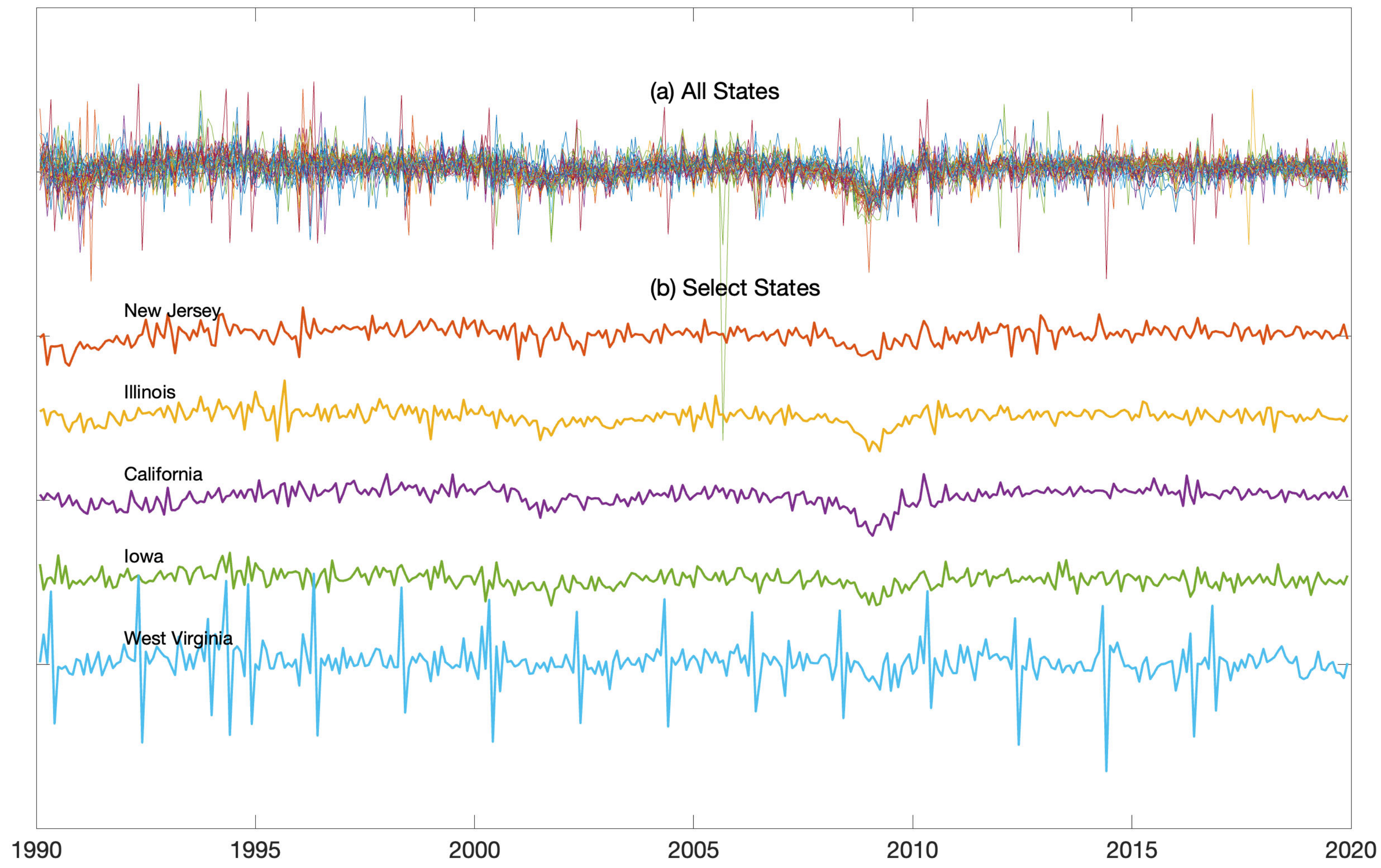
Forecasting Related Time Series

Ulrich K. Müller and Mark W. Watson
Princeton University

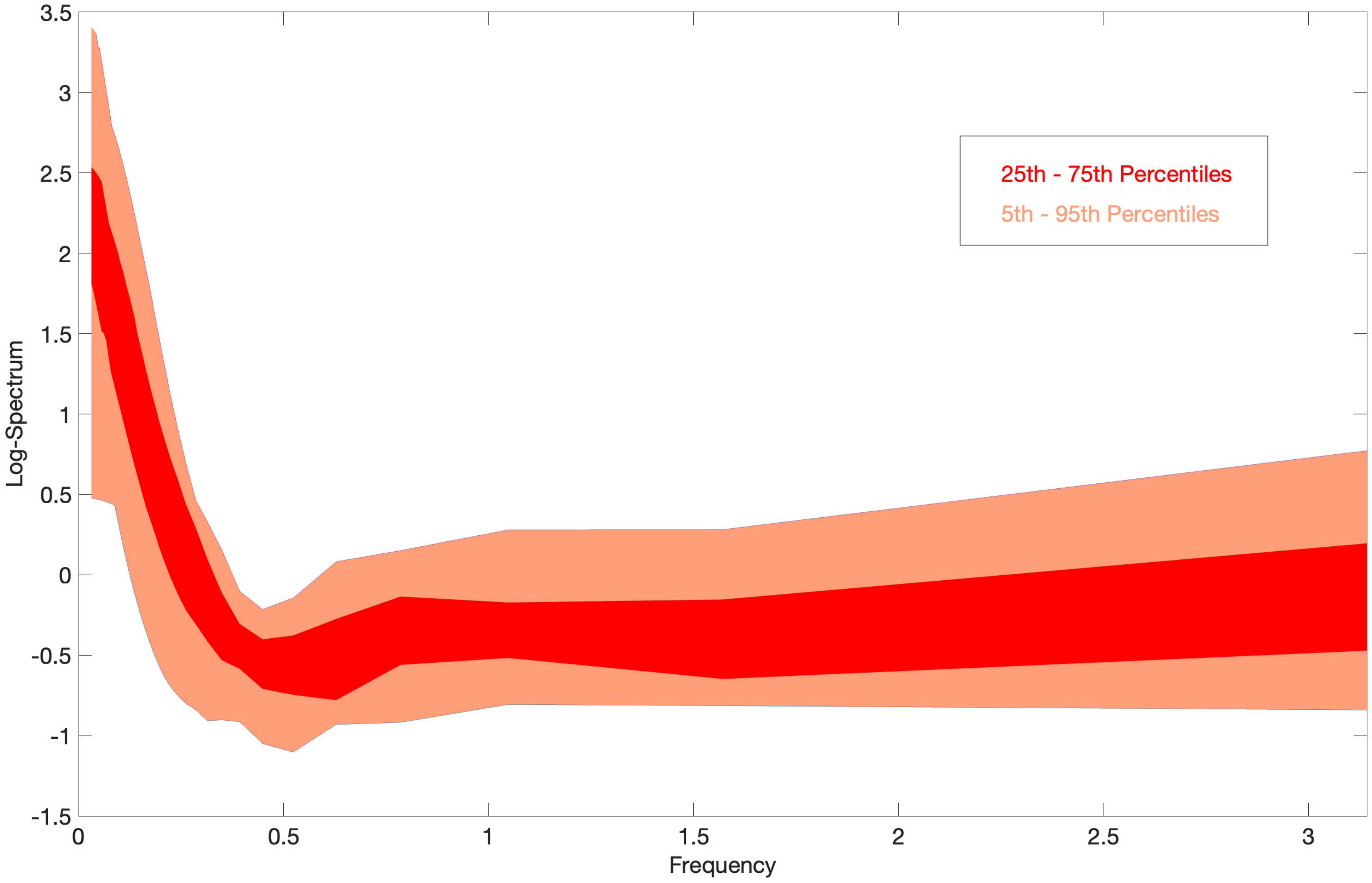
Work in Progress

May 2025

Employment Growth in 50 US States



Estimated Spectra of Employment Growth in 50 US States



This Paper

- Bayesian forecasts of n related time series with
  - Hierarchical priors (=shrink towards common parameter values)
  - Fat-tailed shocks and stochastic volatility (=downweighs outliers and times of turmoil)
  - Time varying parameters (=puts more weight on more recent data when forming forecasts)
  - Common factor (=captures cross sectional dependence)
- Three applications with monthly data
  - Employment in $n = 50$ U.S. states
  - Industrial production in $n = 16$ European countries
  - U.S. inflation series in $n = 17$ sectors

