Locally Stable Time-Varying Multi-Seasonal AR Models

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Collaborators

- **Ganna Fagerberg** (lead author), Stockholm University
- **Robert Kohn**, University of New South Wales

Structure of the talk

- **Time-varying multi-seasonal AR models**
- **Dynamic shrinkage process priors**
- **Bayesian inference**
- **Simulation experiments**
- **Application to US industrial production 1919-2024**

- **Slides**: http://mattiasvillani.com/news.
- **Paper**: on arXiv very soon.

Motivation

- **Box-Jenkins. Global stationarity** is restrictive.
- **Differencing**. Information loss. Overdifferencing. Levels are non-stationary. Poor short-run forecasts.
- **Data in most recent regime**. Wasteful. No probability for regime change.

- **Local stationarity** [1]

- **Time-varying parameters** has a long history. [2], [3]
- Recently: more realistic parameter evolutions. [4], [5], [6], [7]
- Parameters can enter **explosive region** for some periods.

- **Seasonality** is predictable. Model it! Time-varying?
- Time series with **multiple seasonal periods** is now common.

Multi-seasonal AR models

- **Seasonal AR(p, P)** with season s

$$\phi_p(L)\Phi_P(L^s)(x_t - \mu) = \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

- Lag operator: $L^j x_t = x_{t-j}$
- Regular **lag polynomial**

$$\phi_p(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$$

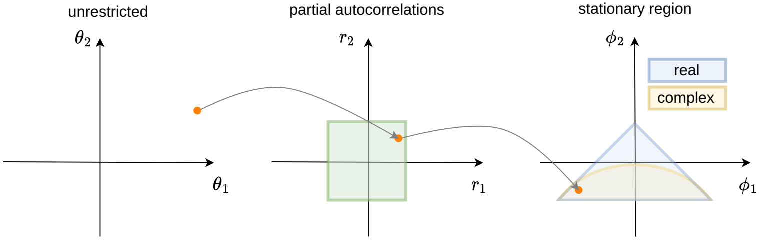
- Seasonal lag polynomial

$$\Phi_p(L^s) = 1 - \Phi_1 L^s - \ldots - \Phi_P L^{Ps}$$

- **Multiple seasonal periods**, e.g. hourly data with daily, weekly and yearly cycles.
- **Multi-seasonal AR** models with M polynomials

$$\prod_{j=1}^{M} \phi_j(L^{s_j})(x_t - \mu) = \varepsilon_t$$

Enforce stability in AR models



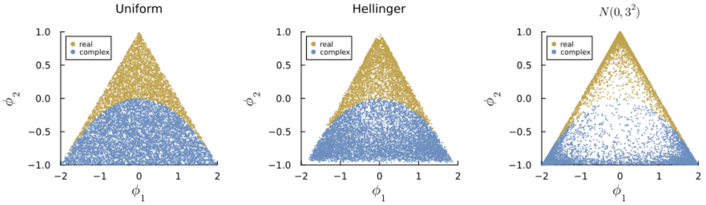
- **Unrestr. params** $\theta_k \in \mathbb{R} \Rightarrow$ **Partial autocorr** $r_k \in (-1, 1)^p$

$$r_k = \frac{\theta_k}{(1 + \theta_k^2)^{1/2}}$$

- Partial autocorrs $\boldsymbol{r} \Rightarrow$ **stable AR coeffs** ϕ by a recursion [8]
- Same parameterization for all polynomial factors.

Uniform distribution over stationary region

- Unrestricted parameters θ_k have no interpretation. Priors?
- **Uniform distribution over stationary region** $\mathbb{S}_p$ for ϕ.



- Lemma: If, independently,

$$\theta_k \sim \begin{cases} t\left(k+1, 0, \frac{1}{\sqrt{k+1}}\right) & \text{if } k \text{ is odd} \\ t_{\text{skew}}\left(\frac{k}{2}, \frac{k+2}{2}, 0, \frac{1}{\sqrt{k+1}}\right) & \text{if } k \text{ is even,} \end{cases}$$

  then $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ is uniformly distributed over $\mathbb{S}_p$.
- The skew-t distribution is the one in [9].

Time-varying multi-seasonal AR

- Time-varying multi-seasonal AR

$$\prod_{j=1}^{M} \phi_{jt}(L^{s_j})(x_t - \mu_t) = \varepsilon_t \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_t^2)$$

ϕ_{jkt} is kth AR parameter in jth polynomial at time t.

- $\text{TVSAR}(\boldsymbol{p}, \boldsymbol{s})$
  - $\boldsymbol{s} = (s_1, \dots, s_M)$ are the seasonal periods
  - $\boldsymbol{p} = (p_1, \dots, p_M)$ lags for the seasonal periods.

