

# Bayesian inference in time-varying parameter models with global-local shrinkage process priors

Mattias Villani

Department of Statistics  
Stockholm University



[mattiasvillani.com](http://mattiasvillani.com)



@matvil



@matvil



[mattiasvillani](http://mattiasvillani)

# HiTEc Cost Action



**COST Action HITEC**  
Text, functional and other  
high-dimensional data in econometrics:  
New models, methods, applications



# HiTEc COST Action



<http://www.hitecation.org/>  
Instagram: hitec\_cost\_action

## Collaborators

- **Ganna Fagerberg**, Stockholm University
- **Robert Kohn**, University of New South Wales
- **Oskar Gustafsson**, Stockholm University
- **Yijie Niu**, Stockholm University

# Plan for the talk

- Setting the stage: Bayesian inference in state-space models
- Motivation: Time-varying multi-seasonal AR models with dynamic shrinkage process prior
- Challenge: Algorithms for Bayesian inference based on Gaussian approximations
- Slides: <http://mattiasvillani.com/news>
- Paper: arXiv: 2409.18640 (ARMA extension soon on arXiv)

# Collaborators

- **Ganna Fagerberg**, Stockholm University
- **Robert Kohn**, University of New South Wales
- **Oskar Gustafsson**, Stockholm University
- **Yijie Niu**, Stockholm University

# Time-varying parameter models

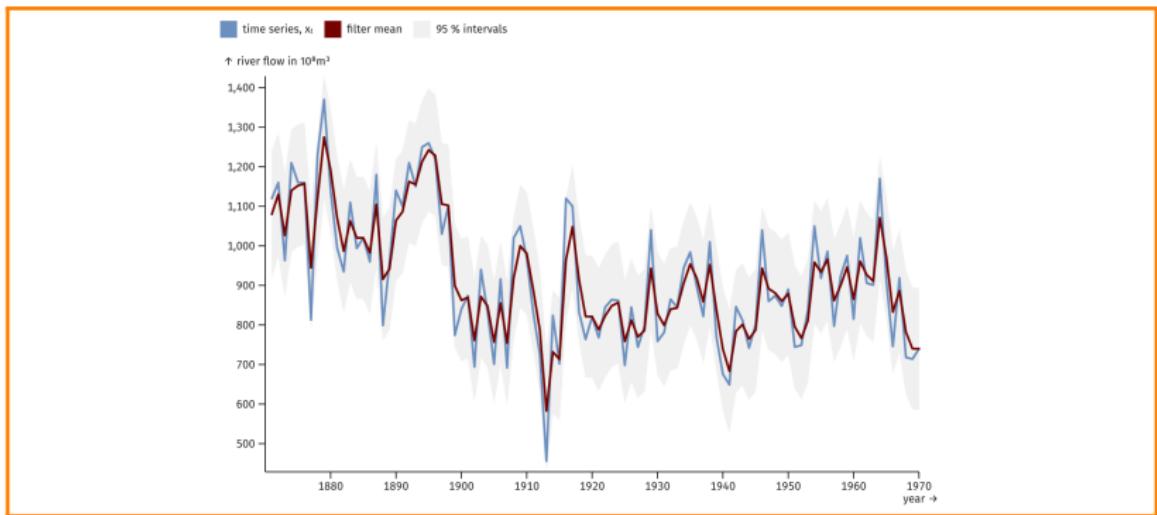
## ■ Regression with time-varying parameters

Measurement model:  $y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$

Transition model:  $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\nu)$

- Linear Gaussian **state-space model** with state:  $\mathbf{z}_t = \boldsymbol{\beta}_t$
- **Filtering posterior:**  $p(\mathbf{z}_t | y_{1:t})$ . Instantaneous posterior.
- **Smoothing posterior:**  $p(\mathbf{z}_t | y_{1:T})$ . Retrospective posterior.
- **Joint smoothing posterior:**  $p(\mathbf{z}_{1:T} | y_{1:T})$ .

# Local level model for the Nile data



# Parameter evolutions matter

## ■ Regression with time-varying parameters

Measurement model:  $y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$

Transition model:  $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\nu)$

- Homoscedastic Gaussian parameter innovations  
 $\boldsymbol{\nu}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\nu)$  often a problem. Over- and under-smoothing.
- Ideal: parameters can stand still, move rapidly and jump. [1]
- **Dynamic global-local priors** is a great step forward. [2, 3, 4]
- ... but can be tricky computationally.

# Taxonomy of state-space models

## ■ Linear Gaussian models

Measurement model:  $\mathbf{y}_t = \mathbf{C}\mathbf{z}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$

Transition model:  $\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\nu})$

## ■ Non-linear (additive) Gaussian models

Measurement model:  $\mathbf{y}_t = \mathbf{c}(\mathbf{z}_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$

Transition model:  $\mathbf{z}_t = \mathbf{a}(\mathbf{z}_{t-1}) + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\nu})$

## ■ General distribution (non-linear and non-Gaussian) models

Measurement model:  $\mathbf{y}_t \sim p(\mathbf{y}_t | \mathbf{z}_t)$

Transition model:  $\mathbf{z}_t \sim p(\mathbf{z}_t | \mathbf{z}_{t-1})$

## ■ Hybrids: nonlinear measurement + linear Gaussian transition.

## Common Bayesian computational approaches

- Kalman filtering and **FFBS sampling** for linear Gaussian models. [5, 6]
- **Particle MCMC** and **SMC** [7]
- **Hamiltonian Monte Carlo** (Stan/Turing.jl)
- **Variational approximations** [8]
- **INLA** [9]
- **Gaussian approximations** [10]

# Multi-seasonal AR models

## ■ Seasonal AR( $p, P$ ) with season $s$

$$\phi_p(L)\Phi_P(L^s)y_t = \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

## ■ Seasonal AR can be written as a **non-linear regression**:

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

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

- ▶ covariates  $\mathbf{x}_t = (y_{t-1}, y_{t-s}, y_{t-(1+s)})^\top$  and
- ▶ regression coefficients  $\tilde{\boldsymbol{\phi}} = (\phi_1, \Phi_1, \phi_1 \Phi_1)^\top$ .

