

Methodological Details of Bayesian State-Space Models with Variable Selection

José Mauricio Gómez Julián

Table of Contents

| | | |
|-------|---|---|
| 1 | State-Space Model Framework for Time Series | 2 |
| 1.1 | General State-Space Representation | 2 |
| 1.1.1 | Mathematical Formulation | 2 |
| 1.1.2 | Advantages of the State-Space Framework | 3 |
| 1.2 | Implemented Structural Components | 3 |
| 1.2.1 | Local Level (LL) | 3 |
| 1.2.2 | Local Linear Trend (LLT) | 3 |
| 1.2.3 | Seasonal Component (Optional) | 3 |
| 1.3 | Bayesian Variable Selection with Spike-and-Slab | 4 |
| 1.3.1 | Economic Motivation | 4 |
| 1.3.2 | Spike-and-Slab Prior | 4 |
| 1.3.3 | Hyperparameter Calibration | 4 |
| 1.4 | Bayesian Inference via MCMC | 4 |
| 1.4.1 | Sampling Algorithm | 4 |
| 1.4.2 | MCMC Configuration | 5 |
| 1.4.3 | Convergence Diagnostics | 5 |
| 1.5 | Temporal Validation with Leave-Future-Out | 5 |
| 1.5.1 | Validation Scheme | 5 |
| 1.5.2 | Procedure per Fold | 5 |
| 1.6 | Predictive Evaluation Metrics | 6 |
| 1.6.1 | Expected Log Predictive Density (ELPD) | 6 |
| 1.6.2 | Point Error Metrics | 6 |
| 1.6.3 | Calibration Metrics | 6 |
| 1.7 | Optimal Structure Selection | 6 |

| | | |
|--------|---|----|
| 1.7.1 | Structure Grid | 6 |
| 1.7.2 | Selection Criterion | 7 |
| 1.8 | Victory and Stability Criterion | 7 |
| 1.8.1 | Victory per Fold | 7 |
| 1.8.2 | Stability Metrics | 7 |
| 1.9 | Computational Implementation and Optimizations | 7 |
| 1.9.1 | Processing Architecture | 7 |
| 1.9.2 | Special Case Management | 7 |
| 1.9.3 | Output Data Structure | 8 |
| 1.10 | Comparative Analysis with Previous Methodologies | 8 |
| 1.10.1 | Comparative Approach Table | 8 |
| 1.10.2 | Distinctive Advantages of BSTS | 9 |
| 1.10.3 | Relative Limitations | 9 |
| 1.11 | Extensions and Future Developments | 9 |
| 1.11.1 | Immediate Improvements | 9 |
| 1.11.2 | Methodological Extensions | 9 |
| 1.12 | Complete Pipeline Pseudocode | 10 |
| 1.13 | Methodological Conclusions | 11 |

1 State-Space Model Framework for Time Series

State-space models constitute a unified framework for time series analysis that explicitly separates the underlying signal from observational noise. This methodology implements full Bayesian inference on structural time series components (Bayesian Structural Time Series - BSTS), combining the flexibility of structural decomposition with automatic variable selection through spike-and-slab priors. The approach allows uncertainty quantification at all model levels while maintaining economic interpretability of the components.

1.1 General State-Space Representation

1.1.1 Mathematical Formulation

A state-space model is defined by two fundamental equations:

Observation Equation:

$$y_t = Z_t' \alpha_t + \beta' x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

State (Transition) Equation:

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t)$$

where: - y_t is the observation at time t - α_t is the latent state vector of dimension m - x_t is the vector of exogenous regressors of dimension k - β are the regression coefficients - Z_t connects the state to observations - T_t is the state transition matrix - R_t and Q_t define the variance structure of the state process - ϵ_t is the observation error

1.1.2 Advantages of the State-Space Framework

1. **Structural decomposition:** Separates trend, seasonality, and irregular components
2. **Natural handling of missing data:** The Kalman filter automatically interpolates
3. **Complete uncertainty:** Posterior distributions for states and forecasts
4. **Modular flexibility:** Components can be added as needed

1.2 Implemented Structural Components

1.2.1 Local Level (LL)

The simplest model includes only a stochastic level:

$$\begin{aligned} y_t &= \mu_t + \beta' x_t + \epsilon_t \\ \mu_{t+1} &= \mu_t + \eta_{\mu,t} \end{aligned}$$

where $\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$ represents innovations in the level. This model is appropriate for series with varying level but no systematic trend.