- $\boldsymbol{\phi}_t = (\phi_{1t}, \dots, \phi_{p_j t})^\top$ parameters in AR polynomial at time t.
- $\boldsymbol{\phi}_t = \boldsymbol{g}(\boldsymbol{\theta}_t)$ map from unrestricted $\boldsymbol{\theta}_t$ to stable $\boldsymbol{\phi}_t$.

Recap: horseshoe prior in linear regression

- **Linear regression**

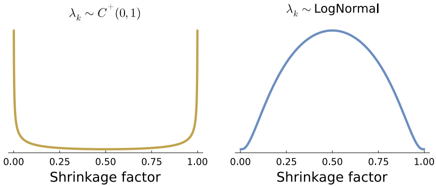
$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

- **Horseshoe prior**

$$\beta_k \sim N(0, \sigma^2 \tau^2 \lambda_k^2)$$
$$\lambda_k \sim C^+(0, 1)$$

- **Global variance** τ^2 is modulated by **local variance** λ_k^2.
- **Shrinkage factor** s_k (orthogonal covariates)

$$\tilde{\beta}_k = (1 - s_k)\hat{\beta}_k, \qquad s_k = \frac{1}{1 + (n/\sigma^2)\tau^2 \lambda_k^2}$$



Dynamic shrinkage process priors for TVSAR

- **Dynamic shrinkage process** prior [6] for TVSAR

$$\boldsymbol{\phi}_t = \boldsymbol{g}(\boldsymbol{\theta}_t)$$

$$\theta_{kt} = \theta_{k,t-1} + \nu_{kt}, \qquad \nu_{kt} \overset{\text{indep}}{\sim} N\left(0, \exp(h_{kt})\right)$$

$$h_{kt} = \mu_k + \kappa_k(h_{k,t-1} - \mu_k) + \eta_{kt}, \qquad \eta_{kt} \overset{\text{iid}}{\sim} Z(1/2, 1/2, 0, 1)$$

- **Global volatility** μ_k
- Persistent **local volatility** η_{kt} from heavy-tailed Z-distr.
- Constant periods, periods of rapid change and jumps.
- Time-series extension of **horseshoe prior** [10], since for $\kappa = 0$

$$\mathrm{Var}(\nu_t) = \exp(h_t) = \exp(\mu)\exp(\eta_t) = \tau^2\lambda_t^2$$

and

$$\eta_t \sim Z(1/2, 1/2, 0, 1) \implies \lambda_t = \exp(\eta_t/2) \sim C^+(0, 1)$$

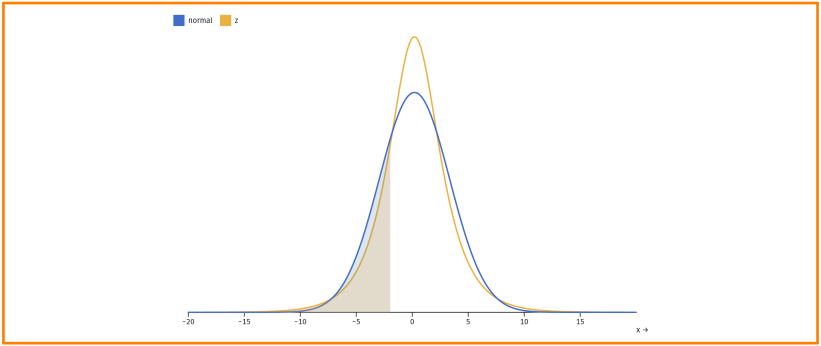
Z-distribution

- Also called **Logistic-Beta distribution** since

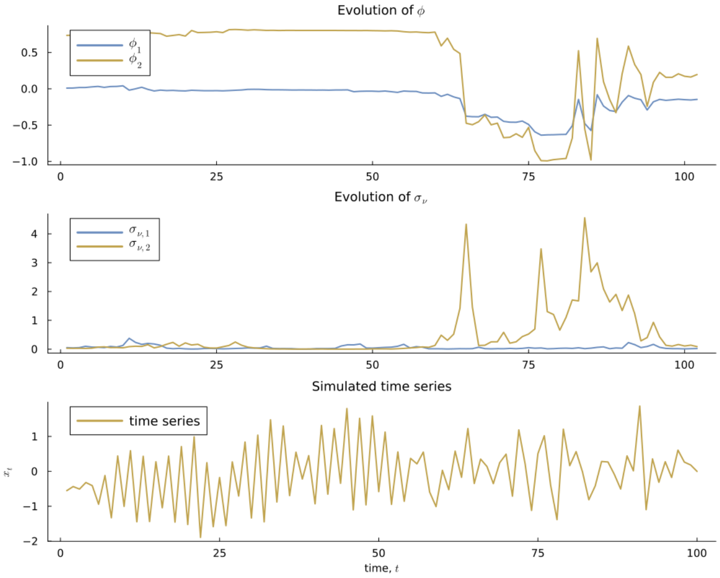
$$X \sim \mathrm{Beta}(\alpha, \beta) \implies \log\left(\frac{X}{1-X}\right) \sim Z(\alpha, \beta, 0, 1)$$

