Forecasting Related Time Series

Ulrich Müller and Mark Watson

IIF MacroFor Seminar
September 16, 2025

Numbers reported here are preliminary

Three Sets of Related Time Series

- Employment Growth Rates for n = 51 U.S. States (+DC)
- Industrial Production Growth Rates for n = 16 European Countries
- Inflation Rates for n = 17 Consumption Sectors

Three Sets of Related Time Series

- Employment Growth Rates for n = 51 U.S. States (+DC)
- Industrial Production Growth Rates for n = 16 European Countries
- Inflation Rates for n = 17 Consumption Sectors

How Are They Related?

1. Similar Series, so similar DGPs
2. Correlated

Three Sets of Related Time Series

- Employment Growth Rates for n = 51 U.S. States (+DC)
- Industrial Production Growth Rates for n = 16 European Countries
- Inflation Rates for n = 17 Consumption Sectors

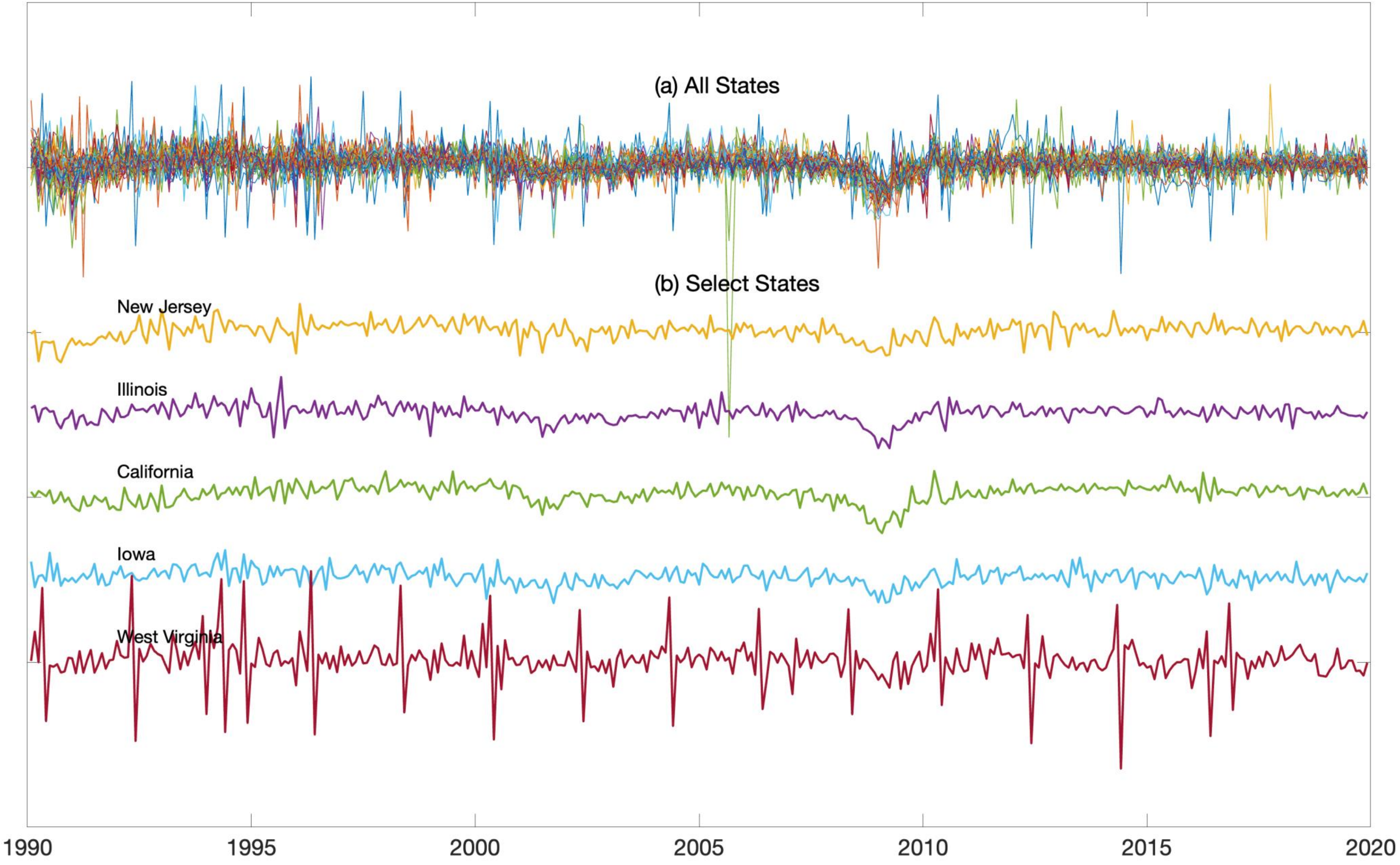
How Are They Related?

1. Similar Series, so similar DGPs
2. Correlated

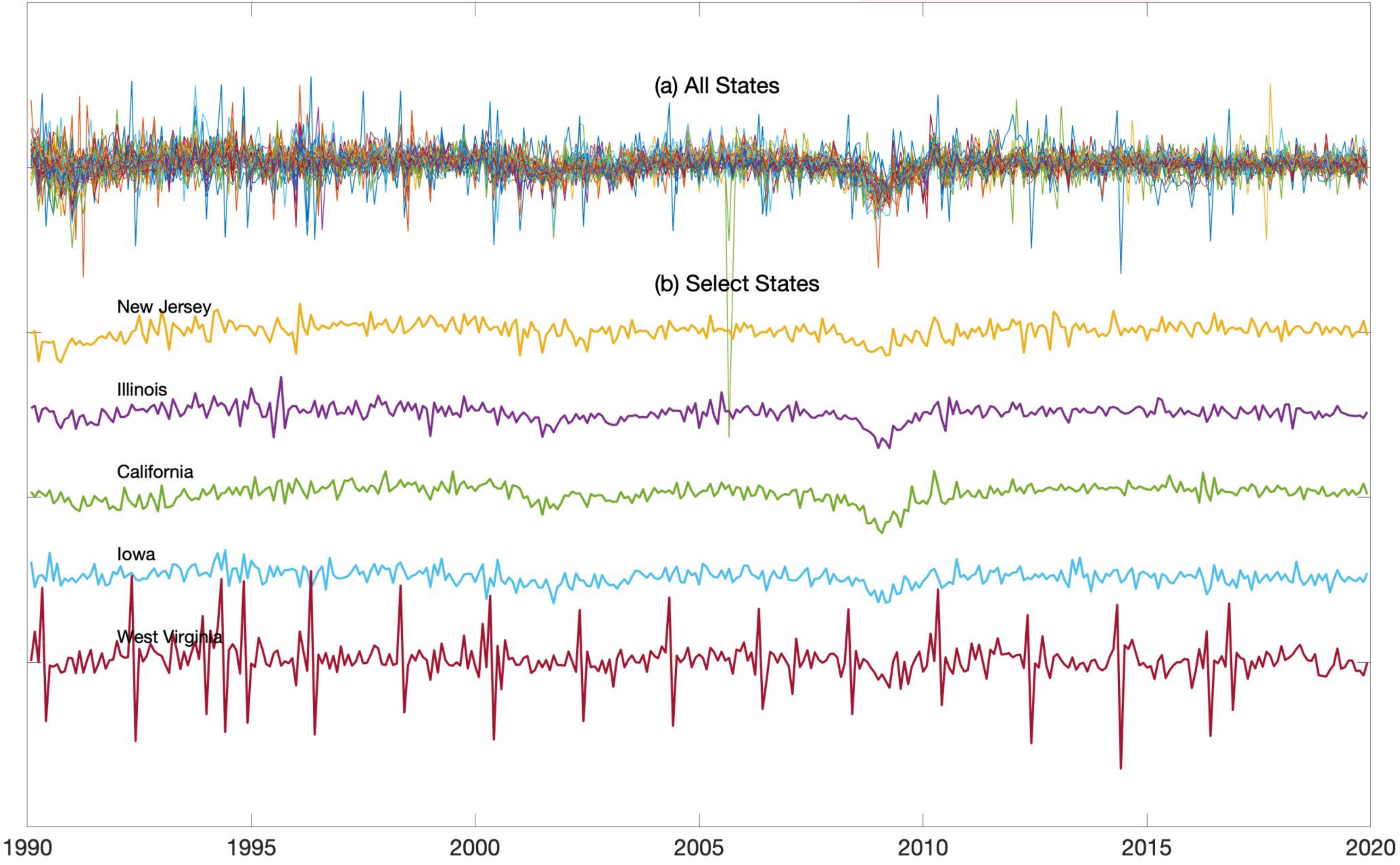
Why is this useful for Forecasting

- ‘Borrow’ information across series for estimating parameter values
- Exploit lead-lag/Granger-causality

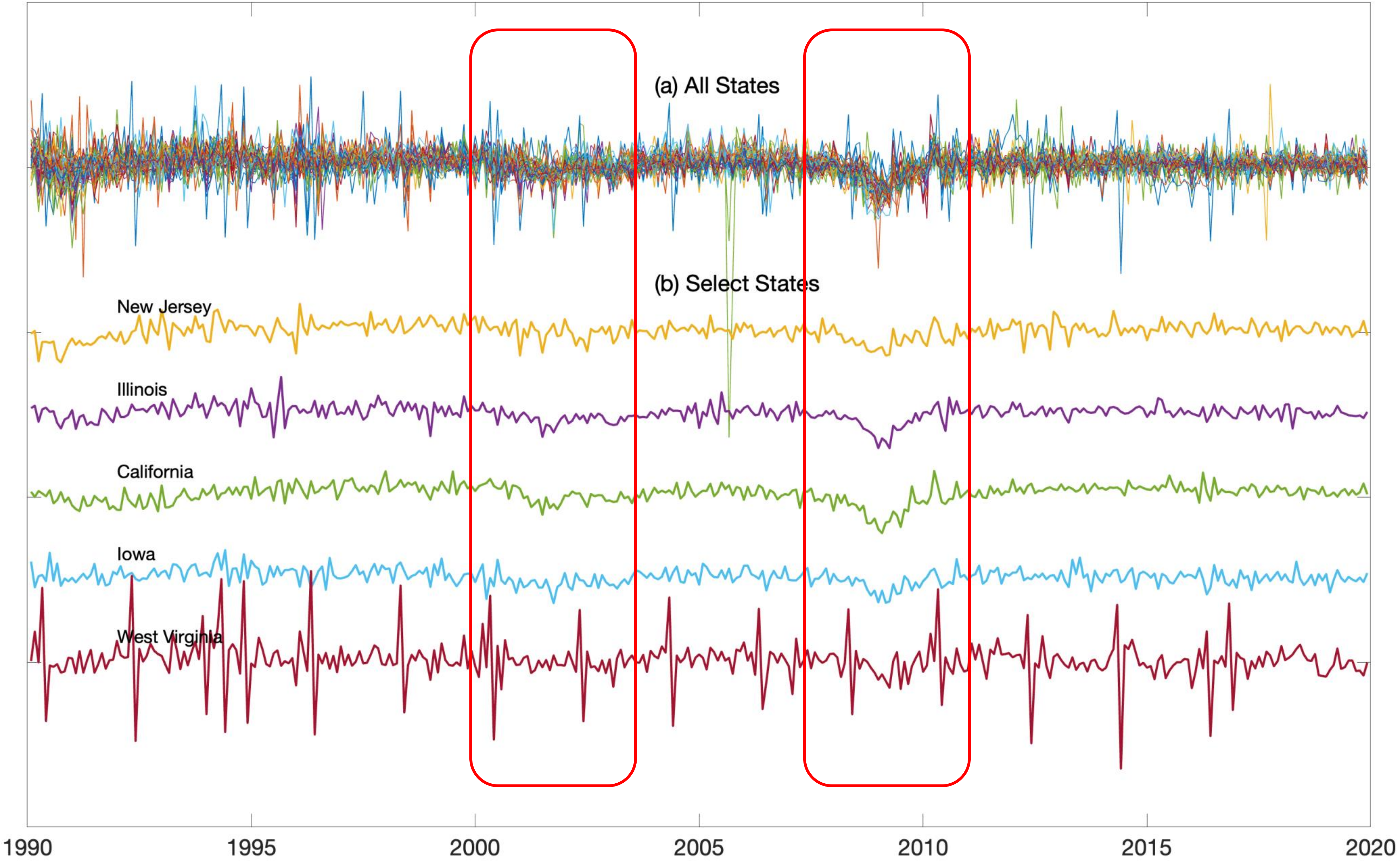
U.S. States Employment Growth Rates (1990m1-2019m12)



U.S. States Employment Growth Rates (1990m1-2019m12)

