Global-local shrinkage processes for more realistic parameter evolution

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Global-local shrinkage priors in linear regression

- **Ridge regression** - one global shrinkage parameter λ

$$\beta_j | \sigma_\varepsilon^2 \overset{\text{iid}}{\sim} N\left(0, \frac{\sigma_\varepsilon^2}{\lambda}\right)$$

- **Lasso regression**

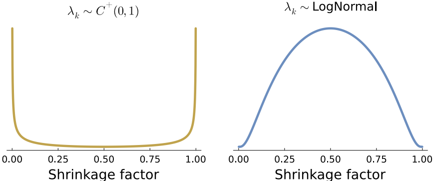
$$\beta_j | \sigma_\varepsilon^2 \overset{\text{iid}}{\sim} \text{Laplace}\left(0, \frac{\sigma_\varepsilon^2}{\lambda}\right)$$

- **Global-Local shrinkage**: global + local shrinkage for each β_j.
- **Horseshoe prior**:

$$\beta_j | \lambda_j^2, \tau^2, \sigma_\varepsilon^2 \sim N\left(0, \tau^2 \lambda_j^2 \sigma_\varepsilon^2\right)$$

$$\lambda_j \overset{\text{iid}}{\sim} C^+(0,1) \quad \text{and} \quad \tau \sim C^+(0,1)$$

- **Shrinkage factor** c_j wrt OLS: $\tilde{\beta}_j = (1 - c_j)\hat{\beta}_j$



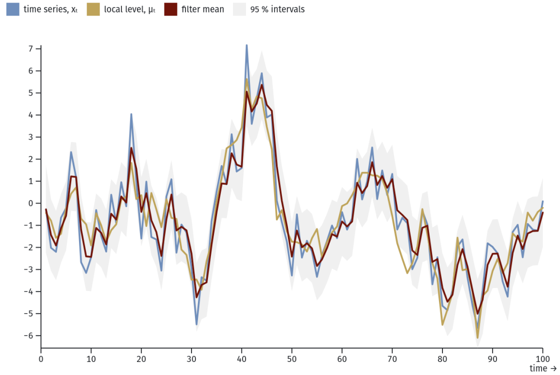
Dynamic horseshoe in state-space models

- **Local level model** (state-space) for time series

$$
\begin{aligned}
x_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\
\mu_t &= \mu_{t-1} + \nu_t, & \nu_t &\sim N(0, \sigma_\nu^2)
\end{aligned}
$$

- **Innovation variance** $\sigma_\nu^2 \Rightarrow$ how fast the mean evolves.
- Same normal $N(0, \sigma_\nu^2)$ for all ν_t. Compare Ridge regression.
- Restrictive parameter evolution. Can't get all of this:
  1. $\nu_t \approx 0$ for a strech of t (parameters stand still)
  2. large ν_t for some t (jumps)
  3. persistent periods of rapid changes

Local level model with Gaussian innovations



Dynamic horseshoe process prior

- **Horseshoe prior** for time series

$$\begin{aligned}\mu_t &= \mu_{t-1} + \nu_t, \qquad \nu_t \sim N(0, \tau^2\lambda_j^2) \\ \lambda_t &\overset{\text{iid}}{\sim} C^+(0,1) \\ \tau &\sim C^+(0,1)\end{aligned}$$

- This gives us Property 1 and 2 above.
- Local variances λ_t^2 are independent. No Property 3.
- **Dynamic horseshoe process** [1]

$$\begin{aligned}\mu_t &= \mu_{t-1} + \nu_t, \qquad \nu_t \sim N(0, \tau^2 \exp(h_t)) \\ h_t &= \phi h_{t-1} + \eta_t, \qquad \eta_t \sim Z(1/2, 1/2, 0, 1) \\ \tau &\sim C^+(0,1)\end{aligned}$$

- The horseshoe prior is the special case with $\phi = 0$

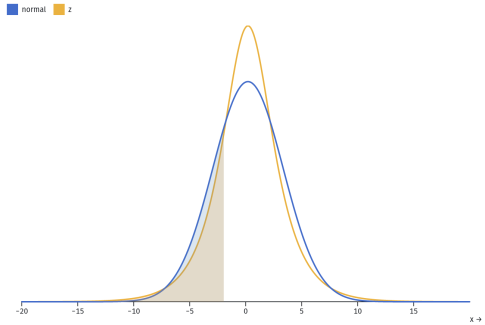
$$\eta_t \sim Z(1/2, 1/2, 0, 1) \Longleftrightarrow \lambda_t = \exp(\eta_t/2) \sim C^+(0,1)$$