Results

- Modern computing power allows for fast and accurate MCMC Bayesian inference

  With $n = 50$ and $T = 400$ estimation of full model takes about 50 seconds (in Fortran)

- Promising empirical performance

  Identical model systematically outperforms univariate OLS AR(12) forecasts in three applications, often by substantial margin

Related Literature

- Large literature on Bayesian time series models

  Doan, Litterman, Sims (1984), Litterman (1986), Sims (1993), Cogley and Sargent (2005), Banbura, Giannone, Reichlin (2010), Giannone, Lenza, Primiceri (2015), Carriero, Clark, Marcellino (2015), Chan (2022), etc.

- Numerically efficient posterior draws of state in linear state space system

  Durbin and Koopman (2002), Chan and Jeliazkov (2009), etc.

Outline of Presentation

1. Base model and its complications, illustrated in U.S. employment application

   Hierarchical priors, Student-t errors, additive outliers, stochastic volatility, time varying parameters, common factors

2. Additional applications

3. Performance of Quantile Forecasts

4. Conclusion

Basic Gaussian AR(12) Model

- Data y_{jt}, $j = 1, \ldots, n$, $t = 1, \ldots, T$

$$\begin{aligned} y_{jt} &= \mu_j + u_{jt} \\ u_{jt} &= \sum_{l=1}^{12} \phi_{jl} u_{j,t-l} + \sigma_j \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim iid\mathcal{N}(0,1) \\ u_{j,-11:0} &\sim \mathcal{N}(0, \sigma_j^2 \Sigma(\phi_j)) \end{aligned}$$

  where $\Sigma(\phi)$ is unconditional covariance of stationary AR(12) with (scaled) coefficient ϕ

  [ignore additional common scale to all series to streamline presentation]

- Priors (all independent)
  - Flat on μ_j
  - Minnesota prior $\phi_{jl} \sim \mathcal{N}(0, \frac{0.2^2}{l^2})$
  - $\ln \sigma_j^2 \sim \mathcal{N}(0,1)$

Algorithm for Posterior Draws

- Recall model

$$
\begin{aligned}
y_{jt} &= \mu_j + u_{jt} \\
u_{jt} &= \sum_{l=1}^{12} \phi_{jl} u_{j,t-l} + \sigma_j \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim iid\mathcal{N}(0,1) \\
u_{j,-11:0} &\sim \mathcal{N}(0, \sigma_j^2 \Sigma(\phi_j))
\end{aligned}
$$

- Algorithm: (i) Draw μ_j and $u_{j,-11:0}$ conditional on (σ_j^2, ϕ_j) by Kalman smoother (using Durbin Koopman (2002) method)

  (ii) Draw $\phi_j|(\mu_j, u_{j,-11:0}, \sigma_j^2)$ from conditional Gaussian [this requires a Metropolis-Hastings correction due to the conditioning on $u_{j,-11:0} \sim \mathcal{N}(0, \sigma_j^2 \Sigma(\phi_j))$]

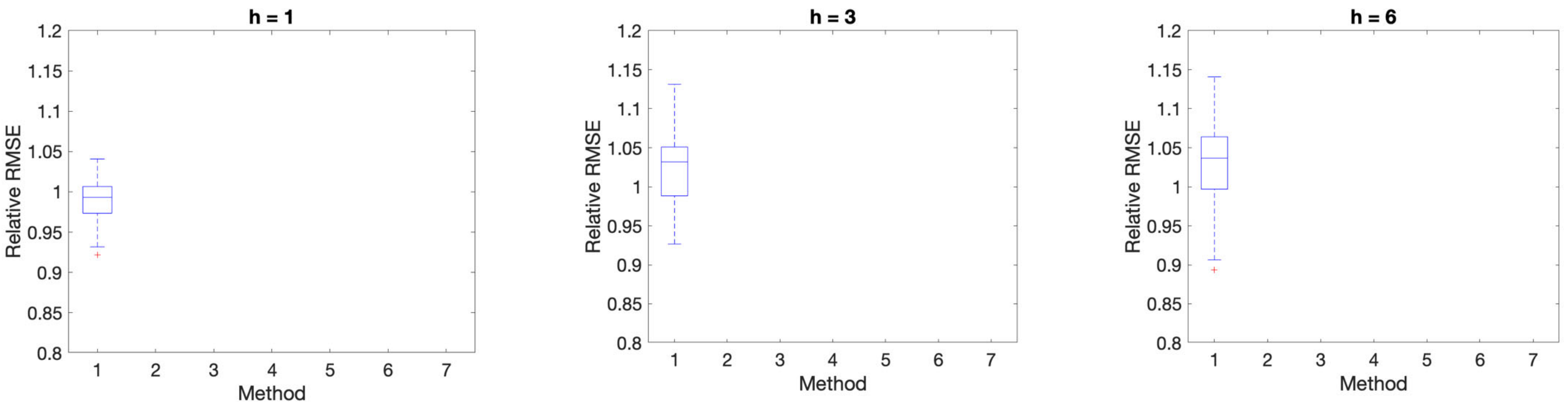
  (iii) Draw $\sigma_j^2|(\mu_j, u_{j,-11:0}, \phi_j)$ via Metropolis algorithm

Out-of-Sample Forecasting Performance, Employment Data

- For $T = 1999{:}12$ to $T = 2019{:}6$, use data $t = 1990{:}2, \ldots, T$ to compute
  - OLS AR(12) forecasts for $t = T + 1, \ldots, t = T + 6$
  - Bayes posterior mean forecasts for $t = T + 1, \ldots, t = T + 6$ (with 1500 MCMC draws, first 500 discarded)
- For horizons, $h = 1, 3, 6$ months, consider average future values
  - compute RMSE relative to OLS benchmark pooled over 50 states
  - for each state $i = 1, \ldots, 50$, compute RMSE relative to OLS benchmark

Results Relative to OLS, Employment Data

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP			
Bayes T			
Bayes AO			
Bayes SV			
Bayes TVP			
Bayes F			



Hierarchical Priors

- Instead of prior $\phi_{jl} \sim iid\mathcal{N}(0, \frac{0.2^2}{l^2})$, use

$$\begin{aligned} \phi_{jl} &\sim iid\mathcal{N}(m_l, V_l) \\ m_l &\sim \mathcal{N}(0, \tfrac{0.2^2}{l^2}) \qquad \ln V_l \sim \mathcal{N}(\ln(\tfrac{0.2^2}{l^2}), 1) \end{aligned}$$

