Sparse Bayesian Estimation of Regime-Switching Time Series

Mario Giacomazzo

6 March 2018

Background: Regime-Switching Time Series Models

- Purpose of Regime-Switching Time Series Models
  - Regime-Dependent Dynamics
  - Understand Volatility Changes
  - Alternative Approach to Handling Nonstationary and Cyclical Phenomenon
  - State-specific Forecasting
  - Characterization and Classification

Background: Regime-Switching Time Series Models

- Popularized for Economic and Financial Time Series
  - Structural Breaks Between Periods of Recession and Expansion
  - Unemployment Rates (Montgomery et al., 1998; Koop and Potter, 1999; Deschamps, 2008)
  - Gross Domestic Product (Teräsvirta, 1995)
  - Stock Prices (Li and Lam, 1995)
  - Agricultural Prices (Zeng et al., 2011)
  - Extensive Study on 215 Macroeconomic Time Series (Stock and Watson, 1998)
- Applications in Other Areas
  - Epidemiology: Epidemic vs Nonepidemic States (Hsin-Min Lu et al., 2010)
  - Psychology: Mood Changes Under Bipolar Disorder (Hamaker et al., 2010)
  - Traffic Management: Free Flow vs Congested States (Kamarianakis et al., 2010)
  - Ecology: Temperature Changes Due to Climate Changes (Battaglia and Protopapas, 2012)

Background: Regime-Switching Time Series Models

- Two Main Types
  - Stochastic Regime Changes
    - Markov Switching Autoregressive Model (MSAR) (Hamilton, 1990)
    - Transitions Determined by Latent States Following a Markov Process
  - Deterministic Regime Changes
    - Threshold Autoregressive Model (TAR) (Tong, 1990, 2012)
    - Smooth Transition Autoregressive Model (STAR) (Terasvirta and Anderson, 1992; Teräsvirta et al., 2010)
    - States are Determined by Evaluation of an Observable Variable at a Previously Known Time Point

Background: Regime-Switching Time Series Models

Given autoregressive order P, let $\boldsymbol{x}_t' = [y_{t-1}, y_{t-2}, \cdots, y_{t-P}]$, $\boldsymbol{\alpha}' = [\alpha_1, \cdots, \alpha_P]$, and $\boldsymbol{\beta}' = [\beta_1, \cdots, \beta_P]$. Consider General 2-Regime Autoregressive Model:

$$y_t = (\mu_\alpha + \boldsymbol{x}_t'\boldsymbol{\alpha} + \epsilon_{\alpha,t})(1 - G(z_t)) + (\mu_\beta + \boldsymbol{x}_t'\boldsymbol{\beta} + \epsilon_{\beta,t})G(z_t)$$

For heteroskedasticity, $\epsilon_{\alpha,t} \sim N(0, \sigma_\alpha^2)$ and $\epsilon_{\beta,t} \sim N(0, \sigma_\beta^2)$.

- Transition Function $G(z_t) : \mathbb{R} \to \mathbb{G} \subseteq [0, 1]$
  - TAR(P): $G(z_t, \delta) = \mathbb{1}_{\{z_t > \delta\}}(z_t)$
  - LSTAR(P): $G(z_t, \gamma^*, \delta) = \{1 + \exp[-(\gamma^*/s_Z)(z_t - \delta)]\}^{-1}$
  - ESTAR(P): $G(z_t, \gamma^*, \delta) = 1 - \exp[-(\gamma^*/s_Z)(z_t - \delta)^2]$

Threshold δ and transition slope $\gamma = \gamma^*/s_Z$ are additional parameters which control location and shape of the transition function.

Background: Regime-Switching Time Series Models

- Regime-switching Dependent on Transition Variable z_t
  - Low Regime: $z_t < \delta$
  - High Regime: $z_t > \delta$
  - Options for z_t Given Delay Parameter d
    - Exogenous Option: $z_t = r_{t-d}$
    - Self-Exciting: $z_t = y_{t-d}$
    - Time: $z_t = t$

- As $\gamma \to \infty$, LSTAR and ESTAR Transition Functions Become Abrupt

- Threshold $\delta \in [q_Z(.15), q_Z(0.85)]$ Where $q_Z(.)$ is the Empirical Quantile Function

Background: Regime-Switching Time Series Models

- Bayesian Estimation Methods
  - TAR(Geweke and Terui, 1993; Chen and Lee, 1995; Koop and Potter, 1999) and STAR(Lubrano, 2000; Lopes and Salazar, 2006; Livingston Jr. and Nur, 2017)
  - Prior Distributions
    - Normal Priors for μ_α, $\boldsymbol{\alpha}$,μ_β, and $\boldsymbol{\beta}$
    - Inverse Gamma Priors for σ_α^2 and σ_β^2
    - Gamma, Log-Normal, Truncated Normal, etc. for γ^*
    - $\delta \sim Uniform[q_Z(0.15), q_Z(0.85)]$
    - $p(d) = \frac{1}{d_{max}}$ for $d \in \{1, 2, \cdots, d_{max}\}$

Background: Regime-Switching Time Series Models

- Bayesian Estimation Methods (Cont.)
  - Exploits the Conditional Linear Nature Given Parameters δ and/or γ
  - MCMC Sampling from Joint Posterior $f(\boldsymbol{\theta}|Data)$ where

$$\boldsymbol{\theta}' = [\mu_\alpha, \boldsymbol{\alpha}, \sigma^2_\alpha, \mu_\beta, \boldsymbol{\beta}, \boldsymbol{\sigma^2_\beta}, \gamma, \delta, d]$$

