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 $\boldsymbol{x}_t' = [y_{t-1}, y_{t-2}, \cdots, y_{t-P}]$, $\boldsymbol{\alpha}' = [\alpha_1, \cdots, \alpha_P]$, and $\boldsymbol{\beta}' = [\beta_1, \cdots, \beta_P]$.

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

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

  - 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 $\gamma \to \infty$, $G(z_t, \gamma^*, \delta)$ 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 $\mu_\beta \sim \mathcal{N}(\cdot,\cdot)$
  - $\sigma^2 \sim \mathcal{IG}(\cdot,\cdot)$
  - $\gamma^* \sim \mathcal{LN}(\cdot,\cdot)$
  - $\delta \sim \mathrm{U}[q_Z(0.15), q_Z(0.85)]$ where $q_Z(\cdot)$ is the empirical quantile function
  - Bayesian Global-Local Shrinkage Priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

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

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

    - 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 $\boldsymbol{d}_t' = [y_{t-1}, y_{t-2}, \cdots, y_{t-d_{max}}]$ and $\boldsymbol{\phi}_t' = [\phi_1, \phi_2, \cdots, \phi_{d_{max}}]$. Reparameterize transition variable $z_t = \boldsymbol{\phi}' \boldsymbol{d}_t$.

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

  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.

Chapter 1: LSTAR Model

- Simulation Study

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

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

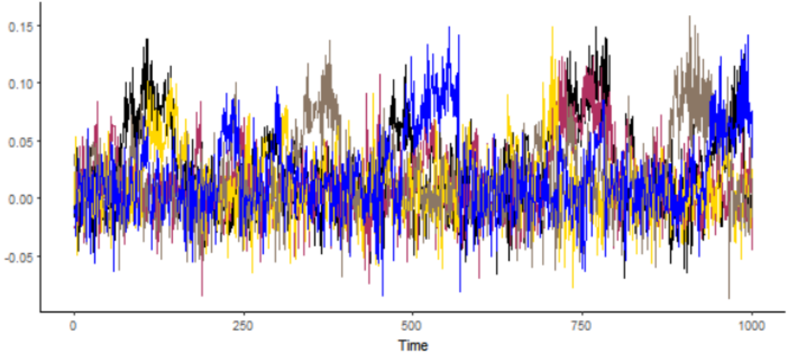
$$\text{where: } G(y_{t-1}) = \left\{1 + \exp\left[-120(\phi'\boldsymbol{d}_t - 0.02)\right]\right\}^{-1}$$

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

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.

Chapter 1: LSTAR Model

Figure: Ten Random Replications



Chapter 1: LSTAR Model

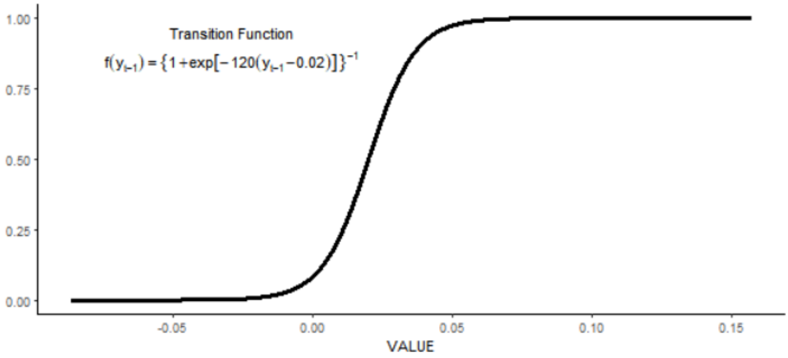
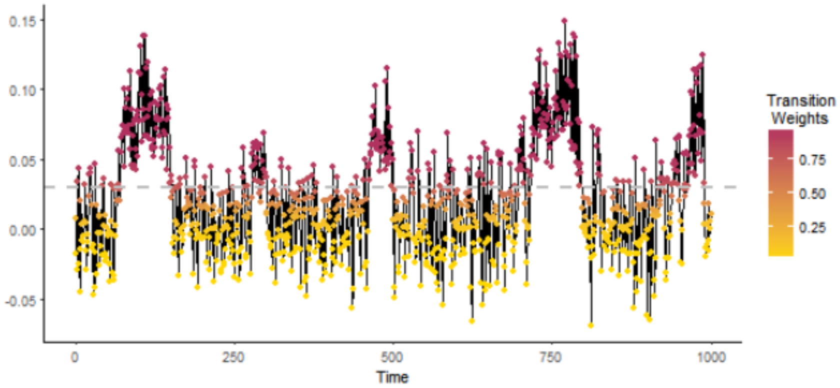