  $\Rightarrow$ Shrink towards common value m_l, with amount of shrinkage also determined by data

- Algorithm: Joint Metropolis step for (m, V), conditional on 'z-scores' $\{z_l\}_{j=1}^n = \{(\phi_{jl} - m_l)/\sqrt{V_l}\}_{j=1}^n$, so evaluate posterior density at

$$\phi_{jl}^p = m_l^p + \sqrt{V_l^p} \frac{\phi_{jl}^c - m_l^c}{\sqrt{V_l^c}} \qquad \text{(so } z_l^p = z_l^c\text{)}$$

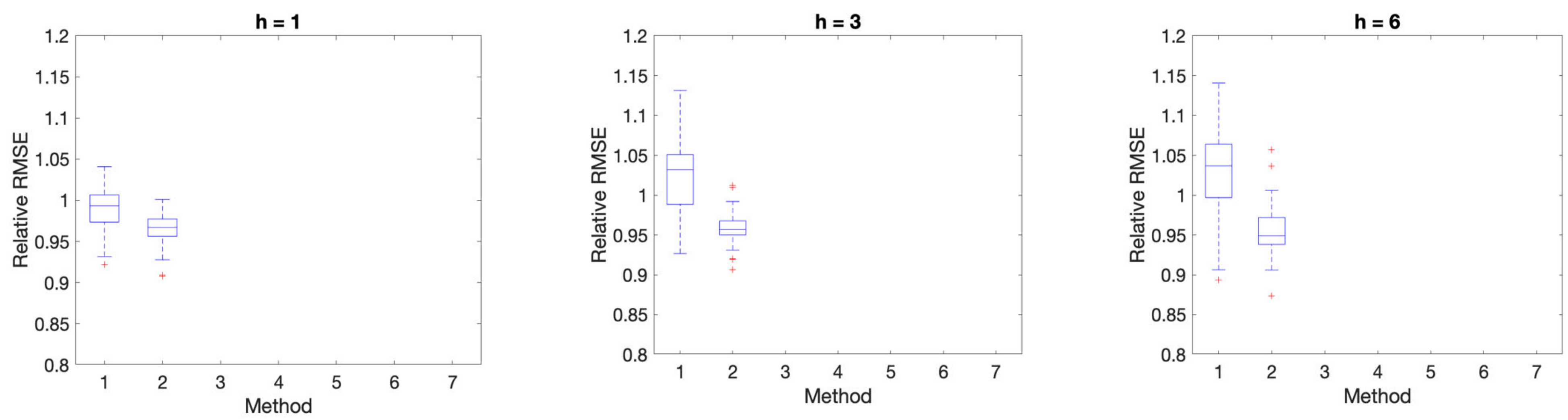
  to see whether we want to jump from $(m^c, V^c, \{\phi_j^c\}_{j=1}^n)$ to $(m^p, V^p, \{\phi_j^p\}_{j=1}^n)$

  [In contrast, textbook algorithm conditions on $\{\phi_j\}_{j=1}^n$ when updating (m, V)]

- Same hierarchical structure for σ_j^2

Results Relative to OLS, Employment Data

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T			
Bayes AO			
Bayes SV			
Bayes TVP			
Bayes F			



Student-t Innovations

- Now

$$u_{jt} = \sum_{l=1}^{12} \phi_{jl} u_{j,t-l} + \sigma_j \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim iid\mathcal{T}(\nu_j)$$

where $\mathcal{T}(\nu_j)$ is student-t with ν degrees of freedom with hierarchical prior $\ln(\nu_j - 2) \sim \mathcal{N}(m_\nu, V_\nu)$, $m_\nu \sim \mathcal{N}(\ln 10, 1)$, $\ln V_\nu \sim \mathcal{N}(\ln 0.5^2, 1)$

- Note that $\varepsilon_{jt} \sim Z_{jt}/\sqrt{S_{jt}}$ where $Z_{jt} \sim iid\mathcal{N}(0,1)$ and $\nu S_{jt} \sim iid\chi^2_\nu$

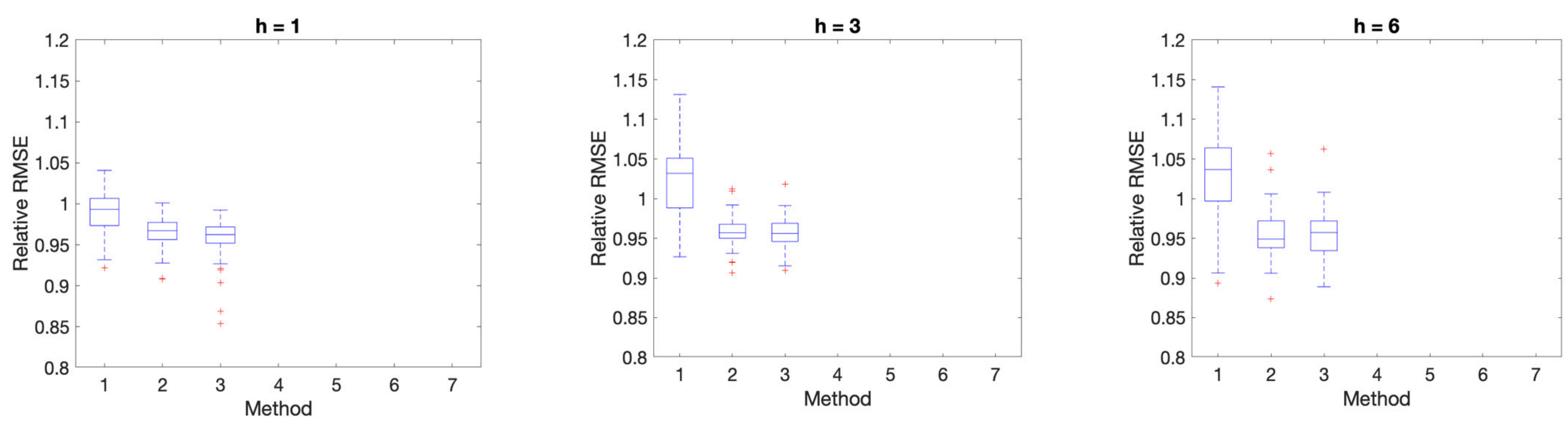
- Algorithm: (i) Draw prior (m_ν, V_ν) via 'z-scores' as above based on student-t likelihood

  (ii) Metropolis step for ν_j conditional on prior based on student-t likelihood

  (iii) Conditional on ν_j, draw $S_{jt}|\varepsilon_{jt}$, then condition on S_{jt} in all other steps so we recover conditionally Gaussian likelihood

Results Relative to OLS, Employment Data

Model	Pooled RMSE $h = 1$	Pooled RMSE $h = 3$	Pooled RMSE $h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T	0.94	0.94	0.96
Bayes AO			
Bayes SV			
Bayes TVP			
Bayes F			



Additive Outliers

- Now

$$y_{jt} = \mu_j + u_{jt} \quad + \kappa_j e_{jt}, \qquad e_{jt} \sim iid\mathcal{T}(2)$$

with hierarchical prior on $\ln \kappa_j^2 \sim \mathcal{N}(m_\kappa, V_\kappa)$, $m_\kappa \sim \mathcal{N}(\ln 0.1^2, 1)$, $\ln V_\kappa \sim \mathcal{N}(\ln 0.3^2, 1)$

