})}, \phi_{j,l,1}\}_{l=1}^{12}$ looping over j: 12-dimensional RW Metropolis step with prior (30) and likelihood computed from same SSS as in Step 3a.

   (c) $\{\{\phi_{j,l,t}\}_{l=1}^{12}, \varepsilon_{j,t}\}_{t=1}^T | \{m_{\phi_l}, v_{\phi_l}, m_{\ln(\gamma^2_{\phi(l)})}, v_{\ln(\gamma^2_{\phi(l)})}, \phi_{j,l,1}, \ln\gamma^2_{\phi(j,l)}\}_{l=1}^{12}$ looping over j: Kalman smoother draw from same SSS as in Step 3a.

4. $\{\{\phi_{j,l,t}\}_{l=1}^{12}, \varepsilon_{j,t}\}_{t=1}^T$ looping over j: Metropolis-Hastings step with proposal generated from Kalman smoother draw from SSS of Step 3a, except that initial state is $\phi_{j,1} \sim \mathcal{N}((m_{\phi_1}, \ldots, m_{\phi_{12}}), \mathrm{diag}(v_{\phi_1}, \ldots, v_{\phi_{12}}))$. The proposal ϕ_j^p is accepted over the current value ϕ_j^c with probability $1 \wedge \frac{L_9(\phi_j^p)}{L_{??}(\phi_j^c)}$, where $L_9(\phi_j)$ is likelihood of (9).

5. $\{\{m_{\phi_l}\}_{l=1}^{12}, \{\{\phi_{j,l,t}\}_{l=1}^{12}, \varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n$: Conditional on $m_\phi = (m_{\phi_1}, \ldots, m_{\phi_{12}})'$, ignoring the likelihood contribution from (9), and integrating out $\{\{\phi_{j,l,t}\}_{l=1}^{12}\}_{t=2}^T$, the log-likelihood for $\{y_{j,t}\}_{t=1}^T$ is quadratic in $\phi_{j,1} = (\phi_{j,1,1}, \ldots, \phi_{j,12,1})'$ with mean $\tilde{\phi}_{j,1}$ and variance $\tilde{P}_j$ that could be computed by the Kalman smoother by initializing the SSS of Step 3a with a diffuse initial

state. Given the Gaussian prior (12) $m_\phi \sim \mathcal{N}(0, \Omega_0)$ with $\Omega_0 = \text{diag}(v_{m_{\phi_1}}, \ldots, v_{m_{\phi_{12}}})$ and $\phi_j | m_\phi \sim \mathcal{N}(m_\phi, \Omega_1)$ with $\Omega_1 = \text{diag}(v_{\phi_1}, \ldots, v_{\phi_{12}})$, we could thus generate a Metropolis-Hastings proposal for m_ϕ and $\{\phi_{j,1}\}_{j=1}^n$ using the algorithm in Section B.1.2, that is $m_\phi \sim \mathcal{N}(V_0 \sum_{j=1}^n (\Omega_1 + \tilde{P}_j)^{-1} \tilde{\phi}_{j,1}, V_0)$ with $V_0^{-1} = \Omega_0^{-1} + \sum_{j=1}^n (\Omega_1 + \tilde{P}_j)^{-1}$ and $\phi_{j,1} | m_\phi \sim \mathcal{N}(V_j \tilde{P}_j^{-1} (\tilde{\phi}_{j,1} - m_\phi) + m_\phi, V_j)$ with $V_j^{-1} = \Omega_1^{-1} + \tilde{P}_j^{-1}$. Furthermore, this proposal can be extended to a proposal for $\{\{\{\phi_{j,l,t}\}_{l=1}^{12}\}_{t=1}^T\}_{j=1}^n$ by taking draws from $\{\{\{\phi_{j,l,t}\}_{l=1}^{12}\}_{t=2}^T$ for $j = 1, \ldots, n$ via the SSS described in Step 3c. The proposal would then be accepted with probability $(1 \wedge \frac{\prod_{j=1}^n L_9(\{\phi_{j,l,1}^p\}_{l=1}^{12})}{\prod_{j=1}^n L_9(\{\phi_{j,l,1}^c\}_{l=1}^{12})})$ in obvious notation.