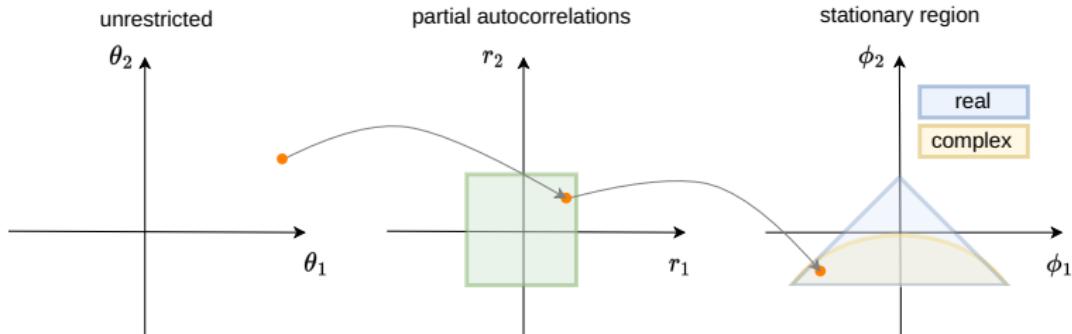
## ■ **Multiple seasonal periods**, e.g. hourly data with daily, weekly and yearly cycles.

## ■ **Multi-seasonal AR** models with $M$ polynomials

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

## ■ Direct extension to **ARMA**. Need to infer the $\varepsilon_t$ .

# Enforce stability/invertibility in ARMA models



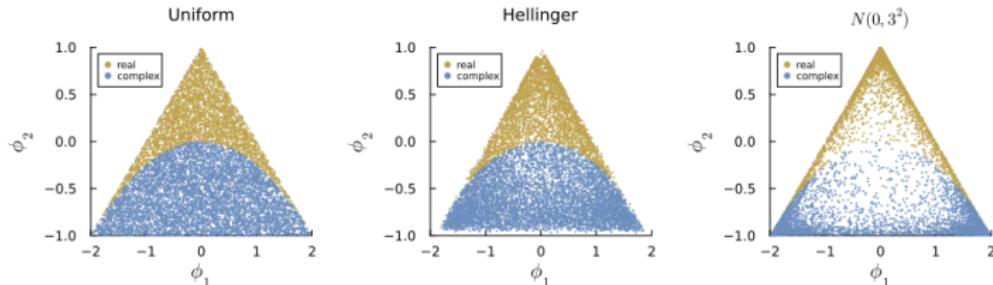
- $\phi = \mathbf{g}(\theta)$  composite map: [11, 12]

unrestricted  $\theta \rightarrow$  partial autocorr  $r \rightarrow$  stable AR  $\phi$

- Same parameterization for all polynomial factors.
- For seasonal models  $\tilde{\phi}_t = \tilde{\mathbf{g}}(\theta_t)$  is the nonlinear mapping from
  - 1 stability restrictions
  - 2 multiplication of polynomials
- **Invertibility in MA** by the same parameterization.

# Uniform distribution over stability region

- Unrestricted parameters  $\theta_k$  have no interpretation. Priors?
- Uniform distribution over stability region  $\mathbb{S}_p$  for  $\phi$ .



- Lemma: If, independently,

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

then  $\phi = (\phi_1, \dots, \phi_p)^\top$  is uniformly distributed over  $\mathbb{S}_p$ .

# Time-varying multi-seasonal AR with dynamic shrinkage process prior

## ■ Time-varying multi-seasonal AR

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

## ■ Dynamic shrinkage process prior [3] for TVSAR

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

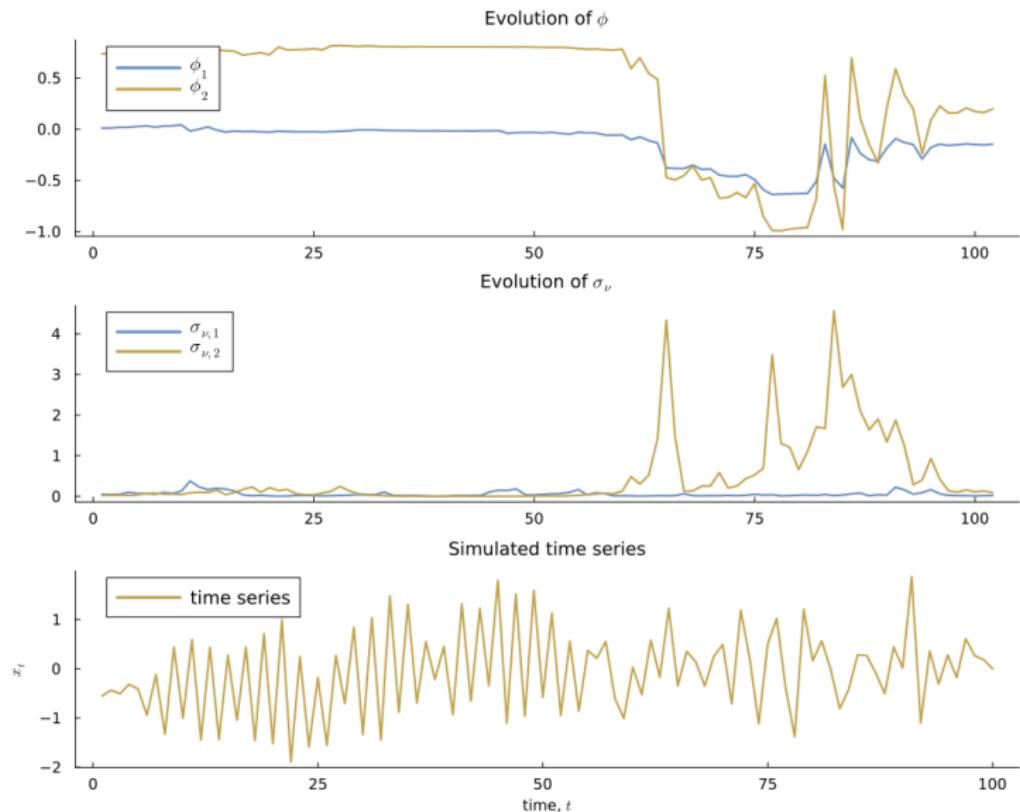
$$\theta_{kt} = \theta_{k,t-1} + \nu_{kt}, \quad \nu_{kt} \stackrel{\text{indep}}{\sim} N(0, \exp(h_{kt}))$$

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

## ■ Global log-variance $\mu_k$ and local log-variance $\eta_{kt}$ .

- Constant periods, periods of rapid change and jumps.
- Time-series extension of the horseshoe prior [13].