    - Gibbs Sampler for Regime-specific Means, Coefficients, and Variances Using Full Conditional Distributions (Gelfand and Smith, 1990)
    - Metropolis-Hastings Approach for γ and δ (Metropolis et al., 1953; Hastings, 1970)
    - Posterior Discrete Uniform Probabilities for $d \in \{1, 2, \cdots, P\}$ Found Using Bayes' Rule

Background: Bayesian Model Selection

- Bayesian Inference for Model Order P
  - Define P as the Maximum Model Order Across Regimes
  - Apply Bayesian Estimation for Many Choices of P and Choose Based on Bayes' Factors
  - Reversible jump Markov Chain Monte Carlo (RJMCMC) Used when P in $\boldsymbol{\theta}$
    - AR models (Troughton and Godsill, 1997; Vermaak et al., 2004)
    - TAR or STAR models (Campbell, 2004; Lopes and Salazar, 2006)
    - Why? $\dim(\boldsymbol{\Theta})$ depends on P
  - Required Assumptions for These Approach
    - Model Order P is the Same in Both Regimes
    - Given P, All Lagged Terms Less than P are Significant
  - Problems Under Assumptions
    - Estimates Inflexible LSTAR Models Solely Determined by P
    - Even if P is Identified Correctly, Overfitting May Occur

Background: Transition

Consider the general matrix representation of the full linear model

$$\underset{T\times 1}{\boldsymbol{y}} = \underset{T\times P}{\boldsymbol{X}} \times \underset{P\times 1}{\boldsymbol{\theta}} + \underset{T\times 1}{\boldsymbol{\epsilon}}$$

Although we consider P covariates, it is possible that $\boldsymbol{\theta}$ contains many 0 entries. For high-dimensional scenarios (Large P), sparse estimation of $\boldsymbol{\theta}$ becomes important to combat overfitting and identify the underlying signal.

See Dellaportas et al. (2002); O'Hara and Sillanpaa (2009); Polson and Scott (2010) for an Overview of Bayesian Approaches to Finding the "Best" Model.

Background: Bayesian Variable Selection

- Binary Indicator Variables Used To Identify Covariate Inclusion. Each submodel $\boldsymbol{m} \in \mathcal{M} = \{0, 1\}^P$ is a $P \times 1$ vector of binary indicators.

$$\boldsymbol{y} = \boldsymbol{X}_m \boldsymbol{\theta}_m + \boldsymbol{\epsilon}$$

- Exploring $\mathcal{M}$ is Time Consuming and Difficult. 2^P Possible Models
- Popular Prior:

$$m_k \sim Bern(\pi_k)$$

  Choice of π_k reflects prior beliefs on the true model complexity. Typically $\pi_k = \pi = 0.5$

- Posterior MCMC Algorithms Incoporate $\boldsymbol{m}$. The Posterior Expectation $E[m_k|Data]$ Updates π_k

Background: Bayesian Variable Selection

- Posterior Probability of Model $\boldsymbol{m}$

$$P(\boldsymbol{m}|Data) = \frac{p(Data|\boldsymbol{m})p(\boldsymbol{m})}{\sum p(Data|\boldsymbol{m}_k)p(\boldsymbol{m}_k)}$$

- Bayes' Factor Comparing Two Candidates $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$

$$BF = \frac{p(Data|\boldsymbol{m}_1)}{p(Data|\boldsymbol{m}_2)} = \frac{p(\boldsymbol{m}_1|Data)p(\boldsymbol{m}_2)}{p(\boldsymbol{m}_2|Data)p(\boldsymbol{m}_1)}$$

- Inclusion Indicators Used in Spike and Slab Mixture Priors i.e. Stochastic Search Variable Selection (George and McCulloch, 1993)

$$p(\theta_k|m_k) = (1 - m_k)N(0, \tau^2) + m_k N(0, g\tau^2)$$

Typically, τ^2 should be small and g should be large.

Background: Bayesian Shrinkage

- Popular Regularization Methods

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \frac{1}{T}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) + \lambda \times \mathit{Penalty}(\boldsymbol{\theta})$$

  - Ridge Regression: $\mathit{Penalty}(\boldsymbol{\theta}) = ||\boldsymbol{\theta}||_2^2$ (Hoerl and Kennard, 1970)
  - Lasso Regression: $\mathit{Penalty}(\boldsymbol{\theta}) = ||\boldsymbol{\theta}||^1$ (Tibshirani, 1996)
  - Elastic Net: $\mathit{Penalty}(\boldsymbol{\theta}) = \alpha||\boldsymbol{\theta}||_2^2 + (1-\alpha)||\boldsymbol{\theta}||^1$ (Zou and Hastie, 2005)

Background: Bayesian Shrinkage

- Bayesian Regularization
  - Priors Represented as Scale Mixtures of Normals with Continuous Mixing Densities
  - Global-Local Hierarchical Representation

$$\theta_k | \lambda_k^2, \lambda^2, \sigma^2 \sim N(0, \sigma^2 \lambda^2 \lambda_k^2)$$

$$\lambda_k^2 \sim p_{Local}(.) \text{ and } \lambda^2 \sim p_{Global}(.)$$

  - Different Priors for Local and Global Shrinkage Parameters Lead to:
    - Different Concentration Around 0
    - Different Tail Behavior

- Optional Bayesian Shrinkage Methods
  - Lasso (Park and Casella, 2008)
  - Horseshoe (Carvalho et al., 2010; Makalic and Schmidt, 2016)
  - Double-Pareto (Armagan et al., 2013)
  - Bridge (Polson et al., 2014)
  - Horsehoe+ (Bhadra et al., 2016)
  - Dirichlet-Laplace (Bhattacharya et al., 2015)