Figure: Transition Function

Chapter 1: LSTAR Model

Figure: Illustration of Regime-switching Behavior

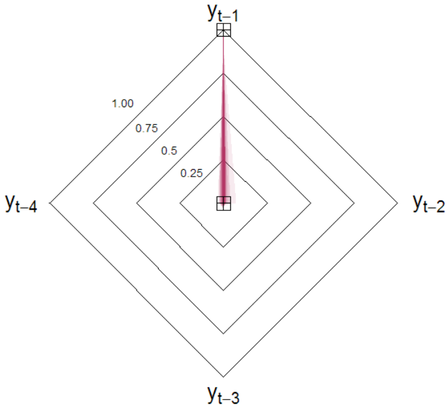


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 1)

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

Figure: Posterior Means of ϕ from 100 Replications

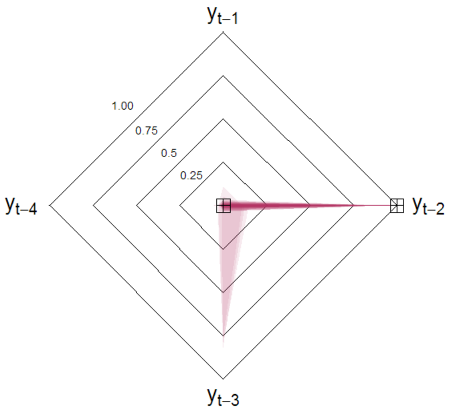


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 2)

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

Figure: Posterior Means of ϕ from 100 Replications

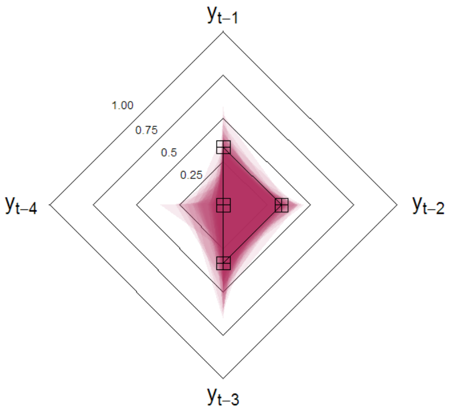


Chapter 1: LSTAR Model

- Bayesian Selection of the Threshold Variable (Scenario 3)

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

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

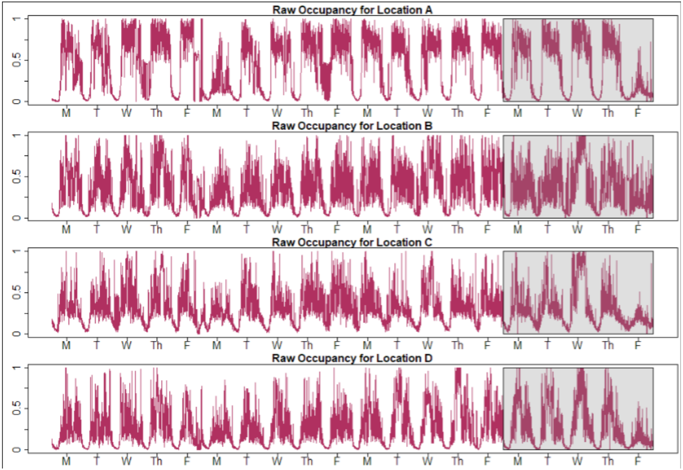
- Need for Traffic Occupancy Models
  - Advanced Traffic Management Systems (ATMS) Monitor Traffic Characteristics in Real Time
  - ATMS Require Fast Short-Term Forecasting to Reduce Congestion
  - Traffic Occupancy is the Percent of Time a Detection Zone is Occupied
  - Different States of Traffic: 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