# Dynamic shrinkage process priors for TVAR(2)



# Bayesian inference for TVSAR

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

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

- Gibbs sampling by data augmentation:
  - ▶ Mixture of normal for  $\log \chi_1^2$  for volatility  $\mathbf{h}_{0:T}$  [14]
  - ▶ Polya-Gamma augmentation for Z-distribution. [3]
- Conditional on  $\mathbf{h}_{1:T}$ , the TVSAR is a state-space model with:
  - ▶ **nonlinear additive Gaussian measurement** model
    - Nonlinearity 1: multiplicative seasonality
    - Nonlinearity 2: stability restrictions
  - ▶ **linear** (heteroscedastic) **Gaussian transition** model

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

## ■ Particle Gibbs with Ancestor Sampling (PGAS) [15]

- ▶ simulation consistent
- ▶ slow
- ▶ **particle degeneracy when model is near-degenerate**  
(parameters standing still for extended periods).
- ▶ Adding an offset fixes particle degeneracy, but parameters can then no longer stand still. [16]

## ■ FFBSx - FFBS with extended Kalman filter

- ▶ **Local linearization** of transition and measurement functions.
- ▶ fast
- ▶ robust to near-degeneracy
- ▶ approximate, but generally accurate for TVSAR
- ▶ **automatic differentiation** of  $\tilde{\phi}_t = \tilde{\mathbf{g}}(\theta_t)$  makes it beautiful.

# Extended Kalman filter

## ■ State-space model

$$\theta_t = A\theta_{t-1} + Bu_t + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta)$$
$$y_t = C(\theta_t) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon)$$