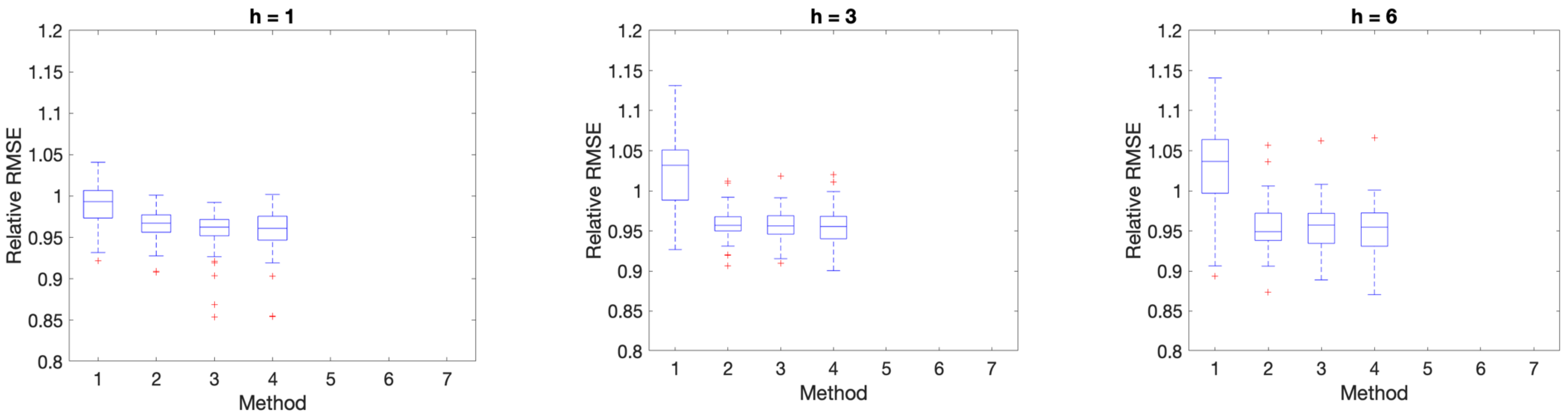
(allows for outliers that do not feed into autoregression)

- Algorithm: (i) Draw prior (m_ν, V_ν) as above based on Kalman filter, so no conditioning on realizations of $\{\varepsilon_{jt}, e_{jt}\}_{t=1}^T$ and μ_j

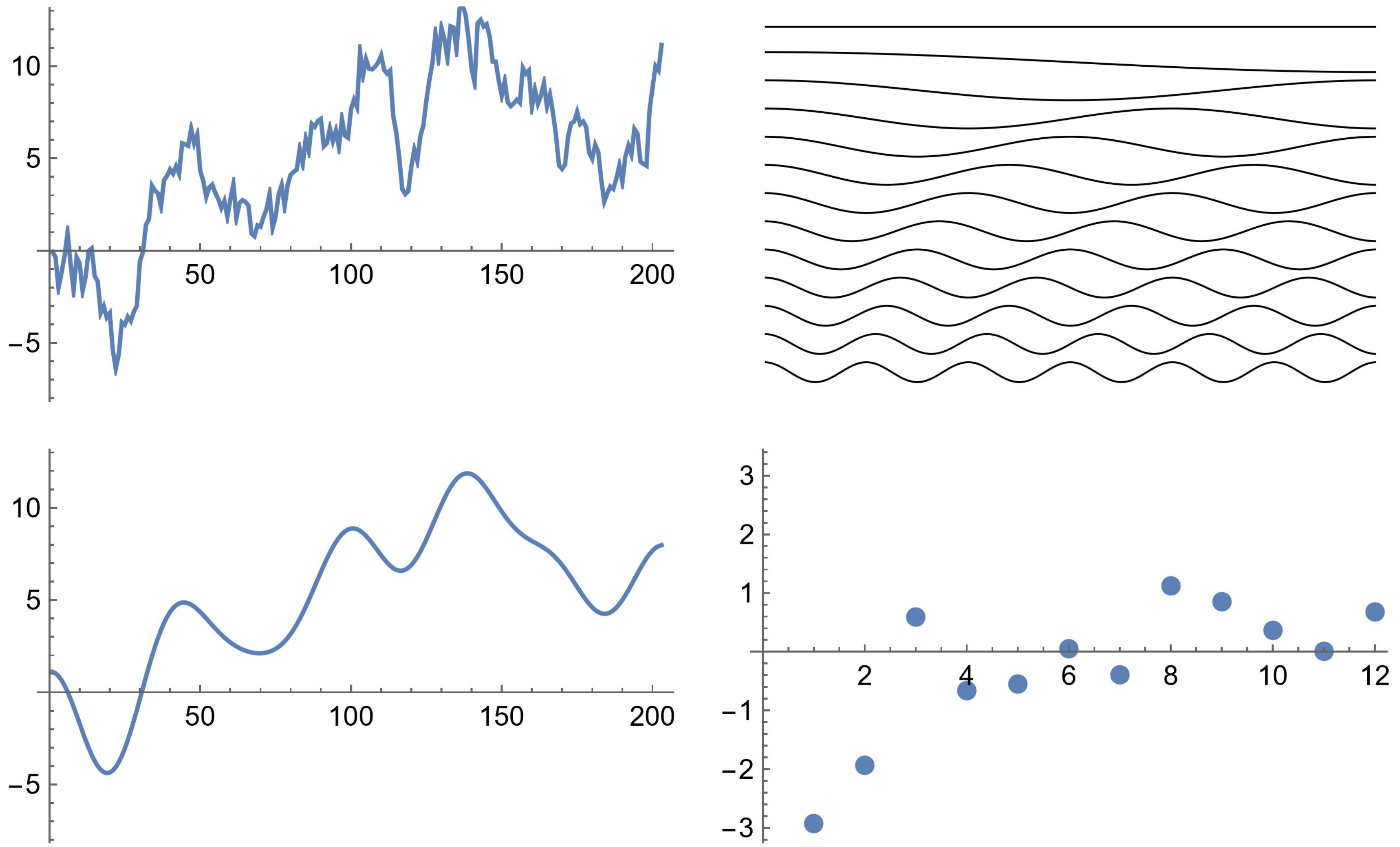
(ii) Draw $\{\mu_j, u_{jt}, e_j\}$ and $u_{j,-11:0}$ conditional on $(\sigma_j^2, \phi_j, \kappa^2)$ via Kalman smoother

Results Relative to OLS, Employment Data

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T	0.94	0.94	0.96
Bayes AO	0.93	0.93	0.95
Bayes SV			
Bayes TVP			
Bayes F			



A Low-Dimensional Approximation to Random Walk



Stochastic Volatility

- Now

$$u_{jt} = \sum_{l=1}^{12} \phi_{jl} u_{j,t-l} + \exp(\tfrac{1}{2} h_{jt}) \varepsilon_{jt}, \quad \varepsilon_{jt} \sim iid\mathcal{T}(\nu_j)$$

where

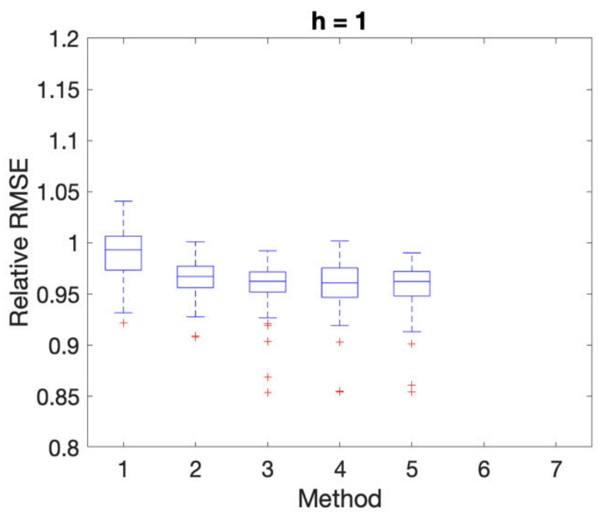
$$h_{jt} = \ln \sigma_j^2 + \sum_{l=1}^{q} \varphi_{lt} \xi_{jl}, \quad \xi_{jl} \sim iid\mathcal{N}(m_l^{\xi}, V_{jl}^{\xi})$$