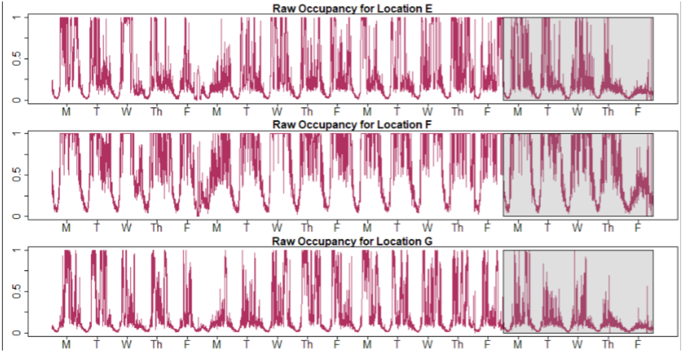
Chapter 2: TAR Model

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

$$\hat{O}_{L,t} \neq \mathrm{logit}^{-1}(\hat{Y}_{L,t})$$

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

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

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

Partitions the Process into $m+1$ regimes such that
$-\infty = \delta_0 < \delta_1 \leq \delta_2 \leq \cdots \leq \delta_m < \delta_{m+1} = \infty$.

Chapter 2: TAR Model

- High Dimensional Linear Representation (Chan et al., 2015, 2017)
  - Let $\boldsymbol{y} = [y_1, \cdots, y_T]'$, $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_T]'$, and define matrix $\boldsymbol{X}$ by

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

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

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

Chapter 2: TAR Model

- High Dimensional Linear Representation (Cont.)
    - Reorder $\boldsymbol{y}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{X}$ According to Transition Variable
      Sorting function $\pi(i) : \{1, \cdots, T\} \to \{1, \cdots, T\}$ where $\pi(i)$ equates to the time index of the ith smallest element in $[y_{1-h}, y_{2-h}, \cdots, y_{T-h}]'$. Now,

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

$$\boldsymbol{\epsilon}_R = [\epsilon_{\pi(1)+h}, \cdots, \epsilon_{\pi(T)+h}]',$$

and

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

Chapter 2: TAR Model

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

    Define the Sample Quantile Function,

    $$q(.) : [0,1] \to [\min\{y_{t-h} : t = 1,2,\cdots,T\}, \max\{y_{t-h} : t = 1,2,\cdots,T\}]$$

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

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

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

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

Chapter 2: TAR Model

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

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

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

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

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

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

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

Chapter 2: TAR Model

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

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

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

Representable as a High Dimensional Linear Regession,

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

- 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 $\boldsymbol{y}_R = \boldsymbol{X}_R\boldsymbol{\theta}_R + \boldsymbol{\epsilon}_R$ 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
    - Adaptive LASSO Used by Chan et al. (2015)
    - BHS^+ Hierarchy for Each θ_i in $\boldsymbol{\theta}_R$ (Bhadra et al., 2016)

$$\begin{aligned} \theta_i|\lambda_i, \tau, \sigma^2 &\sim \mathcal{N}(0, \lambda_i^2\tau^2\sigma^2) \\ \lambda_i &\sim \mathcal{C}^+(0, \eta_i) \\ \eta_i &\sim \mathcal{C}^+(0, 1) \\ \tau &\sim \mathcal{C}^+(0, 1) \end{aligned}$$

    - 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 $\{\boldsymbol{\theta}_R^{(s)}\}_{s=1}^S$ and $\{\sigma^{(s)}\}_{s=1}^S$ from Joint Posterior Distribution
    - Given Candidate Submodel $\mathcal{M}_\perp$, Posterior Samples $\{\boldsymbol{\theta}_\perp^{(s)}\}_{s=1}^S$ and $\{\sigma_\perp^{(s)}\}_{s=1}^S$ Obtained Via Projection
    - Gaussian Linear Models (Piironen and Vehtari, 2015, 2017)

$$\boldsymbol{\theta}_\perp^{(s)} = (\mathbf{X}_\perp'\mathbf{X}_\perp)^{-1}\mathbf{X}_\perp'\mathbf{X}_R\boldsymbol{\theta}_R^{(s)}$$

$$\sigma_\perp^{(s)} = \sqrt{(\sigma_R^{(s)})^2 + \frac{(\mathbf{X}_R\theta_R^{(s)} - \mathbf{X}_\perp\boldsymbol{\theta}_\perp^{(s)})'(\mathbf{X}_R\boldsymbol{\theta}_R^{(s)} - \mathbf{X}_\perp\boldsymbol{\theta}_\perp^{(s)})}{T}}$$