The implementation differs from this conceptually straightforward approach, as the Kalman smoother with diffuse initial state is numerically unstable. Instead, we apply Kalman smoothers with initial state $\phi_{j,1} \sim \mathcal{N}(0, \Omega_1)$ to obtain the smoothed state $\hat{\phi}_{j,1}$ with smoothed covariance matrix $\hat{P}_j$. From Section B.1.2, $\hat{P}_j = (\Omega_1^{-1} + \tilde{P}_j^{-1})^{-1} = V_j$ and $\hat{\phi}_{j,1} = \hat{P}_j \tilde{P}_j^{-1} \tilde{\phi}_{j,1}$. Thus $(\Omega_1 + \tilde{P}_j)^{-1} \tilde{\phi}_{j,1} = \Omega_1^{-1} \hat{\phi}_{j,1}$ and, applying the Woodbury matrix identity, $(\Omega_1 + \tilde{P}_j)^{-1} = \Omega_1^{-1} - \Omega_1^{-1} \hat{P}_j \Omega_1^{-1}$, so the proposal for m_ϕ becomes $m_\phi \sim \mathcal{N}(V_0 \sum_{j=1}^n \Omega_1^{-1} \hat{\phi}_{j,1}, V_0)$ with $V_0^{-1} = \Omega_0^{-1} + \sum_{j=1}^n (\Omega_1^{-1} - \Omega_1^{-1} \hat{P}_j \Omega_1^{-1})$. Finally, to draw the proposals for $\phi_{j,1} | m_\phi$, we exploit that $\hat{P}_j \tilde{P}_j^{-1} (\tilde{\phi}_{j,1} - m_\phi) = \hat{\phi}_{j,1} - \hat{P}_j \tilde{P}_j^{-1} m_\phi = \hat{\phi}_{j,1} - m_\phi + \hat{P}_j \Omega_1^{-1} m_\phi$, where the last equality uses again the Woodbury identity, so $\phi_{j,1} | m_\phi \sim \mathcal{N}(\hat{\phi}_{j,1} + \hat{P}_j \Omega_1^{-1} m_\phi, \hat{P}_j)$.

This step is repeated 3 times for $n \leq 20$, and 5 times for $n > 20$. This is computationally efficient, since acceptance is rare, and most of the computational effort is in the calculation of the Kalman smoothers and the subsequent matrix manipulations, which only need to be performed once.

[This step is not needed given Steps 3-4, but it improves mixing.]

6. Perform Step 2 of Model V and Step 3 of Model IV.

B.2.7 RTS Model

1. $\{\omega, m_{\ln \kappa^2}, m_{\ln \sigma^2}, m_{\ln \gamma_\mu^2}, \ln(\sigma_{n+1}^2), \ln \kappa_{n+1}^2, \ln \gamma_{u(n+1)}^2\}$: Draw $\ln \omega^2 | \ln \omega^2 + m_{\ln \sigma^2}, \ln \omega^2 + m_{\ln \kappa^2}, \ln \omega^2 + m_{\ln \gamma_\mu^2}, \ln \omega^2 + \ln(\sigma_{n+1}^2)$ from conjugate normal obtained from (34), (13), (18), (29), (7), (35) and (36), then update $(m_{\ln \sigma^2}, m_{\ln \kappa^2}, m_{\ln \gamma_\mu^2}, \ln(\sigma_{n+1}^2), \ln \kappa_{n+1}^2, \ln \gamma_{u(n+1)}^2)$ accordingly.

2. $\{\kappa_{n+1}^2, \{\{\eta_{j,t}\}_{t=1}^T\}_{j=1}^{n+1}\}$:

   (a) Draw $\kappa_{n+1}^2 | \{\{\kappa_{n+1} \eta_{n+1,t} + \kappa_{j,t} \eta_{j,t}\}_{t=1}^T\}_{j=1}^n$ by RW Metropolis step based on likelihood of $\{\{\kappa_{n+1} \eta_{n+1,t} + \kappa_{j,t} \eta_{j,t}\}_{t=1}^T\}_{j=1}^n$ induced by $\eta_{j,t} \sim \mathcal{N}(0, S_{j,t}^o)$ independently across $j = 1, \ldots, n+1$ and $t = 1, \ldots, T$.

(b) $\{\eta_{j,t}\}_{j=1}^{n+1}|\kappa_{n+1}^2, \{\kappa_{n+1}\eta_{n+1,t} + \kappa_{j,t}\eta_{j,t}\}_{j=1}^{n}$ looping over t: Draw $\eta_{n+1,t}$ from conjugate Gaussian posterior implied by the likelihood of Step 2a, and update $\{\eta_{j,t}\}_{j=1}^{n}$ accordingly.