Bayesian Shrinkage Estimation of LSTAR Models

- Primary Goals
  - Establish the Efficacy of Bayesian Shrinkage Estimation applied to LSTAR
  - Modify Bayesian Shrinkage Priors to Handle Regime-specific Sparsity
  - Allow for Composite Transition Variable to Be Estimated Using Dirichlet Prior

Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Shrinkage Methods
  - Bayesian Lasso (BLASSO)

$$\alpha_j|\sigma^2, \tau^2_{\alpha_j} \sim N(0, \sigma^2\tau^2_{\alpha_j}),\ \tau^2_{\alpha_j}| \sim EXP(\lambda^2/2)$$
$$\beta_j|\sigma^2, \tau^2_{\beta_j} \sim N(0, \sigma^2\tau^2_{\beta_j}),\ \tau^2_{\beta_j}| \sim EXP(\lambda^2/2)$$

    Hyperparameter λ Controls Global Shrinkage Across Both Regimes. Following Park and Casella (2008), gamma hyperprior for λ leads to inverse-Gaussian full conditional distribution.

  - Regime-Specific Bayesian Lasso (RS-BLASSO)

$$\tau^2_{\alpha_j}| \sim EXP(\lambda_1^2/2),\ \tau^2_{\beta_j}| \sim EXP(\lambda_2^2/2) \tag{1}$$

    Hyperparameters λ_1 and λ_2 control global shrinkage within low and high regimes, respectively.

Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Shrinkage Methods (Cont.)
  - Variable Selection with Bayesian Lasso (VS-BLASSO) Introducing latent binary variables ζ_j and η_j for $j \in \{1, 2, \cdots, p\}$, reparameterize $\alpha_j = \zeta_j \alpha_j^*$ and $\beta_j = \eta_j \beta_j^*$.

$$\zeta_j \sim BERN(0.5),\ \alpha_j^* | \sigma^2 \sim DEXP\left(0, \frac{\sigma^2}{\lambda}\right)$$

$$\eta_j \sim BERN(0.5),\ \beta_j^* | \sigma^2 \sim DEXP\left(0, \frac{\sigma^2}{\lambda}\right)$$

  Method proposed by Lykou and Ntzoufras (2011, 2013). Combines the subset selection approach of Kuo and Mallick (1998) with the Bayesian Lasso of Park and Casella (2008).

Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Shrinkage Methods (Cont.)
  - Bayesian Horseshoe (BHS)

$$
\begin{aligned}
\alpha_j|\lambda_{\alpha_j} &\sim N(0, \lambda_{\alpha_j}),\ \beta_j|\lambda_{\beta_j} \sim N(0, \lambda_{\beta_j}) \\
\lambda_{\alpha_j} &\sim C^+(0, \lambda),\ \lambda_{\beta_j} \sim C^+(0, \lambda) \\
\lambda|\sigma^2 &\sim C^+(0, \sigma)
\end{aligned}
$$

Although hyperparameter λ provides global shrinkage, the additional hyperparameters allow for finer shrinkage locally.

Bayesian Shrinkage Estimation of LSTAR Models

- Estimation of Delay Parameter (Self-Exciting Case)
  - Let $\boldsymbol{d}_t' = [y_{t-1}, y_{t-2}, \cdots, y_{t-d_{max}}]$ and $\phi_t' = [\phi_1, \phi_2, \cdots, \phi_{d_{max}}]$. Reparameterize transition variable $z_t = \phi' \boldsymbol{d}_t$.

$$\phi \sim Dir\left(\left[\frac{1}{d_{max}}, \frac{1}{d_{max}}, \cdots, \frac{1}{d_{max}}\right]'\right)$$

    Now, z_t is a weighted average of all considered transition variables. Because the weights are constrained to sum to 1, the prior for δ does not require modification.
  - Advantages
    - Allows for a composite transition variable
    - Estimates a more encompassing LSTAR model.

Bayesian Shrinkage Estimation of LSTAR Models

- Initial Simulation Studies
    - Evaluate Bayesian Shrinkage Methods on LSTAR
    - Examine Effect of Increased Noise
    - Examine Performance for Different Sparsity Patterns
    - Assume the Delay Parameter d is known *a priori*
    - All Replications are of Length $T = 1000$
    - Assume Maximum Model Order $P = 4$
    - Evaluation Based on $RMSE(\theta) = \sqrt{\sum(\hat{\theta} - \theta)^2 / N}$ Where N Represents the Number of Replications

Bayesian Shrinkage Estimation of LSTAR Models

- Computational Considerations
  - MCMC Sampling Using JAGS through R
  - 3 Chains with Random Starting Values
  - Burn-in=15,000, Thinning=10, Starting Sample=1,000
  - Addtional Samples of 1000 Are Obtained Until...
    - Potential Scale Reduction Factor (PSRF) Indicates Convergence (< 1.05) for All Monitored Parameters
    - Effective Sample Size (ESS) Is Large Enough (> 150) for Alll Monitored Parameters
    - Maximum of 20 Updates
  - Evaluate Methods and Calculate RMSE for Converged Replications