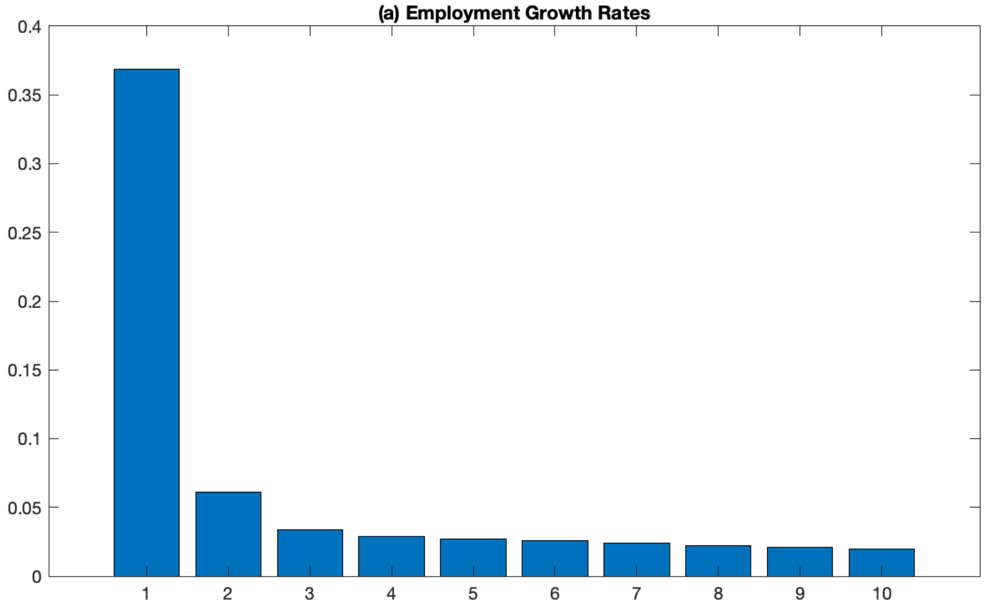


Series are correlated

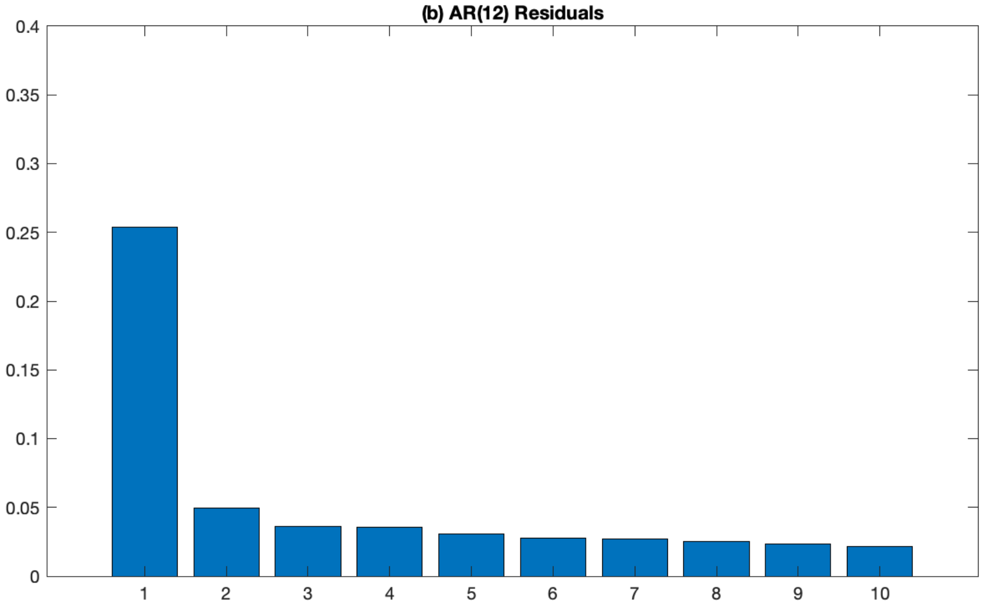


Covariation: Scree Plots

Growth Rates

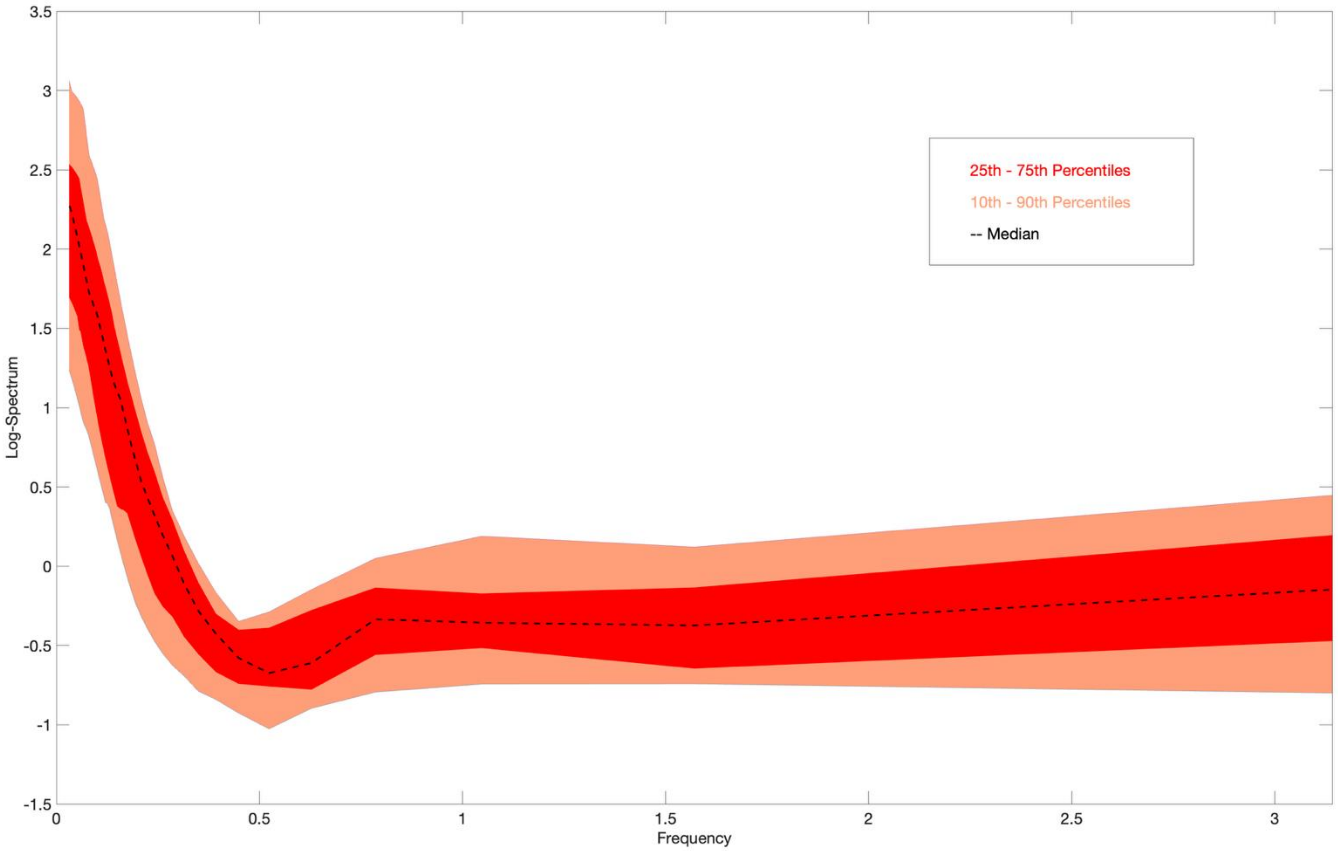


AR(12) Residuals



Cross Section Percentile Across States

Spectra



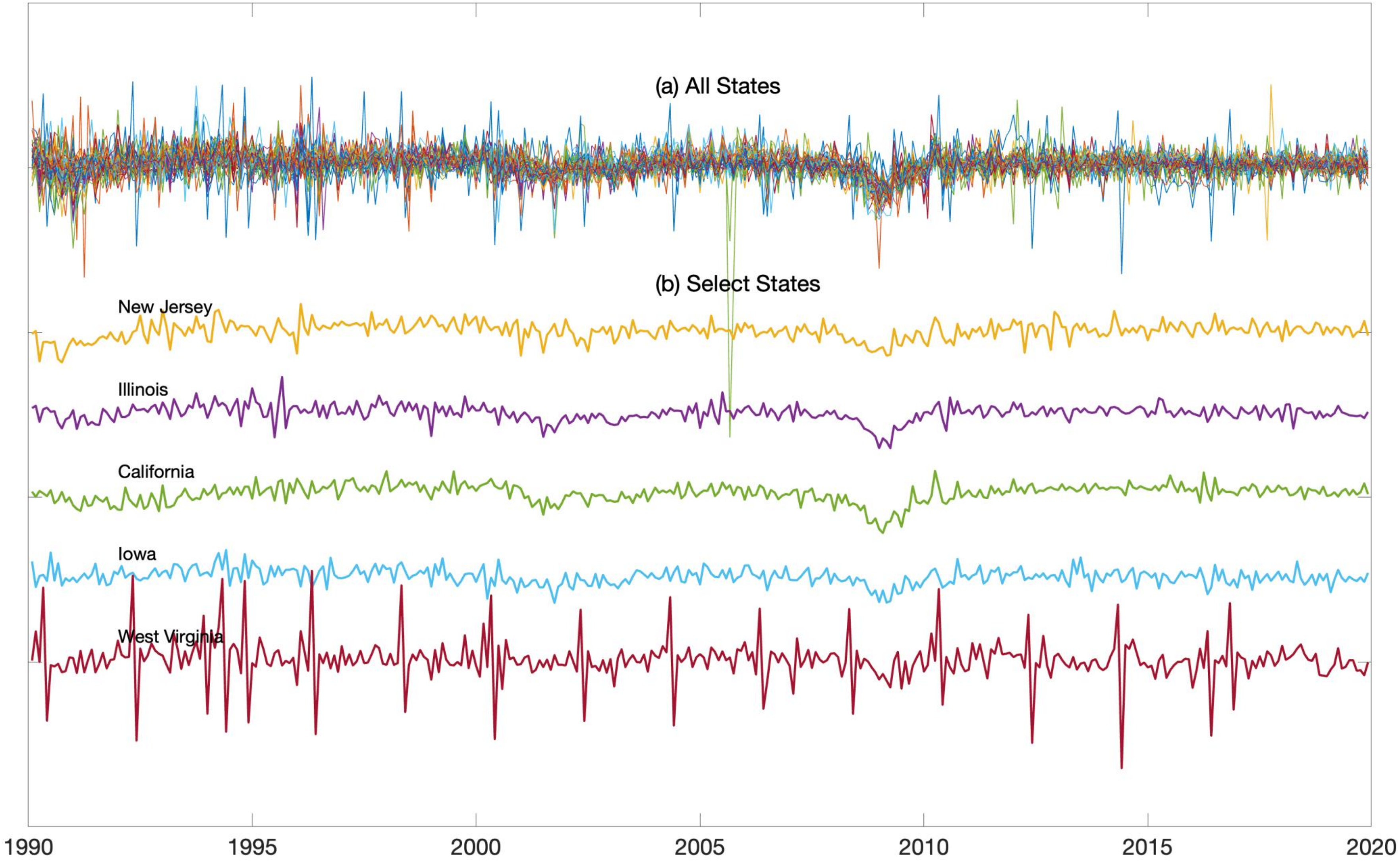
AR(12) innovation standard deviations

Percentile	0.10	0.25	0.50	0.75	0.90
σ	2.18	2.42	2.89	3.42	4.44

Similar Dynamics

Other Features

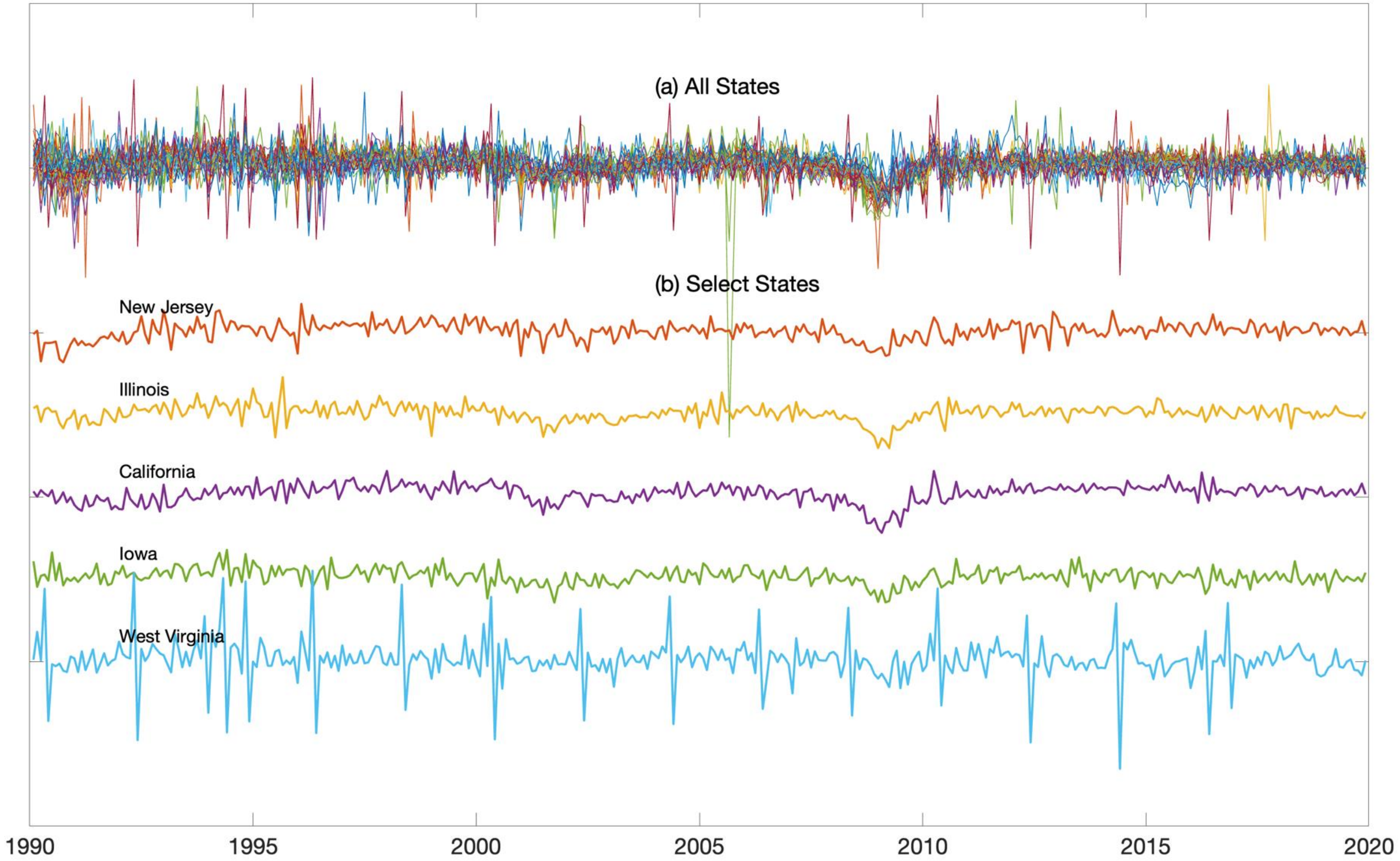
Kurtosis/Outliers



Other Features

Kurtosis/Outliers

Stoch. Volatility (squint)



Other Features

Kurtosis/Outliers   Stoch. Volatility (squint)