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 $p(y_{T+1}|\mathcal{M}_R, \mathbf{y}_\perp, \mathbf{X}_\perp)$
    - KL Divergence for a Particular Sample

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

    - Overall Discrepancy

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

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

      $$\boldsymbol{\theta}_{\perp}^{(1)} = [\boldsymbol{\theta}_1', \mathbf{0}', \mathbf{0}', \cdots, \mathbf{0}']',$$

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

      For each $j \in \{2, \cdots, m+1\}$, $\boldsymbol{\theta}_j$ is Added to $\boldsymbol{\theta}_{\perp}^{(1)}$ 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, $\cdots$

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

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

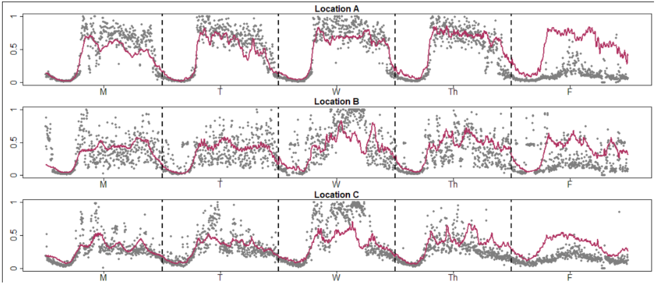
Chapter 2: TAR Model

- Three-Step Procedure For Sparse Estimation (Cont.)
  - Step 3: Final Subset Selection
    - Let $\mathcal{J} = \{j : \boldsymbol{\theta}_j \neq 0\}$ indicate the AR(P) parameter groups in $\boldsymbol{\theta}_R$ Selected Via Forward Algorithm
    - Let $\theta_{i,j}$ Represent the ith Parameter in the jth Vector $\boldsymbol{\theta}_j$ for $i \in \{1, 2, \cdots, P+1\}$ and $j \in \{1, 2, \cdots, m+1\}$.
    - The Set $\mathcal{I} = \{\theta_{i,j} : i = 1, \cdots, P+1 \text{ and } j \in \mathcal{J}\}$ Contains Potentially Relevant Parameters in $\boldsymbol{\theta}_R$
    - 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)

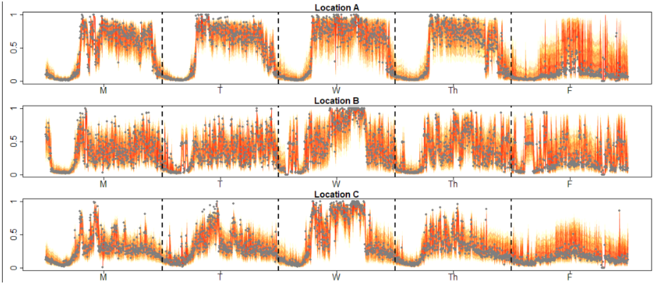
Figure: Forecasts Based on Seasonal Profiles for Westbound Detectors



Chapter 2: TAR Model

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

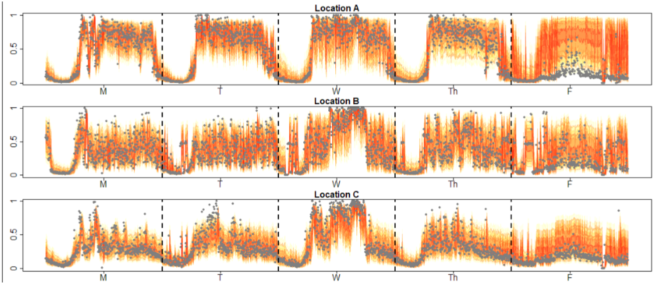
Figure: 1-Step Ahead Density TAR Forecasts for Westbound Detectors



