

Three Essays on Shrinkage Estimation and Model Selection of Linear and Nonlinear Time Series Models

Mario Giacomazzo

9 May 2018



Overview and Motivation

- Parametric Time Series Models

- Logistic Smooth Transition Autoregressive Model (LSTAR) (Terasvirta, 1994)
- Threshold Autoregressive Model (TAR) (Tong, 1990)
- Autoregressive Moving Average Model (ARMA) (Box and Pierce, 1970)

- Model Differences

- ARMA = Popularized for Weakly Stationary Time Series
- LSTAR and TAR = Handling Nonlinear Behavior
 - Changes in Level, Dynamics, and Volatility
 - Alternative Approach to Handling Nonstationary and Cyclical Phenomenon
 - Useful for Clustering and Classification of Realizations



Overview and Motivation

- Purpose
 - Primary Goal is Forecasting
 - Economic Variables: Past Realizations Assist in Prediction (Mann and Wald, 1943)
 - Estimation and Fitting of ARMA (Durbin, 1960)
 - Practical Procedures Developed for Industrial Engineering (Box and Pierce, 1970)
 - Asymmetries Noticed in US Unemployment (Rothman, 1998; Montgomery et al., 1998; Koop and Potter, 1999)



Overview and Motivation

- Model Complexity
 - Determined by Order Parameters
 - Autoregressive (AR)
 - Moving Average (MA)
 - Distributed Lag (DL)
 - Regime-specific Model Orders
 - Frequentist Approach: Select Based on Information Criteria or Forecasting Performance
 - Bayesian Approach: Reversible jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995)
 - AR(P) (Troughton and Godsill, 1997)
 - TAR(P) (Campbell, 2004; Nieto et al., 2013)
 - STAR(P) (Lopes and Salazar, 2006)



Overview and Motivation

- Model Complexity (Cont.)

- Problem

- Difficult When Multiple Order Parameters Must Be Chosen
 - Often Leads to Inflexible Representations
 - Overfitting Can Still Occur

- Solution

- Intentionally Fix Orders to Be Large
 - Restructure Time Series Model as a High Dimensional Linear Regression
 - Apply Penalized Estimation Methods Aimed at Sparse Models



Introduction: Overview and Motivation

Dissertation Theme

Bayesian Automatic Estimation and Variable
Selection Procedures for Flexible Subset Models



Chapter 1: LSTAR Model

- Gaussian LSTAR(P) Model With 2-Regimes

- Given autoregressive order P , let $\mathbf{x}'_t = [y_{t-1}, y_{t-2}, \dots, y_{t-P}]$, $\boldsymbol{\alpha}' = [\alpha_1, \dots, \alpha_P]$, and $\boldsymbol{\beta}' = [\beta_1, \dots, \beta_P]$.

$$y_t = (\mu_\alpha + \mathbf{x}'_t \boldsymbol{\alpha})(1 - G(z_t)) + (\mu_\beta + \mathbf{x}'_t \boldsymbol{\beta})G(z_t) + \epsilon_t$$

where $\epsilon_t \sim \text{ i.i.d. } N(0, \sigma^2)$ and $G(z_t) : \mathbb{R} \rightarrow \mathbb{G} \subseteq [0, 1]$.

- For LSTAR, consider transition function $G(\cdot)$ such that

$$G(z_t, \gamma^*, \delta) = \{1 + \exp[-(\gamma^*/s_Z)(z_t - \delta)]\}^{-1}$$

- Threshold Parameter δ Determines When Transition Occurs
- Slope Parameter $\gamma = \gamma^*/s_Z$ Determines the Rate of Transition
- As $\gamma \rightarrow \infty$, $G(z_t, \gamma^*, \delta)$ Becomes a Step Function



Chapter 1: LSTAR Model

- Prior Distributions

- TAR(Geweke and Terui, 1993; Chen and Lee, 1995; Koop and Potter, 1999), STAR(Lubrano, 2000; Lopes and Salazar, 2006; Livingston Jr. and Nur, 2017)
- $\mu_\alpha \sim \mathcal{N}(\cdot, \cdot)$ and $\mu_\beta \sim \mathcal{N}(\cdot, \cdot)$
- $\sigma^2 \sim \mathcal{IG}(\cdot, \cdot)$
- $\gamma^* \sim \mathcal{LN}(\cdot, \cdot)$
- $\delta \sim U[q_Z(0.15), q_Z(0.85)]$ where $q_Z(\cdot)$ is the empirical quantile function
- Bayesian Global-Local Shrinkage Priors for α and β

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

$$\lambda_k^2 \sim \pi_{Local}(\cdot) \text{ and } \lambda^2 \sim \pi_{Global}(\cdot)$$

- LASSO (Park and Casella, 2008; Hans, 2009; Schmidt and Makalic, 2013)
- Regime-Specific LASSO (Separate Global Tuning Parameters)
- Variable Selection LASSO (Lykou and Ntzoufras, 2013)
- Horseshoe (Carvalho et al., 2010; Makalic and Schmidt, 2016)



Chapter 1: LSTAR Model

- Transition Variable

- Change-Point Option: $z_t = t$
- Exogenous Option: $z_t = x_{t-d}$
- Endogenous Option: $z_t = y_{t-d}$ (Self-Exciting)
- Let $\mathbf{d}'_t = [y_{t-1}, y_{t-2}, \dots, y_{t-d_{max}}]$ and $\phi'_t = [\phi_1, \phi_2, \dots, \phi_{d_{max}}]$. Reparameterize transition variable $z_t = \phi' \mathbf{d}_t$.

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

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.



Chapter 1: LSTAR Model

- Simulation Study

Let $\mathbf{d}_t = [y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]'$ and $\phi' = [\phi_1, \phi_2, \phi_3, \phi_4]$.

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

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

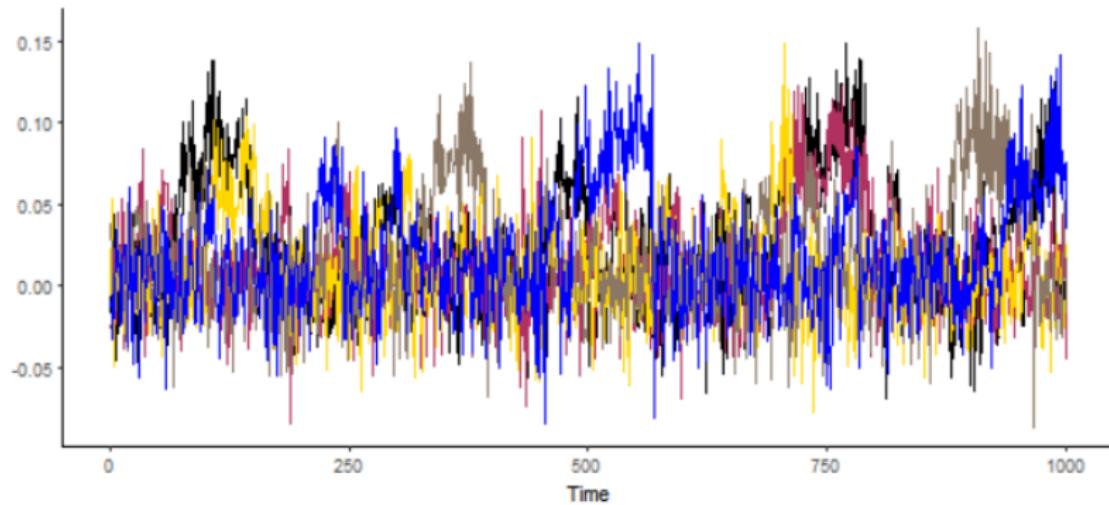
$$\text{and } \epsilon_t \sim \text{i.i.d. } N(0, 0.02^2)$$

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.