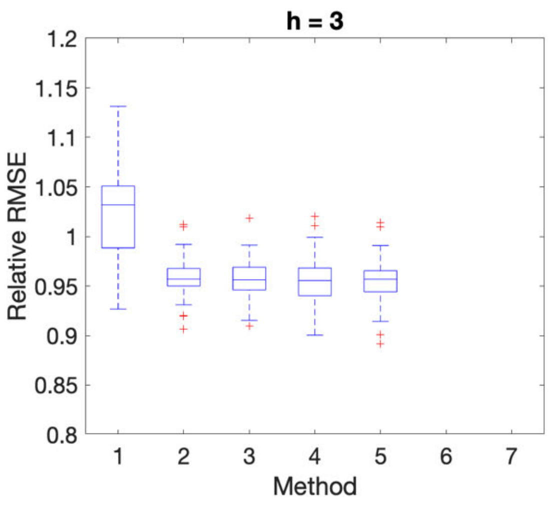
with φ_{lt} suitably scaled sinusoidal functions, $q = \lfloor T/36 \rfloor$ and hierarchical prior on V_{jl}^{ξ}

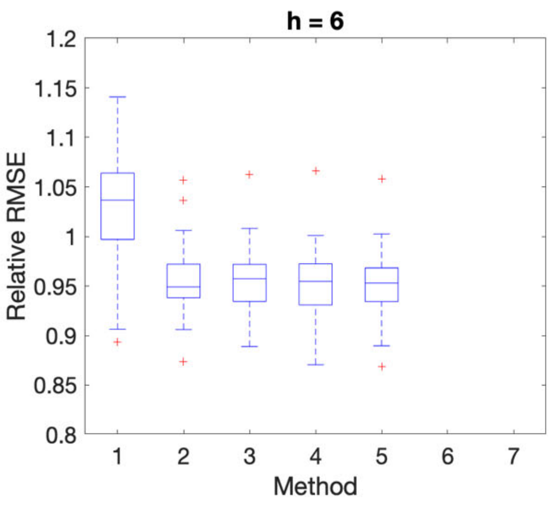
- Algorithm: as above, vector version of drawing (prior on) σ_j^2

Results Relative to OLS, Employment Data

Model	Pooled RMSE $h=1$	$h=3$	$h=6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T	0.94	0.94	0.96
Bayes AO	0.93	0.93	0.95
Bayes SV	0.93	0.93	0.95
Bayes TVP			
Bayes F			







Time Varying Parameters

- Now

$$
\begin{aligned}
y_{jt} &= \mu_{jt} + u_{jt} + \kappa_j e_{jt} \\
u_{jt} &= \sum_{l=1}^{12} \phi_{jlt} u_{j,t-l} + \exp(\tfrac{1}{2} h_{jt})\varepsilon_{jt}, \qquad \varepsilon_{jt} \sim iid\mathcal{T}(\nu_j)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{jt}|\mu_{j,t-1} &\sim \mathcal{N}(\mu_{jt}, V_j^{\Delta\mu}) \qquad \phi_{jlt}|\phi_{jl,t-1} \sim \mathcal{N}(\phi_{jl,t-1}, V_{jl}^{\Delta\phi}) \\
\mu_{j0} &\sim \mathcal{N}(0, \infty) \qquad \phi_{jl0} \sim \mathcal{N}(m_{jl}^{\phi}, V_{jl}^{\phi})
\end{aligned}
$$

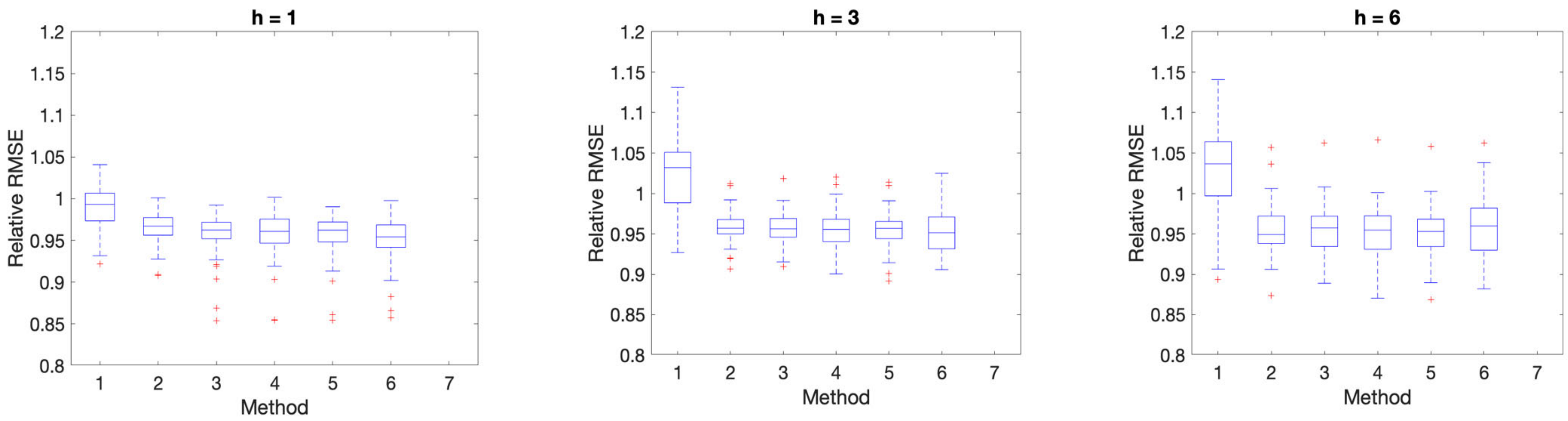
and hierarchical prior on $(V_j^{\Delta\mu}, V_{jl}^{\Delta\phi})$

- Algorithm for $\{\phi_{jlt}\}$: (i) For Metropolis step for (prior on) $(m_{jl}^{\phi}, V_{jl}^{\phi}, V_{jl}^{\Delta\phi})$, use Kalman filter to obtain likelihood, that is integrate out possible paths

  (ii) Then draw paths $\{\phi_{jlt}\}_{t=1}^{T}$ conditional on $(m_{jl}^{\phi}, V_{jl}^{\phi}, V_{jl}^{\Delta\phi})$ by Kalman smoothing, as before

Results Relative to OLS, Employment Data

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T	0.94	0.94	0.96
Bayes AO	0.93	0.93	0.95
Bayes SV	0.93	0.93	0.95
Bayes TVP	0.93	0.93	0.95
Bayes F			



Common Factor

- Now

$$
\begin{aligned}
y_{jt} &= \mu_{jt} + u_{jt} + \kappa_j e_{jt} + \sum_{l=0}^{5} \lambda_{jlt} f_{t-l} + \mu_t^g + \kappa^g e_t^g \\
f_t &\sim u_{n+1,t}, \;\; \mu_t^g \sim \mu_{n+1,t}, \;\; \kappa^g e_t^g \sim \kappa_{n+1} e_{n+1,t}
\end{aligned}
$$