## Standard Kalman filter

```
# Prior propagation step
```

$$\bar{\mu} = A*\mu .+ B*u$$

$$\bar{\Omega} = A*\Omega*A' + \Sigma_n$$

```
# Measurement update
```

$$K = \bar{\Omega}*C' / (C*\bar{\Omega}*C' .+ \Sigma_e)$$

$$\mu = \bar{\mu} + K*(y .- C*\bar{\mu})$$

$$\Omega = (I - K*C)*\bar{\Omega}$$

## Extended Kalman filter

```
# Prior propagation step
```

$$\bar{\mu} = A*\mu + B*u$$

$$\bar{\Omega} = A*\Omega*A' + \Sigma_n$$

$$\bar{C} = \partial C(\bar{\mu}, \text{Cargs})$$

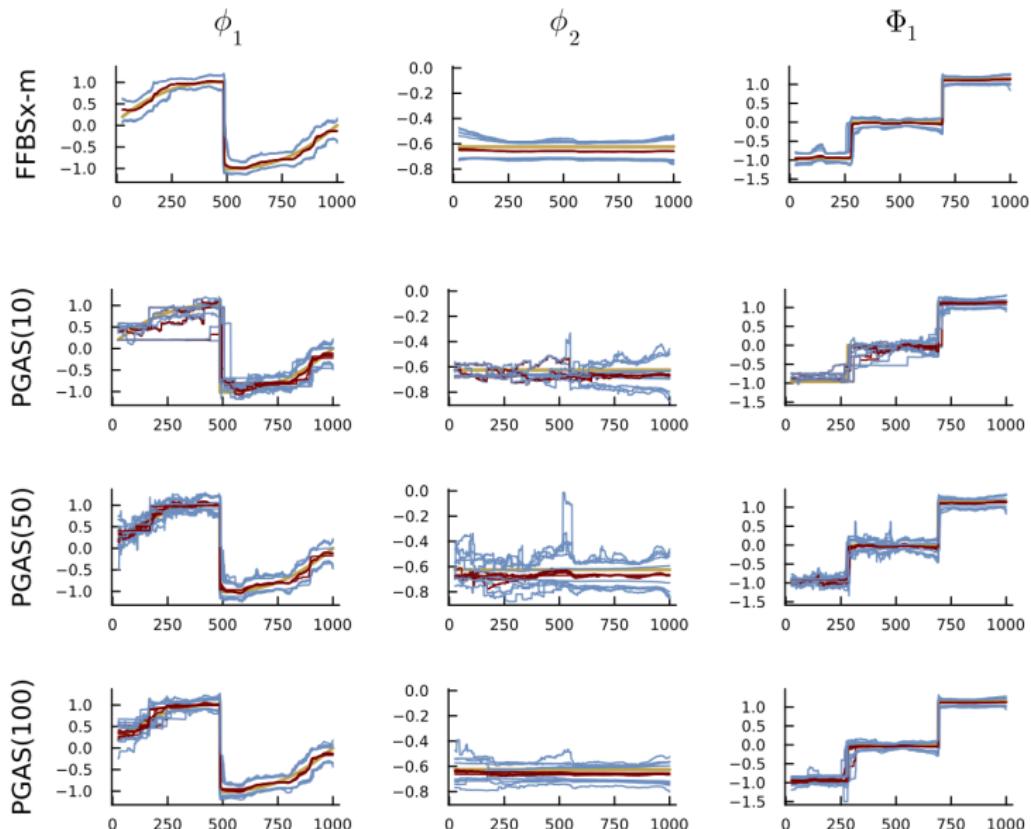
```
# Measurement update
```

$$K = \bar{\Omega}*\bar{C}' / (\bar{C}*\bar{\Omega}*\bar{C}' .+ \Sigma_e)$$

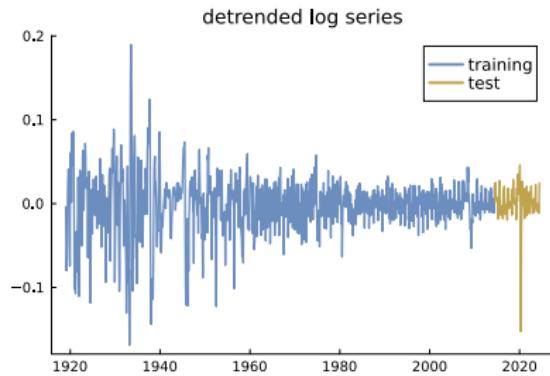
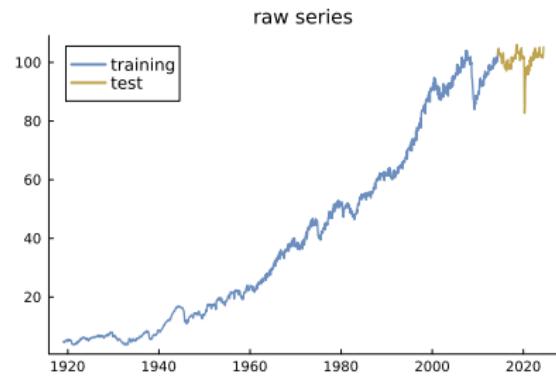
$$\mu = \bar{\mu} + K*(y .- C(\bar{\mu}, \text{Cargs}))$$

$$\Omega = (I - K*\bar{C})*\bar{\Omega}$$

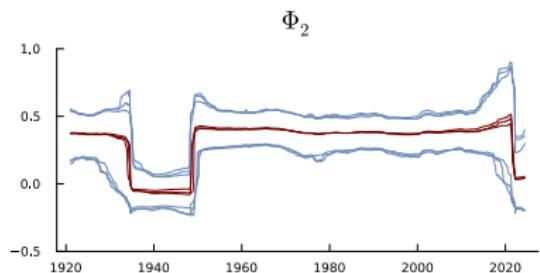
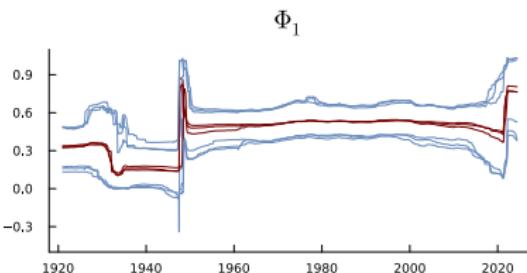
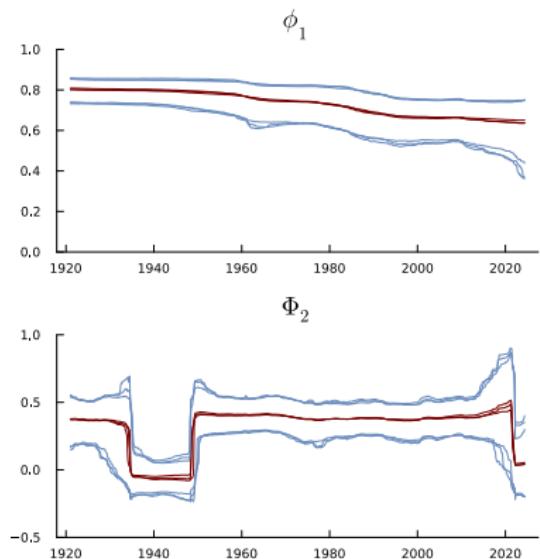
# Simulation TVSAR with $p = (2, 2)$ and $s = (1, 12)$



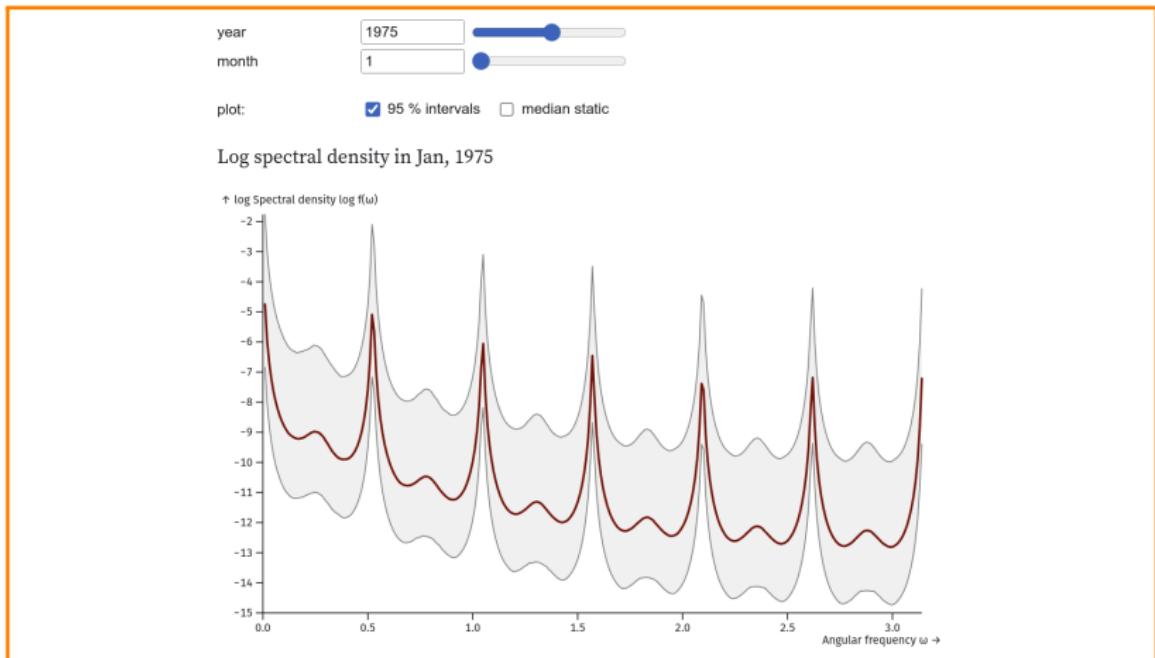
# US industrial production 1919-2024



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



# SAR(1,2) spectral density snapshots



# Approximate Gaussian Kalman-based samplers

- Key for **prior propagation update**: approximate the joint distribution  $p(\mathbf{z}_t, \mathbf{z}_{t-1} | \mathbf{y}_{1:t-1})$