- Four-parameter version by location-scale

$$X \sim Z(\alpha, \beta, 0, 1) \implies \mu + \sigma X \sim Z(\alpha, \beta, \mu, \sigma)$$



Dynamic shrinkage process priors for TVSAR



Conditional likelihood function

- Multi-seasonal TVSAR can be expressed as **regression** with **nonlinear restrictions** by multiplying out polynomials in

$$\prod_{j=1}^{M} \phi_{jt}(L^{s_j})x_t = \varepsilon_t$$

- Example (static case):

$$(1-\phi_1 L)(1-\Phi_1 L^s)x_t = (1-\phi_1 L - \Phi_1 L^s + \phi_1\Phi_1 L^{1+s})x_t$$

$$y_t = \boldsymbol{z}_t^\top \tilde{\boldsymbol{\phi}} + \varepsilon_t$$

  - $y_t = x_t$
  - $\boldsymbol{z}_t = (x_{t-1}, x_{t-s}, x_{t-(1+s)})^\top$
  - $\tilde{\boldsymbol{\phi}} = (\phi_1, \Phi_1, \phi_1\Phi_1)^\top$.
  - Lags $2, \ldots, s-1$ have zero coefficients.

Bayesian inference

- TVSAR with conditional likelihood:

$$y_t = \boldsymbol{z}_t^\top \tilde{\boldsymbol{g}}(\boldsymbol{\theta}_t) + \varepsilon_t \qquad \varepsilon_t \overset{\text{indep}}{\sim} N(0, \sigma_t^2)$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \overset{\text{indep}}{\sim} N\left(0, \mathrm{Diag}(\exp(\boldsymbol{h}_t))\right)$$

$$\boldsymbol{h}_t = \boldsymbol{\mu} + \boldsymbol{\kappa}(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \qquad \eta_{kt} \overset{\text{iid}}{\sim} Z(1/2, 1/2, 0, 1)$$

- Aim: posterior conditional on time series $y_{1:T}$

$$p(\phi_{0:T}, \boldsymbol{h}_{0:T}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \sigma_{1:T} | y_{1:T})$$

- Gibbs sampling by data augmentation:
  - Mixture of normal for $\log \chi_1^2$ for volatility $\boldsymbol{h}_{0:T}$ [11]
  - Polya-Gamma augmentation for Z-distribution. [6]

$$\eta_t \overset{\text{iid}}{\sim} Z(1/2, 1/2, 0, 1)$$

$$\Longleftrightarrow$$

$$\eta_t | \xi_t \overset{\text{indep}}{\sim} N(0, \xi_t^{-1}), \quad \xi_t \overset{\text{iid}}{\sim} \mathrm{PG}(1, 0)$$

Sampling from conditional posterior for $\theta_{0:T}$

- Full conditional posterior for $\boldsymbol{\theta}_{0:T}$:
  - state-space model with state vector $\boldsymbol{\theta}_t$.
  - **nonlinear Gaussian observation model**.
  - Linear (heteroscedastic) Gaussian state transition.

- **Particle Gibbs with Ancestor Sampling** (**PGAS**) [12]
  - simulation consistent
  - slow
  - particle degeneracy when model is near-degenerate (parameters standing still for extended periods).

- **FFBSx - FFBS with extended Kalman filter**
  - fast
  - robust to near-degeneracy
  - approximate, but shown to be very accurate for TVSAR
  - **automatic differentiation** $\tilde{\phi}_t = \tilde{\boldsymbol{g}}(\boldsymbol{\theta}_t)$ makes it all beautiful.

Extended Kalman filter

- State-space model

$$
\begin{aligned}
\boldsymbol{\theta}_t &= \boldsymbol{A}\boldsymbol{\theta}_{t-1} + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{\eta}_t, \quad & \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\
y_t &= C(\boldsymbol{\theta}_t) + \varepsilon_t, & \varepsilon_t \sim N(0, \boldsymbol{\Sigma}_\varepsilon)
\end{aligned}
$$

- Yesterday's posterior: $\boldsymbol{\theta}_{t-1}|y_{1:t-1} \sim N(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Omega}_{t-1|t-1})$
- Today's prior: $\boldsymbol{\theta}_t|y_{1:t-1} \sim N(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Omega}_{t|t-1})$
- Today's posterior: $\boldsymbol{\theta}_t|y_{1:t} \sim N(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Omega}_{t|t})$