Chapter 1: LSTAR Model

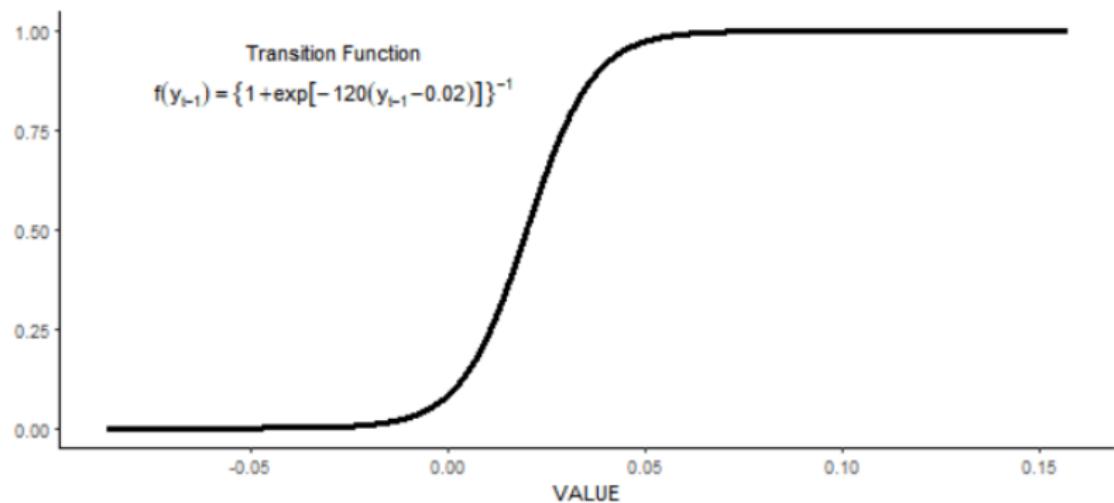
Figure: Ten Random Replications





Chapter 1: LSTAR Model

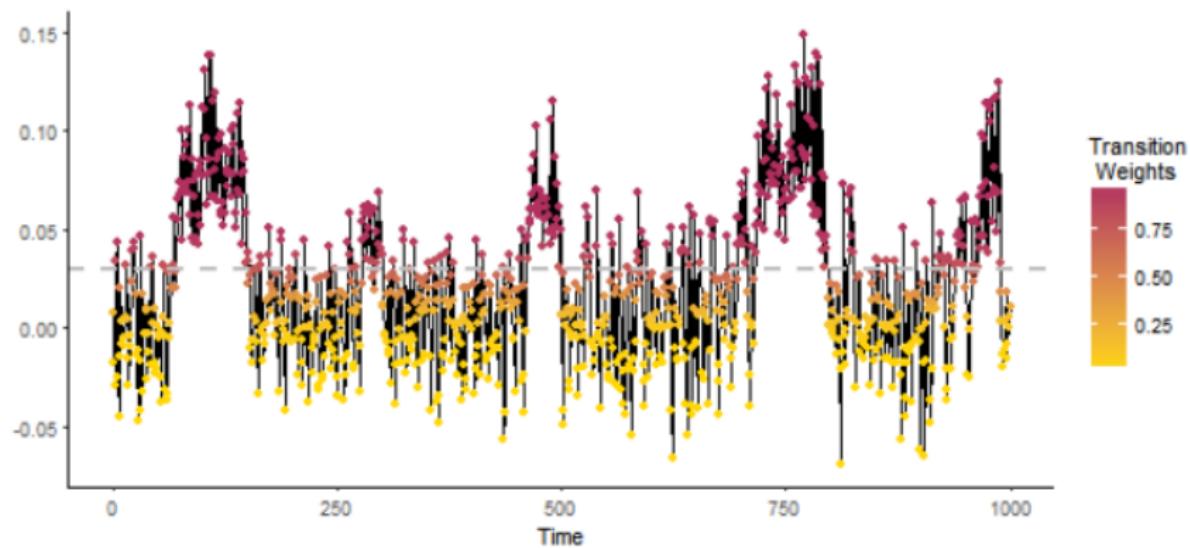
Figure: Transition Function





Chapter 1: LSTAR Model

Figure: Illustration of Regime-switching Behavior



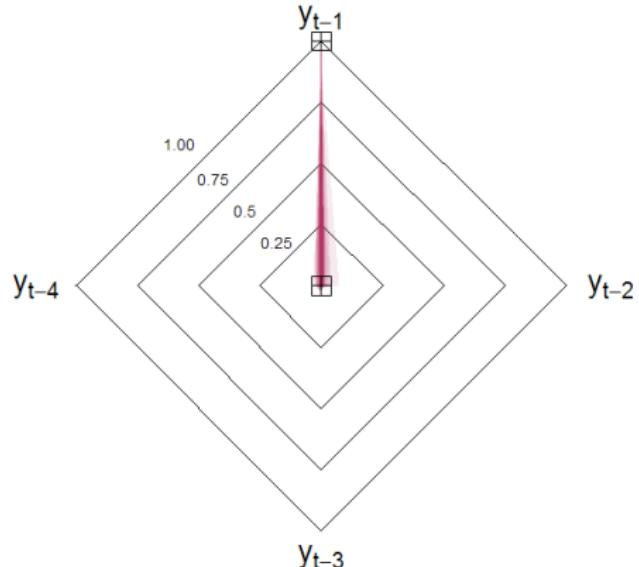


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 1)

Consider $z_{1,t} = y_{t-1} = [1, 0, 0, 0] \mathbf{d}_t$.

Figure: Posterior Means of ϕ from 100 Replications



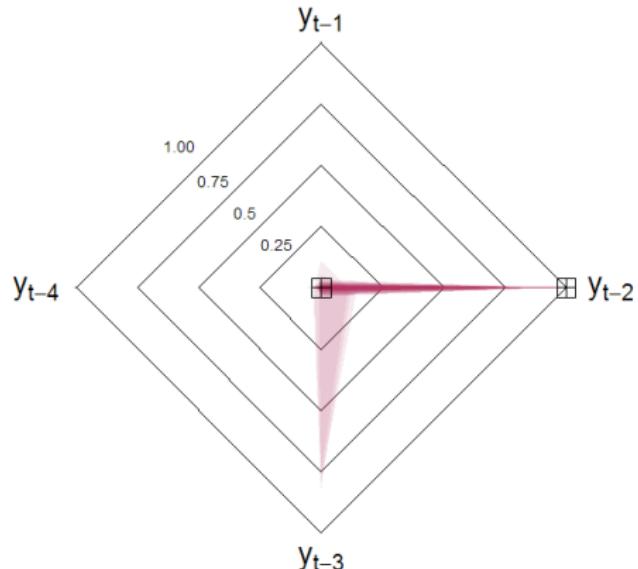


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 2)

Consider $z_{2,t} = y_{t-2} = [0, 1, 0, 0] \mathbf{d}_t$.

Figure: Posterior Means of ϕ from 100 Replications



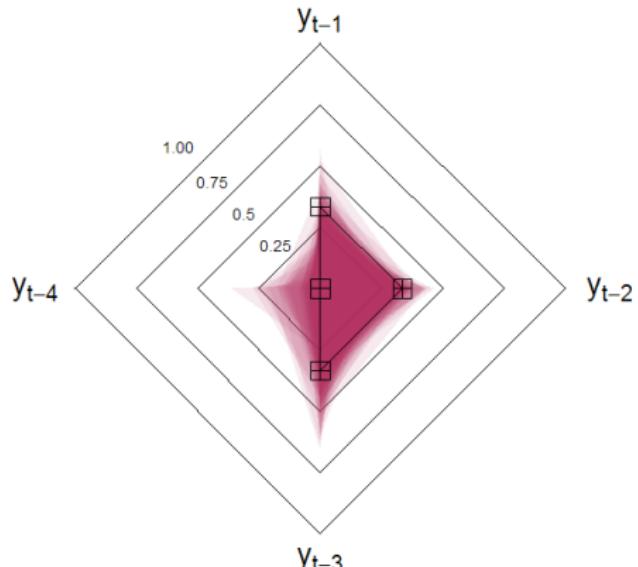


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 3)

Consider $z_{3,t} = \frac{y_{t-1} + y_{t-2} + y_{t-3}}{3} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0] d_t$.

Figure: Posterior Means of ϕ from 100 Replications





Chapter 1: LSTAR Model

- Application to Sunspot Data (Granger, 1957; Teräsvirta et al., 2010)
- Application to Daily Maximum Water Temperatures (Kamarianakis et al., 2016)
 - Data Used From 31 Rivers in Spain
 - Models to Forecast Daily Maximum Water Temperature
 - Inclusion of Exogenous Distributed Lag Terms from Known Air Temperatures
 - Horizon-Specific Models Targeting 3-step and 7-step Ahead Forecasts
 - Nonlinear Models Improved Forecasting Accuracy for Some Rivers



Chapter 1: LSTAR Model

- Contribution and Novelty
 - Efficacy of Bayesian Regularization for Nonlinear Time Series Models
 - Use of the Dirichlet Prior Leads to More Encompassing LSTAR Specification
 - Regime-Specific Tuning Parameters Influences Convergence in MCMC
 - Detailed **R** Code Provided for Reproducibility
- Feedback from *International Journal of Forecasting*
 - Focus on Dirchlet Priors for Estimating Transition Variable
 - Better Forecasting Application
 - Consider Density Forecasts Along with Point Forecasts