Bayesian Shrinkage Estimation of LSTAR Models

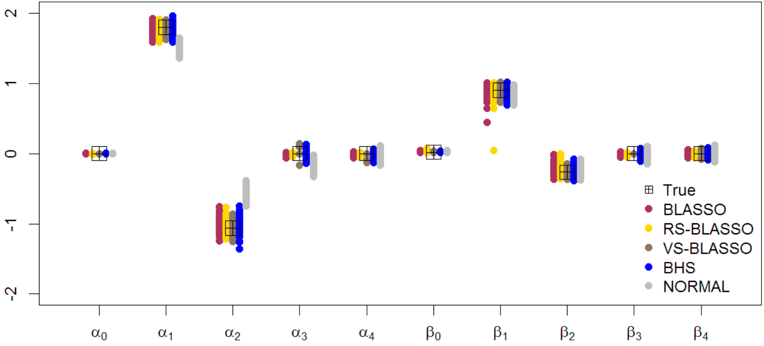
- Simulation Study 1: Well-Behaved LSTAR

$$
\begin{aligned}
y_t &= (1.8y_{t-1} - 1.06y_{t-2})[1 - G(y_{t-2})] \\
&\quad + (0.02 + 0.9y_{t-1} - 0.265y_{t-2})[G(y_{t-2})] + \epsilon_t \\
&\text{where: } G(y_{t-2}) = \left\{1 + \exp\left[-100(y_{t-2} - 0.02)\right]\right\}^{-1} \\
&\text{and } \epsilon_t \sim \text{i.i.d. } N(0, 0.02^2).
\end{aligned}
$$

Used in Lopes and Salazar (2006)

Bayesian Shrinkage Estimation of LSTAR Models

- Simulation Study 1: Well-Behaved LSTAR (Cont.)



Bayesian Shrinkage Estimation of LSTAR Models

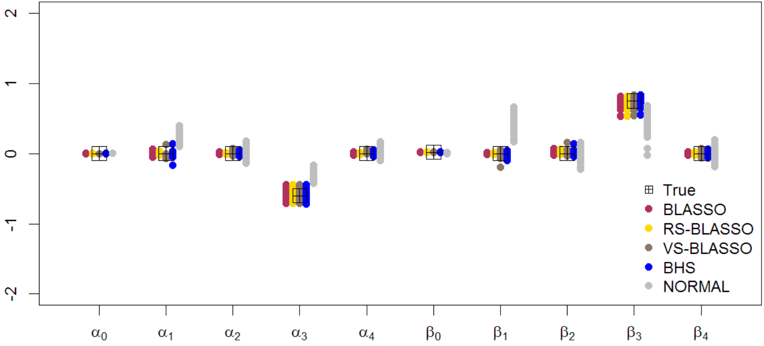
- Simulation Study 2: LSTAR With Gaps and Increased Noise

$$
\begin{aligned}
y_t &= (-0.6 y_{t-3})[1 - G(y_{t-1})] \\
&\quad + (0.02 + 0.75 y_{t-3})[G(y_{t-1})] + \epsilon_t \\
&\text{where: } G(y_{t-1}) = \Big\{ 1 + \exp\big[-120(y_{t-1} - 0.02) \big] \Big\}^{-1} \\
&\text{and } \epsilon_t \sim \text{i.i.d. } N(0, \sigma_k^2)
\end{aligned}
$$

We evaluate results for $\sigma_j = 0.02j \; \forall j \in \{1, 2, \cdots, 5\}$

Bayesian Shrinkage Estimation of LSTAR Models

- Simulation Study 2: LSTAR With Gaps and Increased Noise (Cont.)



Bayesian Shrinkage Estimation of LSTAR Models

- Simulation Study 2: LSTAR With Gaps and Increased Noise (Cont.)

  Naturally, increases in σ will cause increases in s_y. Recall the reparameterized transition slope $\gamma = \frac{\gamma^*}{s_y}$. For a fixed $\gamma = 120$, increases in σ indirectly increase γ^*. To ensure γ^* stays constant, we target $\gamma^* \approx 4$ and simulate data with $\gamma_j \approx \frac{4}{s_y}$ for each proposed $\sigma_j = 0.02j$.

Bayesian Shrinkage Estimation of LSTAR Models

- Simulation Study 2: LSTAR With Gaps and Increased Noise (Cont.) Table Gives RMSE for the Different Regime-Specific Coefficients.

Method	Parameter	Fixed Transition Slope		Modified Transition Slope	
	Choice of σ	0.02	0.1	0.02	0.1
	Choice of γ	120		109.60	30.02
BLASSO	α_0	0.0009	0.0032	0.001	0.0049
	α_1	0.0068	0.013	0.0091	0.0258
	α_2	0.0125	0.0102	0.0121	0.0136
	α_3	0.0479	0.0328	0.0501	0.0543
	α_4	0.0119	0.0106	0.012	0.0099
	β_0	0.0019	0.0059	0.0019	0.0069
	β_1	0.006	0.0177	0.0069	0.0218
	β_2	0.0091	0.0151	0.0127	0.02
	β_3	0.0494	0.0403	0.0579	0.0707
	β_4	0.008	0.0204	0.0093	0.0193
HS	α_0	0.0011	0.0049	0.0011	0.0065
	α_1	0.0054	0.0251	0.0063	0.0345
	α_2	0.0075	0.0181	0.0074	0.0223
	α_3	0.0485	0.0334	0.0508	0.0568
	α_4	0.0079	0.0174	0.008	0.02
	β_0	0.0018	0.0058	0.0018	0.0067
	β_1	0.004	0.0242	0.0039	0.0258
	β_2	0.0068	0.0237	0.0071	0.0261
	β_3	0.0524	0.042	0.0608	0.0739
	β_4	0.0062	0.0258	0.0063	0.0259