$$\begin{pmatrix} \mathbf{z}_t \\ \mathbf{z}_{t-1} \end{pmatrix} | \mathbf{y}_{1:t-1} \stackrel{\text{approx}}{\sim} N \left[ \begin{pmatrix} \mathbb{E}(\mathbf{z}_t) \\ \mathbb{E}(\mathbf{z}_{t-1}) \end{pmatrix}, \begin{pmatrix} \mathbb{V}(\mathbf{z}_t) & \mathbb{C}(\mathbf{z}_t, \mathbf{z}_{t-1}) \\ \mathbb{C}(\mathbf{z}_t, \mathbf{z}_{t-1})^\top & \mathbb{V}(\mathbf{z}_{t-1}) \end{pmatrix} \right]$$

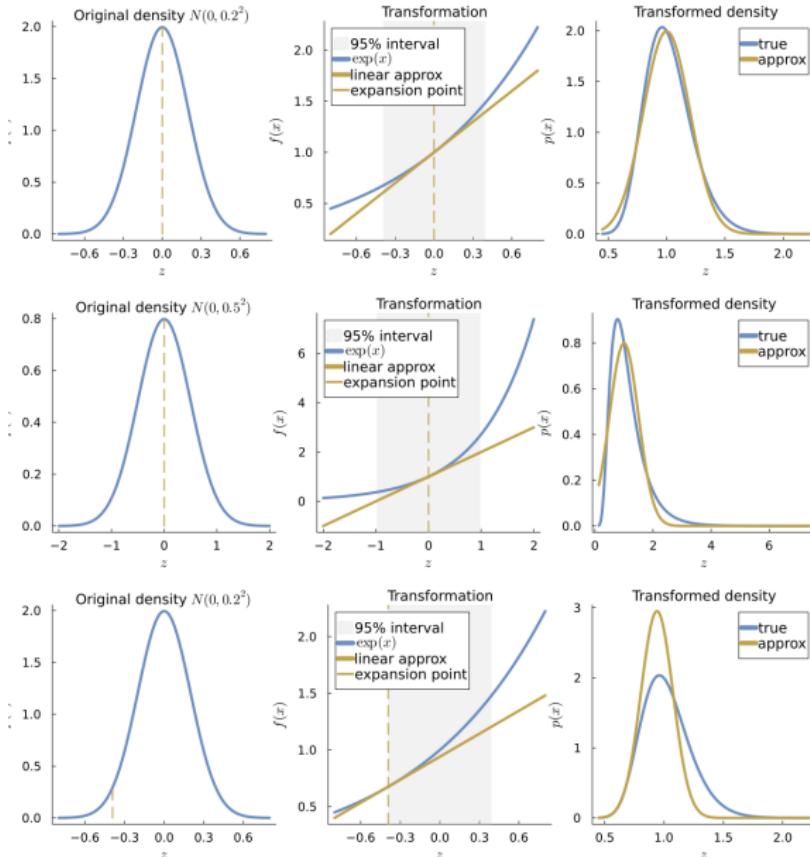
- Key for **measurement update**: approximate the joint distribution  $p(\mathbf{y}_t, \mathbf{z}_t | \mathbf{y}_{1:t})$

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} | \mathbf{y}_{1:t} \stackrel{\text{approx}}{\sim} N \left[ \begin{pmatrix} \mathbb{E}(\mathbf{y}_t) \\ \mathbb{E}(\mathbf{z}_t) \end{pmatrix}, \begin{pmatrix} \mathbb{V}(\mathbf{y}_t) & \mathbb{C}(\mathbf{y}_t, \mathbf{z}_t) \\ \mathbb{C}(\mathbf{y}_t, \mathbf{z}_t)^\top & \mathbb{V}(\mathbf{z}_t) \end{pmatrix} \right]$$

- **Extended Kalman filter (EKF)**

- ▶ linearizes transition model around posterior mean at  $t - 1$ .
- ▶ linearizes measurement model around prior mean at  $t$ .