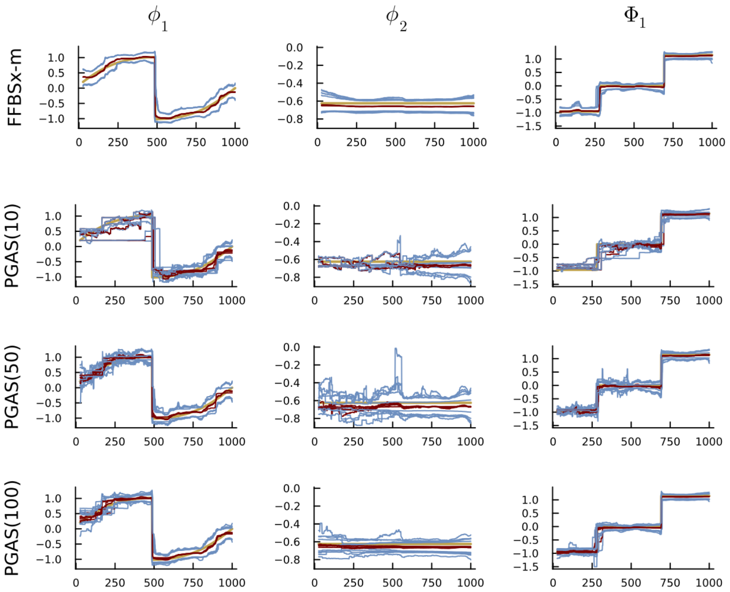
```
# Prior propagation step
μ̄ = A*μ .+ B*u
Ω̄ = A*Ω*A' + Σₙ

# Measurement update
K = Ω̄*C' / (C*Ω̄*C' .+ Σₑ)
μ = μ̄ + K*(y .- C*μ̄)
Ω = (I - K*C)*Ω̄
```

```
# Prior propagation step
μ̄ = A*μ + B*u
Ω̄ = A*Ω*A' + Σₙ
C̄ = ∂C(μ̄, Cargs)

# Measurement update
K = Ω̄*C̄' / (C̄*Ω̄*C̄' .+ Σₑ)
μ = μ̄ + K*(y .- C(μ̄, Cargs))
Ω = (I - K*C̄)*Ω̄
```

Simulation TVSAR with $p = (2, 2)$ and $s = (1, 12)$



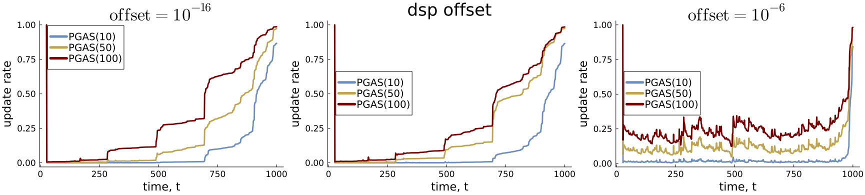
PGAS - larger offset improves convergence

- Volatility models

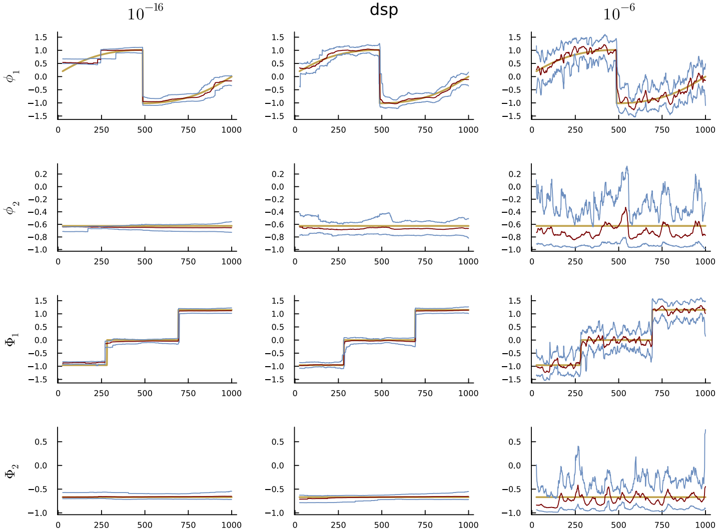
$$\nu_t = \exp(h_t/2)\epsilon_t, \qquad \epsilon_t \sim N(0,1)$$