GARCH(1,1)-t models for the AR(12) residuals

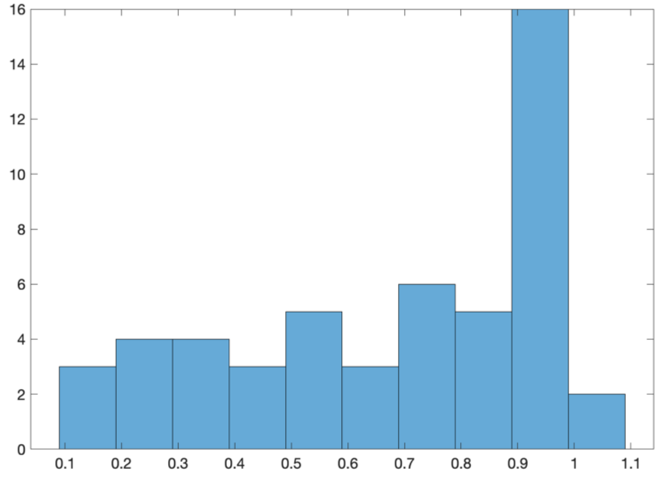
$$e_{j,t} = \sigma_{j,t}\epsilon_{j,t}$$

$$\text{with } \sigma^2_{j,t} = \alpha_j + \beta_{1,j}\sigma^2_{j,t-1} + \beta_{2,j}e^2_{j,t-1}$$

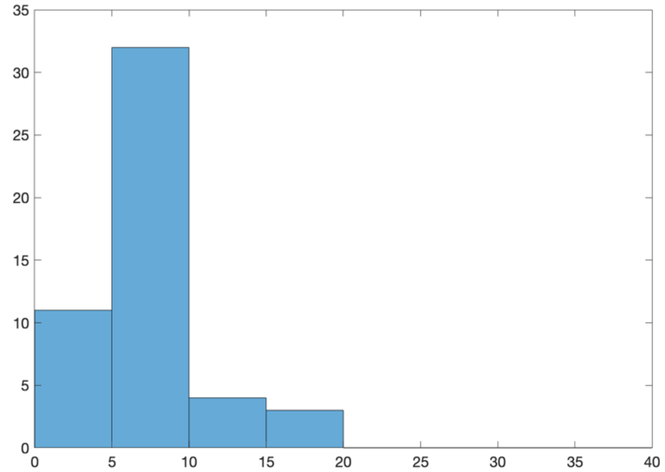
$$\text{and } \epsilon_{j,t} \sim t_{\nu_j}$$

distribution across 51 states

Sum of GARCH Coefficients



t-Degrees of Freedom

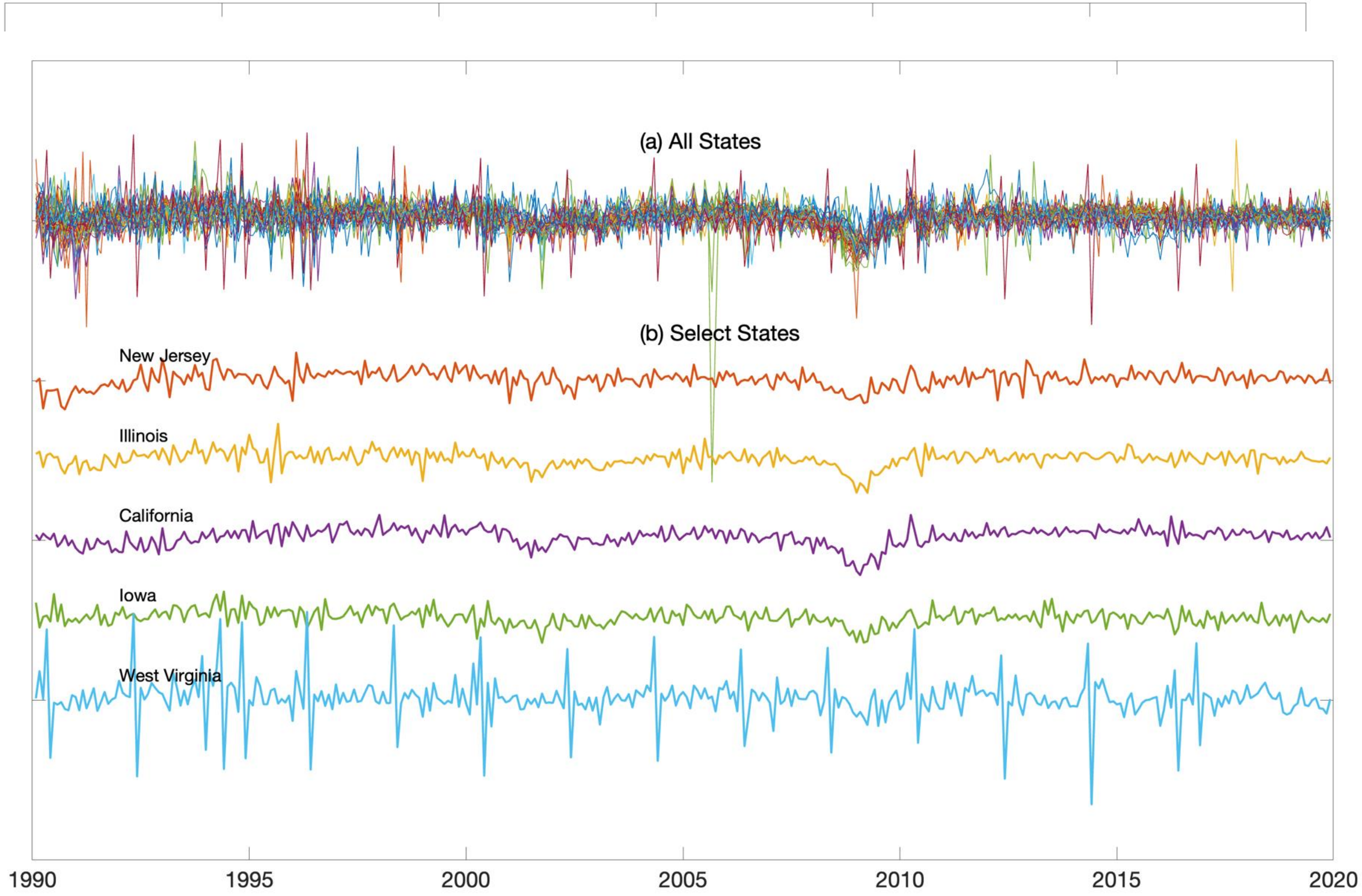


Other Features

Kurtosis/Outliers

Stoch. Volatility (squint)

Time Varying Level ??

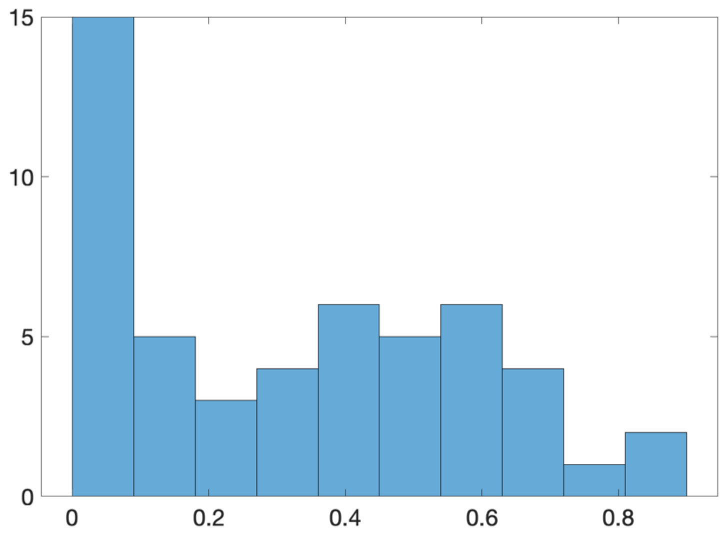


Other Features

p-values of Nyblom Tests

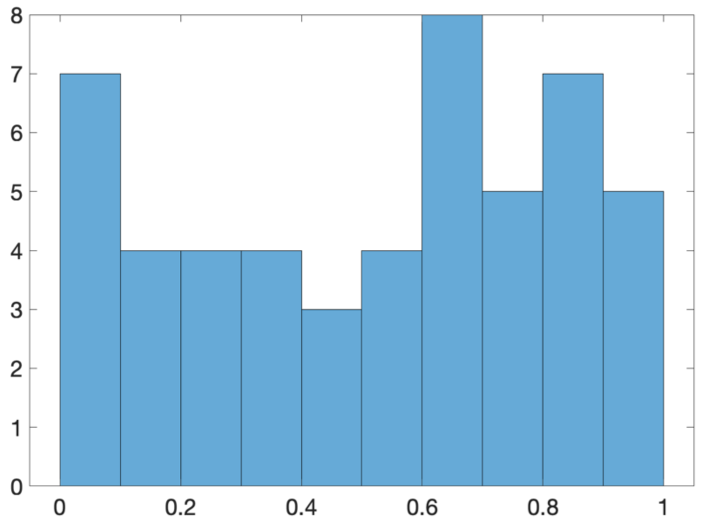
Kurtosis/Outliers

Level



Persistence

(sum of AR(12) coefficients)



Stoch. Volatility (squint)

Time Varying Level ??

What do we do in this paper?

- Propose a sequence of tractable models that incorporate these features

- Examine the empirical performance of these models using 3 datasets.

What do we do in this paper?

- Propose a sequence of **tractable** models that incorporate these features
  - Linear and (conditionally) Gaussian
  - Estimated using Bayes methods
    - Careful attention to algorithms
    - Estimation in seconds (Fortran, typical workstation)
- Examine the empirical performance of these models using 3 datasets.

Some Building Blocks

MCMC Bayes Methods: Gelman et al (2004), Geweke (2004), Meng and Wong (1996)…

Econometric methods and models:

Almuzara and Sbordonne (2022), Antolin-Diaz, Dreschel and Petrella (2024), Atkeson and Ohanian (2001), Bai and Ng (2002), Banbura, Giannone and Reichlin (2010), Carriero, Clark and Marecellino (2015), Carriero, Clark, Marecellino and Mertens (2015), Carriero, Pettenuzzo and Shekhar (2024), Chan (2022), Chan and Jeliazkov (2009), Cogley and Sargent (2005), D'Agostino and Giannone (2012), del Negro and Otrok (2008), Doan, Litterman and Sims (1984), Durbin and Koopman (2002), Giannone, Lenza and Primiceri (2015), Kim, Shephard and Chib (1998), Litterman (1986), Omori, Chib, Shephand and Nakajima (2007), Primiceri (2005), Sims (1993), Stock and Watson (several), ….

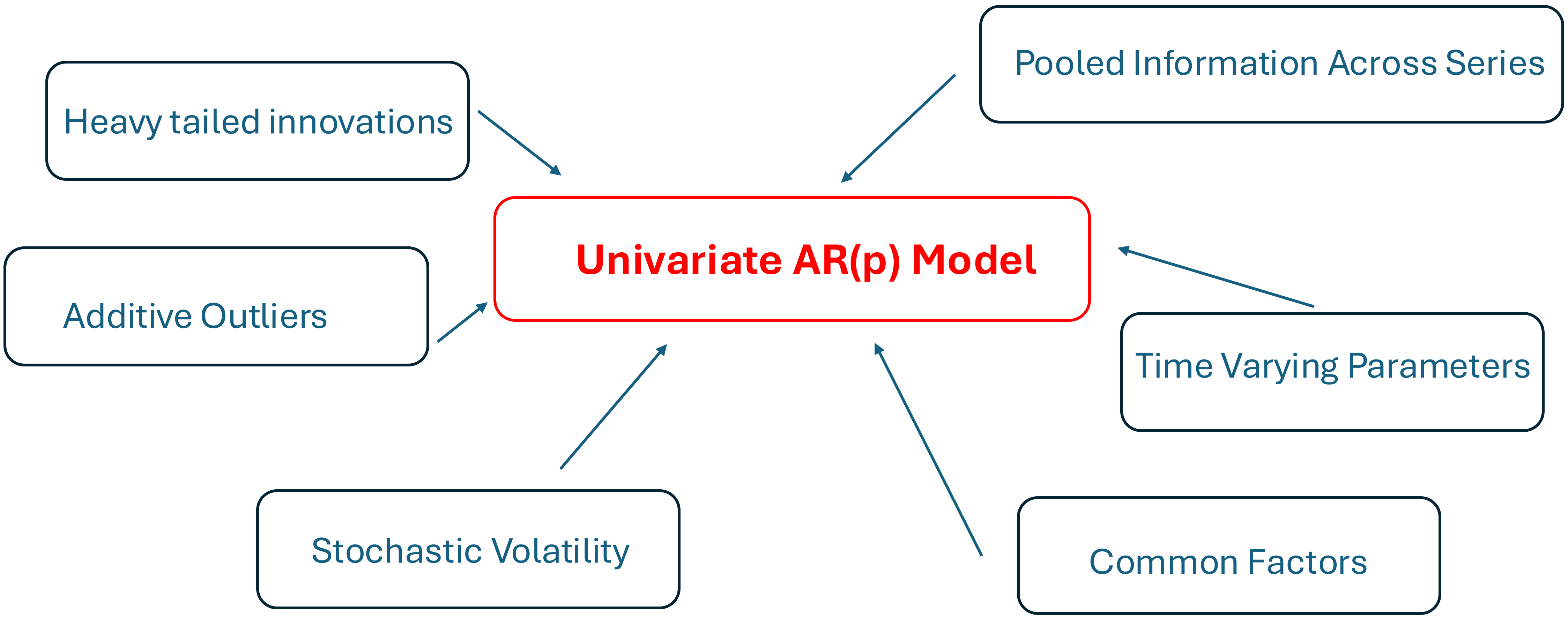
and many, many others.

Simple Tools

Posterior Simulators:

- Gibbs
- Metropolis
- Geweke Code Checking
- Bayes Factors via Bridge Sampling

Related Time Series (RTS) Model



8 Models

0. AR(12) estimated by OLS

1. Same as (0), but estimated by Bayes Methods (shrinkage)

2. Same as (1), but using Hierarchical Priors (shrink toward common parameter values

3. Same as (2), but add outliers (Student-t innovations)

4. Same as (3), but add additive outliers

5. Same as (4), but add time varying volatility

6. Same as (5), but add time varying conditional mean parameters

7. Same as (6), but add common factors. ← **RTS Model**

Model Evaluation

Pseudo-out-of-sample forecasts: horizons *h* = 1, 3, 6

- Forecasting h-period growth rates: $y^h_{t+h} = (1200/h) \times \ln(Emp_{i,t+h}/Emp_{i,t})$
- Root mean square error
- Quantiles and Prediction Intervals

Full Sample: Bayes Factors

Model Evaluation

Squared error loss
(I will show root mean squared error)

$$l(y, \hat{y}^{Model}) = (y - \hat{y}^{Model})^2$$

q^{th} -Quantile loss

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

90-10 interval loss

$$l(y, \hat{q}_{10}, \hat{q}_{90}) = (\hat{q}_{90} - \hat{q}_{10}) + \frac{1}{.10}((\hat{q}_{10} - y) \times 1(y < \hat{q}_{10}) + (y - \hat{q}_{90}) \times 1(y > \hat{q}_{90}))$$

Model 0: AR(12)

(Notation is overly-complicated here – but will come into play in later models)

Observations: $y_{j,t}$ for $j = 1, ..., n$ and $t = 1, ..., T$

$$y_{j,t} = \omega v_{j,t} \tag{1}$$

$$v_{j,t} = \mu_j + u_{j,t} \tag{2}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t} \tag{3}$$

$$\epsilon_{j,t} = \sigma_j \varepsilon_{j,t} \tag{4}$$

$$\varepsilon_{j,t} \sim iid\mathcal{N}(0,1) \tag{5}$$

Estimation: OLS, separately for each variable j using observations $t = 13, ..., T^*$.