Chapter 2: TAR Model

- Need for Traffic Occupancy Models
 - Advanced Traffic Management Systems (ATMS) Monitor Traffic Characteristics in Real Time
 - ATMS Require Fast Short-Term Forecasting to Reduce Congestion
 - Traffic Occupancy is the Percent of Time a Detection Zone is Occupied
 - Different States of Traffic: Free-Flow, Congested, Transitional
 - Factors Influencing Regime Changes : Weekly Work Patterns, Accidents, Weather, etc.
- Traffic Data Considered
 - Major Athens' Arterial: Alexandras Ave.
 - Time Period: April 2000
 - Obtained by National Technical University of Athens
 - Provided for 2013 TRANsportation Data FORecasting Competition (TRANSFOR) Developed by the Traffic Research Board (TRB) for Annual Meeting Workshop (Kamarianakis, 2014)
 - Measured on 90s Interval, but Mean Aggregated to 3min Interval



Chapter 2: TAR Model

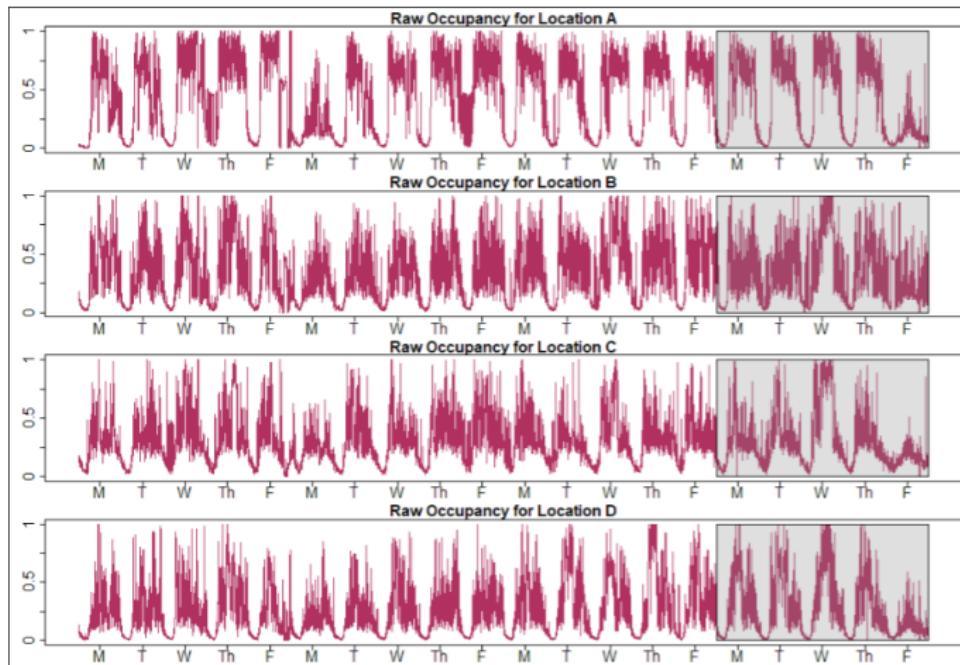
Figure: Map of Traffic Network in Athens, Greece





Chapter 2: TAR Model

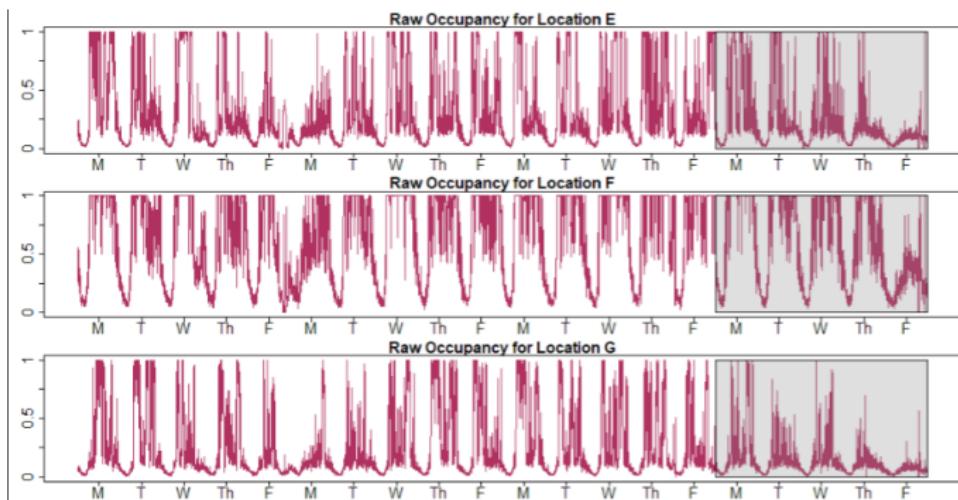
Figure: Raw Traffic Occupancy From Westbound Detectors





Chapter 2: TAR Model

Figure: Raw Traffic Occupancy From Eastbound Detectors





Chapter 2: TAR Model

- (L, D, h) -Specific Models

- Location $L \in \{A, B, C, D, E, F, G\}$
- Work Day $D \in \{M, T, W, Th, F\}$
- Horizon $h \in \{1, 3, 5\}$

- Data Transformation

- Let O_t Represent the Traffic Occupancy at Time t
- $Y_t = \text{logit}(O_t) = \log[O_t / (1 - O_t)]$
- Raw Data Adjusted at the Boundary so $\text{logit}(\cdot)$ Is Defined
- Forecasts Evaluated on Original Scale, but

$$\hat{O}_t \neq \text{logit}^{-1}(\hat{Y}_t)$$

- Density Forecasts Produced from $\{\text{logit}^{-1}(\hat{Y}_t^{(s)})\}_{s=1}^S$ where $\{\hat{Y}_t^{(s)}\}_{s=1}^S$ are S posterior samples obtained from the posterior predictive distribution $f(\hat{Y}_t | \mathcal{I}_t^*)$ where $\mathcal{I}_t^* = \{y_k\}_{k=t-h}^{t-h-P+1}$



Chapter 2: TAR Model

- Horizon-Specific Gaussian TAR(P) Model with $(m + 1)$ -regimes

$$y_t = \phi_0^{(j)} + \sum_{i=1}^P \phi_i^{(j)} y_{t-h-i+1} + \sigma \epsilon_t, \text{ for } \delta_{j-1} < y_{t-h} \leq \delta_j$$

where $\sigma > 0$, $j \in \{1, 2, \dots, m + 1\}$, $h \in \mathbb{N}$, and $\epsilon_t \sim \mathcal{N}(0, 1)$.

Vector of Thresholds $\boldsymbol{\delta} = [\delta_1, \dots, \delta_m]$.

Partitions the Process into $m + 1$ regimes such that

$$-\infty = \delta_0 < \delta_1 \leq \delta_2 \leq \dots \leq \delta_m < \delta_{m+1} = \infty.$$



Chapter 2: TAR Model

- High Dimensional Linear Representation (Chan et al., 2015, 2017)

- Let $\mathbf{y} = [y_1, \dots, y_T]', \epsilon = [\epsilon_1, \dots, \epsilon_T]',$ and define matrix \mathbf{X} by

$$\mathbf{X} = \begin{bmatrix} 1 & y_{1-h} & y_{1-h-1} & \dots & y_{1-h-P+1} \\ 1 & y_{2-h} & y_{2-h-1} & \dots & y_{2-h-P+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{T-h} & y_{T-h-1} & \dots & y_{T-h-P+1} \end{bmatrix}.$$

Second Column of \mathbf{X} Contains the h -Specific Transition Variable.

Model Matrix \mathbf{X} Often Used in Linear AR(P) Regressions.



Chapter 2: TAR Model

- High Dimensional Linear Representation (Cont.)

- Reorder \mathbf{y} , ϵ , and \mathbf{X} According to Transition Variable

Sorting function $\pi(i) : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ where $\pi(i)$ equates to the time index of the i th smallest element in $[y_{1-h}, y_{2-h}, \dots, y_{T-h}]'$. Now,

$$\mathbf{y}_R = [y_{\pi(1)+h}, \dots, y_{\pi(T)+h}]',$$

$$\epsilon_R = [\epsilon_{\pi(1)+h}, \dots, \epsilon_{\pi(T)+h}]',$$

and

$$\mathbf{X}_1 = \begin{bmatrix} 1 & y_{\pi(1)} & y_{\pi(1)-1} & \cdots & y_{\pi(1)-P+1} \\ 1 & y_{\pi(2)} & y_{\pi(2)-1} & \cdots & y_{\pi(2)-P+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{\pi(T)} & y_{\pi(T)-1} & \cdots & y_{\pi(T)-P+1} \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_{\pi(1)} \\ \mathbf{y}'_{\pi(2)} \\ \vdots \\ \mathbf{y}'_{\pi(T)} \end{bmatrix}$$



Chapter 2: TAR Model

- High Dimensional Linear Representation (Cont.)