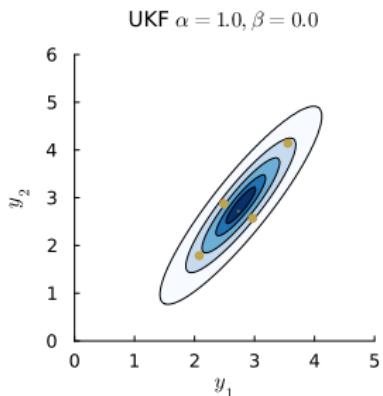
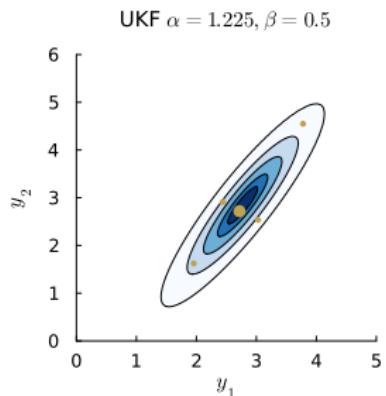
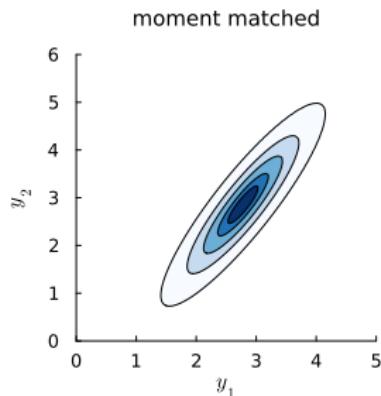
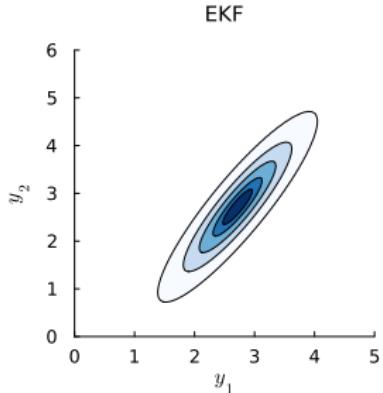
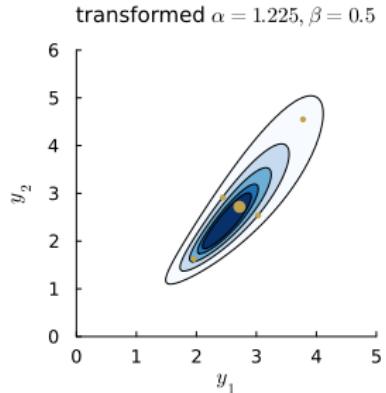
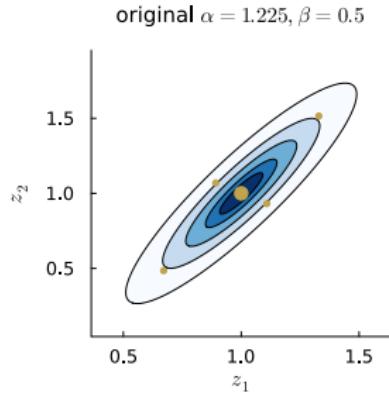
# Extended Kalman filter



## Alternative Gaussian Kalman-based samplers

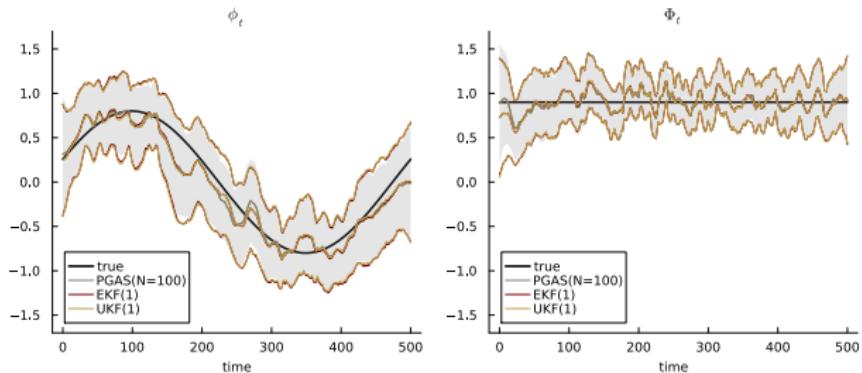
- **Moment matching** by numerical integration. [10]
- **Unscented Kalman Filter (UKF)**: map sigma points through nonlinear transition and measurement functions. [17]
- **Iterated EKF** and **Iterated UKF**: expands around posterior mean instead of prior mean. [18, 19]
- Iterated EKF/UKF with **line search**. [18, 19]
- **Prior Linearization filter**. Linearizes  $\mathbf{y} = \mathbf{g}(\mathbf{x})$  by minimizing  $\mathbb{E}_{\text{prior}} \left( \|\mathbf{g}(\mathbf{x}) - (\mathbf{a} + \mathbf{Cx})\|^2 \right)$ . [20]
- **Posterior linearization filter** minimizes  $\mathbb{E}_{\text{post}} \left( \|\mathbf{g}(\mathbf{x}) - (\mathbf{a} + \mathbf{Cx})\|^2 \right)$  by iterating. [20]
- **Laplace approximation**: approximates by posterior mode and inverse negative Hessian. [21]