Pseudo-out-off-sample forecasting from $T^* = 100$ to T.

Model 0: AR(12)

(Notation is overly-complicated here – but will come into play in later models)

Observations: $y_{j,t}$ for $j = 1, ..., n$ and $t = 1, ..., T$

$$y_{j,t} = \omega v_{j,t} \tag{1}$$

$$v_{j,t} = \mu_j + u_{j,t} \tag{2}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t} \tag{3}$$

AR(12) model with constant

$$\epsilon_{j,t} = \sigma_j \varepsilon_{j,t} \tag{4}$$

$$\varepsilon_{j,t} \sim iid\mathcal{N}(0,1) \tag{5}$$

Estimation: OLS, separately for each variable j using observations $t = 13, ..., T^*$.

Pseudo-out-off-sample forecasting from $T^* = 1999\text{m}12$ to $T = 2019\text{m}12 - h$.

Model 0: AR(12)

(Notation is overly-complicated here – but will come into play in later models)

Observations: $y_{j,t}$ for $j = 1, ..., n$ and $t = 1, ..., T$

$$y_{j,t} = \omega v_{j,t} \tag{1}$$

overall scale = 1 in this model

$$v_{j,t} = \mu_j + u_{j,t} \tag{2}$$

mean

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t} \tag{3}$$

AR coefficients

$$\epsilon_{j,t} = \sigma_j \varepsilon_{j,t} \tag{4}$$

innovation st. dev. (relative value)

$$\varepsilon_{j,t} \sim iid\mathcal{N}(0, 1) \tag{5}$$

Estimation: OLS, separately for each variable j using observations $t = 13, ..., T^*$.

Pseudo-out-off-sample forecasting from $T^* = 1999\text{m}12$ to $T = 2019\text{m}12 - h$.

Model 0: AR(12)

(Notation is overly-complicated here – but will come into play in later models)

Observations: $y_{j,t}$ for $j = 1, ..., n$ and $t = 1, ..., T$

$$y_{j,t} = \omega v_{j,t} \tag{1}$$

$$v_{j,t} = \mu_j + u_{j,t} \tag{2}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t} \tag{3}$$

$$\epsilon_{j,t} = \sigma_j \varepsilon_{j,t} \tag{4}$$

$$\varepsilon_{j,t} \sim iid\mathcal{N}(0, 1) \tag{5}$$

Benchmark for predictive distributions

Estimation: OLS, separately for each variable j using observations $t = 13, ..., T^*$.

Pseudo-out-off-sample forecasting from $T^* = 1999\text{m}12$ to $T = 2019\text{m}12 - h$.

POOS Root Mean Squared Error – Distribution across the 51 States

		Quantile					
	min	0.10	0.25	0.50	0.75	0.90	max
$h = 1$	1.82	2.16	2.43	2.78	3.28	4.18	5.78
$h = 3$	1.25	1.50	1.58	1.77	2.12	2.49	4.25
$h = 6$	1.11	1.30	1.43	1.62	1.87	2.16	2.83

Large ?

Model 1

Model: Same as Model 0

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors

- Diffuse priors for ω and $\{\mu_j\}_{j=1}^n$. (Induces scale and location equivariance)

$$\ln(\omega^2) \sim N(0,\infty) \text{ and } \mu_j \sim N(0,\infty)$$

- Minnesota-like priors for AR coefficients:

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

- σ_j:

$$\ln(\sigma_j^2) \sim N(0,1)$$

- Initial conditions: $\{u_{j,t}\}_{t=-11}^{0} | (\phi_j, \sigma_j)$ from a stationary Gaussian AR model:

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

  and the constant $c \leq 1$ is chosen so that largest root of companion matrix is no larger than 0.98.

Model 1

Model: Same as Model 0

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors

- Diffuse priors for ω and $\{\mu_j\}_{j=1}^{n}$. (Induces scale and location equivariance)

$$\ln(\omega^2) \sim N(0,\infty) \text{ and } \mu_j \sim N(0,\infty)$$

- Minnesota-like priors for AR coefficients:

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

- σ_j:

$$\ln(\sigma_j^2) \sim N(0,1)$$

- Initial conditions: $\{u_{j,t}\}_{t=-11}^{0} | (\phi_j, \sigma_j)$ from a stationary Gaussian AR model:

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

and the constant $c \leq 1$ is chosen so that largest root of companion matrix is no larger than 0.98.

Model 1

Model: Same as Model 0

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors

- Diffuse priors for ω and $\{\mu_j\}_{j=1}^n$. (Induces scale and location equivariance)

$$\ln(\omega^2) \sim N(0,\infty) \text{ and } \mu_j \sim N(0,\infty)$$

- Minnesota-like priors for AR coefficients:

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

- σ_j:

$$\ln(\sigma_j^2) \sim N(0,1)$$

- Initial conditions: $\{u_{j,t}\}_{t=-11}^{0} | (\phi_j, \sigma_j)$ from a stationary Gaussian AR model:

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

and the constant $c \leq 1$ is chosen so that largest root of companion matrix is no larger than 0.98.

Model 1

Model: Same as Model 0

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors

- Diffuse priors for ω and $\{\mu_j\}_{j=1}^n$. (Induces scale and location equivariance)

$$\ln(\omega^2) \sim N(0,\infty) \text{ and } \mu_j \sim N(0,\infty)$$

- Minnesota-like priors for AR coefficients:

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

- σ_j:

$$\ln(\sigma_j^2) \sim N(0,1)$$

- Initial conditions: $\{u_{j,t}\}_{t=-11}^{0} | (\phi_j, \sigma_j)$ from a stationary Gaussian AR model:

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

and the constant $c \leq 1$ is chosen so that largest root of companion matrix is no larger than 0.98.

Model 1

Model: Same as Model 0

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors

- Diffuse priors for ω and $\{\mu_j\}_{j=1}^{n}$. (Induces scale and location equivariance)

$$\ln(\omega^2) \sim N(0,\infty) \text{ and } \mu_j \sim N(0,\infty)$$

- Minnesota-like priors for AR coefficients:

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

- σ_j:

$$\ln(\sigma_j^2) \sim N(0,1)$$

- Initial conditions: $\{u_{j,t}\}_{t=-11}^{0} | (\phi_j, \sigma_j)$ from a stationary Gaussian AR model:

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

and the constant $c \leq 1$ is chosen so that largest root of companion matrix is no larger than 0.98.

Posterior Draws: Algorithm 1

Standard, but with a few wrinkles

Common scale factor ω

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 (9) and (6), 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.]

Metropolis for σ_j

2. $\{\sigma_j, \mu_j, u_{j,-11:0}, \{\epsilon_{j,t}\}_{t=1}^T\}$ looping over j:
   (a) Draw σ_j: Metropolis step with prior (9) 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 (2) and state evolution (3), initial state drawn from $\mu_j \sim \mathcal{N}(0, \infty)$ (approximated by using large but finite variance) and (7)
   (b) Draw $\{\mu_j, u_{j,-11:0}, \{\epsilon_{j,t}\}_{t=1}^T\} | \sigma_j$: Kalman smoother draw from same state-space-system as in Step 2a.

Metropolis correction for initial values

3. $\{\phi_j, \{\epsilon_{j,t}\}_{t=1}^T\}$ looping over j: Metropolis-Hastings step with proposal generated from Kalman smoother draw from linear SSS with state $(\phi_{j,1}, \ldots, \phi_{j,12})$ and measurement equation (3) and initial state drawn from (8). The proposal ϕ_j^p is accepted over the current value ϕ_j^c with probability $1 \wedge \frac{L_7(\phi_j^p)}{L_7(\phi_j^c)}$, where $L_7(\phi_j)$ is likelihood of (7).

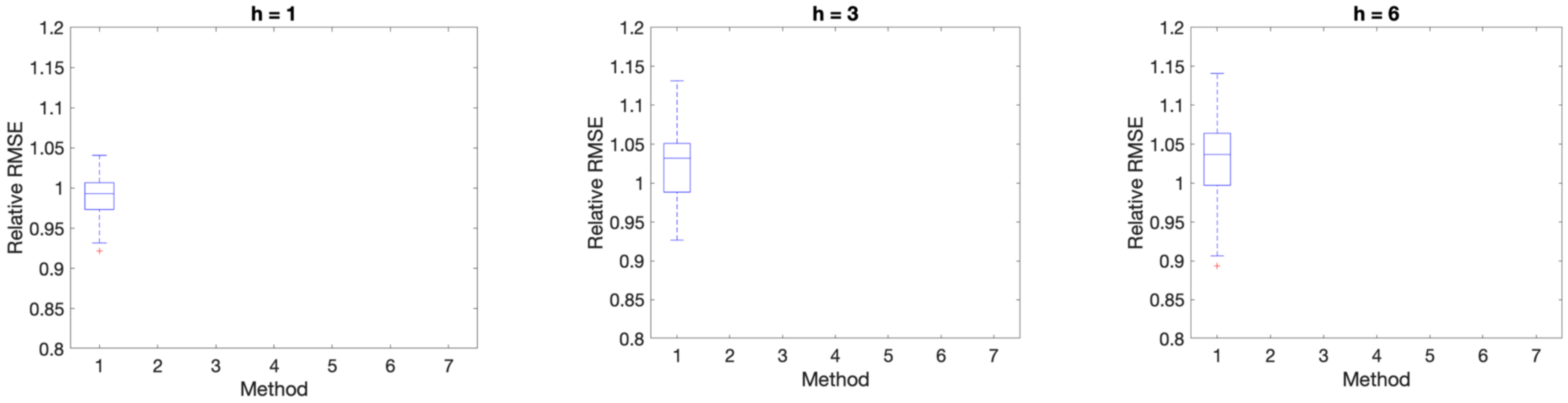
Forecasting Performance: State Employment

Relative Root Mean Squared Forecast Error
T = 1999:12 through 2019:6

(a) Pooled Over States

Forecasting Model	Forecast of Employment Growth Rate from T to T+h		
	h = 1	**h = 3**	**h = 6**
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2)			
(3)			
(4)			
(5)			
(6)			
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk (Pooled Over States)

$h = 3$

Forecasting Model	Quantile				90-10
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.98	1.02
(2)					
(3)					
(4)					
(5)					
(6)					
(7)					

Model 2

Model: Same as Model 1

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Hierarchical Priors

Notation: Hierarchical Normal Distribution

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

$$w_j|(m_w, v_w) \sim iid\mathcal{N}(m_w, v_w)$$

with

$$m_w \sim \mathcal{N}(m_{m_w}, v_{m_w})$$

and

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

We write this as

$$\{w_j\}_{j=1}^n \sim \mathcal{HN}(m_{m_w}, v_{m_w}, m_{\ln(v_w)}, v_{\ln(v_w)})$$

Model 2

Model: Same as Model 1

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Priors: Same as Model 1

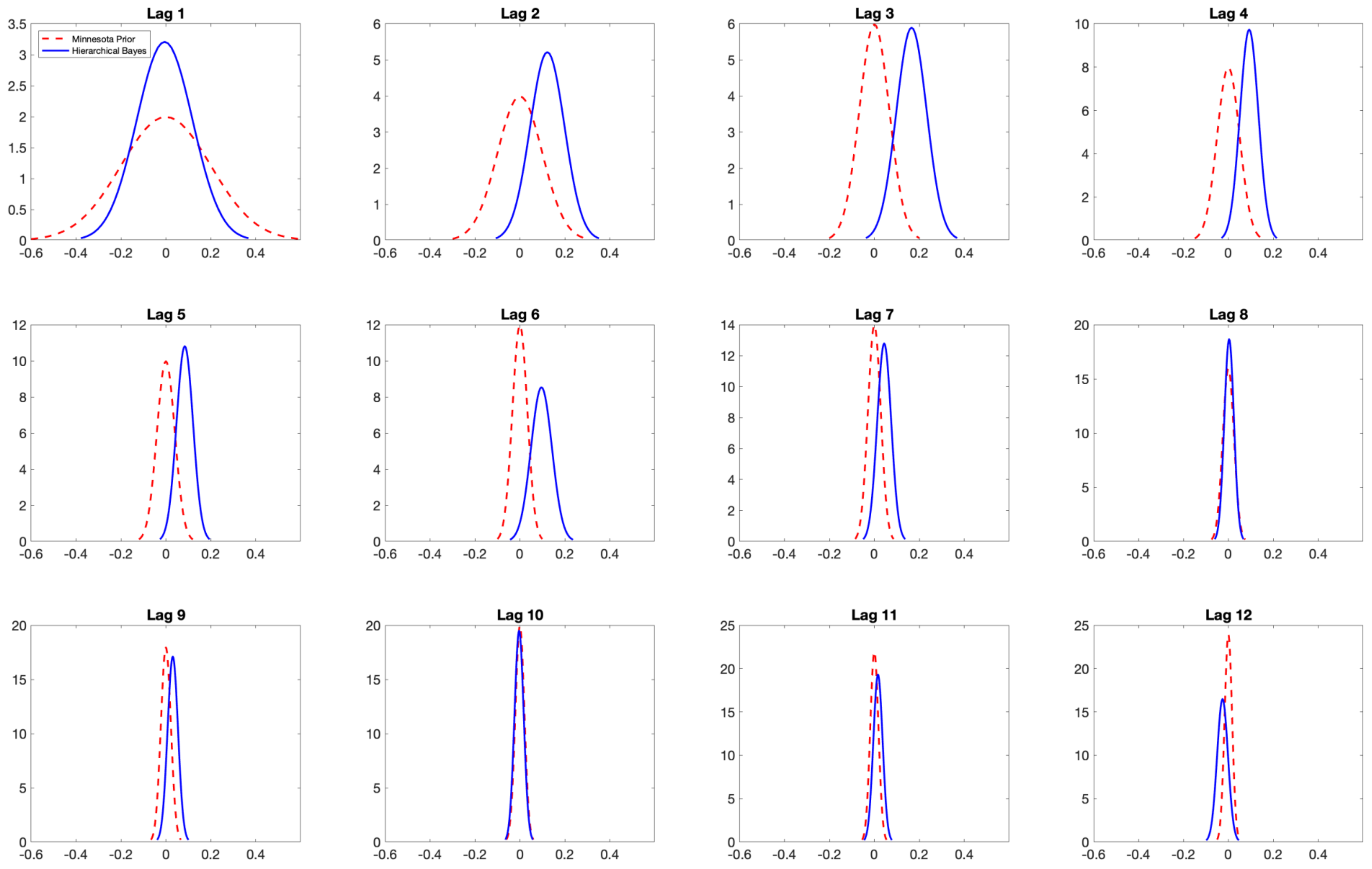
(Shrink toward common values)

except now shrinkage location and strength estimated from data

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

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

Minnesota versus Hierarchical Priors (shrinkage)