- The usual trick

$$\log(\nu_t^2 + \text{offset}) = h_t + \log \epsilon_t^2$$



PGAS - but makes parameters more wiggly



Spectral density

- **Autocovariance function**

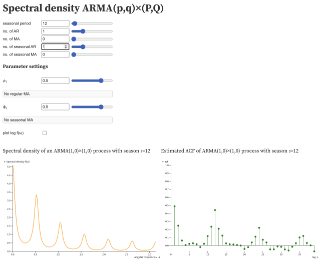
$$\gamma_\tau = \mathbb{E}\left[(x_t - \mu)(x_{t-\tau} - \mu)\right], \quad \tau = 0, 1, \ldots$$

- **Spectral density**

$$f(\omega) \equiv \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_\tau \exp(-\mathrm{i}\omega\tau) \ \text{ for } \omega \in (-\pi, \pi].$$

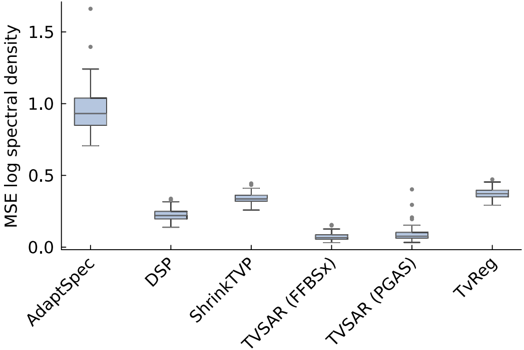
- Seasonal AR

$$f(\omega) = \frac{\sigma^2}{\pi} \frac{1}{|\phi_p(e^{-i\omega})|^2} \frac{1}{|\Phi_P(e^{-is\omega})|^2} \ \text{ for } \omega \in (0, \pi)$$

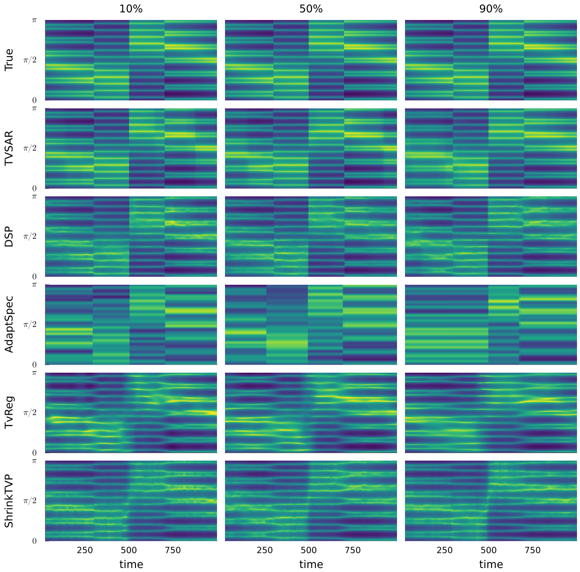


Fitting the time-varying spectral density

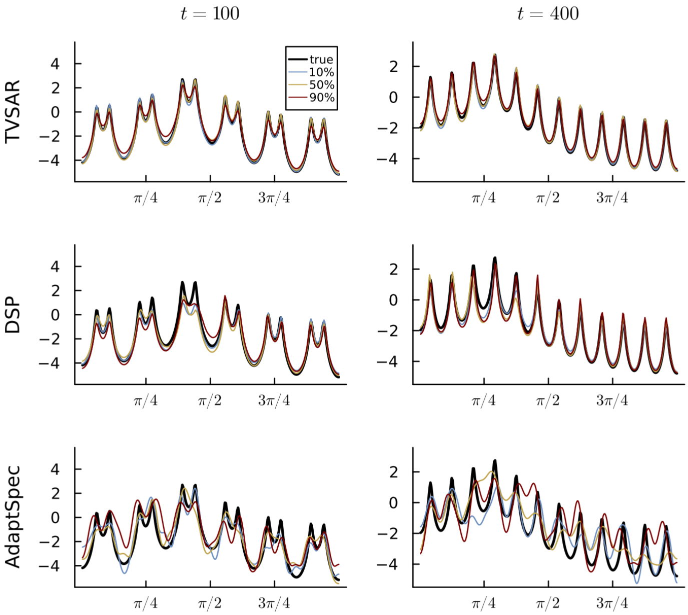
- Benchmark models are linear:
  - Fit $y_t = \boldsymbol{z}_t^\top \tilde{\boldsymbol{\phi}}_t + \varepsilon_t$ with zero restrictions
  - Coeff on lag $x_{t-(1+s)}$ fitted without multiplicative restriction.
- **Time-varying spectral density** $f(\omega, t)$. Spectrogram.
- MSE of log spectrogram estimate: $\widehat{\log f(\omega, t)}$ integrated over frequency $\omega \in (0, \pi]$ and time $t = 1, \ldots, 1000$.