Bayesian Shrinkage Estimation of LSTAR Models

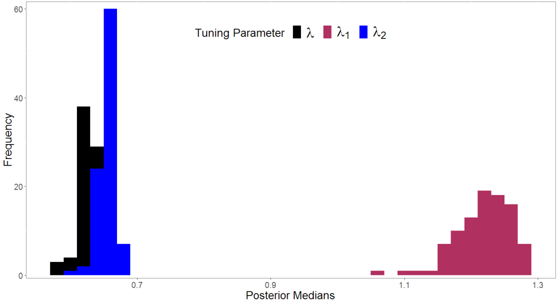
- Simulation Study 3: LSTAR With Regime-Specific Sparsity

$$
\begin{aligned}
y_t &= (-0.7y_{t-3})[1 - G(y_{t-1})] \\
&\quad + (0.06 + 0.4y_{t-1} - 0.35y_{t-2} + 0.2y_{t-3})[G(y_{t-1})] + \epsilon_t \\
&\text{where: } G(y_{t-1}) = \Big\{1 + \exp\big[-120(y_{t-1} - 0.03)\big]\Big\}^{-1} \\
&\text{and } \epsilon_t \sim \text{i.i.d. } N(0, 0.02^2)
\end{aligned}
$$

Bayesian Shrinkage Estimation of LSTAR Models

- Simulation Study 3: LSTAR With Regime-Specific Sparsity (Cont.)

  Comparison of Posterior Distributions for Shrinkage Parameters from BLASSO (λ) and RS-BLASSO (λ_1 and λ_2)



For BLASSO, 75% of replications converged compared to 94% for RS-BLASSO.

Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Selection of the Threshold Variable

Let $\boldsymbol{d}_t = [y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]'$ and $\phi' = [\phi_1, \phi_2, \phi_3, \phi_4]$. Consider Reparameterized Model From Simulation 2 With Modified Threshold Variable $z_t = \phi' \boldsymbol{d}_t$,

$$y_t = (-0.6 y_{t-3})[1 - G(y_{t-1})] + (0.02 + 0.75 y_{t-3})[G(y_{t-1})] + \epsilon_t$$

$$\text{where: } G(y_{t-1}) = \left\{ 1 + \exp\left[-120(\phi' \boldsymbol{d}_t - 0.02) \right] \right\}^{-1}$$

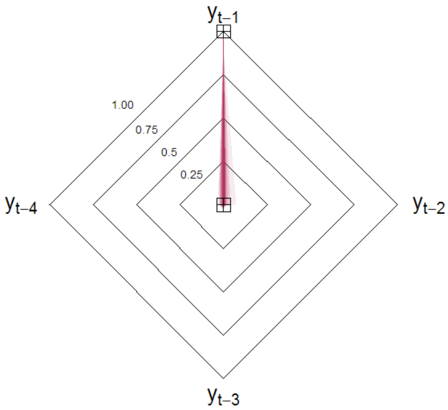
$$\text{and } \epsilon_t \sim \text{i.i.d. } N(0, 0.02^2)$$

Under prior $\phi \sim Dir([0.25, 0.25, 0.25, 0.25]')$, we conduct posterior sampling for three different threshold variables $\{z_{1,t}, z_{2,t}, z_{3,t}\}$ defined through ϕ. BHS priors are used for autoregressive coefficients.

Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Selection of the Threshold Variable (Scenario 1)

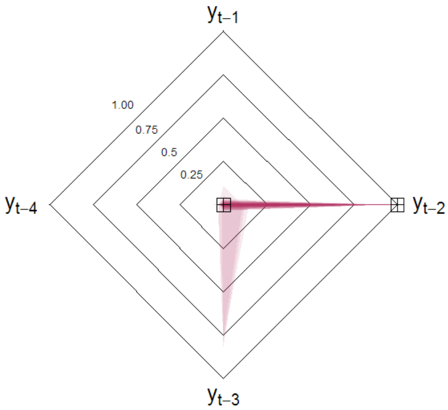
  Consider Original Choice $z_{1,t} = y_{t-1} = [1, 0, 0, 0]\boldsymbol{d}_t$. Posterior Means of ϕ from 100 Replications are Plotted Below.



Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Selection of the Threshold Variable (Scenario 2)

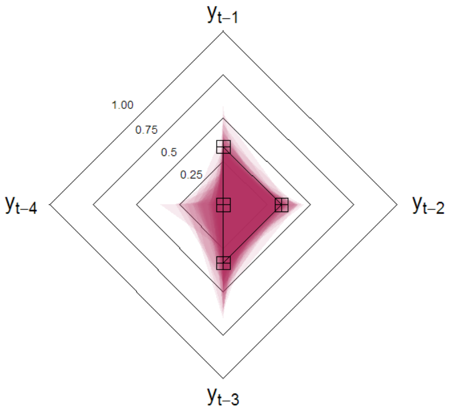
  Consider $z_{2,t} = y_{t-2} = [0,1,0,0]\boldsymbol{d}_t$. Posterior Means of ϕ from 100 Replications are Plotted Below.



Bayesian Shrinkage Estimation of LSTAR Models

- Bayesian Selection of the Threshold Variable (Scenario 3)

Consider $z_{3,t} = \frac{y_{t-1}+y_{t-2}+y_{t-3}}{3} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0]\boldsymbol{d}_t$. Posterior Means of ϕ from 100 Replications are Plotted Below.