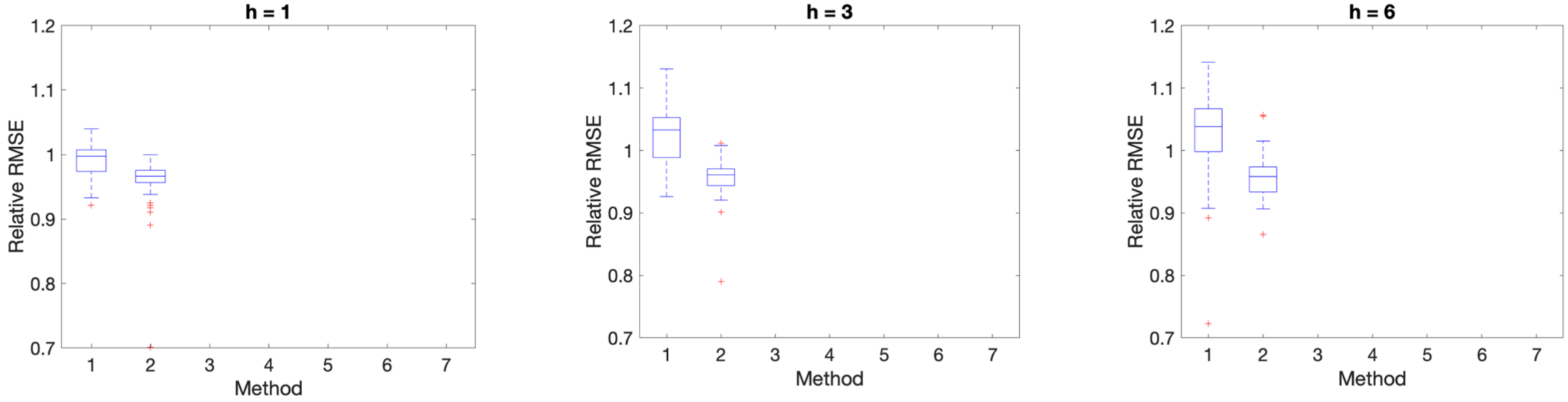


Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	**Forecast of Employment Growth Rate from T to T+h**		
	h = 1	**h = 3**	**h = 6**
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.94	0.94	0.96
(3)			
(4)			
(5)			
(6)			
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

$h = 3$

Forecasting Model	**Quantile**				**90-10**
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3)					
(4)					
(5)					
(6)					
(7)					

Model 3

Model: Same as Model 2

$$y_{j,t} = \omega v_{j,t}$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

except

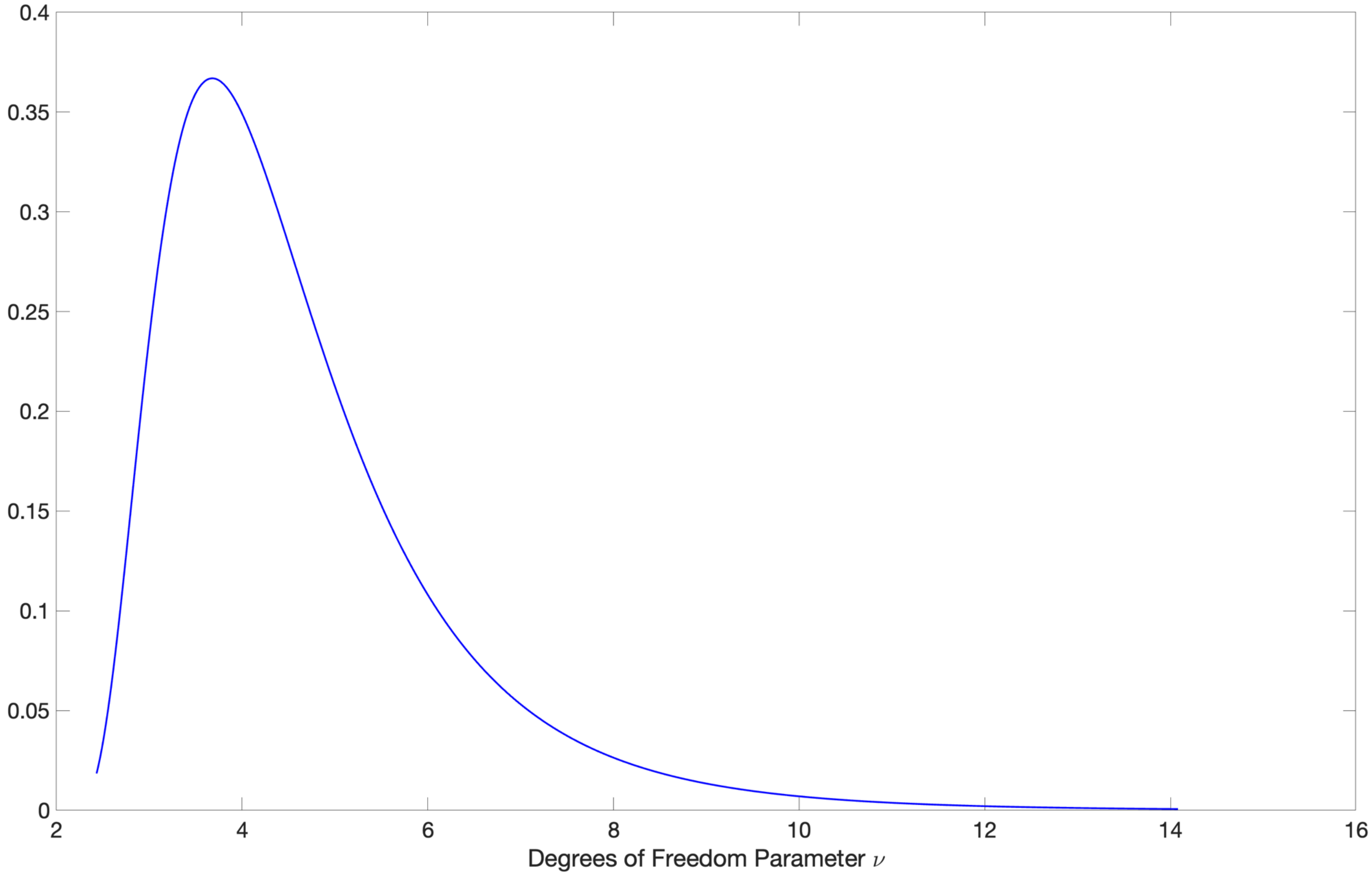
$$\varepsilon_{j,t} \sim iid\mathcal{T}(\nu_j)$$ ← Student-t errors

Priors: Same as Model 2

additionally with

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

‘Prior’ distribution of degrees of freedom (evaluated at posterior mean)

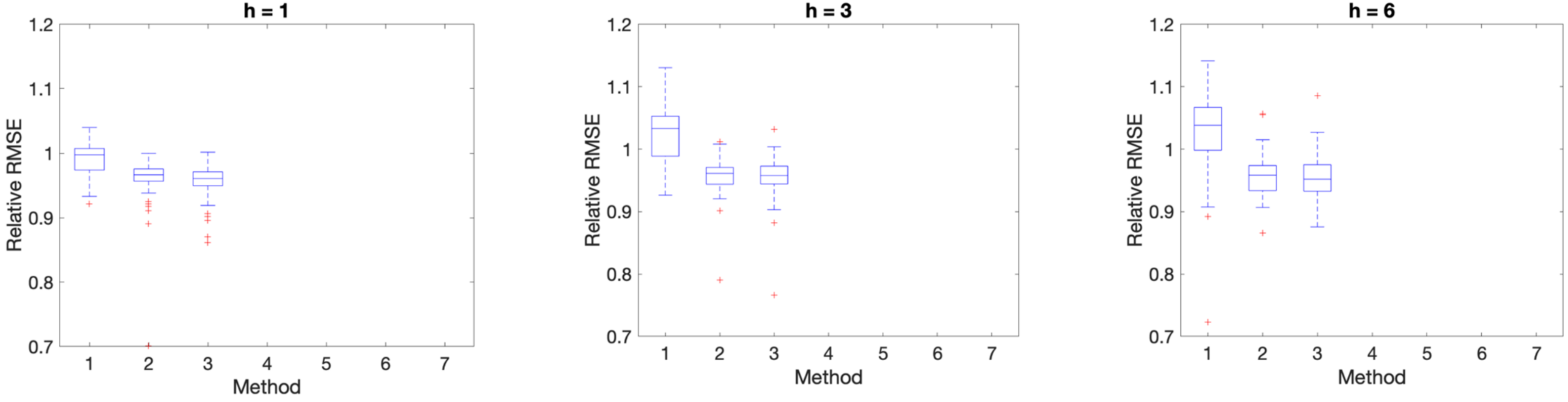


Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	Forecast of Employment Growth Rate from T to T+h		
	h = 1	h = 3	h = 6
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.95	0.94	0.96
(3) t-innovations	0.94	0.94	0.95
(4)			
(5)			
(6)			
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

h = 3

Forecasting Model	**Quantile**				**90-10**
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3) t-innovations	0.92	0.95	0.95	0.98	0.95
(4)					
(5)					
(6)					
(7)					

Model 4

Model: include additive outliers

$$y_{j,t} = \omega(v_{j,t} + o_{j,t})$$

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$o_{j,t}$ are additive outliers