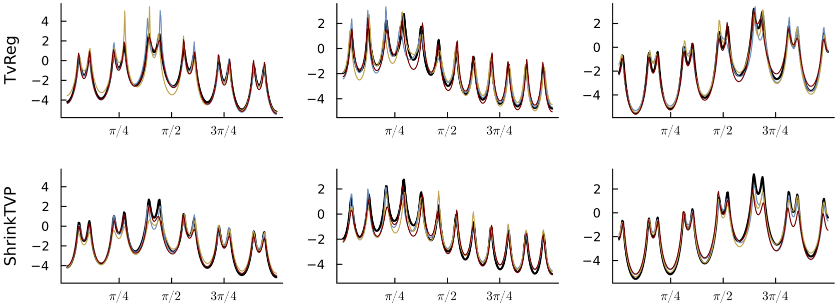
Fitted log spectrogram



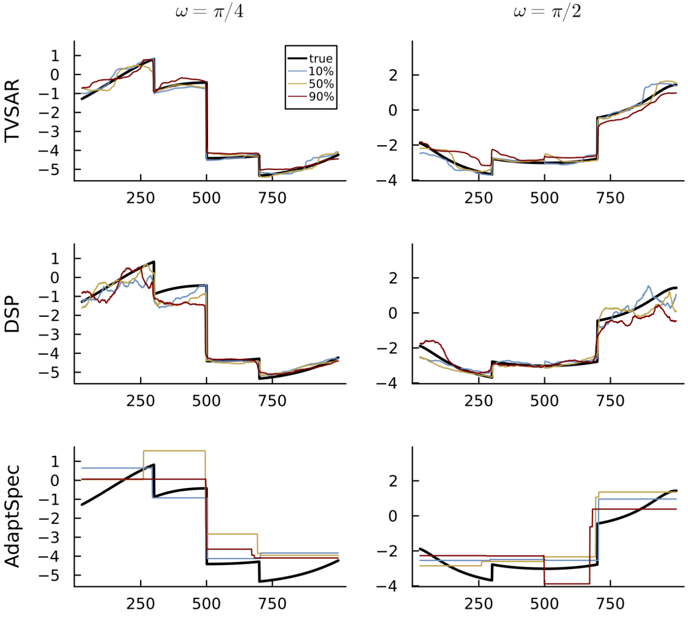
Fitted log spectrogram snapshots



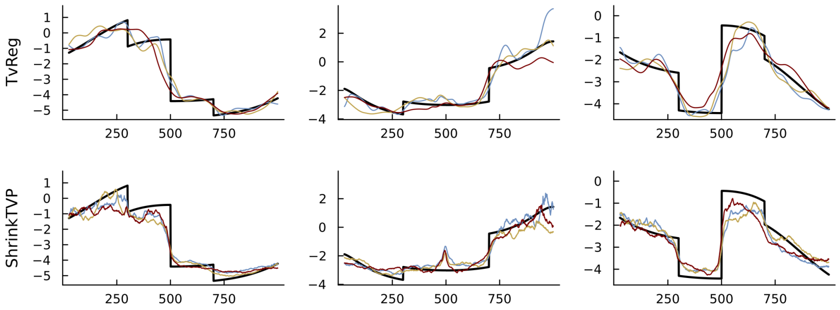
Fitted log spectrogram snapshots



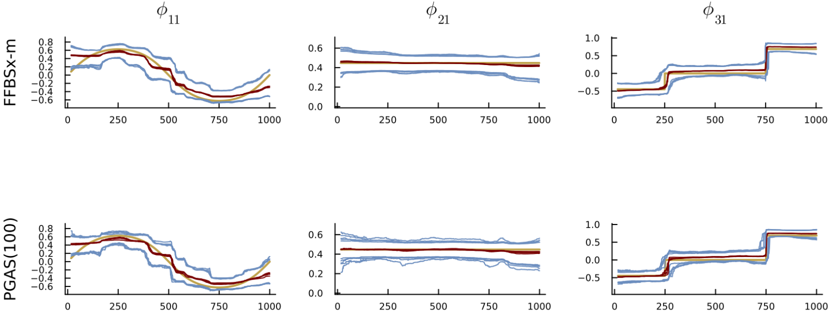
Fitted log spectral density over time



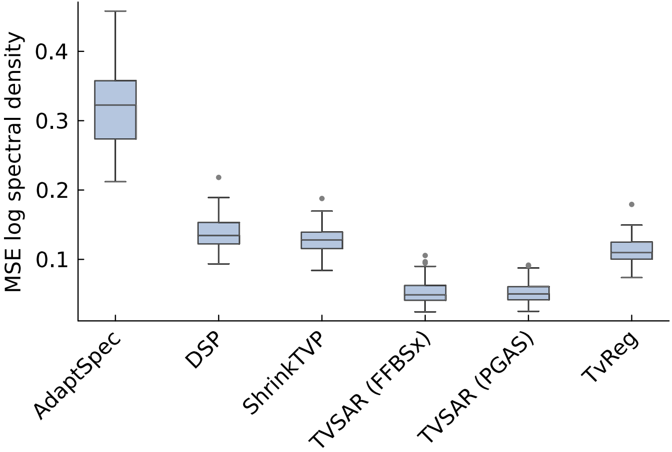