Z-distribution

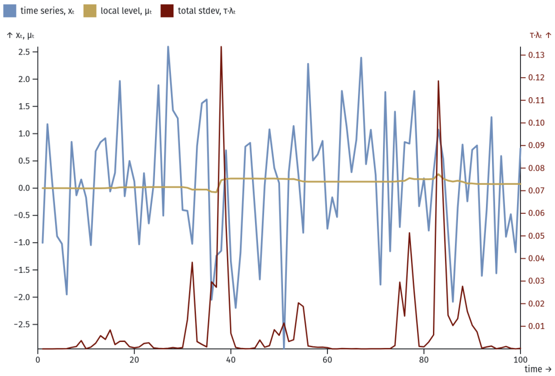
- Also called **Logistic-Beta distribution** since

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

- The $Z(\alpha, \beta, 0, 1)$ distribution is heavy tailed.
- Linearly decaying log density.



Local level with dynamic shrinkage process

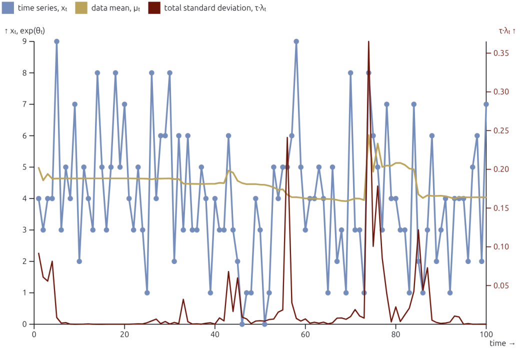


Poisson time series with dynamic shrinkage process

time series length	100
initial mean, μ_0	5
global std, τ	0.0025
persistence, ϕ	0.7
α	0.5
β	0.5
σ_η	1

plot: ☑ time series, x_t ☑ data mean, μ_t ☑ total stdev, $\tau \cdot \lambda_t$

Simulate new data



Bayesian inference for models with DSP priors

- Aim: posterior conditional on time series $\boldsymbol{y}_{1:T}$

$$p(\boldsymbol{z}_{0:T}, \boldsymbol{h}_{1:T}, \boldsymbol{\tau}, \phi | \boldsymbol{y}_{1:T})$$

- Gibbs sampling by data augmentation:
  - Mixture of normal for $\log \chi_1^2$ for volatility $\boldsymbol{h}_{1:T}$ [2]
  - Polya-Gamma augmentation for Z-distribution. [1]
- Draw from full conditional of $\boldsymbol{h}_{1:T}$ by tri-diagonal multivariate Gaussian.
- Conditional on $\boldsymbol{h}_{1:T}$ and Polya-Gamma draws, draw the state $\boldsymbol{z}_{0:T}$ by sampling from posterior in a state-space model with:
  - **nonlinear/non-Gaussian measurement** model
  - **linear** (heteroscedastic) **Gaussian transition** model.

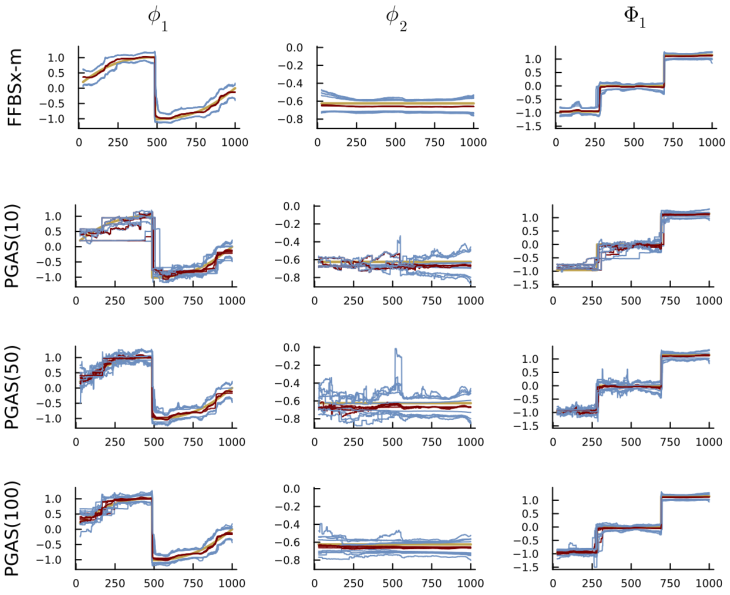
Stable seasonal AR with DSP prior

- **Time-varying Seasonal AR(p, P)** with season s [3]