Note: no dynamics

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

Same as Model 3

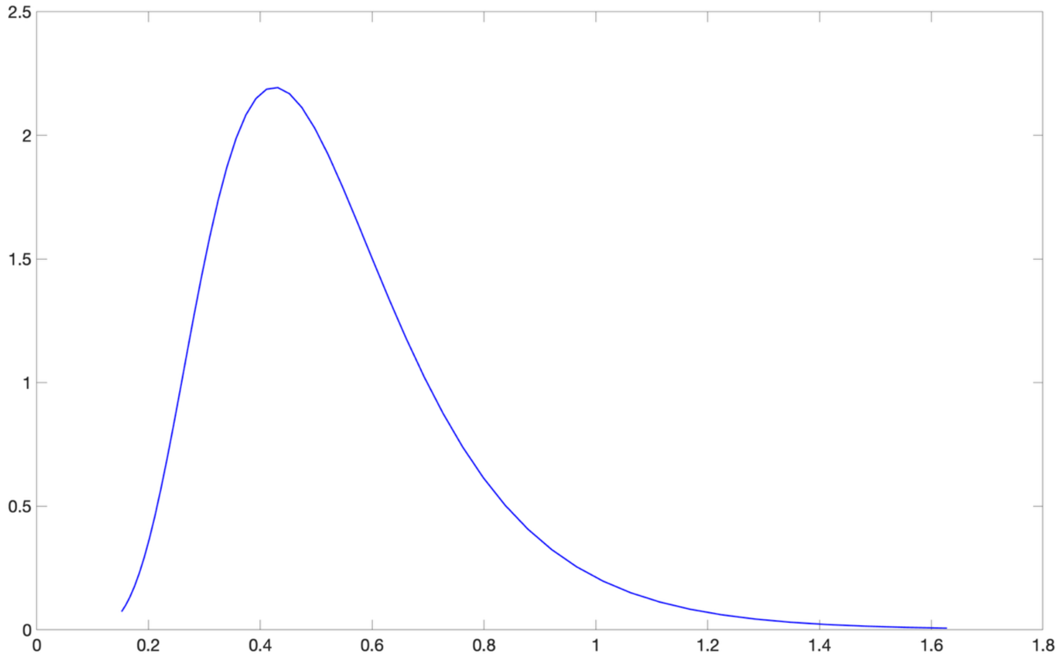
Priors: Same as Model 3, but including

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

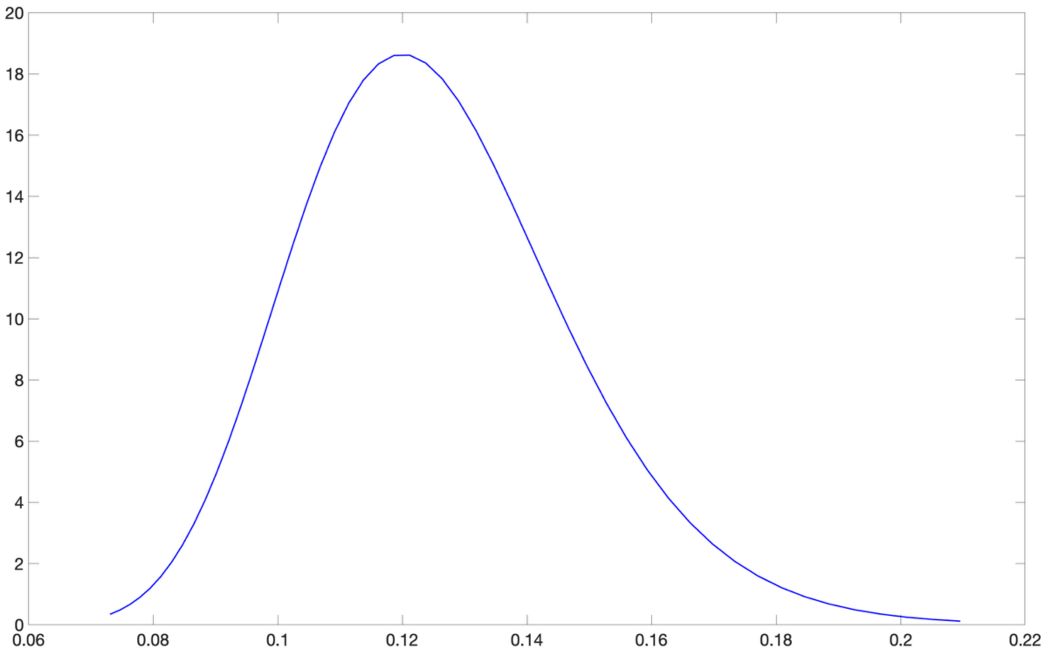
HN Prior for outlier scale

Posterior distribution of σ_j and κ_j

σ_j



κ_j



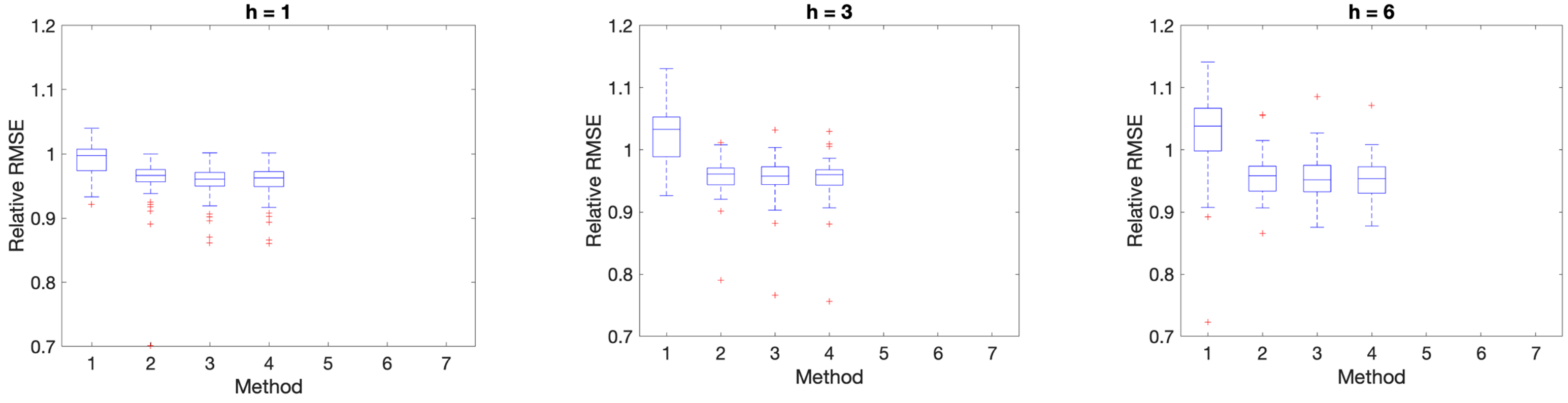
$$\kappa_j^2/(\sigma_j^2 + \kappa_j^2) \approx \left(\frac{0.12^2}{0.5^2 + 0.12^2}\right) = 0.05$$

Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	**Forecast of Employment Growth Rate from T to T+h**		
	h = 1	**h = 3**	**h = 6**
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.95	0.94	0.96
(3) t-innovations	0.94	0.94	0.96
(4) A-outliers	0.94	0.94	0.95
(5)			
(6)			
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

$h = 3$

Forecasting Model	Quantile				90-10
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3) t-innovations	0.92	0.95	0.95	0.98	0.95
(4) a-outliers	0.92	0.95	0.95	0.97	0.95
(5)					
(6)					
(7)					

Model 5

Model: include stochastic volatility

$$y_{j,t} = \omega(v_{j,t} + o_{j,t})$$

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$$v_{j,t} = \mu_j + u_{j,t}$$

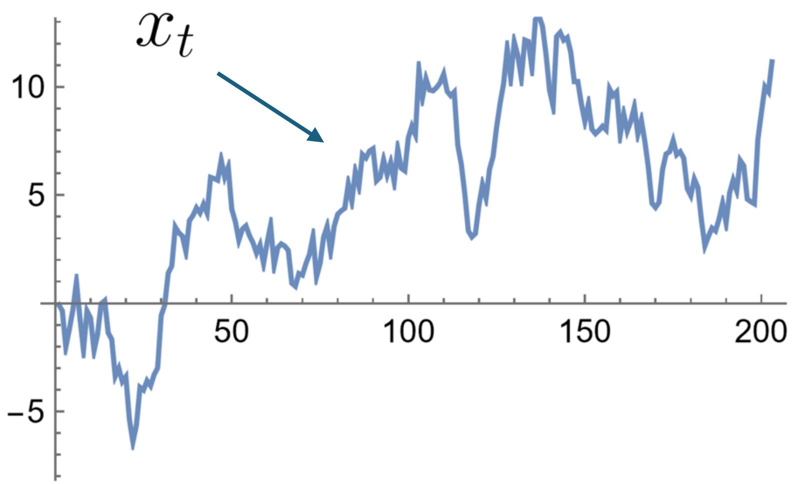
$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

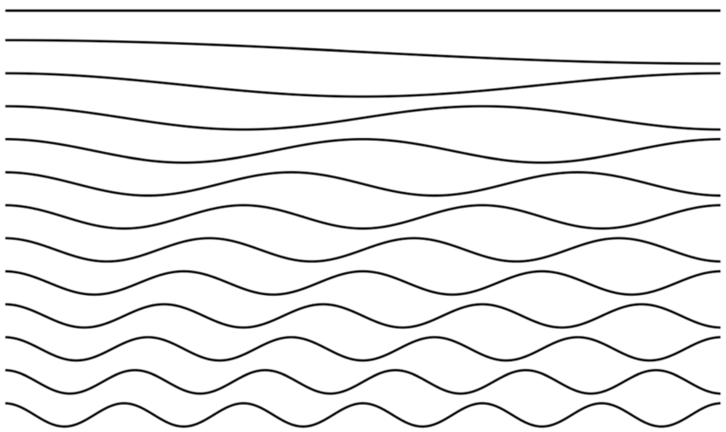
$$\epsilon_{j,t} = \sigma_{j,t} \varepsilon_{j,t}$$

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

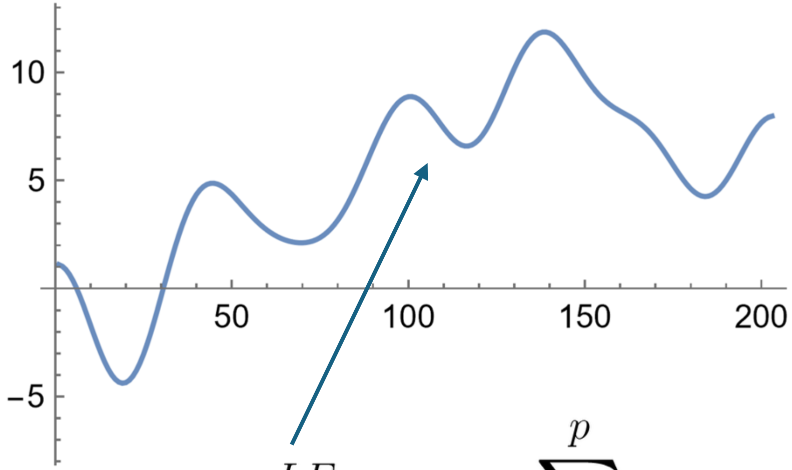
$\ln(\sigma_{j,t})$ evolves as 'random walk'

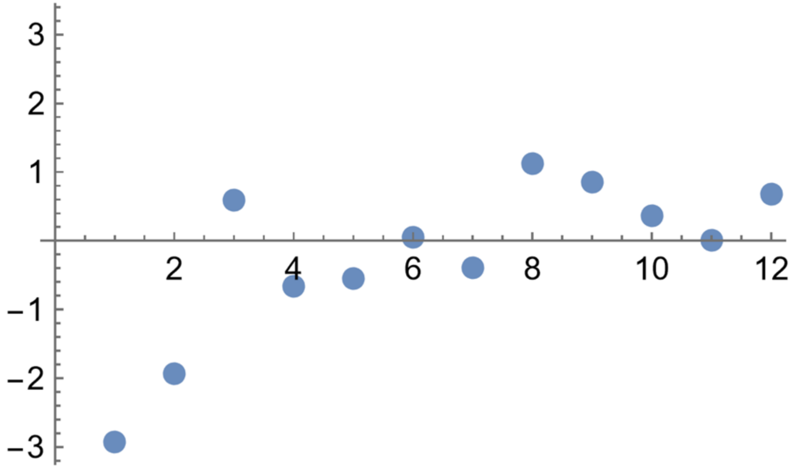
Low Frequency approximation to random walk





$$\varphi_{l,t},\ \ l = 1, \ldots, p$$





$$\xi_l,\ \ l = 1, \ldots, p$$

$$x_t^{LF} = \xi_0 + \sum_{l=1}^{p} \varphi_{l,t}\xi_l$$

Model 5

Model: include stochastic volatility

$$y_{j,t} = \omega(v_{j,t} + o_{j,t})$$

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

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

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

$$\ln(\sigma^2_{j,t}) = \ln(\sigma^2_j) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

Priors: Same as Model 4, but including

$$\xi_{j,l} \sim \mathcal{N}(m_{\xi_l}, v_\xi)$$

$$\text{with } m_{w_l} \sim \mathcal{N}(0, 0.01^2) \text{ and } \ln(v_\xi) \sim \mathcal{N}(\ln(0.01^2), 1)$$

Model 5

Model: include stochastic volatility

$$y_{j,t} = \omega(v_{j,t} + o_{j,t})$$

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$$v_{j,t} = \mu_j + u_{j,t}$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l} u_{j,t-l} + \epsilon_{j,t}$$

$$\epsilon_{j,t} = \sigma_{j,t} \varepsilon_{j,t} \qquad \ln(\sigma^2_{j,t}) = \ln(\sigma^2_j) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

Priors: Same as Model 4, but including

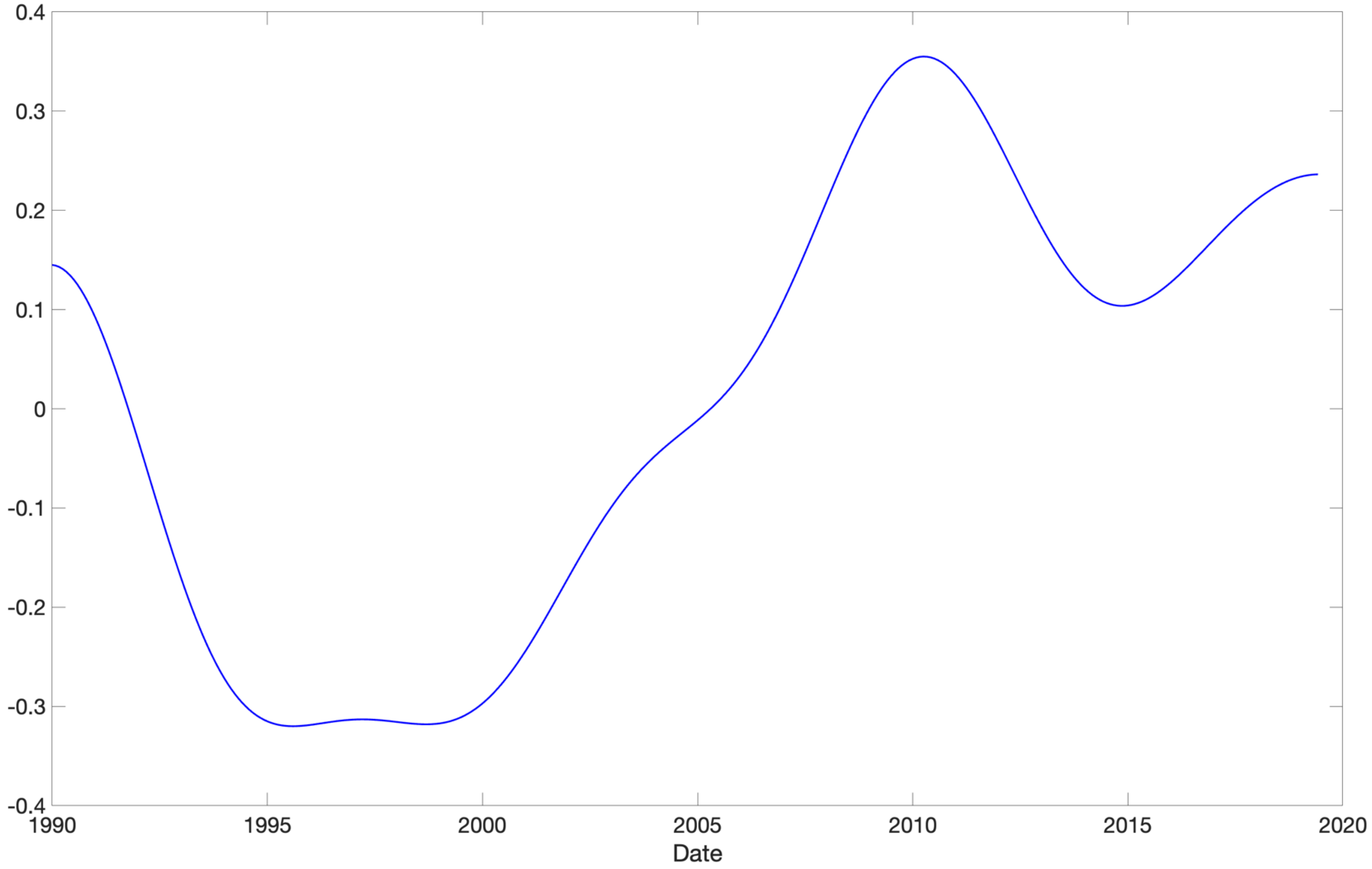
$$\xi_{j,l} \sim \mathcal{N}(m_{\xi_l}, v_\xi)$$

$$\text{with } m_{w_l} \sim \mathcal{N}(0, 0.01^2) \text{ and } \ln(v_\xi) \sim \mathcal{N}(\ln(0.01^2), 0.55^2)$$