Bayesian Shrinkage Estimation of LSTAR Models

- Application to Annual Sunspot Numbers
  - Textbook Example for Nonlinear Models Since Granger (1957)
  - Gathered and Updated by the World Data Center SILSO, Royal Observatory of Belgium, Brussels
  - Square Root Transformation: $y_t = 2[\sqrt{1 + x_t} - 1]$ (Ghaddar and Tong, 1981)
  - In Teräsvirta et al. (2010), LSTAR Outperformed Other AR, TAR,STAR, and Aritificial Neural Net (AR-NN) Models. Sparsity Achieved Via Stepwise Frequentist Procedure Using AIC
  - Training Period (1700-1979) and Testing Period (1980-2006)

Bayesian Shrinkage Estimation of LSTAR Models

- Application to Annual Sunspot Numbers (Cont.)
  - Terasvirta's Best LSTAR Model (F_T)
    LSTAR(10) Model With $d = 2$
  - Frequentist Estimation of Full Saturated LSTAR(10) (F_S)
  - BHS Estimated Linear Model AR(10) (B_L)
  - BHS Estimated LSTAR(10) with $d = 2$ (B_2)
  - BHS Estimated LSTAR(10) Applying Dirichlet Prior (B_D)
  - BHS Estimated LSTAR(10) with $d = 3$ (B_3)

Bayesian Shrinkage Estimation of LSTAR Models

- Application to Annual Sunspot Numbers (Cont.)

  Compare Models on RMSFE(h) for Horizons $h \in \{1, 2, 3, 4, 5\}$

  Bootstrap Method Used for Multi-step Ahead Forecasts for 1980-2006

Model	Horizon 1	2	3	4	5
F_T	1.42	2	2.36	2.51	2.35
F_S	1.86	3.21	3.7	3.63	3.16
B_L	1.73	2.3	2.54	2.53	2.56
B_2	1.42	1.96	2.29	2.19	2.19
B_D	1.77	2.83	3.38	3.5	3.29
B_3	1.86	3.11	3.58	3.62	3.58

Bayesian Shrinkage Estimation of LSTAR Models

- Application to Daily Maximum Water Temperatures
  - Data Used From 31 Rivers in Spain
  - Models Estimated to Forecast Daily Maximum Water Temperature Using Previously Known Daily Maximum Water Temperatures and Daily Maximum Air Temperatures
  - Combination of BHS and Dirichlet Priors for Estimation of Linear and Nonlinear Models Under Assumption $P = 6$
  - Horizon Specific Models Targeting 3-step and 7-step Ahead Forecasts
  - Nonlinear Models Improved Forecasting Accuracy for a Couple of Rivers (Details Provided in Paper)

Bayesian Model Selection of TAR Models

- Consider General TAR(P) Model

$$y_t = \begin{cases} \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_P y_{t-P} + \epsilon_{\alpha,t} & \text{if } z_t < \delta \\ \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_P y_{t-P} + \epsilon_{\beta,t} & \text{if } z_t > \delta \end{cases}$$
$$= \begin{cases} \boldsymbol{\alpha}' \boldsymbol{y_{t-1}} + \epsilon_{\alpha,t} & \text{if } z_t < \delta \\ \boldsymbol{\beta}' \boldsymbol{y_{t-1}} + \epsilon_{\beta,t} & \text{if } z_t > \delta \end{cases}$$

where $\epsilon_{\alpha,t} \sim N(0, \sigma_\alpha^2)$ and $\epsilon_{\beta,t} \sim N(0, \sigma_\beta^2)$

For future reference, we abbreviate the full TAR(P) model $\mathcal{M}_R$.

Model $\mathcal{M}_R$ is fully defined by parameter vector $\boldsymbol{\theta}'_{\mathcal{M}_R} = [\boldsymbol{\alpha}', \boldsymbol{\beta}', \sigma_\alpha, \sigma_\beta, \delta]$.

Furthermore, let $\mathcal{M}_{LR}$ and $\mathcal{M}_{HR}$ represent the low-regime and high-regime models, respectively.

From the $2^{2(P+1)}$ submodels, we aim to identify the best submodel $\mathcal{M}_R^*$

Bayesian Model Selection of TAR Models

- Methodology Step 1: Bayesian Shrinkage

  Apply Regime-Specific BHS for Initial Sparse Estimation

$$
\begin{aligned}
&\alpha_j \sim N(0, \sigma_\alpha^2 \lambda_{\alpha_j}^2 \lambda_\alpha^2),\ \ \lambda_{\alpha_j} \sim C^+(0,1),\ \ \lambda_\alpha \sim C^+(0,1) \\
&\beta_j \sim N(0, \sigma_\beta^2 \lambda_{\beta_j}^2 \lambda_\beta^2),\ \ \lambda_{\beta_j} \sim C^+(0,1),\ \ \lambda_\beta \sim C^+(0,1) \\
&\quad \forall j \in \{1, 2, 3, \cdots, P\}
\end{aligned}
$$

Bayesian Model Selection of TAR Models

- Methodology Step 2: Projection Predictive Variable Selection
  - Kullback-Leibler (KL) Divergence
    - Asymmetric Measure of Distance Between 2 Probability Distributions
    - Developed by Goutis and Robert (1998); Dupuis and Robert (2003); Piironen and Vehtari (2017) for Generalized Linear Models
    - Measure Discrepancy Between Full Model $\mathcal{M}_R$ and Proposed Submodel $\mathcal{M}_R^{\perp}$
  - Necessary Definitions

$$\text{Matrix } \boldsymbol{Y_{t-1}} \text{ where } k\text{th row is } \boldsymbol{y'_{k-1}} = [1, y_{k-1}, y_{k-2}, \cdots, y_{k-P}]$$

$$\text{Regime-Specific Full Reference Models: } \mathcal{M}_{LR} \text{ and } \mathcal{M}_{HR}$$

$$\text{Regime-Specific Submodels: } \mathcal{M}_{LR}^{\perp} \text{ and } \mathcal{M}_{HR}^{\perp}$$