Fitted log spectral density over time



Simulation TVSAR with $p = (1, 1, 1)$ and $s = (1, 4, 12)$

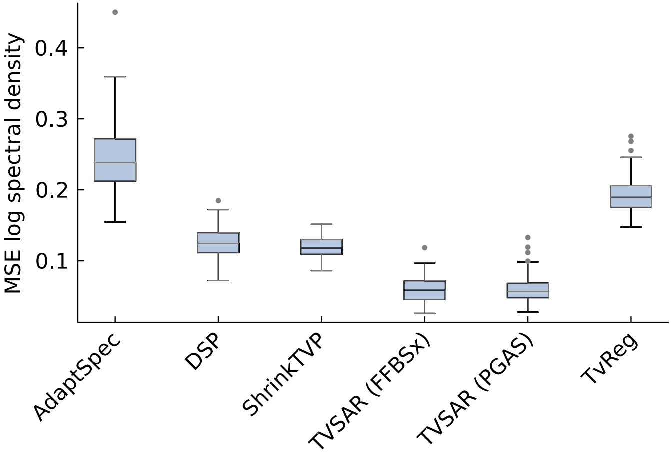


MSE for TVSAR $p = (1, 1, 1)$ and $s = (1, 4, 12)$

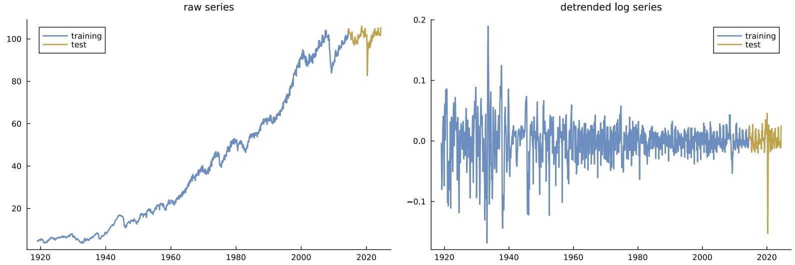


MSE log spectrogram - redundant lags

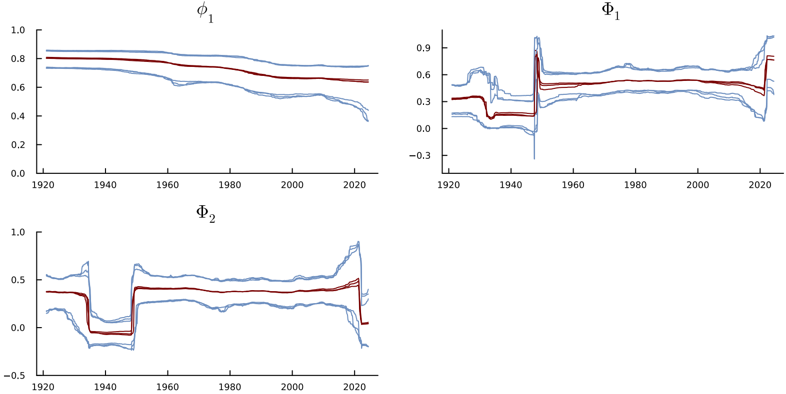
- Simulate from TVSAR$(1,1)$ fit TVSAR$(2,2)$.