Common Volatility Path

$$\sum_{l=1}^{q} \varphi_l\left(\frac{t-1/2}{T}\right)\hat{m}_{\xi_l}$$

Posterior mean

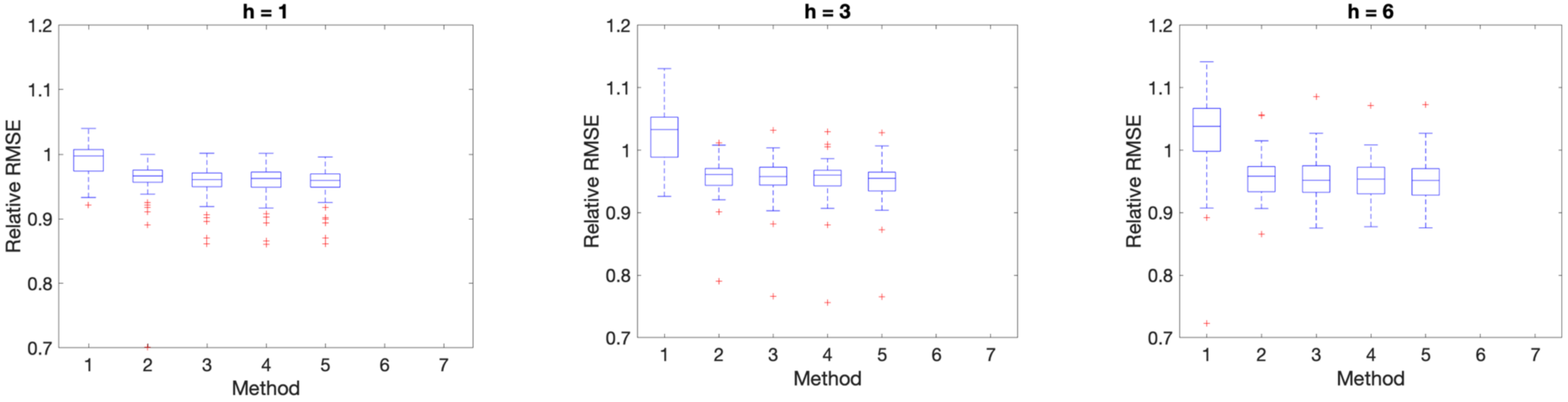


Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	Forecast of Employment Growth Rate from T to T+h		
	h = 1	h = 3	h = 6
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.95	0.94	0.96
(3) t-innovations	0.94	0.94	0.96
(4) A-outliers	0.93	0.93	0.95
(5) Stoch. Vol.	0.93	0.93	0.95
(6)			
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

$h = 3$

Forecasting Model	Quantile				90-10
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3) t-innovations	0.92	0.95	0.95	0.98	0.95
(4) A-outliers	0.92	0.95	0.95	0.97	0.95
(5) Stoch. Vol.	0.92	0.94	0.93	0.90	0.92
(6)					
(7)					

Model 6

Model: include time varying conditional mean parameters

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

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t}$$

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

$$\ln(\sigma_{j,t}^2) = \ln(\sigma_j^2) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

Evolve as random walks

$$\mu_{j,t} \sim RW(\mu_j, \gamma_{\mu(j)}^2) \text{ and } \phi_{j,l,t} \sim RW(\phi_{j,l}, \gamma_{\phi(j,l)}^2)$$

Model 6

Model: include time varying conditional mean parameters

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

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t}$$

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

$$\ln(\sigma_{j,t}^2) = \ln(\sigma_j^2) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

$$\mu_{j,t} \sim RW(\mu_j, \gamma_{\mu(j)}^2) \text{ and } \phi_{j,l,t} \sim RW(\phi_{j,l}, \gamma_{\phi(j,l)}^2)$$

Priors: Same as Model 5, but including

$$\{\ln(\gamma_{\mu(j)}^2)\}_{j=1}^n \sim HN(\ln(0.005^2), 4, \ln(0.3^2), 1)$$

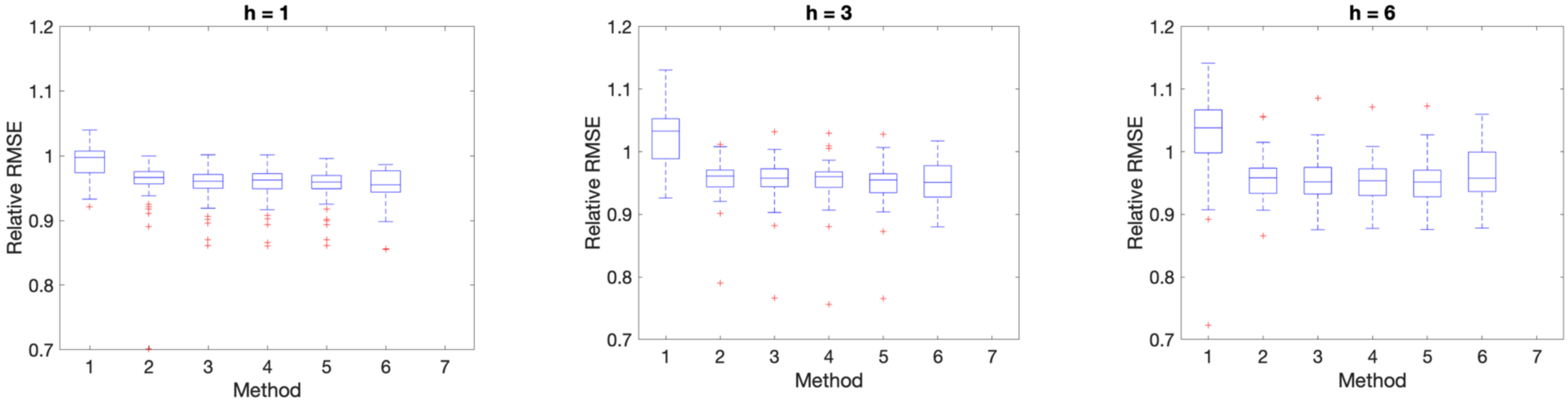
$$\{\ln(\gamma_{\phi(j,l)}^2\}_{j=1}^n \sim HN(\ln((0.005/l)^2), 1, \ln(0.3^2), 1)$$

Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	Forecast of Employment Growth Rate from T to T+h		
	h = 1	**h = 3**	**h = 6**
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.95	0.94	0.96
(3) t-innovations	0.94	0.94	0.96
(4) A-outliers	0.93	0.93	0.95
(5) Stoch. Vol.	0.93	0.93	0.95
(6) TV mean parms	0.93	0.92	0.94
(7)			

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

$h = 3$

Forecasting Model	Quantile				90-10
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3) t-innovations	0.92	0.95	0.95	0.98	0.95
(4) A-outliers	0.92	0.95	0.95	0.97	0.95
(5) Stoch. Vol.	0.92	0.94	0.93	0.90	0.92
(6) TV mean parms	0.87	0.93	0.94	0.91	0.91
(7)					

Model 7: Include Common Factors

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

Common factors:
- c is distributed lag of an AR(12)
- μ is a common RW trend
- o is an outlier

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

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

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t}$$

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

$$\ln(\sigma^2_{j,t}) = \ln(\sigma^2_j) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

Same as Model 6, but now for j=1,...,n+1

Model 7

Model

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

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

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

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t}$$

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

$$\ln(\sigma^2_{j,t}) = \ln(\sigma^2_j) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

Priors: Same as Model 6 but now including

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

$$\{\lambda_{j,l}\}_{j=1}^{n} \sim \mathcal{HN}(0, (0.5/(l+1))^2, \ln(0.3^2), 1) \text{ for } j > 0$$

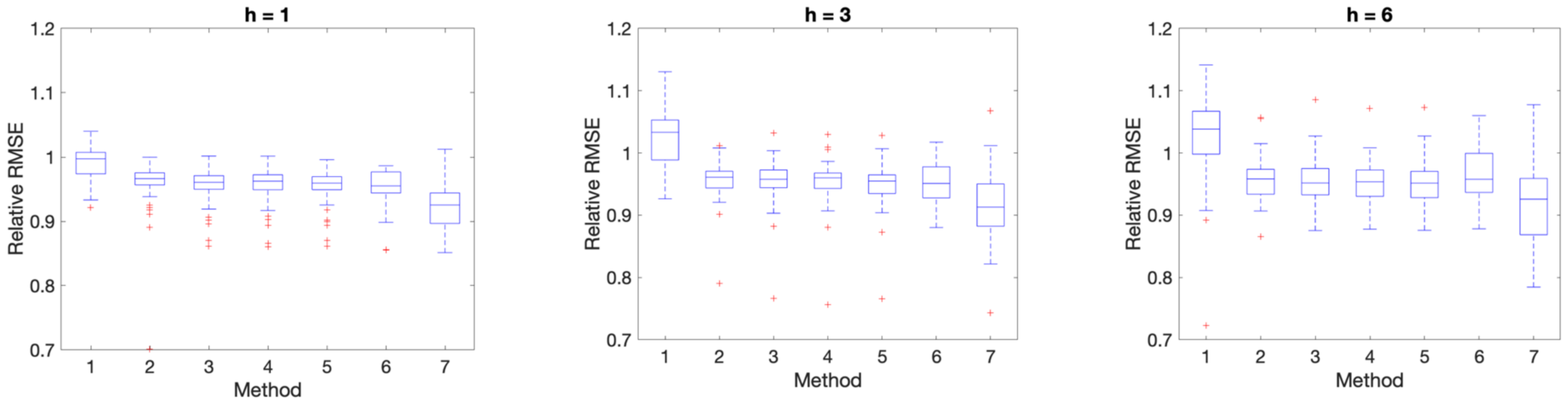
$$\{\ln(\gamma^2_{\lambda(j,l)})\}_{j=1}^{n} \sim \mathcal{HN}(\ln((0.0005/(l+1))^2), 4, \ln(0.3^2), 1)$$

Forecasting Performance: State Employment

(a) Pooled Over States

Forecasting Model	**Forecast of Employment Growth Rate from T to T+h**		
	h = 1	**h = 3**	**h = 6**
Benchmark	1.00	1.00	1.00
(1) Bayes	0.97	0.99	1.02
(2) H.Prior	0.95	0.94	0.96
(3) t-innovations	0.94	0.94	0.96
(4) A-outliers	0.93	0.93	0.95
(5) Stoch. Vol.	0.93	0.93	0.95
(6) TV mean parms	0.93	0.93	0.95
(7) Com. Factors	0.90	0.89	0.91

(b) Distribution Across States



Quantile and 90-10 Relative Risk
(Pooled Over States)

$h = 3$

Forecasting Model	**Quantile**				**90-10**
	0.05	**0.25**	**0.75**	**0.95**	
Benchmark	1.00	1.00	1.00	1.00	1.00
(1) Bayes	1.04	1.03	1.00	0.99	1.02
(2) H.Prior	0.93	0.96	0.97	0.97	0.96
(3) t-innovations	0.92	0.95	0.95	0.98	0.95
(4) A-outliers	0.92	0.95	0.95	0.97	0.95
(5) Stoch. Vol.	0.92	0.94	0.93	0.90	0.92
(6) TV mean parms	0.87	0.93	0.94	0.91	0.91
(7) Com. Factors	0.86	0.90	0.90	0.85	0.87

Log-Bayes Factors – Model 7 versus

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

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

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

$$u_{j,t} = \sum_{l=1}^{12} \phi_{j,l,t} u_{j,t-l} + \epsilon_{j,t}$$

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

$$\ln(\sigma^2_{j,t}) = \ln(\sigma^2_j) + \sum_{l=1}^{p} \varphi_{l,t} \xi_{j,l} \text{ with } p = \lfloor T/36 \rfloor$$

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

$$o_{j,t} = \kappa_j \eta_{j,t} \text{ and } \eta_{j,t} \sim \mathcal{T}(2)$$

Restricted Model	$\ln(BF)$
Reduce hierarchical priors for (ϕ, σ)	86
Increase ν in $\mathcal{T}(\nu_j)$ for $\varepsilon_{j,t}$	14
Reduce outliers $o_{j,t}$ and $o_{n+1,t}$	11
Reduce stochastic volatility	31
Reduce time variation in ϕ	18
Reduce time variation in μ	6
Reduce variance of $c_{j,t}$ factor	3.2
Reduce time variation in λ	18

Summary of Empirical Analysis: State Employment

Forecasting

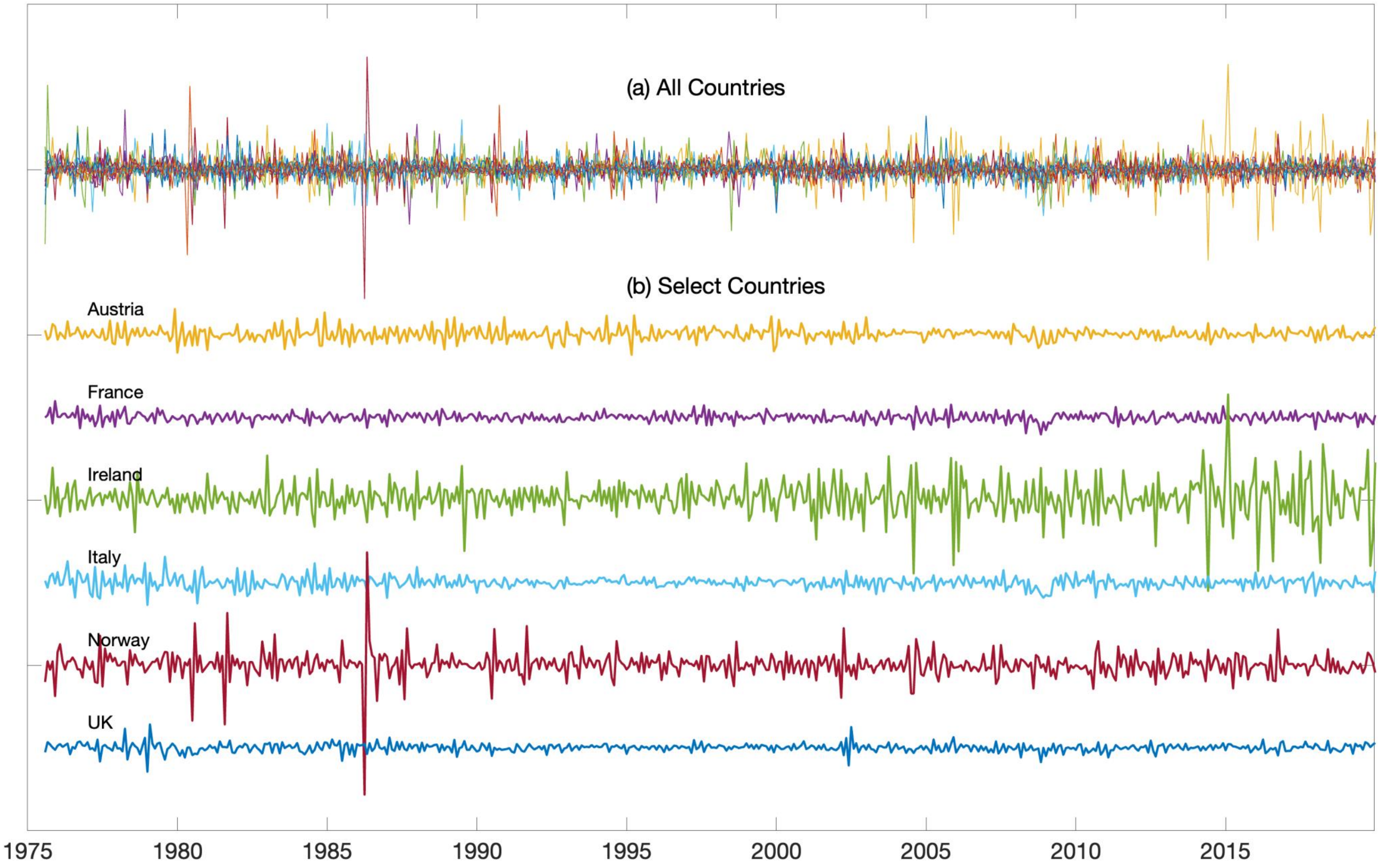
- Gains from Hierarchical Priors
  - Exploit similar DGPs
- Gains for covariance/lead-lag
- Predictive quantile gains gains from
  - fat tails
  - stochastic volatility
  - time-varying mean parameters