- Finite Set of m Thresholds for an $(m + 1)$ -Regime TAR(P)

Define the Empirical Quantile Function,

$$q(\cdot) : [0, 1] \rightarrow [\min\{y_{t-h} : t = 1, 2, \dots, T\}, \max\{y_{t-h} : t = 1, 2, \dots, T\}]$$

Select Sequence $\{p_k\}_{k=1}^m$ of m Evenly Spaced Percentiles where

$$p_{\min} = p_1 < \dots < p_m = p_{\max}$$

For a Fully Saturated TAR Model Limited to $(m + 1)$ Regimes, Fix *a priori*

$$\delta = [q(p_1), q(p_2), \dots, q(p_m)]'$$



Chapter 2: TAR Model

- High Dimensional Linear Representation (Cont.)
 - Finite Set of m Thresholds for an $(m + 1)$ -Regime TAR(P) (Cont.)

For $j \in \{2, \dots, m + 1\}$, Let k_j Represent the Number of Elements in $[y_{1-h}, y_{2-h}, \dots, y_{T-h}]'$ Less than $q(p_{j-1})$ and Define

$$\mathbf{X}_j = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 1 & y_{\pi(k_j+1)} & y_{\pi(k_j+1)-1} & \dots & y_{\pi(k_j+1)-P+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_{\pi(T)} & y_{\pi(T)-1} & \dots & y_{\pi(T)-P+1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{y}'_{\pi(k_j+1)} \\ \vdots \\ \mathbf{y}'_{\pi(T)} \end{bmatrix}.$$

- Fully Saturated $(m + 1)$ -Regime TAR(P) as a Linear Regression

$$\mathbf{y}_R = \mathbf{X}_R \boldsymbol{\theta}_R + \boldsymbol{\epsilon}_R$$

$\mathbf{X}_R = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{m+1}]$ is a $T \times (P + 1)(m + 1)$ Matrix

$\boldsymbol{\theta}_R = [\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_{m+1}]'$ is a $(P + 1)(m + 1) \times 1$ Vector of Grouped Reparameterized Coefficients



Chapter 2: TAR Model

- Baseline (L, D) -Specific Seasonal Model (Cont.)

$$y_t = \mu + \sum_{j=1}^H \left[\alpha_j \sin\left(\frac{2\pi t j}{480}\right) + \beta_j \cos\left(\frac{2\pi t j}{480}\right) \right] + \sigma \epsilon_t$$

where $\sigma > 0$, $H \in \mathbb{N}$, and $\epsilon_t \sim \mathcal{N}(0, 1)$.

Representable as a High Dimensional Linear Regression,

$$\mathbf{y}_F = \mathbf{X}_F \boldsymbol{\theta}_F + \boldsymbol{\epsilon}_F$$

- Considerations for Traffic Occupancy

- Maximum AR Order $P = 7$
- Maximum Number of Thresholds $m = 50$
- Set $p_{min} = 0.15$ and $p_{max} = 0.85$
- Saturated 51-Regime TAR(7) Model with 408 Parameters in $\boldsymbol{\theta}_R$
- Maximum Number of Sine/Cosine Pairs $H = 150$
- Saturated Seasonal Harmonic Regression Model with 301 Parameters in $\boldsymbol{\theta}_F$



Chapter 2: TAR Model

- Three-Step Procedure For Automatic Estimation and Selection
 - Full Model $\mathbf{y}_R = \mathbf{X}_R \boldsymbol{\theta}_R + \boldsymbol{\epsilon}_R$ Nests 6.61×10^{122} Different $(m^* + 1)$ -Regime Subset TAR(P) Models where $0 \leq m^* \leq m$
 - Step 1: Sparse Estimation Using Horseshoe+ Shrinkage Priors
 - Group LASSO Used by Chan et al. (2015)
 - BHS⁺ Hierarchy for Each θ_i in $\boldsymbol{\theta}_R$ (Bhadra et al., 2016)

$$\theta_i | \lambda_i, \tau, \sigma^2 \sim \mathcal{N}(0, \lambda_i^2 \tau^2 \sigma^2)$$

$$\lambda_i \sim \mathcal{C}^+(0, \eta_i)$$

$$\eta_i \sim \mathcal{C}^+(0, 1)$$

$$\tau \sim \mathcal{C}^+(0, 1)$$

- Modified Hierarchy Required for Gibbs Sampling (Makalic and Schmidt, 2016)

If $\lambda_i^2 | \nu_i \sim \mathcal{IG}(\frac{1}{2}, \frac{1}{\nu_i})$ and $\nu_i \sim \mathcal{IG}(\frac{1}{2}, 1)$, then $\lambda_i^2 \sim \mathcal{C}^+(0, 1)$ (Wand et al., 2011).



Chapter 2: TAR Model

- Three-Step Procedure For Sparse Estimation (Cont.)
 - Step 2: Regime Identification
 - Good Starting Point Under Full Saturated Model \mathcal{M}_R
 - Samples $\{\theta_R^{(s)}\}_{s=1}^S$ and $\{\sigma^{(s)}\}_{s=1}^S$ from Joint Posterior Distribution
 - Given Candidate Submodel \mathcal{M}_\perp , Posterior Samples $\{\theta_\perp^{(s)}\}_{s=1}^S$ and $\{\sigma_\perp^{(s)}\}_{s=1}^S$ Obtained Via Projection
 - Gaussian Linear Models (Piironen and Vehtari, 2015, 2017)

$$\theta_\perp^{(s)} = (\mathbf{X}'_\perp \mathbf{X}_\perp)^{-1} \mathbf{X}'_\perp \mathbf{X}_R \theta_R^{(s)}$$
$$\sigma_\perp^{(s)} = \sqrt{(\sigma^{(s)})^2 + \frac{(\mathbf{X}_R \theta_R^{(s)} - \mathbf{X}_\perp \theta_\perp^{(s)})' (\mathbf{X}_R \theta_R^{(s)} - \mathbf{X}_\perp \theta_\perp^{(s)})}{T}}$$



Chapter 2: TAR Model

- Three-Step Procedure For Sparse Estimation (Cont.)
 - Step 2: Regime Identification (Cont.)
 - Kullback-Leibler (KL) Divergence (Kullback and Leibler, 1951) Measures the Overall Discrepancy Between the Posterior Predictive Distributions $p(y_{T+1}|\mathcal{M}_R, \mathbf{y}_R, \mathbf{X}_R)$ and $p(y_{T+1}|\mathcal{M}_\perp, \mathbf{y}_\perp, \mathbf{X}_\perp)$
 - KL Divergence for a Particular Sample

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

- Overall Discrepancy

$$D(\mathcal{M}_R || \mathcal{M}_\perp) = \frac{1}{S} \sum_{s=1}^S d_\perp^{(s)}(\boldsymbol{\theta}_R^{(s)}, \sigma^{(s)})$$



Chapter 2: TAR Model

- Three-Step Procedure For Sparse Estimation (Cont.)
 - Step 2: Regime Identification (Cont.)

- Forward Stepwise Selection Algorithm (Peltola et al., 2014)

Begin with Linear AR(P) Model, Denoted $\mathcal{M}_{\perp}^{(1)}$, where

$$\theta_{\perp}^{(1)} = [\theta'_1, \mathbf{0}', \mathbf{0}', \dots, \mathbf{0}']',$$

with initial discrepancy $D(\mathcal{M}_R || \mathcal{M}_{\perp}^{(1)})$

For each $j \in \{2, \dots, m+1\}$, θ_j is Added to $\theta_{\perp}^{(1)}$ and the Best 2-Regime TAR(P) Model $\mathcal{M}_{\perp}^{(2)}$ Minimizes the Discrepancy $D(\mathcal{M}_R || \mathcal{M}_{\perp}^{(2)})$.

Continue to Identify the Best 3-Regime TAR, 4-Regime TAR, ...