# Gaussian approximations to bivariate log-normal

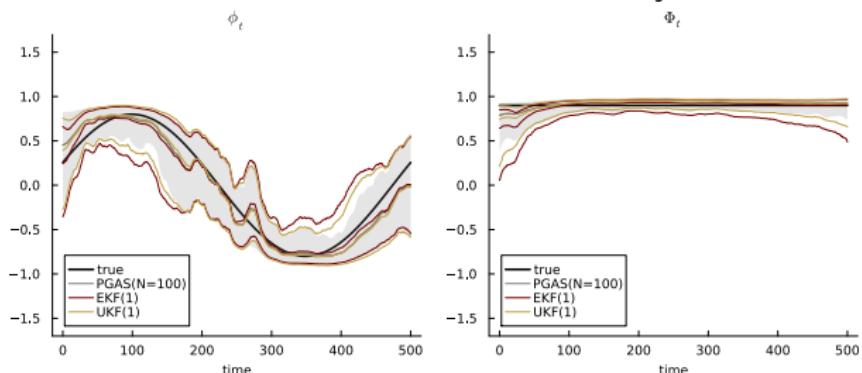


# UKF improves on EKF

Near non-stable DGP - fit without stability restrictions



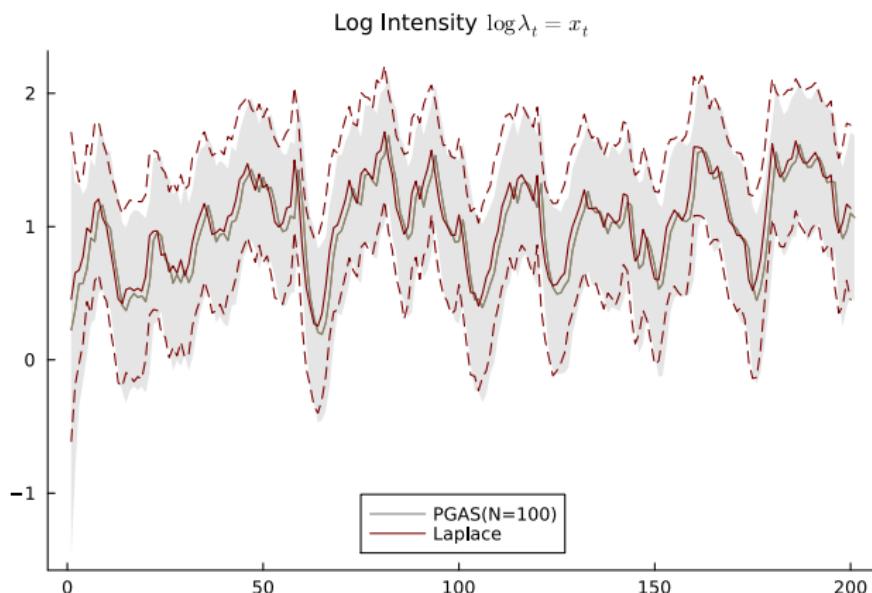
Near non-stable DGP - fit with stability restrictions



# Laplace approximation illustration

$$y_t | x_t \sim \text{Poisson}(\exp(x_t))$$

$$x_t = \mu + a(x_{t-1} - \mu) + \nu_t, \quad \nu_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\nu^2)$$



# Summary

- **Dynamic global-local shrinkage priors** have interesting parameter evolutions.
- The **near-constancy of parameters** makes particle methods problematic.
- **Gaussian approximations work well** in many settings.
- **Multi-seasonal ARMA** models with **time-varying** parameters following a **dynamic shrinkage process**
  - ▶ can be made stable at every time period
  - ▶ the extended Kalman filter works well in many settings
  - ▶ large changes in seasonality in US industrial production data
  - ▶ Julia package for arbitrary number of polynomials is coming

-  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.
-  C. K. Carter and R. Kohn, "On gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, 1994.

-  S. Frühwirth-Schnatter, "Data augmentation and dynamic linear models," *Journal of time series analysis*, vol. 15, no. 2, pp. 183–202, 1994.
-  N. Chopin, O. Papaspiliopoulos, et al., *An introduction to sequential Monte Carlo*, vol. 4. Springer, 2020.
-  L. S. Tan and D. J. Nott, "Gaussian variational approximation with sparse precision matrices," *Statistics and Computing*, vol. 28, no. 2, pp. 259–275, 2018.
-  H. Rue, S. Martino, and N. Chopin, "Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 71, no. 2, pp. 319–392, 2009.
-  S. Särkkä and L. Svensson, *Bayesian filtering and smoothing*, vol. 17. Cambridge university press, 2023.

- ❑ O. Barndorff-Nielsen and G. Schou, “On the parametrization of autoregressive models by partial autocorrelations,” *Journal of Multivariate Analysis*, vol. 3, no. 4, 1973.
- ❑ J. F. Monahan, “A note on enforcing stationarity in autoregressive-moving average models,” *Biometrika*, vol. 71, no. 2, pp. 403–404, 1984.
- ❑ 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.

- ❑ G. Fagerberg, M. Villani, and R. Kohn, “Time-varying multi-seasonal ar models,” *arXiv preprint arXiv:2409.18640*, 2024.
- ❑ S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- ❑ M. A. Skoglund, G. Hendeby, and D. Axehill, “Extended kalman filter modifications based on an optimization view point,” in *2015 18th International Conference on Information Fusion (Fusion)*, pp. 1856–1861, IEEE, 2015.
- ❑ M. A. Skoglund, F. Gustafsson, and G. Hendeby, “On iterative unscented kalman filter using optimization,” in *2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–8, IEEE, 2019.
- ❑ Á. F. García-Fernández, L. Svensson, M. R. Morelande, and S. Särkkä, “Posterior linearization filter: Principles and

implementation using sigma points," *IEEE transactions on signal processing*, vol. 63, no. 20, pp. 5561–5573, 2015.

-  S. Koyama, L. Castellanos Pérez-Bolde, C. R. Shalizi, and R. E. Kass, "Approximate methods for state-space models," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 170–180, 2010.

# Bayesian inference TVSARMA

■ **TVSARMA** with exact likelihood:

$$y_t = \mathbf{x}_t^\top \tilde{\phi}_t + \mathbf{z}_t^\top \tilde{\psi}_t + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{indep}}{\sim} N(0, \sigma_t^2)$$
$$(\tilde{\phi}_t, \tilde{\psi}_t) = \tilde{\mathbf{g}}(\boldsymbol{\theta}_t)$$
$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \stackrel{\text{indep}}{\sim} N(0, \text{Diag}(\exp(\mathbf{h}_t)))$$
$$\mathbf{h}_t = \boldsymbol{\mu} + \boldsymbol{\kappa}(\mathbf{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_{kt} \stackrel{\text{iid}}{\sim} Z(1/2, 1/2, 0, 1)$$

■ Add Gibbs update steps for

- ▶ past disturbances:  $\mathbf{z}_t = (\varepsilon_{t-1}, \varepsilon_{t-s}, \varepsilon_{t-(1+s)})^\top$
- ▶ pre-sample lags:  $y_0, y_{-1}, y_{-2}, \dots$

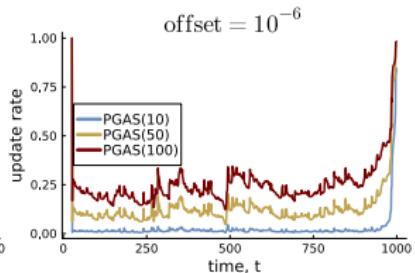
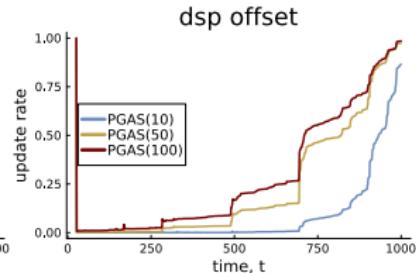
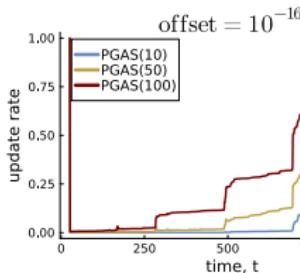
# PGAS - larger offset improves convergence