Bayesian Model Selection of TAR Models

- Methodology Step 2: Projection Predictive Variable Selection (Cont.)
  - Regime-Specific Projection Approach

$$f(\boldsymbol{\theta}_{\mathcal{M}_R}|\mathcal{M}_R, Data) \to f(\boldsymbol{\theta}_{\mathcal{M}_R^{\perp}}|\mathcal{M}_R^{\perp}, Data)$$

$$\boldsymbol{\theta}^{(s)} \to \boldsymbol{\theta}^{\perp(s)}$$

$$[\boldsymbol{\alpha}^{(s)}, \sigma_{\alpha}^{(s)}, \boldsymbol{\beta}, \sigma_{\beta}^{(s)}, \delta^{(s)}] \to [\boldsymbol{\alpha}^{\perp(s)}, \sigma_{\alpha}^{\perp(s)}, \boldsymbol{\beta}^{\perp(s)}, \sigma_{\beta}^{\perp(s)}, \delta^{(s)}]$$

  - Regime-Specific Projection Process
    Split $\boldsymbol{Y_{t-1}}$ into $\boldsymbol{Y_{LR,t-1}}$ and $\boldsymbol{Y_{HR,t-1}}$ based on $\delta^{(s)}$.
    Identify $\boldsymbol{Y^{\perp}_{LR,t-1}}$ and $\boldsymbol{Y^{\perp}_{HR,t-1}}$ based on proposed $\mathcal{M}^{\perp}_{LR}$ and $\mathcal{M}^{\perp}_{HR}$

$$\boldsymbol{\alpha}^{\perp(s)} = (\boldsymbol{Y^{\perp\prime}_{LR,t-1}} \boldsymbol{Y^{\perp}_{LR,t-1}})^{-1} \boldsymbol{Y^{\perp\prime}_{LR,t-1}} \boldsymbol{Y_{LR,t-1}} \boldsymbol{\alpha}^{(s)}$$

$$\sigma_{\alpha}^{\perp(s)} = \sqrt{\sigma_{\alpha}^{(s)} + \frac{(\boldsymbol{Y_{LR,t-1}}\boldsymbol{\alpha}^{(s)} - \boldsymbol{Y^{\perp}_{LR,t-1}}\boldsymbol{\alpha}^{\perp(s)})'(\boldsymbol{Y_{LR,t-1}}\boldsymbol{\alpha}^{(s)} - \boldsymbol{Y^{\perp}_{LR,t-1}}\boldsymbol{\alpha}^{\perp(s)})}{T}}$$

$$\boldsymbol{\beta}^{\perp(s)} = (\boldsymbol{Y^{\perp\prime}_{HR,t-1}} \boldsymbol{Y^{\perp}_{HR,t-1}})^{-1} \boldsymbol{Y^{\perp\prime}_{HR,t-1}} \boldsymbol{Y_{HR,t-1}} \boldsymbol{\beta}^{(s)}$$

$$\sigma_{\beta}^{\perp(s)} = \sqrt{\sigma_{\beta}^{(s)} + \frac{(\boldsymbol{Y_{HR,t-1}}\boldsymbol{\beta}^{(s)} - \boldsymbol{Y^{\perp}_{HR,t-1}}\boldsymbol{\beta}^{\perp(s)})'(\boldsymbol{Y_{HR,t-1}}\boldsymbol{\beta}^{(s)} - \boldsymbol{Y^{\perp}_{HR,t-1}}\boldsymbol{\beta}^{\perp(s)})}{T}}$$

Bayesian Model Selection of TAR Models

- Methodology Step 2: Projection Predictive Variable Selection (Cont.)
  - Regime-Specific Projection Process (Cont.)
    For each $\boldsymbol{\theta}^{(s)}$, we obtain regime-specific KL divergences,

$$d_{LR}^{(s)}(\boldsymbol{\alpha}^{(s)}, \sigma_{\alpha}^{(s)}) = \frac{1}{2}\log\left(\frac{\sigma_{\alpha}^{\perp(s)}}{\sigma_{\alpha}^{(s)}}\right)$$

$$d_{HR}^{(s)}(\boldsymbol{\beta}^{(s)}, \sigma_{\beta}^{(s)}) = \frac{1}{2}\log\left(\frac{\sigma_{\beta}^{\perp(s)}}{\sigma_{\beta}^{(s)}}\right)$$

    Finally, we measure regime-specific discrepancies by

$$D(\mathcal{M}_{LR}||\mathcal{M}_{LR}^{\perp}) = \frac{1}{S}\sum_{s=1}^{S} d_{LR}^{(s)}(\boldsymbol{\alpha}^{(s)}, \sigma_{\alpha}^{(s)})$$

$$D(\mathcal{M}_{HR}||\mathcal{M}_{HR}^{\perp}) = \frac{1}{S}\sum_{s=1}^{S} d_{HR}^{(s)}(\boldsymbol{\beta}^{(s)}, \sigma_{\beta}^{(s)})$$

Bayesian Model Selection of TAR Models

- Methodology Step 2: Projection Predictive Variable Selection (Cont.)
  - Forward Stepwise Selection Algorithm

    For Intercept-only Models $\mathcal{M}^0_{LR}$ and $\mathcal{M}^0_{HR}$, calculate initial discrepancies $D(\mathcal{M}_{LR}||\mathcal{M}^0_{LR})$ and $D(\mathcal{M}_{HR}||\mathcal{M}^0_{HR})$

    For each level of flexibility $p \in \{1, 2, \cdots, P\}$, we identify the best regime-specific submodels $\mathcal{M}^p_{LR}$ and $\mathcal{M}^p_{HR}$.

    Algorithm conducted such that for $j < k$ $\mathcal{M}^j_{HR}$ is nested in $\mathcal{M}^k_{HR}$.