Stopping Rule Based on Relative Explanatory Power ($ReIE$) from Dupuis and Robert (2003)

$$ReIE(\mathcal{M}_{\perp}) = 1 - \frac{D(\mathcal{M} || \mathcal{M}_{\perp})}{D(\mathcal{M} || \mathcal{M}_{\perp}^{(1)})}$$



Chapter 2: TAR Model

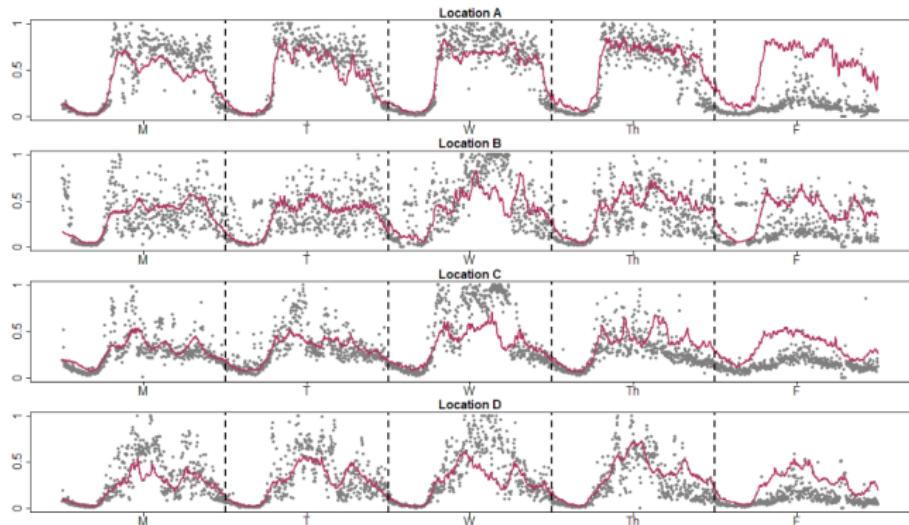
- Three-Step Procedure For Sparse Estimation (Cont.)
 - Step 3: Final Subset Selection
 - Let $\mathcal{J} = \{j : \theta_j \neq 0\}$ indicate the AR(P) parameter groups in θ_R Selected Via Forward Algorithm
 - Let $\theta_{i,j}$ Represent the i th Parameter in the j th Vector θ_j for $i \in \{1, 2, \dots, P+1\}$ and $j \in \{1, 2, \dots, m+1\}$.
 - The Set $\mathcal{I} = \{\theta_{i,j} : i = 1, \dots, P+1 \text{ and } j \in \mathcal{J}\}$ Contains Potentially Relevant Parameters in θ_R
 - Repeat Forward Stepwise Algorithm Across \mathcal{I} . The Intercept Only Model Utilized for $\mathcal{M}_{\perp}^{(1)}$.
 - Result: Final Choice \mathcal{M}_* is a $(m^* + 1)$ -Regime Subset TAR(P) Model where m^* is the Number of Parameter Groups with at Least 1 Selected Parameter.



Chapter 2: TAR Model

- Results
 - (L, D) -Specific Seasonal Profiles (Quickly Forecasts at All Horizons)

Figure: Forecasts Based on Seasonal Profiles for Westbound Detectors

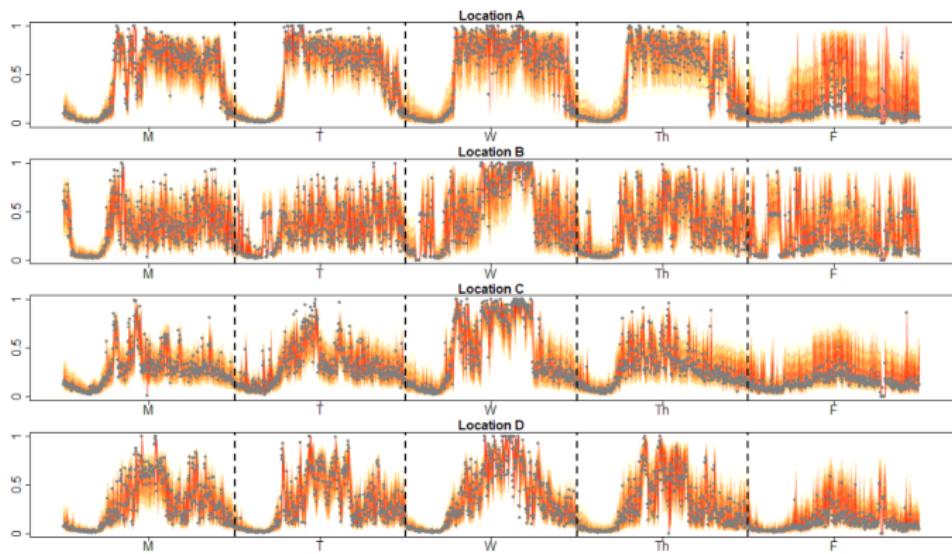




Chapter 2: TAR Model

- Results (Cont.)
 - $(L, D, 1)$ -Specific Final Subset TAR(7) Models

Figure: 1-Step Ahead Density TAR Forecasts for Westbound Detectors

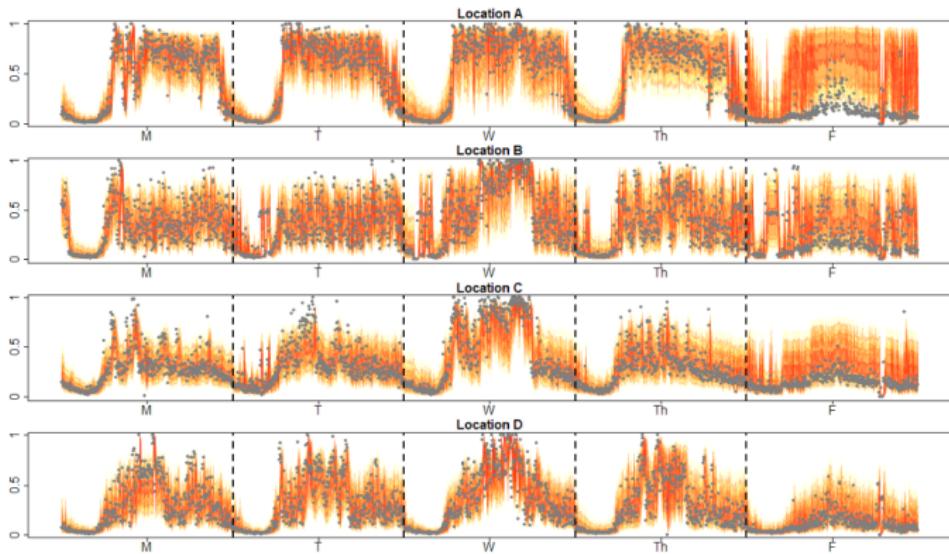




Chapter 2: TAR Model

- Results (Cont.)
 - $(L, D, 3)$ -Specific Final Subset TAR(7) Models

Figure: 3-Step Ahead Density TAR Forecasts for Westbound Detectors

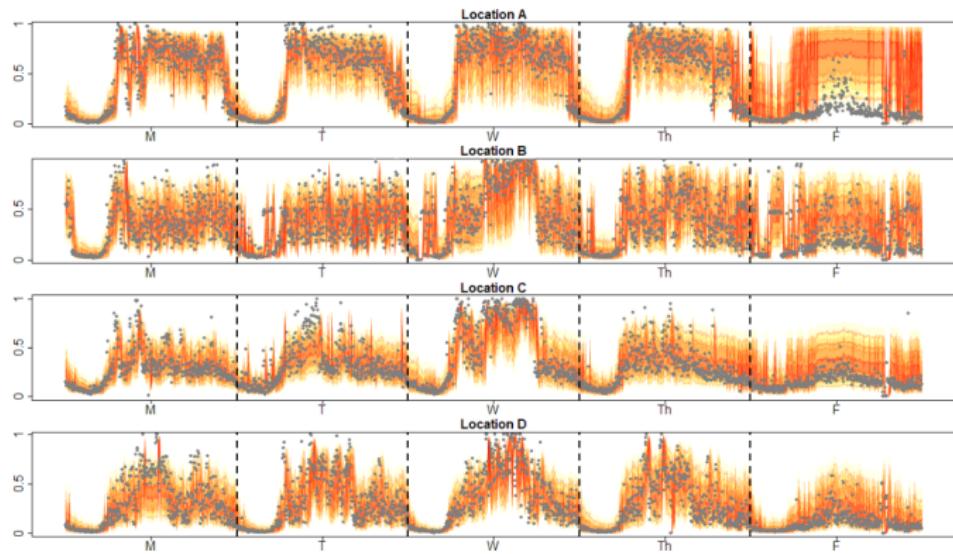




Chapter 2: TAR Model

- Results (Cont.)
 - $(L, D, 5)$ -Specific Final Subset TAR(7) Models

Figure: 5-Step Ahead Density TAR Forecasts for Westbound Detectors





Chapter 2: TAR Model

- Results (Cont.)
 - Comparison of Point Forecasts
 - Evaluated on Mean Absolute Scaled Forecast Error (Hyndman and Koehler, 2006)
- $$\text{MASFE}(h) = \frac{1}{T_h} \sum_{t=P+h}^{480} \left| \frac{O_t - \hat{O}_t}{\text{MAE}_{RW}(h)} \right|$$
- $\text{MAE}_{RW}(h)$ is the Fitted MAE from Naive Random Walk where $\hat{O}_t = O_{t-h}$

Table: 1-Step Ahead MASFE Forecast Comparison

Day	Model	Location					
		A	B	C	D	E	F
M	TAR	1.02	1.10	0.88	1.07	1.87	0.81
	SEAS	1.80	1.47	1.15	1.43	4.02	1.57
T	TAR	0.90	1.05	1.04	0.98	1.36	1.03
	SEAS	1.35	1.36	1.22	1.46	3.36	1.65
W	TAR	1.04	1.11	0.91	0.97	2.27	1.86
	SEAS	1.39	2.01	2.18	1.61	4.65	2.90
Th	TAR	0.93	0.89	0.82	0.92	1.48	1.52
	SEAS	1.44	1.43	1.51	1.42	3.98	2.74
F	TAR	1.80	1.08	1.01	0.85	1.45	2.40
	SEAS	4.77	2.23	1.98	1.83	4.37	6.24



Chapter 2: TAR Model

- Results (Cont.)
 - Comparison of Point Forecasts (Cont.)