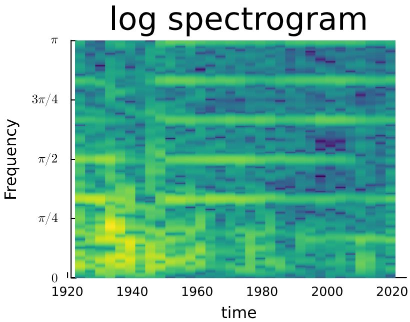
US industrial production 1919-2024

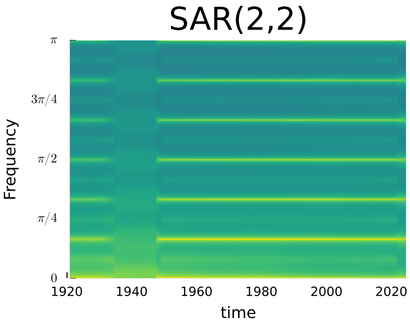


SAR(1,2) - FFBSx three random initial values

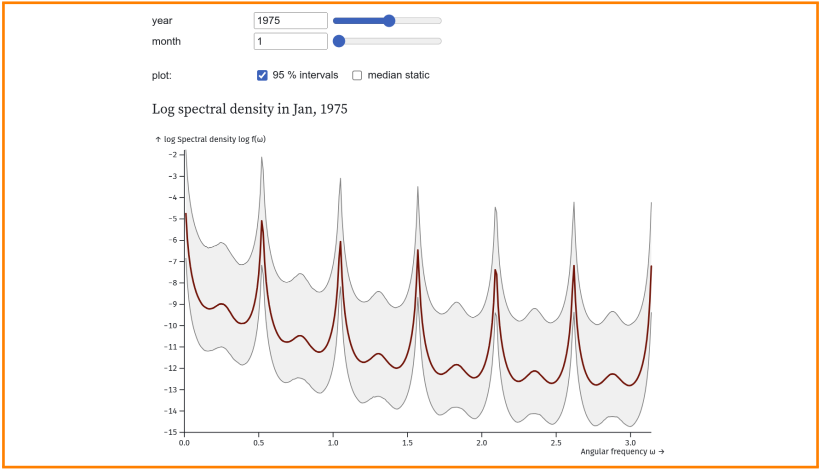


SAR(1,2) spectral density heatmap

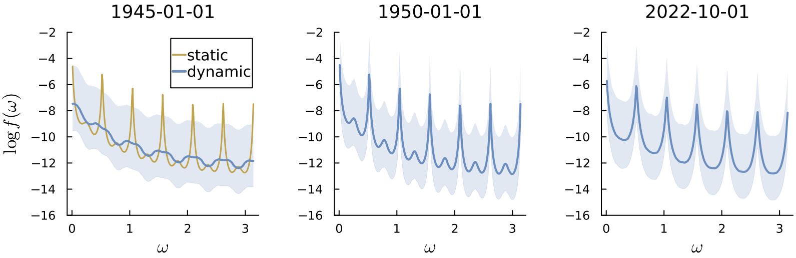




SAR(1,2) spectral density snapshots



SAR(1,2) spectral density snapshots



Conclusions

- **Multi-seasonal AR** model with **time-varying** parameters.
- Stability restrictions at every time point. **Locally stable**.
- Parameter evolution by **dynamic shrinkage processes**.
- Parameters can stand still, move rapidly or jump.
- Challenging inference problem:
  - **non-linear** - stability restrictions and multiplicative seasonality
  - dynamic shrinkage prior tends to give **near-degeneracy**.
- Gibbs sampler with **particle/extended Kalman update** step.
- In progress:
  - Extension to ARMA and exact likelihood.
  - Seasonal VARMA with Ansley-Kohn stability restrictions.
  - Models directly in the spectral domain.
- julia package for arbitrary number of polynomials is coming.

R. Dahlhaus, "Locally stationary processes," in *Handbook of statistics*, vol. 30, pp. 351–413, Elsevier, 2012.

T. Doan, R. Litterman, and C. Sims, "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, vol. 3, no. 1, pp. 1–100, 1984.

G. E. Primiceri, "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, vol. 72, no. 3, pp. 821–852, 2005.

R. Gerlach, C. Carter, and R. Kohn, "Efficient bayesian inference for dynamic mixture models," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 819–828, 2000.

M. Kalli and J. E. Griffin, "Time-varying sparsity in dynamic regression models," *Journal of Econometrics*, vol. 178, no. 2, pp. 779–793, 2014.

D. R. Kowal, D. S. Matteson, and D. Ruppert, "Dynamic shrinkage processes," *Journal of the Royal Statistical Society:*

Series B (Statistical Methodology), vol. 81, no. 4, pp. 781–804, 2019.

P. Knaus, A. Bitto-Nemling, A. Cadonna, and S. Frühwirth-Schnatter, "Shrinkage in the time-varying parameter model framework using the R package shrinkTVP," *Journal of Statistical Software*, vol. 100, no. 13, 2021.

O. Barndorff-Nielsen and G. Schou, "On the parametrization of autoregressive models by partial autocorrelations," *Journal of Multivariate Analysis*, vol. 3, no. 4, 1973.

M. C. Jones and M. Faddy, "A skew extension of the t-distribution, with applications," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 65, no. 1, pp. 159–174, 2003.

C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.

S. Kim, N. Shephard, and S. Chib, "Stochastic volatility: likelihood inference and comparison with arch models," *The review of economic studies*, vol. 65, no. 3, pp. 361–393, 1998.

F. Lindsten, M. Jordan, and T. B. Schön, "Particle gibbs with ancestor sampling," *Journal of Machine Learning Research*, vol. 15, p. 2145—2184, 06 2014.