Bayesian Model Selection of TAR Models

- Methodology Step 3: Final Model Selection
  - Choose $\mathcal{M}^*_{LR}$ and $\mathcal{M}^*_{HR}$ Based on Relative Explanatory Power

$$RelE(\mathcal{M}^p) = 1 - \frac{D(\mathcal{M}||\mathcal{M}^p)}{D(\mathcal{M}||\mathcal{M}^0)}$$

  - Choose Based On Minimization of RMSFE for Time Period Intentionally Ignored

$$RMSFE(\mathcal{M}^p) = \sqrt{\frac{1}{T}\sum(y_t - \hat{y}_t)^2}$$

Current Work

Autoregressive Moving Average Model: ARMA(p,q)

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) y_t = \left(1 + \sum_{j=1}^{q} \theta_j B^j\right) \epsilon_t$$

We utilize the backshift operator B where $B^k y_t = y_{t-k}$, and assume $\epsilon_t \sim N(0, \sigma^2)$

Questions

References I

Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double pareto shrinkage. *Statistica Sinica*, 23(1):119.

Battaglia, F. and Protopapas, M. K. (2012). An analysis of global warming in the alpine region based on nonlinear nonstationary time series models. *Statistical Methods & Applications*, 21(3):315–334.

Bhadra, A., Datta, J., Polson, N. G., Willard, B., et al. (2016). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*.

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). Dirichletlaplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490. PMID: 27019543.

Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of time series analysis.*, 25(4):467–482.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chen, C. W. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, 16(5):483–492.

Dellaportas, P., Forster, J., and Ntzoufras, I. (2002). On bayesian model and variable selection using mcmc. *Statistics and Computing*, 12(1):27–36.

Deschamps, P. J. (2008). Comparing smooth transition and Markov switching autoregressive models of US unemployment. *Journal of Applied Econometrics*, 23(4):435–462.

Dupuis, J. A. and Robert, C. P. (2003). Variable selection in qualitative models via an entropic explanatory power. *Journal of Statistical Planning and Inference*, 111(1):77–94.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

References II

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14(5):441–454.

Ghaddar, D. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *Applied Statistics*, 30(3).

Goutis, C. and Robert, C. P. (1998). Model choice in generalised linear models: A bayesian approach via kullback-leibler projections. *Biometrika*, 85(1):29–37.

Granger, C. W. J. (1957). A statistical model for sunspot activity. *Astrophysical Journal*, 126:152.

Hamaker, E. L., Grasman, R. P. P. P., and Kamphuis, J. H. (2010). Regime-switching models to study psychological processes. American Psychological Association,.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics*, 45(1-2):39–70.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hsin-Min Lu, D., Zeng, D., and Hsinchun Chen, D. (2010). Prospective infectious disease outbreak detection using markov switching models. *Knowledge and Data Engineering, IEEE Transactions on*, 22(4):565–577.

Kamarianakis, Y., Gao, H. O., and Prastacos, P. (2010). Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions. *Transportation Research Part C: Emerging Technologies*, 18(5):821–840.

References III

Koop, G. and Potter, S. M. (1999). Dynamic asymmetries in us unemployment. *Journal of Business & Economic Statistics*, 17(3):298–312.

Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *The Indian Journal of Statistics, Series B (1960-2002)*, 60(1):65–81.

Li, W. and Lam, K. (1995). Modelling asymmetry in stock returns by a threshold autoregressive conditional heteroscedastic model. *The Statistician*, pages 333–341.

Livingston Jr., G. and Nur, D. (2017). Bayesian inference for smooth transition autoregressive (star) model: A prior sensitivity analysis. *Communications in Statistics - Simulation and Computation*, 46(7):5440–5461.

Lopes, H. F. and Salazar, E. (2006). Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis*, 27(1):99–117.

Lubrano, M. (2000). Bayesian analysis of nonlinear time series models with a threshold.

Lykou, A. and Ntzoufras, I. (2011). WinBUGS: a tutorial. *WIREs Comp Stat*, 3(5):385–396.

Lykou, A. and Ntzoufras, I. (2013). On bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, 23(3):361–390.

Makalic, E. and Schmidt, D. F. (2016). High-dimensional bayesian regularised regression with the bayesreg package. *arXiv preprint arXiv:1611.06649*.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998). Forecasting the u.s. unemployment rate. *Journal of the American Statistical Association*, 93(442):478–493.

O'Hara, R. B. and Sillanpaa, M. J. (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian Anal.*, 4(1):85–117.

References IV

Park, T. and Casella, G. (2008). The bayesian lasso.

Piironen, J. and Vehtari, A. (2017). Comparison of bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian Statistics*, 9:501–538.

Polson, N. G., Scott, J. G., and Windle, J. (2014). The bayesian bridge. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):713–733.

Stock, J. H. and Watson, M. W. (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series.

Teräsvirta, T. (1995). Modelling nonlinearity in us gross national product 1889–1987. *Empirical Economics*, 20(4):577–597.

Terasvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7(S1).

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling nonlinear economic time series*. Oxford University Press Oxford.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tong, H. (1990). Non-linear time series. *A Dynamical System Approach*.

Tong, H. (2012). *Threshold models in non-linear time series analysis*, volume 21. Springer Science & Business Media.

Troughton, P. T. and Godsill, S. J. (1997). A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves.

References V

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. J. (2004). Reversible jump markov chain monte carlo strategies for bayesian model selection in autoregressive processes. *Journal of Time Series Analysis*, 25(6):785–809.

Zeng, J.-H., Lee, C.-C., and Chang, C.-P. (2011). Are fruit and vegetable prices non-linear stationary? evidence from smooth transition autoregressive models. *Economics Bulletin*, 31(1):189–207.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.