Table: 3-Step Ahead MASFE Forecast Comparison

Day	Model	Location					
		A	B	C	D	E	F
M	TAR	0.94	1.06	0.88	1.13	1.85	0.88
	SEAS	1.36	1.17	0.93	1.14	2.91	1.19
T	TAR	0.87	1.04	0.96	1.03	1.46	1.15
	SEAS	1.04	1.06	0.91	1.10	2.24	1.22
W	TAR	1.09	1.15	1.10	0.99	1.75	2.00
	SEAS	1.15	1.61	1.69	1.21	2.94	2.24
Th	TAR	0.90	0.96	0.82	0.93	1.88	1.47
	SEAS	1.15	1.09	1.14	1.03	2.69	1.96
F	TAR	3.04	1.09	1.00	0.66	1.14	2.47
	SEAS	3.53	1.57	1.42	1.33	3.12	4.35



Chapter 2: TAR Model

- Results (Cont.)
 - Comparison of Point Forecasts (Cont.)

Table: 5-Step Ahead MASFE Forecast Comparison

Day	Model	Location					
		A	B	C	D	E	F
M	TAR	0.94	0.99	0.89	1.12	1.97	0.86
	SEAS	1.24	1.06	0.88	1.06	2.46	1.08
T	TAR	0.81	1.05	0.95	1.00	1.47	1.13
	SEAS	0.95	0.95	0.85	0.99	1.85	1.07
W	TAR	1.02	1.12	1.04	0.99	1.67	1.94
	SEAS	1.01	1.44	1.56	1.12	2.30	1.95
Th	TAR	0.84	0.98	0.82	0.87	1.51	1.48
	SEAS	1.05	1.03	1.07	0.94	2.13	1.73
F	TAR	2.85	1.20	0.88	0.70	1.26	2.52
	SEAS	3.07	1.48	1.33	1.19	2.59	3.82
							1.00
							2.01



Chapter 2: TAR Model

- Contribution and Novelty
 - Advances Methodology for Estimating $\text{TAR}(P)$ Models with Potentially Many Regimes
 - Shows Relevancy in an Industry Needing Short-term Forecasting
 - Easy 3-Step Bayesian Approach Capable of Selecting Regimes and AR Parameters Within Regimes (More Flexible Final Models)
 - Appendix Provides Defense for Horseshoe+ Hierarchy
- Future Developments
 - Modifications for Student t Distributed Errors (Shows Promise)
 - Quality of Forecast Credible Regions
 - Other Techniques for Modeling Heteroskedasticity
 - Look at Using Different Transition Variables (Composite, Difference, etc.)



Chapter 3: ARMA Model

- ARMA(p, q) Model

- Classic Parametric Form

$$\phi(B)y_t = \theta(B)\epsilon_t$$

where

$$\phi(B) = 1 - \sum_{j=1}^p \phi_j B^j \text{ and } \theta(B) = 1 + \sum_{k=1}^q \theta_k B^K$$

and B Represents the Backshift Operator such that $B^k y_t = y_{t-k}$

- Autoregressive Order p and Moving Average Order q
- Stationary: Roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$
- Invertible: Roots of $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$
- Multiplicative Seasonal ARMA (Box and Jenkins, 1976) are Subset ARMA Processes
- Presence of MA Terms Poses an Estimation Problem (Hamilton, 1994; Cryer and Chan, 2008)



Chapter 3: ARMA Model

- Fast Estimation of ARMA(p, q)

- Residuals $\{\hat{\epsilon}_t : t = p' + 1, \dots, T\}$ of a Long AR(p') Process Approximate the Unobserved $\{\epsilon_t\}$ (Hannan and Rissanen, 1982; Brockwell and Davis, 2016)
- Let $\mathbf{y} = [y_m, \dots, y_T]', \boldsymbol{\epsilon} = [\epsilon_m, \dots, \epsilon_T]', \boldsymbol{\beta} = [\boldsymbol{\phi}', \boldsymbol{\theta}']' = [\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q]', m = p' + \max\{p, q\} + 1$, and

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_m \\ \mathbf{x}'_{m+1} \\ \vdots \\ \mathbf{x}'_T \end{bmatrix} = \begin{bmatrix} y_{m-1} & \cdots & y_{m-p} & \hat{\epsilon}_{m-1} & \cdots & \hat{\epsilon}_{m-q} \\ y_m & \cdots & y_{m-p+1} & \hat{\epsilon}_m & \cdots & \hat{\epsilon}_{m-q+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{T-1} & \cdots & y_{T-p} & \hat{\epsilon}_{T-1} & \cdots & \hat{\epsilon}_{T-q} \end{bmatrix}.$$

The ARMA(p, q) Model is Expressed by $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

- Consider $p' = 10 \log_{10}(T)$. Hannan and Kavalieris (1984) and Chen and Chan (2011) Suggest Using Information Criteria to Select p' . Reduces the Loss of Data, but Substantially Effects Results.



Chapter 3: ARMA Model

- Subset ARMA(p, q) Penalized Estimation Methods

- Classic Techniques

- Adaptive LASSO (Zou, 2006)

$$\hat{\beta}_{AL}(\lambda) = \underset{\beta}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_{i=1}^{p+q} \hat{w}_i |\beta_i|$$

- Adaptive Elastic Net (Zou and Zhang, 2009)

$$\hat{\beta}_{AE}(\lambda, \alpha) = \underset{\beta}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \left[(1 - \alpha) \sum_{i=1}^{p+q} \beta_i^2 + \alpha \sum_{i=1}^{p+q} \hat{w}_i |\beta_i| \right]$$

- Vector of Weights $\hat{\mathbf{w}} = |\hat{\beta}_L + 1/T|^{-\eta}$ where

$$\hat{\beta}_L(\lambda) = \underset{\beta}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_{i=1}^{p+q} |\beta_i|.$$

$\hat{\beta}_L$ is the Original LASSO Estimate of Tibshirani (1996), and $\eta = 2$, as Recommended by Zou (2006); Chen and Chan (2011)



Chapter 3: ARMA Model

- Subset ARMA(p, q) Penalized Estimation Methods

- Classic Techniques (Cont.)

- Optimal Selection of Tuning Parameters λ and α

For ADLASSO and ADENET, Model Selection Determined by λ and α . Two Stages Require Tuning. Estimate Prediction Error for Grid of λ and α

Time Series Studies Often Perform Selection Based on Out-of-Sample (OOS) Forecasting

Penalized Information Criteria Such as AIC or BIC (Chen and Chan, 2011)

Cross-Validated (CV) Measures of Error (Stone, 1974; Hastie et al., 2009)

Appropriateness of CV Questioned in Time Series Analysis. Blocked Approaches Used for Dependent Data (Burman et al., 1994; Racine, 2000; Arlot et al., 2010; Bergmeir and Benítez, 2012)

Regular K-Fold CV Consistently Outperforms OOS in Estimating Prediction Error (Bergmeir et al., 2018).



Chapter 3: ARMA Model

- Subset ARMA(p, q) Penalized Estimation Methods (Cont.)
 - Bayesian Techniques
 - Bayesian Horseshoe (Carvalho et al., 2009, 2010) and Bayesian Horseshoe+ (Bhadra et al., 2016) Hierarchies Considered for Initial Estimation
 - Projection Method with Forward Selection Algorithm Used to Identify the Best Model at Every Level of Flexibility from Intercept-Only to Fully Saturated ARMA(p, q)
 - Final Model Selected Based on Relative Explanatory Power (*ReE*) or Out-of-Sample Forecasting Results (Piironen and Vehtari, 2017)



Chapter 3: ARMA Model

- Overview of Methods

ADLASSO and ADENET Variants Denoted AL_m and AE_m where $m \in \{1, 2, \dots, 11\}$

Table: Summary of ADLASSO and ADENET Variants

Method (m)	Initial Weights (Stage 1)	Final Model (Stage 2)
1	AIC	AIC
2	AIC	BIC
3	BIC	BIC
4		OOS
5		depOOS
6		CV-5
7		CV-10
8		LOOCV
9		BCV-5
10		BCV-10
11		LOBOCV



Chapter 3: ARMA Model

- Overview of Methods (Cont.)