Chapter 2: TAR Model

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

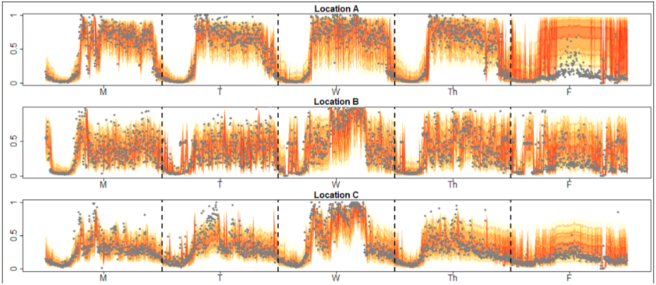
Figure: 3-Step Ahead Density TAR Forecasts for Westbound Detectors



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)

$$\mathrm{MASFE}(h) = \frac{1}{T_h} \sum_{t=P+h}^{480} \left| \frac{O_{L,t} - \widehat{O}_{L,t}}{\mathrm{MAE}_{RW}(h)} \right|$$

    - $\mathrm{MAE}_{RW}(h)$ is the Fitted MAE from Naive Random Walk where $\widehat{O}_t = O_{t-h}$

Table: 1-Step Ahead MASFE Forecast Comparison

Day	Model	A	B	C	D	E	F	G
				Location				
M	TAR	1.02	1.10	0.88	1.07	1.87	0.81	1.48
	SEAS	1.80	1.47	1.15	1.43	4.02	1.57	3.66
T	TAR	0.90	1.05	1.04	0.98	1.36	1.03	1.95
	SEAS	1.35	1.36	1.22	1.46	3.36	1.65	3.29
W	TAR	1.04	1.11	0.91	0.97	2.27	1.86	1.55
	SEAS	1.39	2.01	2.18	1.61	4.65	2.90	2.80
Th	TAR	0.93	0.89	0.82	0.92	1.48	1.52	1.83
	SEAS	1.44	1.43	1.51	1.42	3.98	2.74	4.07
F	TAR	1.80	1.08	1.01	0.85	1.45	2.40	1.16
	SEAS	4.77	2.23	1.98	1.83	4.37	6.24	3.78

Chapter 2: TAR Model

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

Table: 3-Step Ahead MASFE Forecast Comparison

		Location						
Day	Model	A	B	C	D	E	F	G
M	TAR	0.94	1.06	0.88	1.13	1.85	0.88	1.50
	SEAS	1.36	1.17	0.93	1.14	2.91	1.19	2.57
T	TAR	0.87	1.04	0.96	1.03	1.46	1.15	1.21
	SEAS	1.04	1.06	0.91	1.10	2.24	1.22	2.15
W	TAR	1.09	1.15	1.10	0.99	1.75	2.00	1.26
	SEAS	1.15	1.61	1.69	1.21	2.94	2.24	1.71
Th	TAR	0.90	0.96	0.82	0.93	1.88	1.47	1.14
	SEAS	1.15	1.09	1.14	1.03	2.69	1.96	2.68
F	TAR	3.04	1.09	1.00	0.66	1.14	2.47	0.92
	SEAS	3.53	1.57	1.42	1.33	3.12	4.35	2.42

Chapter 2: TAR Model

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

Table: 5-Step Ahead MASFE Forecast Comparison

Day	Model	A	B	C	D	E	F	G
				Location				
M	TAR	0.94	0.99	0.89	1.12	1.97	0.86	1.68
	SEAS	1.24	1.06	0.88	1.06	2.46	1.08	2.17
T	TAR	0.81	1.05	0.95	1.00	1.47	1.13	1.15
	SEAS	0.95	0.95	0.85	0.99	1.85	1.07	1.77
W	TAR	1.02	1.12	1.04	0.99	1.67	1.94	1.26
	SEAS	1.01	1.44	1.56	1.12	2.30	1.95	1.44
Th	TAR	0.84	0.98	0.82	0.87	1.51	1.48	1.22
	SEAS	1.05	1.03	1.07	0.94	2.13	1.73	2.24
F	TAR	2.85	1.20	0.88	0.70	1.26	2.52	1.00
	SEAS	3.07	1.48	1.33	1.19	2.59	3.82	2.01

Chapter 2: TAR Model

- Contribution and Novelty
  - Advances Methodology for Estimating 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

-

Questions?

References I

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 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.

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.

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.

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.

References II

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4):835–845.

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.

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.

References III

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.

Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, 80(1):164–168.

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.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, 89(425):208–218.

References IV

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

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.