3. $\{\gamma_{\mu(n+1)}^2, \{\mu_{n+1,t}\}_{t=1}^{T}, \{\{\mu_{j,t}\}_{t=1}^{T}\}_{j=1}^{n}\}$:

   (a) Draw $\gamma_{\mu(n+1)}^2|\{\{\mu_{n+1,t} + \mu_{j,t}\}_{t=1}^{T}\}_{j=1}^{n}$ by RW Metropolis step based on likelihood of $\{\{\Delta\mu_{n+1,t} + \Delta\mu_{j,t}\}_{t=1}^{T}\}_{j=1}^{n}$ induced by $\Delta\mu_{j,t} \sim \mathcal{N}(0, \gamma_{\mu(j)}^2)$ independently across $j = 1, \ldots, n+1$ and $t = 2, \ldots, T$.

   (b) $\mu_{n+1,t}|\gamma_{\mu(n+1)}, \{\mu_{n+1,t} + \mu_{j,t}\}_{j=1}^{n}$ looping over t: Draw $\mu_{n+1,t}$ from conjugate Gaussian posterior implied by the likelihood of Step 3a, and update $\{\mu_{j,t}\}_{j=1}^{n}$ accordingly.

4. $\{\{m_{\lambda_l}, v_{\lambda_l}, m_{\ln(\gamma_{\lambda(l)}^2)}, v_{\ln(\gamma_{\lambda(l)}^2)}\}_{l=0}^{5}, \{\{\lambda_{j,l,t}, \gamma_{\lambda(j,l)}^2\}_{l=0}^{5}, \{\varepsilon_{j,t}\}_{t=1}^{T}\}_{j=1}^{n}\}$:

   (a) Draw $\{m_{\lambda_l}, v_{\lambda_l}, m_{\ln(\gamma_{\lambda(l)}^2)}, v_{\ln(\gamma_{\lambda(l)}^2)}\}_{l=0}^{5}|\{\{(\ln(\gamma_{\lambda(j,l)}^2) - m_{\ln(\gamma_{\lambda(l)}^2)})/\sqrt{v_{\ln(\gamma_{\lambda(l)}^2)}}\}_{l=0}^{5}\}_{j=1}^{n}$: 24-dimensional RW Metropolis step with prior (39) and likelihood computed from n Kalman filters with state $(u_{j,t-1}, \ldots, u_{j,t-12}, \lambda_{j,0,t}, \ldots \lambda_{j,5,t})$, evolution $\lambda_{j,t}|\lambda_{j,t-1} \sim \mathcal{N}(\lambda_{j,t-1}, \mathrm{diag}(\gamma_{\lambda(0)}^2, \ldots, \gamma_{\lambda(5)}^2))$, measurement $\sum_{l=1}^{12}\phi_{j,t,l}u_{j,t-l} + \sum_{l=0}^{5}\lambda_{j,l,t}u_{n+1,t-l} + \epsilon_{j,t}$ and initial state $(u_{j,0}, \ldots, u_{j,-11})$, $\lambda_{j,1} \sim \mathcal{N}((m_{\lambda(0)}, \ldots, m_{\lambda(5)})', \mathrm{diag}(v_{\lambda(0)}, \ldots, v_{\lambda(5)}))$. Update $\{\{\gamma_{\lambda(j,l)}^2\}_{l=0}^{5}\}_{j=1}^{n}$ if accepted.

   (b) $\{\ln\gamma_{\lambda(j,l)}^2\}_{l=0}^{5}|\{m_{\lambda_l}, v_{\lambda_l}, m_{\ln(\gamma_{\lambda(l)}^2)}, v_{\ln(\gamma_{\lambda(l)}^2)}\}_{l=0}^{5}$ looping over j: 6-dimensional RW Metropolis step with prior (39) and likelihood computed same SSS as in Step 3a.

   (c) $\{\{\lambda_{j,l,t}\}_{l=0}^{5}, \varepsilon_{j,t}\}_{t=1}^{T}|\{m_{\lambda_l}, v_{\lambda_l}, m_{\ln(\gamma_{\lambda(l)}^2)}, v_{\ln(\gamma_{\lambda(l)}^2)}, \lambda_{j,l,1}, \ln\gamma_{\lambda(j,l)}^2\}_{l=0}^{5}$ looping over j: Kalman smoother draw from same SSS as in Step 3.

5. $\{m_{\ln\kappa^2}, v_{\ln\kappa^2}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^{q}, v_\xi, m_{\ln\gamma_\mu^2}, v_{\ln\gamma_\mu^2}, \{\mu_j, u_{j,-11:0}, \sigma_j, \{\xi_{j,l}\}_{l=1}^{q}, \gamma_{\mu(j)}^2, \kappa_j, v_{j,t}, \{\eta_{j,t}, \varepsilon_{j,t}\}_{t=1}^{T}\}_{j=1}^{n}\}$: Same as Step 2 of Model VI, except that in Step 2a, there is an additional contribution to the posterior from $\xi_{l,n+1} \sim \mathcal{N}(m_{\xi_l}, v_\xi)$, $l = 1, \ldots, q$, and in Steps 2b and 2c there are now $n+1$ processes.