Bayesian Variants Denoted BHS_m and BHS_m^+ where $m \in \{1, 2, \dots, 4\}$

Table: Summary of BHS and BHS⁺ Variants

Method (m)	Final Model Selection
1	$RelE(\cdot) > 0.90$
2	$RelE(\cdot) > 0.95$
3	$RelE(\cdot) > 0.98$
4	OOS



Chapter 3: ARMA Model

- Simulation Study
 - Gaussian ARMA Processes (Chen and Chan, 2011)

$$y_{1,t} = 0.8y_{1,t-1} + 0.7y_{1,t-6} - 0.56y_{1,t-7} + \epsilon_{1,t}$$

$$\begin{aligned}y_{2,t} = & 0.8y_{2,t-1} + 0.7y_{2,t-6} - 0.56y_{2,t-7} \\& + 0.8\epsilon_{2,t-1} + 0.7\epsilon_{2,t-6} + 0.56\epsilon_{2,t-7} + \epsilon_{2,t}\end{aligned}$$

$$y_{3,t} = 0.8\epsilon_{3,t-1} + 0.7\epsilon_{3,t-6} + 0.56\epsilon_{3,t-7} + \epsilon_{3,t}$$

Standard Gaussian Errors $\{\epsilon_{1,t}\}$, $\{\epsilon_{2,t}\}$, and $\{\epsilon_{3,t}\}$. Abbreviated Models I, II, and III, Respectively. Samples of Length $T \in \{120, 240, 360\}$.

Data Generating Processes are Subset ARMA(7, 7). Maximum ARMA Orders $P = Q = 14$.

- Evaluating Subset ARMA Selection
 - C: Relative Frequency of Selecting All Relevant Parameters
 - I: Relative Frequency of Identifying the True Model
 - -: False Negative Rate (Probability of Missing a Relevant Parameter)
 - +: False Positive Rate (Probability of Selecting an Irrelevant Parameter)



Chapter 3: ARMA Model

- Simulation Study (Cont.)

- Sensitivity: Order Selection for Long $AR(p')$ Process

Table: Effect of Using AIC to Select p' on ADLASSO Subset ARMA(14, 14) Estimation of Model I Based on 500 Replications

T	C	Long $AR(p')$			Short $AR(p')$		
		I	$-$	$+$	C	I	$-$
AL ₁	120	0.19	0.01	0.36	0.28	0.02	0.00
	240	0.40	0.05	0.24	0.27	0.02	0.00
	360	0.46	0.07	0.21	0.26	0.03	0.00
AL ₂	120	0.16	0.04	0.40	0.17	0.02	0.00
	240	0.36	0.13	0.27	0.18	0.01	0.00
	360	0.45	0.17	0.22	0.18	0.03	0.01
AL ₃	120	0.05	0.01	0.44	0.12	0.01	0.00
	240	0.15	0.05	0.35	0.17	0.01	0.00
	360	0.21	0.09	0.31	0.20	0.01	0.01



Chapter 3: ARMA Model

- Simulation Study (Cont.)
 - Results for Model I

Table: Subset ARMA(14, 14) Results from 200 Replications of Model I

m	T	C	AL _{m}			AE _{m}		
			I	-	+	I	-	+
1	360	0.50	0.08	0.20	0.24	0.54	0.08	0.17
2	360	0.42	0.16	0.23	0.19	0.50	0.20	0.19
3	360	0.20	0.10	0.32	0.19	0.20	0.10	0.30
4	360	0.28	0.12	0.30	0.14	0.70	0.00	0.10
5	360	0.24	0.12	0.32	0.15	0.66	0.00	0.12
6	360	0.36	0.16	0.27	0.16	0.52	0.18	0.19
7	360	0.44	0.16	0.23	0.15	0.54	0.16	0.18
8	360	0.42	0.21	0.24	0.15	0.53	0.17	0.19
9	360	0.44	0.20	0.23	0.12	0.60	0.04	0.15
10	360	0.36	0.16	0.26	0.13	0.54	0.04	0.17
11	360	0.46	0.24	0.23	0.11	0.62	0.05	0.14
m	T	C	BHS _{m}			BHS _{m} ⁺		
			I	-	+	I	-	+
1	360	0.70	0.60	0.18	0.04	0.66	0.57	0.21
2	360	0.88	0.62	0.08	0.04	0.88	0.60	0.07
3	360	0.92	0.42	0.05	0.06	0.90	0.49	0.06
4	360	0.92	0.30	0.04	0.09	0.92	0.32	0.05



Chapter 3: ARMA Model

- Simulation Study (Cont.)

- Results for Model II

Table: Subset ARMA(14, 14) Results from 200 Replications of Model II

m	T	AL_m				AE_m			
		C	I	-	+	C	I	-	+
1	360	0.26	0.01	0.15	0.38	0.42	0.00	0.13	0.38
2	360	0.26	0.02	0.16	0.32	0.36	0.02	0.14	0.33
3	360	0.20	0.01	0.18	0.30	0.23	0.02	0.17	0.31
4	360	0.06	0.03	0.28	0.08	0.70	0.00	0.05	0.78
5	360	0.06	0.04	0.27	0.08	0.67	0.00	0.06	0.78
6	360	0.18	0.05	0.18	0.25	0.30	0.00	0.14	0.32
7	360	0.16	0.04	0.19	0.24	0.26	0.01	0.16	0.30
8	360	0.16	0.02	0.20	0.26	0.26	0.01	0.16	0.31
9	360	0.06	0.04	0.29	0.06	0.34	0.00	0.14	0.36
10	360	0.08	0.06	0.26	0.07	0.26	0.00	0.16	0.37
11	360	0.04	0.03	0.27	0.06	0.31	0.00	0.14	0.38
m	T	BHS_m				BHS_m^+			
		C	I	-	+	C	I	-	+
1	360	0.13	0.02	0.24	0.09	0.12	0.03	0.25	0.09
2	360	0.57	0.18	0.10	0.09	0.51	0.16	0.12	0.10
3	360	0.89	0.05	0.03	0.15	0.88	0.08	0.04	0.14
4	360	0.84	0.09	0.05	0.17	0.87	0.10	0.04	0.15



Chapter 3: ARMA Model

- Simulation Study (Cont.)
- Results for Model III

Table: Subset ARMA(14, 14) Results from 200 Replications of Model III

m	T	AL _{m}				AE _{m}			
		C	I	-	+	C	I	-	+
1	360	0.45	0.03	0.26	0.33	0.47	0.02	0.21	0.34
2	360	0.36	0.04	0.33	0.21	0.40	0.07	0.27	0.18
3	360	0.78	0.19	0.10	0.11	0.78	0.18	0.10	0.11
4	360	0.49	0.18	0.24	0.21	0.86	0.00	0.05	0.61
5	360	0.52	0.20	0.23	0.20	0.90	0.00	0.03	0.61
6	360	0.46	0.22	0.25	0.12	0.48	0.12	0.22	0.18
7	360	0.44	0.23	0.28	0.12	0.50	0.14	0.23	0.16
8	360	0.48	0.22	0.24	0.10	0.48	0.12	0.23	0.17
9	360	0.55	0.14	0.20	0.15	0.60	0.04	0.16	0.29
10	360	0.56	0.21	0.18	0.14	0.64	0.02	0.14	0.25
11	360	0.42	0.08	0.27	0.19	0.48	0.01	0.22	0.34
m	T	BHS _{m}				BHS _{m} ⁺			
		C	I	-	+	C	I	-	+
1	360	0.26	0.03	0.46	0.14	0.26	0.04	0.46	0.14
2	360	0.40	0.00	0.34	0.22	0.38	0.00	0.36	0.21
3	360	0.62	0.00	0.18	0.39	0.59	0.00	0.20	0.34
4	360	0.40	0.02	0.35	0.25	0.35	0.02	0.36	0.24



Chapter 3: ARMA Model

- Application

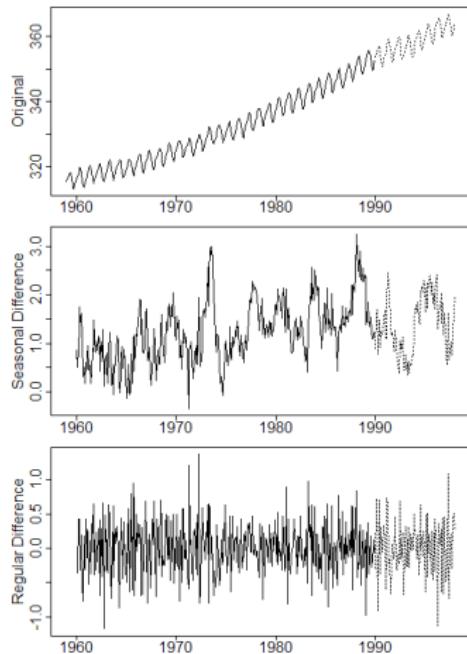
- Monthly CO₂ Levels Measured in Mauna Loa, Hawaii
- Popular Textbook Example for Seasonal ARMA Processes with *Period* = 12
- Fitting Period: Jan 1994 - Dec 1989 (468 Time Points)
- Forecasting Period: Jan 1990 - Dec 1997 (96 Time Points)
- Apply Seasonal and Regular Differencing



Chapter 3: ARMA Model

- Application (Cont.)