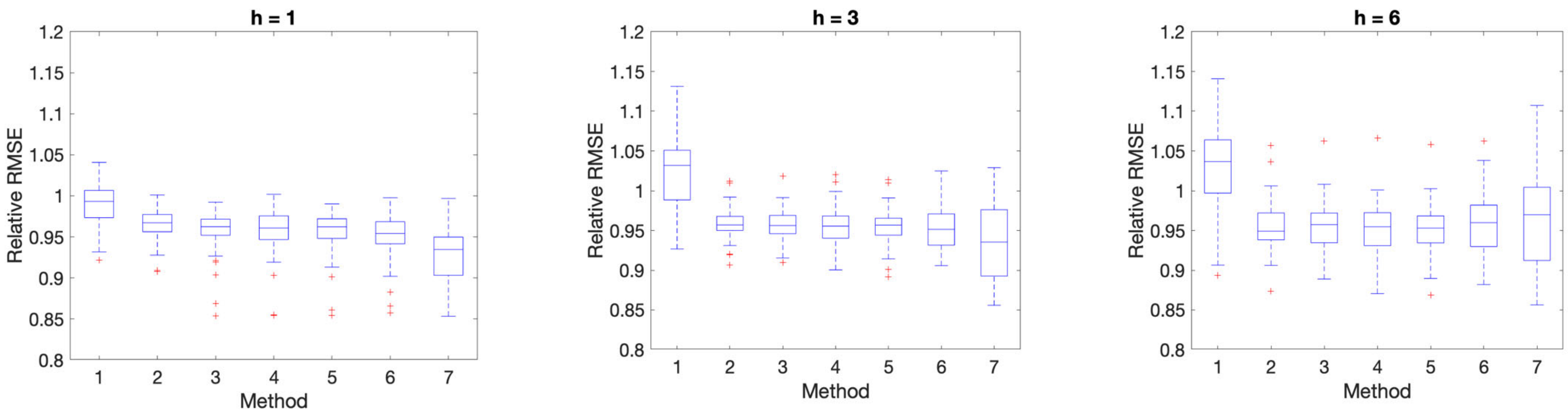
  $\Rightarrow$ common factor, common random walk and common additive outliers modelled as additional individual series in previous model (including sharing the same hierarchical prior). Hierarchical priors on random walks $\{\lambda_{jlt}\}_{t=1}^{T}$ that shrink towards 1 for contemporaneous ($l = 0$) loading, and towards zero otherwise.

- Algorithm: (i) For $\{f_t\}_{t=1}^{T}$, exploit that conditional on loadings $\{\lambda_{jlt}\}_{t=1}^{T}$ posterior is Gaussian with $T \times T$ band precision matrix, so use specialized linear algebra routines to generate appropriate draws (cf. Chan and Jeliazkov (2009))

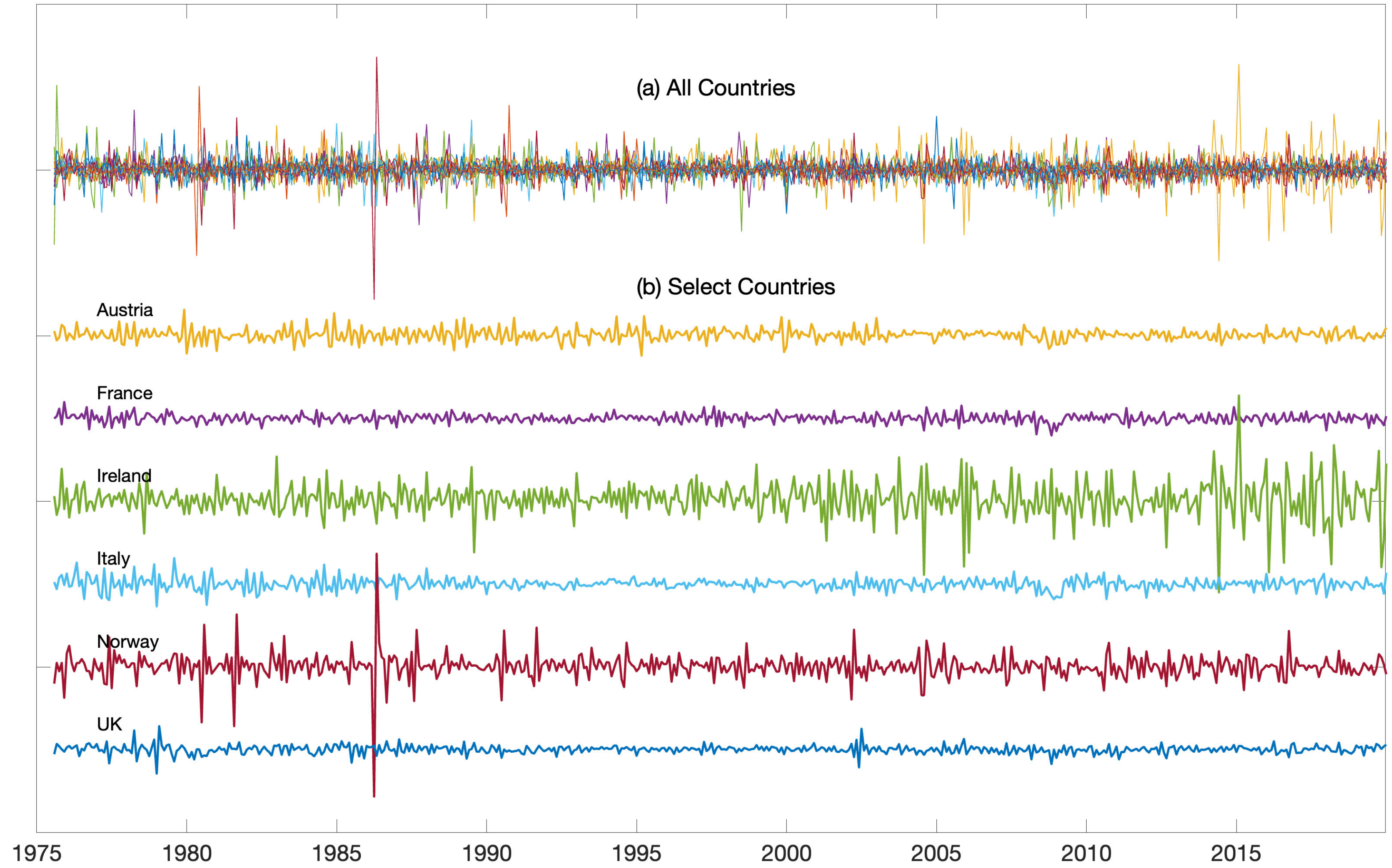
  (ii) Use Kalman filter likelihood to update variances of $\{\lambda_{jlt}\}_{t=1}^{T}$ conditional on $\{f_t\}_{t=1}^{T}$, followed by draw from time-path $\{\lambda_{jlt}\}_{t=1}^{T}$

Results Relative to OLS, Employment Data

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.97	0.99	1.02
Bayes HP	0.95	0.94	0.96
Bayes T	0.94	0.94	0.96
Bayes AO	0.93	0.93	0.95
Bayes SV	0.93	0.93	0.95
Bayes TVP	0.93	0.93	0.95
Bayes F	0.90	0.90	0.95