6. $\{\{m_{\phi_l}, v_{\phi_l}, m_{\ln(\gamma_{\phi(l)}^2)}, v_{\ln(\gamma_{\phi(l)}^2)}\}_{l=1}^{12}, \{\{\gamma_{\phi(j,l)}^2\}_{l=1}^{12}, \{\{\phi_{j,l,t}\}_{l=1}^{12}, \varepsilon_{j,t}\}_{t=1}^{T}\}_{j=1}^{n+1}\}$: Same as Step 3 of Model VI, except that there are now $n+1$ processes.

7. $\{\sigma_{n+1}^2, \{\xi_{n+1,l}\}_{l=1}^{q}, \{u_{j,-11:0}, \{\varepsilon_{j,t}\}_{t=1}^{T}\}_{j=1}^{n+1}\}$: Conditional on $\{\sigma_{n+1}^2, \{\xi_{n+1,l}\}_{l=1}^{q}\}$, (26) and (9) imply an a priori mean-zero Gaussian distribution for $(u_{n+1,-11}, \ldots, u_{n+1,T})'$ with a band diagonal precision matrix with bandwidth 13. Furthermore, with $u_{j,t}^0 = (y_{j,t} - \mu_{j,t} - \mu_{n+1,t} - \omega(o_{j,t} + o_{n+1,t}))/\omega = u_{j,t} + \sum_{l=0}^{5}\lambda_{j,l,t}u_{n+1,t-l}$, $u_{j,t}^0 - \sum_{l=1}^{12}\phi_{j,l,t}u_{j,t-l}^0 = \sum_{l=0}^{5}\lambda_{j,l,t}u_{n+1,t-l} - \sum_{l=1}^{12}\phi_{j,l,t}\sum_{\ell=0}^{5}\lambda_{j,\ell,t-l}u_{n+1,t-\ell-l} + \epsilon_{j,t}$ for $t = 1, \ldots, T$, $j = 1, \ldots, n$ are independent Gaussian measurements of a linear combination of 18 consecutive elements of $u_{n+1,t}$. The precision matrix of the Gaussian posterior of $\{u_{n+1,t}\}_{t=-11}^{T}|\{\{u_{j,t}^0\}_{t=-11}^{T}\}_{j=1}^{n}$ is thus band diagonal with

bandwidth 18, and specialized linear algebra routines that exploit this band diagonal structure can be employed to efficiently generate a draw from the posterior. See for details. The same matrix calculations also enable to efficiently compute the likelihood of $\{\{u^0_{j,t}\}_{t=-11}^T\}_{j=1}^n$ as a function of $\{\ln\sigma^2_{n+1}, \{\xi_{n+1,l}\}_{l=1}^q\}$ without conditioning on $\{u_{n+1,t}\}_{t=-11}^T$, which we employ in a RW Metropolis step that precedes the draw from $\{u_{n+1,t}\}_{t=-11}^T|\{\{u^0_{j,t}\}_{t=-11}^T\}_{j=1}^n$.

8. $\{\kappa^2_{n+1}, \{\eta_{n+1,t}\}_{t=1}^T, \{\{\varepsilon_{j,t}\}_{t=1}^T\}_{j=1}^n\}$: With $u^0_{j,t} = (y_{j,t} - \mu_{j,t} - \mu_{n+1,t} - \omega(c_{j,t} + o_{j,t}))/\omega = u_{j,t} + o_{n+1,t}$, $u^0_{j,t} - \sum_{l=1}^{12}\phi_{j,l,t}u^0_{j,t-l} = o_{n+1,t} - \sum_{l=1}^{12}\phi_{j,l,t}o_{n+1,t-l} + \epsilon_{j,t}$ for $t = 1,\ldots,T$, $j = 1,\ldots,n$ with $o_{n+1,t} = \kappa_{n+1}\eta_{n+1,t}$ are independent Gaussian measurements of a linear combination of 13 consecutive elements of $o_{n+1,t}$. Furthermore, conditional on κ^2_{n+1}, $\{o_{n+1,t}\}_{t=1}^T$ has a diagonal prior precision matrix. We can thus use the same approach as in Step 7 to update κ^2_{n+1} and $\{\eta_{n+1,t}\}_{t=1}^T$ given $\{\{u^0_{j,t}\}_{t=1}^T\}_{j=1}^n$. [Step 8 is not necessary given Step 2, but it improves mixing.]