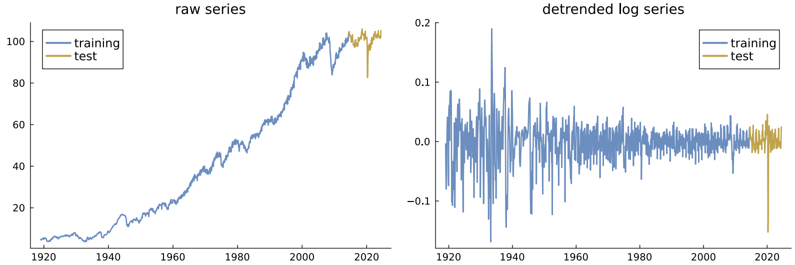
$$\phi_{p,t}(L)\Phi_{P,t}(L^s)y_t = \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_t^2)$$

- Parameters $\phi_{1,t}, \ldots, \phi_{p,t}$ and $\Phi_{1,t}, \ldots, \Phi_{P,t}$ follow DSP priors.
- Can be directly extended to multi-seasonal AR.
- Two non-linearities:
  - **Multiplicative structure** of regular and seasonal polynomials.
  - **Stability restrictions** at every t (non-linear recursive).

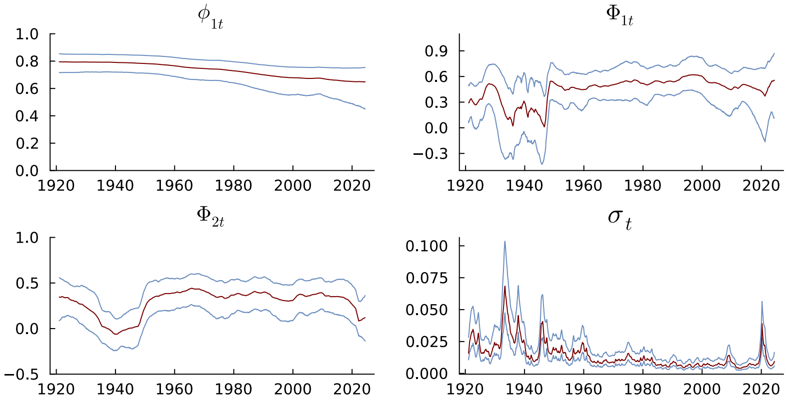
PGAS performs poorly due to near-degeneracy



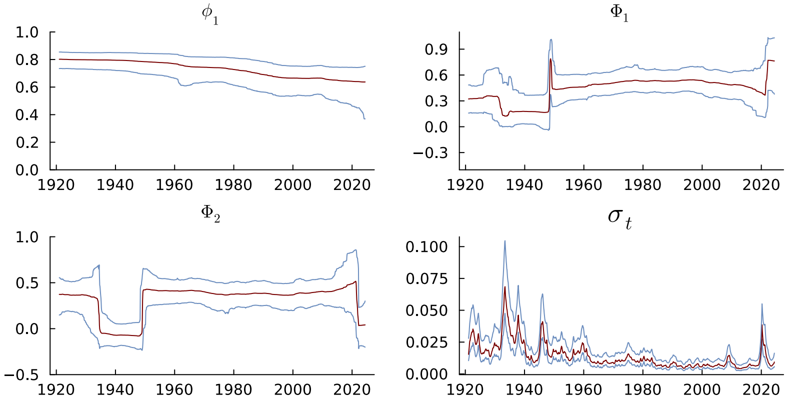
US industrial production 1919-2024



Posterior with homoscedastic Gaussian evolution



Posterior with DSP evolution



- 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.
- 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.
- G. Fagerberg, M. Villani, and R. Kohn, “Time-varying multi-seasonal ar models,” *Journal of Computational and Graphical Statistics*, 2026.