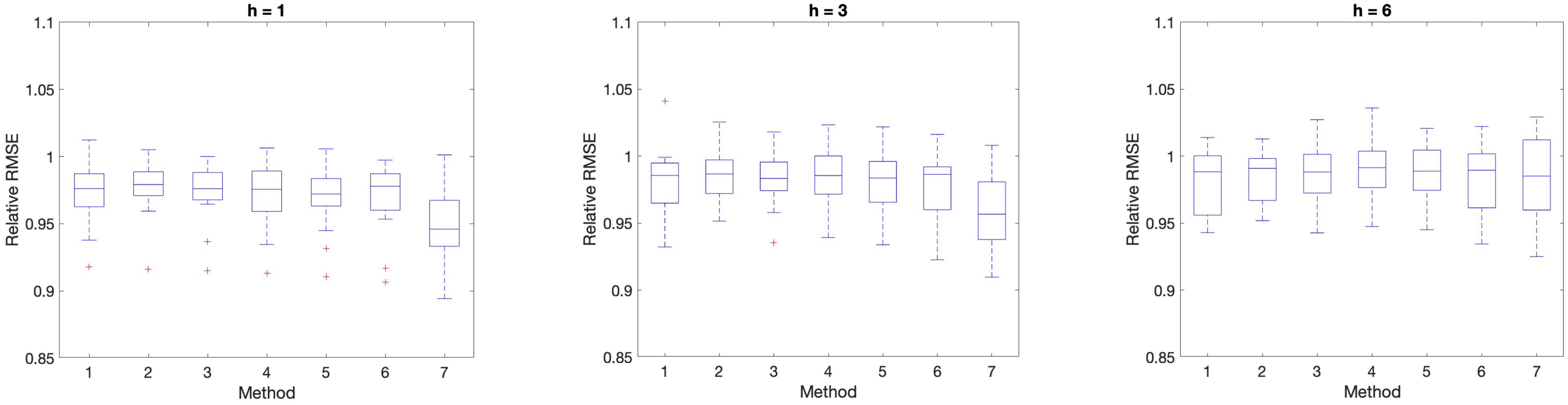


Industrial Production in 16 European Countries

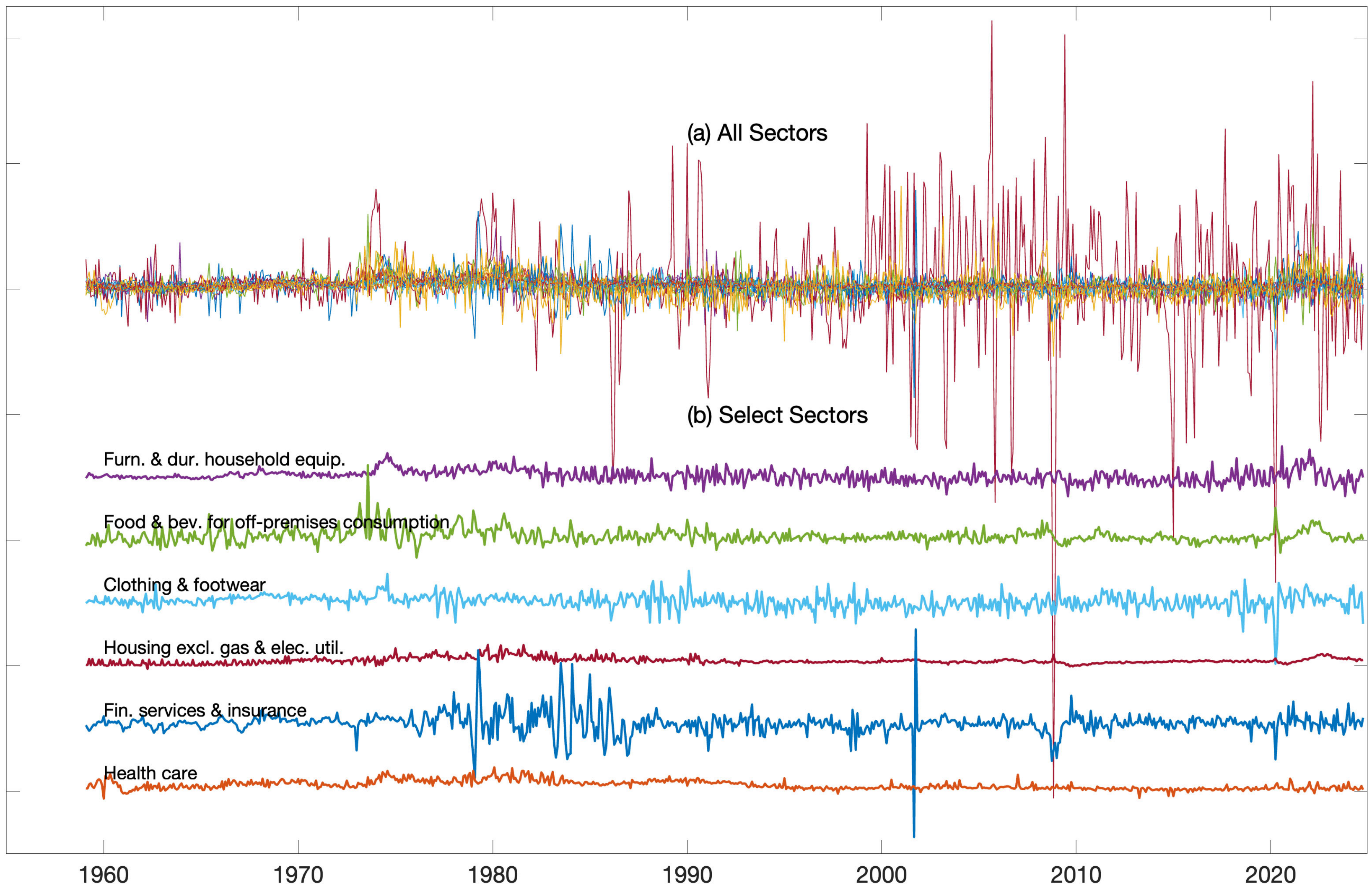


Results Relative to OLS, European IP Data 1985:6-2019:6

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.98	0.99	0.98
Bayes HP	0.98	0.99	0.98
Bayes T	0.97	0.99	0.99
Bayes AO	0.97	0.98	0.99
Bayes SV	0.97	0.99	0.99
Bayes TVP	0.97	0.98	0.98
Bayes F	0.95	0.96	0.98