- Volatility models

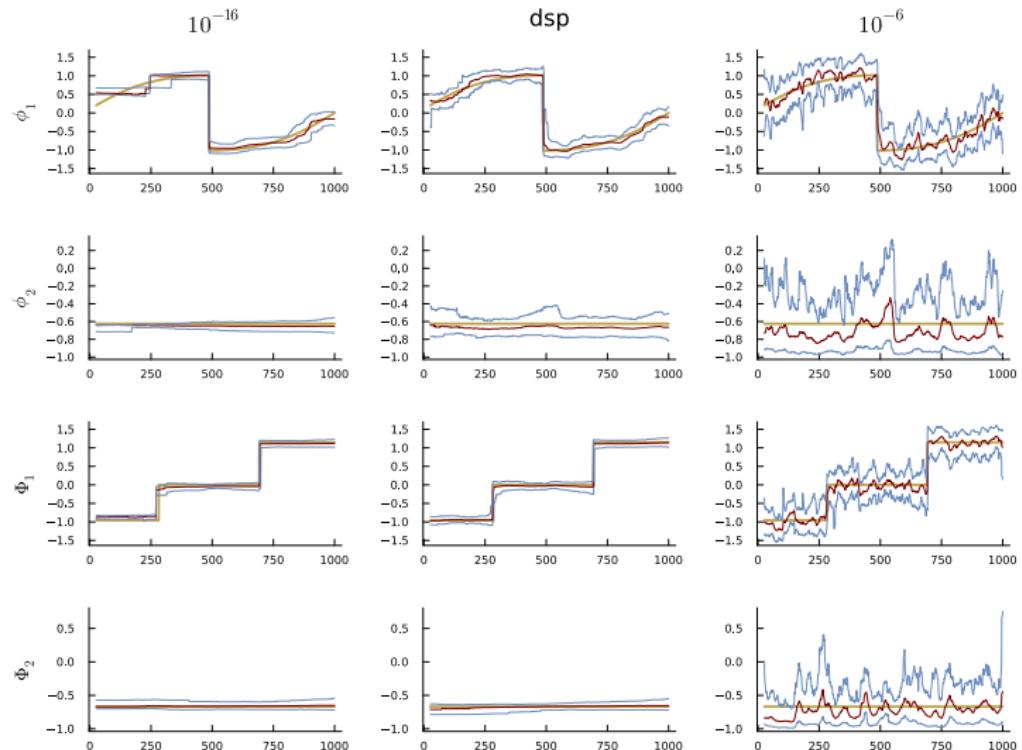
$$\nu_t = \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- The usual **square-and-log trick**

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



... but makes parameters more “wiggly”



# Spectral density Seasonal ARMA

## Spectral density ARMA(p,q)×(P,Q)

seasonal period

no. of AR

no. of MA

no. of seasonal AR

no. of seasonal MA

### Parameter settings

$\phi_1$

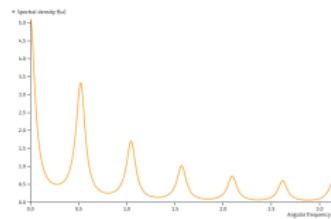
No regular MA

$\Phi_1$

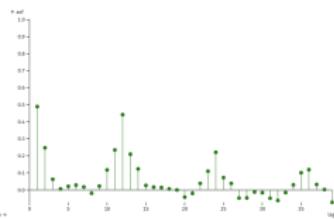
No seasonal MA

plot log f(u)

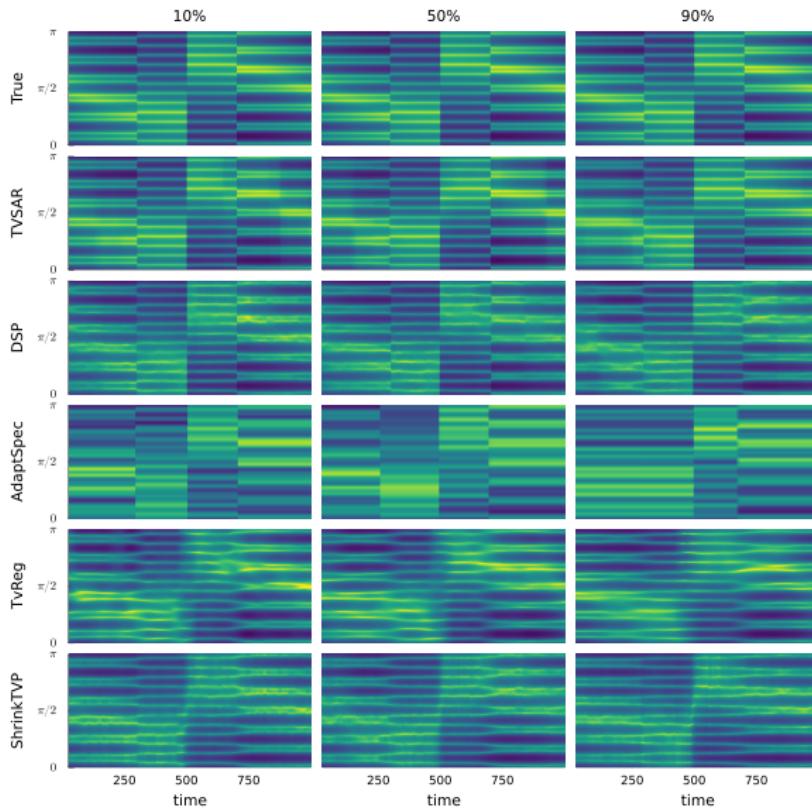
Spectral density of an ARMA(1,0)×(1,0) process with season s=12



Estimated ACF of ARMA(1,0)×(1,0) process with season s=12



# Fitted log spectrogram



# Quantifying the spectral density fit