9. $\{\gamma^2_{\mu(n+1)}, \{\mu_{n+1,t}\}_{t=2}^T, \{\{\varepsilon_{j,t}\}_{t=2}^T\}_{j=1}^n\}$: With $u^0_{j,t} = (y_{j,t} - \mu_{j,t} - \omega(o_{j,t} + o_{n+1,t} + c_{j,t}))/\omega = u_{j,t} + \mu_{n+1,t}/\omega$, $u^0_{j,t} - \sum_{l=1}^{12}\phi_{j,l,t}u^0_{j,t-l} = \mu_{n+1,t}/\omega - \sum_{l=1}^{12}\phi_{j,l,t}\mu_{n+1,t-l} + \epsilon_{j,t}$ for $t = 2,\ldots,T$, $j = 1,\ldots,n$ are independent Gaussian measurements of a linear combination of 13 consecutive elements of $\mu_{n+1,t}/\omega$. Furthermore, conditional on $\gamma^2_{\mu(n+1)}$, $\{\mu_{n+1,t}\}_{t=2}^T$ has a band diagonal precision matrix with bandwidth 2. We can thus use the same approach as in Step 7 to update $\gamma^2_{\mu(n+1)}$ and $\{\mu_{n+1,t}\}_{t=2}^T$ given $\{\{u^0_{j,t}\}_{t=2}^T\}_{j=1}^n$. [Step 9 is not necessary given Step 3, but it improves mixing.]

10. $\{m_{\ln(\nu-2)}, v_{\ln(\nu-2)}, m_{\ln\sigma^2}, v_{\ln\sigma^2}, \{m_{\xi_l}\}_{l=1}^q, v_\xi, \{\nu_j, \sigma^2_j, \{\xi_{j,l}\}_{l=1}^q, \{S_{j,t}\}_{t=1}^T\}_{j=1}^{n+1}\}$: Same as Step 2 of Model V, except that there are now $n+1$ processes, and in Step 2a, σ^2_{n+1} does not follow the hierarchical prior.

11. $\{m_{\ln(\nu^o-2)}, v_{\ln(\nu^o-2)}, m_{\ln\kappa^2}, v_{\ln\kappa^2}, \{\nu^o_j, \kappa_j, \{S^o_{j,t}\}_{t=1}^T\}_{j=1}^{n+1}\}$: Same as Step 3 of Model IV, except that there are now $n+1$ processes, and in Step 4a, κ^2_{n+1} does not follow the hierarchical prior.

B.3 Bayes Factors

The Bayes factors were obtained by using the bridge sampling approach of Meng and Wong (1996): For two models A and B with the same parameter space Θ, priors π_A and π_B and likelihood $f_A(y|\theta)$ and $f_B(y|\theta)$, respectively, define $\mathrm{LR}_{A/B}(\theta) = \frac{f_A(y|\theta)\pi_A(\theta)}{f_B(y|\theta)\pi_B(\theta)}$ and $\mathrm{LR}_{B/A}(\theta) = \frac{f_B(y|\theta)\pi_B(\theta)}{f_A(y|\theta)\pi_A(\theta)}$. Then the Bayes factor BF can be written as

$$\mathrm{BF} = \frac{\int f_A(y|\theta)\pi_A(\theta)d\theta}{\int f_B(y|\theta)\pi_B(\theta)d\theta} = \frac{\mathbb{E}_B[\frac{\mathrm{LR}_{A/B}(\theta)}{\mathrm{LR}_{A/B}(\theta)+\mathrm{BF}}]}{\mathbb{E}_A[\frac{\mathrm{LR}_{B/A}(\theta)}{1+\mathrm{BF}\cdot\mathrm{LR}_{B/A}(\theta)}]}$$

where $E_A[\cdot]$ and $E_B[\cdot]$ are expectations with respect to the posterior of θ under model A and B, respectively. The Bayes factors can then be obtained by replacing the posterior expectations by averages from MCMC output, and by iterating the above to convergence.

It is useful to "bridge" the gap between the baseline model A_0 and a fairly distinct alternative model A_k by intermediate models via the identity $\mathrm{BF}_{A_0/A_k} = \prod_{i=1}^{k} \mathrm{BF}_{A_{i-1}/A_i}$ in obvious notation, as this improves numerical stability of the estimate. In our application, we set $k = 3$, so there is one intermediate model.