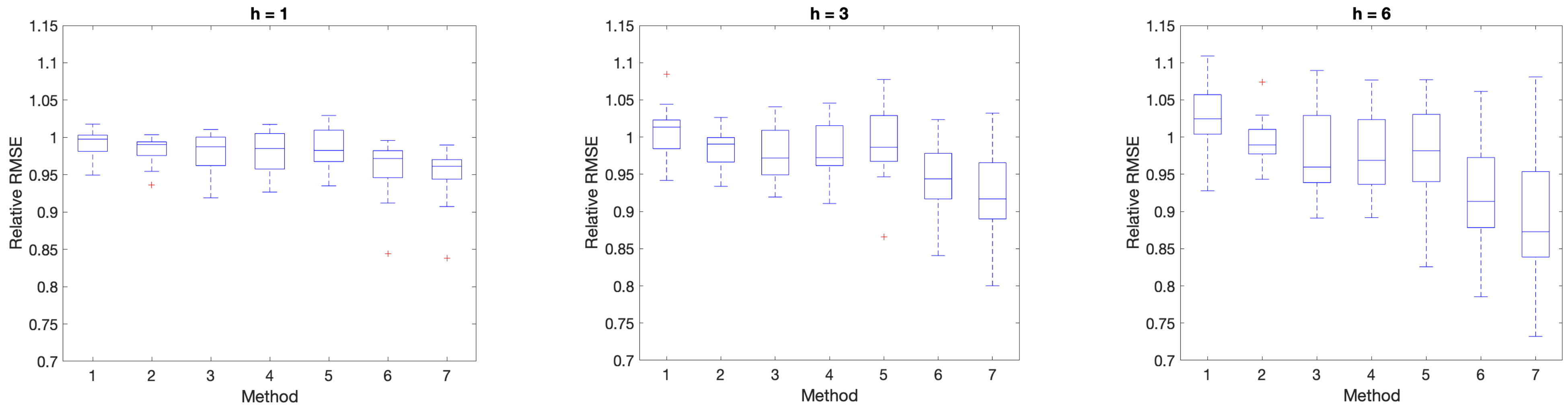


US Inflation in 17 Sectors



Results Relative to OLS, US Inflation Data 1984:12-2019:6

	Pooled RMSE		
Model	$h = 1$	$h = 3$	$h = 6$
Bayes B	0.96	0.96	0.97
Bayes HP	0.96	0.98	1.01
Bayes T	0.94	0.97	1.01
Bayes AO	0.94	0.97	1.01
Bayes SV	0.96	1.00	1.02
Bayes TVP	0.95	0.95	0.96
Bayes F	0.95	0.96	0.96



Results Relative to OLS, US Inflation Data 1984:12-2019:6

Model	$h = 1$	$h = 3$	$h = 6$
		Pooled RMSE	
Bayes F	0.95	0.96	0.96
NYFed	0.95	1.03	1.15
	Aggregate Inflation ('Headline' – All Sectors)		
Bayes F	0.94	0.92	0.87
NYFed	1.00	0.99	1.01
	Aggregate Inflation ('Core' – excl. Food and Energy)		
Bayes F	0.90	0.78	0.69
NYFed	0.93	0.83	0.76

Performance of Bayes Quantile Forecasts

- Compute OLS quantile forecasts of average future values

  $\Rightarrow$ Assume $iid\mathcal{N}(0, \sigma^2)$ innovations with σ^2 estimated by sample variance of residuals

- For each draw of parameters of Bayes model, iterate model forward to obtain posterior distribution of average future values, then compute quantiles

- Compare quality of quantile forecasts by average of quantile loss function

$$\ell_q(y, \hat{q}) = \begin{cases} q|y - \hat{q}| & \text{for } y \geq q \\ (1-q)|y - \hat{q}| & \text{for } y < q \end{cases}$$

  $\Rightarrow$ average loss minimizing $\hat{q}$ for this loss function is the qth quantile of distribution of y

- Same sample periods as for RMSE comparisons

Pooled Quantile Losses Relative to OLS, Employment Data

	$h=1$					$h=3$				
Quantile	0.05	0.33	0.50	0.67	0.95	0.05	0.33	0.50	0.67	0.95
Bayes B	1.00	0.99	0.98	0.99	0.99	1.04	1.02	1.00	1.00	0.98
Bayes HP	0.98	0.97	0.96	0.96	0.97	0.93	0.96	0.96	0.96	0.96
Bayes T	0.97	0.95	0.95	0.94	0.96	0.92	0.95	0.96	0.96	0.97
Bayes AO	0.97	0.95	0.95	0.94	0.96	0.92	0.95	0.96	0.95	0.96
Bayes SV	0.96	0.95	0.95	0.93	0.93	0.92	0.95	0.95	0.94	0.91
Bayes TVP	0.95	0.94	0.94	0.93	0.93	0.89	0.94	0.95	0.94	0.92
Bayes F	0.91	0.91	0.92	0.91	0.92	0.83	0.91	0.94	0.95	0.93
	$h=6$									
Bayes B	1.08	1.04	1.00	0.98	0.93					
Bayes HP	0.89	0.96	0.96	0.96	0.94					
Bayes T	0.88	0.95	0.96	0.97	0.96					
Bayes AO	0.87	0.95	0.96	0.96	0.94					
Bayes SV	0.89	0.95	0.95	0.94	0.89					
Bayes TVP	0.86	0.94	0.96	0.96	0.90					
Bayes F	0.84	0.94	0.98	0.99	0.90					

Pooled Quantile Losses Relative to OLS, European IP Data

	$h=1$					$h=3$				
Quantile	0.05	0.33	0.50	0.67	0.95	0.05	0.33	0.50	0.67	0.95
Bayes B	0.99	0.99	0.98	0.98	0.99	1.01	0.99	0.99	1.00	1.01
Bayes HP	0.99	0.99	0.98	0.98	0.99	1.01	0.99	0.99	0.99	0.99
Bayes T	0.98	0.97	0.98	0.96	0.96	1.00	0.99	0.99	0.98	0.99
Bayes AO	0.97	0.97	0.98	0.96	0.96	0.99	0.99	0.99	0.98	0.98
Bayes SV	0.96	0.97	0.98	0.96	0.94	0.98	0.99	0.99	0.98	0.96
Bayes TVP	0.96	0.97	0.97	0.96	0.93	0.99	0.98	0.98	0.97	0.94
Bayes F	0.92	0.95	0.96	0.95	0.91	0.94	0.96	0.96	0.95	0.93
	$h=6$									
Bayes B	1.00	0.99	0.99	1.01	1.02					
Bayes HP	1.00	0.99	0.99	1.00	1.00					
Bayes T	0.98	0.99	1.00	1.00	1.00					
Bayes AO	0.98	1.00	1.00	1.00	0.99					
Bayes SV	0.98	0.99	1.00	0.99	0.96					
Bayes TVP	0.98	0.99	0.99	0.98	0.96					
Bayes F	0.94	0.97	0.97	0.97	0.94					

Pooled Quantile Losses Relative to OLS, US Inflation

	$h = 1$					$h = 3$				
Quantile	0.05	0.33	0.50	0.67	0.95	0.05	0.33	0.50	0.67	0.95
Bayes B	1.01	0.99	0.99	0.99	1.00	1.00	0.99	1.00	1.00	0.99
Bayes HP	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99
Bayes T	1.05	0.99	0.99	0.98	1.06	1.01	1.00	0.99	0.98	1.03
Bayes AO	1.04	0.99	0.99	0.98	1.06	1.01	1.00	0.99	0.98	1.03
Bayes SV	0.90	0.99	0.99	0.98	0.88	0.94	1.01	1.00	0.98	0.92
Bayes TVP	0.86	0.96	0.97	0.95	0.88	0.88	0.96	0.96	0.94	0.88
Bayes F	0.83	0.95	0.96	0.94	0.85	0.86	0.96	0.95	0.92	0.84
	$h = 6$									
Bayes B	0.99	1.00	1.02	1.02	1.00					
Bayes HP	0.99	1.00	1.00	1.01	1.02					
Bayes T	0.99	1.00	1.00	0.99	1.09					
Bayes AO	0.98	1.00	1.00	0.99	1.07					
Bayes SV	0.95	1.02	1.00	0.99	1.02					
Bayes TVP	0.89	0.96	0.95	0.93	0.94					
Bayes F	0.86	0.96	0.93	0.89	0.88					

Summary

- Developed fully-fledged generic Bayesian model for related time series

- Promising MSE and quantile forecast performance in three examples

Thank you!