Which model fit better?
Full-Sample Bayes Factors

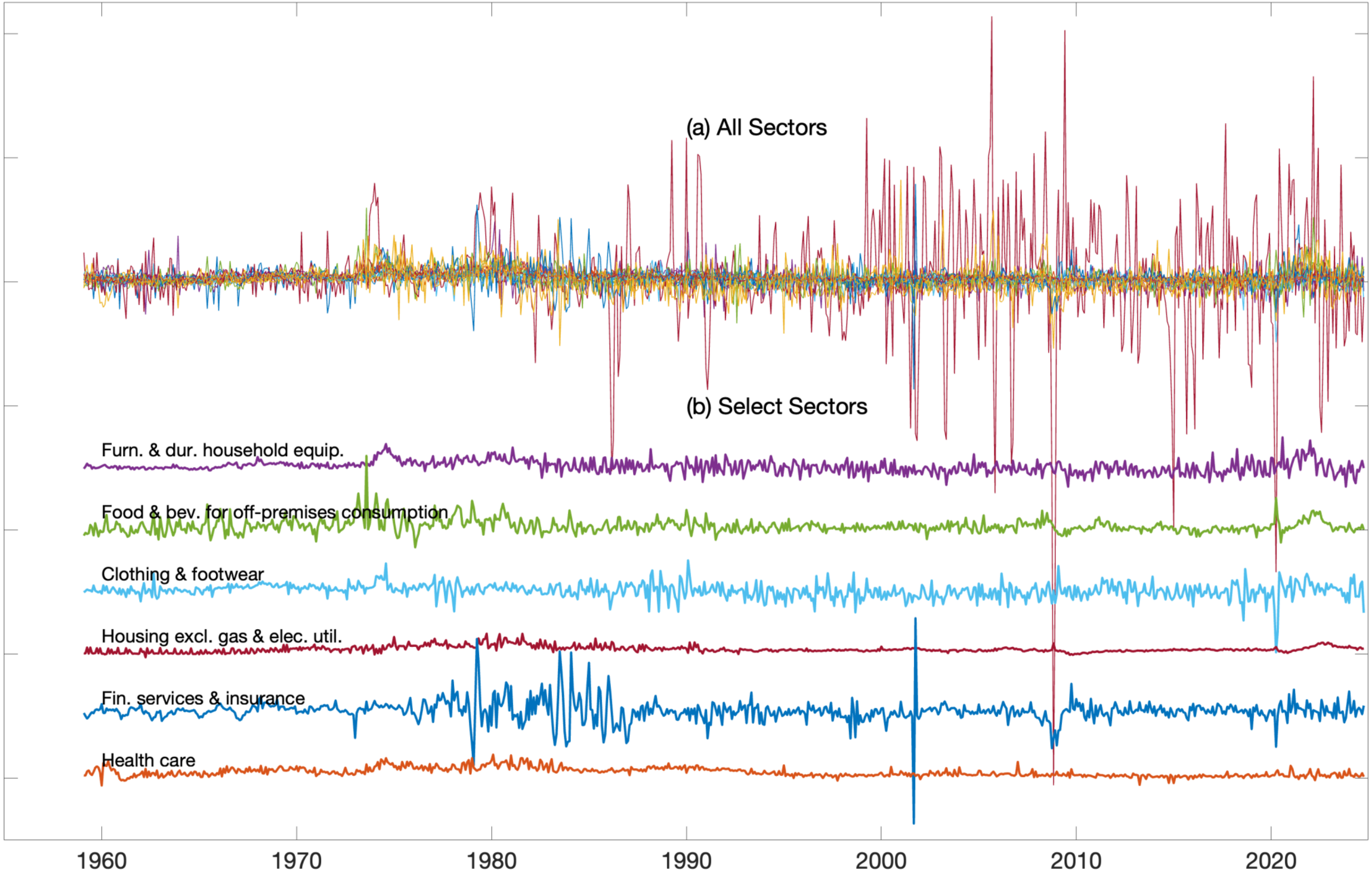
- All features are important

Two New Datasets

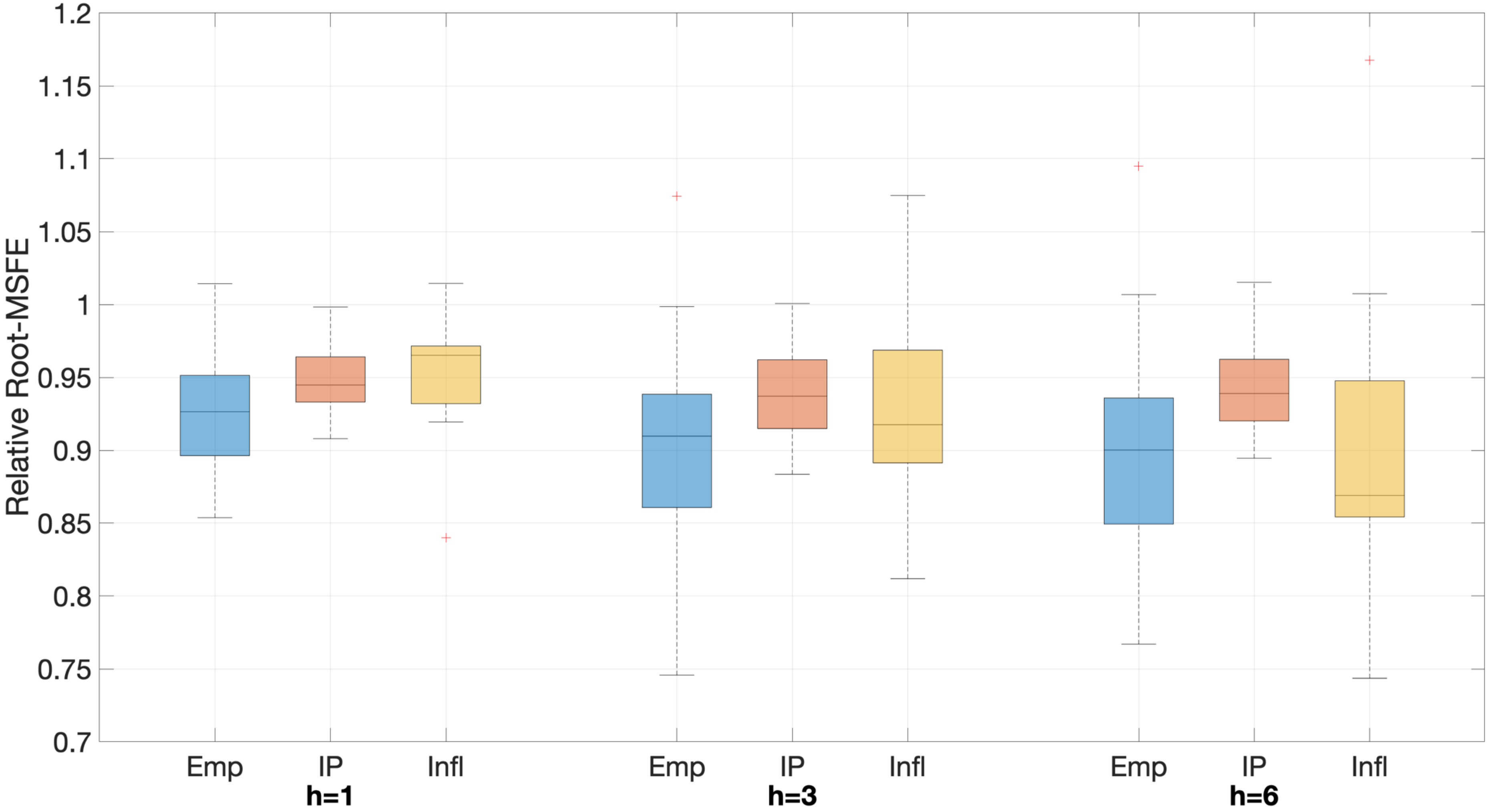
IP Growth: 16 European Countries



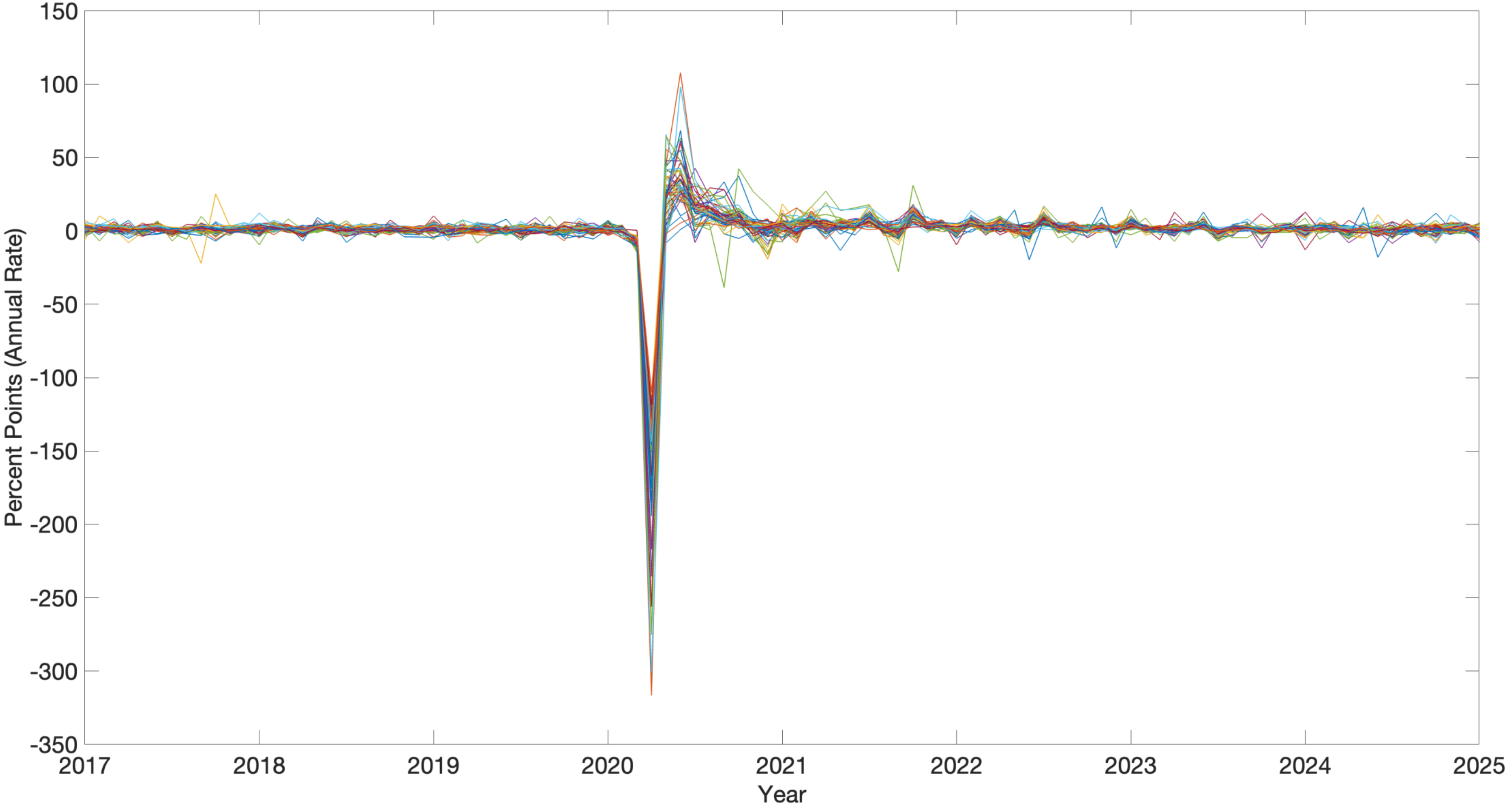
U.S. PCE Inflation: 17 Sectors



Forecasting Performance of RTS model – through 2019m6



Forecasting in the Aftermath of Covid



Forecasting in the Aftermath of Covid

POOS forecast period	**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) of the benchmark AR(12) model		
199m12-2019m6	3.1	2.0	1.7
2020m6-2024m4	27.7	17.4	8.1
2021m6-2024m4	4.1	2.4	1.8
	(a) Relative root-MSFE of Model 7		
199m12-2019m4	0.90	0.88	0.89
2020m6-2024m4	0.19	0.19	0.28
2021m6-2024m4	0.86	0.73	0.69

Wrapping Up

What have we done ?

- 7 Increasingly Rich Models to Capture Characteristics of Related Macroeconomic Time Series
- Algorithms for efficiently computing forecasts
- 3 Empirical Applications

What have we learned?

- Hierarchical Models help
- Common Factors Help
- Other features help for predictive quantiles, but not so much for point forecasts

Open Questions?

- Comparative performance of alternative models
- Why are predictive R^2 so low?