Figure: Plots of Mauna Loa CO₂ Levels





Chapter 3: ARMA Model

- Application (Cont.)

- Forecasting Metrics

- Root Mean Squared Error

$$RMSE = \frac{1}{96} \sum_{j=1}^{96} (y_j - \hat{y}_j)^2$$

- Mean Absolute Scaled Error
 - Mean Bias

$$MB = \frac{1}{96} \sum_{j=1}^{96} (y_j - \hat{y}_j)$$

- Mean Directional Bias

$$MDB = \frac{1}{96} \sum_{j=1}^{96} \text{sgn}(y_j - \hat{y}_j)$$

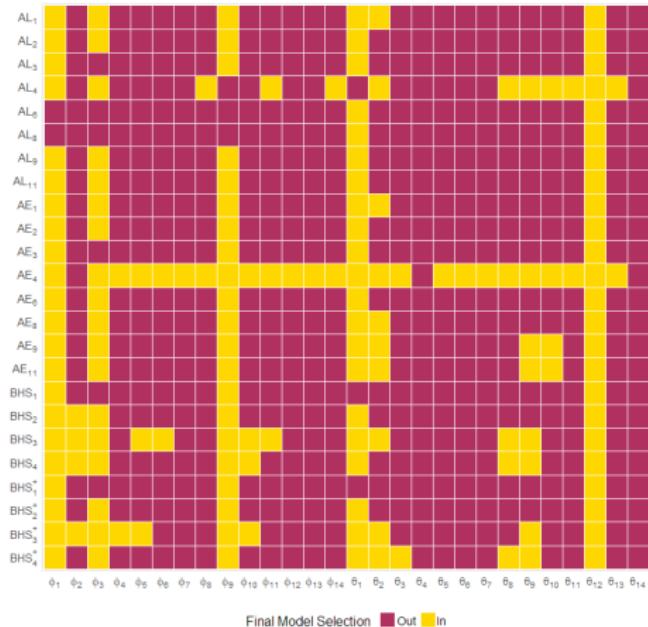
where $\text{sgn}(x) = 1$ if $x > 0$ and $\text{sgn}(x) = -1$ if $x < 0$



Chapter 3: ARMA Model

- Application (Cont.)
 - Final Model Selection

Figure: Subset ARMA Models for Mauna Loa CO₂





Chapter 3: ARMA Model

- Application (Cont.)
 - Out-of-Sample Prediction

Table: One-Step Ahead Forecasting Results for Mauna Loa CO₂

m	RMSE		MASE		MB		MDB	
	AL _m	AE _m	AL _m	AE _m	AL _m	AE _m	AL _m	AE _m
1	0.34	0.34	0.53	0.53	-0.13	-0.13	-0.31	-0.29
2	0.33	0.33	0.52	0.52	-0.10	-0.10	-0.21	-0.21
3	0.34	0.34	0.52	0.52	-0.10	-0.10	-0.21	-0.21
4					Not Invertible (NI)			
6	0.34	0.33	0.53	0.52	-0.10	-0.09	-0.17	-0.23
8	0.34	0.34	0.53	0.54	-0.10	-0.14	-0.19	-0.31
9	0.34	0.36	0.52	0.57	-0.10	-0.18	-0.23	-0.44
11	0.33	0.36	0.52	0.56	-0.10	-0.17	-0.21	-0.44
m	BHS _m	BHS _m ⁺	BHS _m	BHS _m ⁺	BHS _m	BHS _m ⁺	BHS _m	BHS _m ⁺
1	0.31	0.31	0.49	0.49	-0.01	-0.01	0.04	0.06
2	0.32	0.32	0.50	0.50	-0.02	-0.02	0.00	-0.02
3	0.32	0.32	0.51	0.51	-0.02	-0.02	0.00	0.00
4	0.32	0.32	0.50	0.50	-0.01	-0.01	0.02	0.02
RW	0.64		1.03		0.00		-0.02	
ARMA	0.37		0.60		0.01		0.13	
SARMA	0.30		0.49		-0.04		-0.02	



Chapter 3: ARMA Model

- Contribution and Novelty
 - Highlight Issues with Previous Usage of Adaptive LASSO in Subset ARMA Selection
 - Extends Subset ARMA Estimation to Adaptive Elastic Net
 - Demonstrate the Appropriateness of Multiple Cross Validation Techniques with Penalized Regression for ARMA Models
 - Multi-step Bayesian Approach that Outperforms in Estimation, Selection, and Forecasting
 - Provides Detailed R Code for Reproducibility
- Future Developments
 - Extend Theoretical Asymptotic Results for Adaptive Elastic Net
 - More Focus on Bayesian Predictive Posterior Projection Method
 - Consider Alternative Bayesian Handling of Moving Average Terms

Questions?

References |

- Arlot, S., Celisse, A., et al. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79.
- Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213.
- Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.
- Bhadra, A., Datta, J., Polson, N. G., Willard, B., et al. (2016). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*.
- Box, G. E. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control, Revised Ed.* Holden-Day.
- Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.
- Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.
- Burman, P., Chow, E., and Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, 81(2):351–358.
- 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. (2009). Handling sparsity via the horseshoe. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 73–80, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA. PMLR.

References II

- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.
- Chan, N. H., Ing, C.-K., Li, Y., and Yau, C. Y. (2017). Threshold estimation via group orthogonal greedy algorithm. *Journal of Business & Economic Statistics*, 35(2):334–345.
- Chan, N. H., Yau, C. Y., and Zhang, R.-M. (2015). Lasso estimation of threshold autoregressive models. *Journal of Econometrics*, 189(2):285–296.
- Chen, C. W. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, 16(5):483–492.
- Chen, K. and Chan, K.-S. (2011). Subset arma selection via the adaptive lasso. *Statistics and its Interface*, 4(2):197–205.
- Cryer, J. D. and Chan, K.-S. (2008). Seasonal models. *Time series analysis: with applications in R*, pages 227–246.
- 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.
- Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, pages 233–244.
- Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14(5):441–454.
- Granger, C. W. J. (1957). A statistical model for sunspot activity. *Astrophysical Journal*, 126:152.
- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination.
- Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.

References III

- Hannan, E. J. and Kavalieris, L. (1984). A method for autoregressive-moving average estimation. *Biometrika*, 71(2):273–280.
- Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1):81–94.
- Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning 2nd edition.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.
- Kamarianakis, Y. (2014). Space-time modeling of traffic variables with adaptive lasso. In *JSM Proceedings*, Section on Statistical Learning and Data Mining, pages 2732–2738. American Statistical Association.
- Kamarianakis, Y., Ayuso, S. V., Rodríguez, E. C., and Velasco, M. T. (2016). Water temperature forecasting for spanish rivers by means of nonlinear mixed models. *Journal of Hydrology: Regional Studies*, 5:226–243.
- Koop, G. and Potter, S. M. (1999). Dynamic asymmetries in us unemployment. *Journal of Business & Economic Statistics*, 17(3):298–312.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.
- 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.

References IV

- Lubrano, M. (2000). Bayesian analysis of nonlinear time series models with a threshold.
- 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). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.
- Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica, Journal of the Econometric Society*, pages 173–220.
- 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.
- Nieto, F. H., Zhang, H., and Li, W. (2013). Using the reversible jump mcmc procedure for identifying and estimating univariate tar models. *Communications in Statistics - Simulation and Computation*, 42(4):814–840.
- Park, T. and Casella, G. (2008). The bayesian lasso.
- Peltola, T., Havulinna, A. S., Salomaa, V., and Vehtari, A. (2014). Hierarchical bayesian survival analysis and projective covariate selection in cardiovascular event risk prediction. In *Proceedings of the Eleventh UAI Conference on Bayesian Modeling Applications Workshop - Volume 1218, BMAW'14*, pages 79–88, Aachen, Germany, Germany. CEUR-WS.org.
- Piironen, J. and Vehtari, A. (2015). Projection predictive model selection for gaussian processes.
- Piironen, J. and Vehtari, A. (2017). Comparison of bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735.
- Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of econometrics*, 99(1):39–61.
- Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, 80(1):164–168.

References V

- Schmidt, D. F. and Makalic, E. (2013). Estimation of stationary autoregressive models with the bayesian lasso. *Journal of Time Series Analysis*, 34(5):517–531.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, pages 111–147.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, 89(425):208–218.
- 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*.
- 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.
- Wand, M. P., Ormerod, J. T., Padoan, S. A., and Frühwirth, R. (2011). Mean field variational bayes for elaborate distributions. *Bayesian Analysis*, 6(4):847–900.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.
- Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*, 37(4):1733.