1.2.2 Local Linear Trend (LLT)

Extends the LL model by adding a stochastic trend:

$$\begin{aligned} y_t &= \mu_t + \beta' x_t + \epsilon_t \\ \mu_{t+1} &= \mu_t + \delta_t + \eta_{\mu,t} \\ \delta_{t+1} &= \delta_t + \eta_{\delta,t} \end{aligned}$$

where: - μ_t is the level at time t - δ_t is the slope (rate of change) at time t - $\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$ are level innovations - $\eta_{\delta,t} \sim N(0, \sigma_\delta^2)$ are slope innovations

1.2.3 Seasonal Component (Optional)

For series with seasonal patterns:

$$\gamma_{t+1} = - \sum_{s=1}^{S-1} \gamma_{t-s+1} + \eta_{\gamma,t}$$

where S is the seasonal period (e.g., 12 for monthly data) and $\eta_{y,t} \sim N(0, \sigma_y^2)$.

1.3 Bayesian Variable Selection with Spike-and-Slab

1.3.1 Economic Motivation

In high-dimensional contexts with multiple potential lags, variable selection is crucial for: - Avoiding overfitting - Maintaining interpretability - Identifying genuine predictors

1.3.2 Spike-and-Slab Prior

For each coefficient β_j we define a hierarchical prior:

$$\beta_j | \gamma_j \sim \begin{cases} N(0, \tau^2 v_j) & \text{if } \gamma_j = 1 \text{ (slab)} \\ \delta_0 & \text{if } \gamma_j = 0 \text{ (spike)} \end{cases}$$

where: - $\gamma_j \in \{0,1\}$ is the inclusion indicator - $\tau^2 v_j$ is the slab variance (non-zero component) - δ_0 is a point mass at zero

The prior on the indicators is:

$$\gamma_j \sim \text{Bernoulli}(\pi_j)$$

with π_j calibrated according to expected model size.

1.3.3 Hyperparameter Calibration

Hyperparameters are configured via:

1. **Expected model size:** $E[\sum \gamma_j] = \max(1, \min(5, k))$ where k is the number of predictors
2. **Prior information weight:** $w = 0.01$ (weakly informative prior)
3. **Diagonal shrinkage:** $\kappa = 0.5$ for moderate regularization

This configuration balances flexibility with parsimony, allowing data to dominate inference while maintaining smooth regularization.

1.4 Bayesian Inference via MCMC

1.4.1 Sampling Algorithm

Inference is performed through a hybrid Gibbs sampling scheme:

1. **Latent states** ($\alpha_{1:T}$): Forward Filtering Backward Sampling (FFBS)
2. **Variances** ($\sigma_\epsilon^2, \sigma_\mu^2, \sigma_\delta^2$): Gibbs with conjugate inverse-gamma priors
3. **Coefficients and indicators** (β, γ): Stochastic Search Variable Selection (SSVS)

1.4.2 MCMC Configuration

- **Total iterations:** 2000
- **Burn-in:** 500 (25%)
- **Thinning:** Not applied (all post-burn samples)
- **Chains:** 1 (parallelization at pair level)
- **Seed:** Fixed for reproducibility

1.4.3 Convergence Diagnostics

While not explicitly reported for computational efficiency, standard diagnostics include: - Visual inspection of traces - Chain autocorrelation - Effective sample size for key parameters - Stable inclusion probabilities for γ_j

1.5 Temporal Validation with Leave-Future-Out

1.5.1 Validation Scheme

We implement Leave-Future-Out (LFO) with rolling origin:

1. **Initial set:** 80% of data or minimum 30 observations
2. **Forecast horizon:** $h = 6$ months
3. **Step between origins:** 6 months
4. **Window type:** Expanding (accumulates history)

This more conservative scheme (6-month horizon and step vs 12 in ECM-MARS) allows greater temporal resolution in stability evaluation.

1.5.2 Procedure per Fold

For each fold f :

1. **Data preparation:**
 - Split into train/test according to split
 - Scale regressors per fold using train statistics
2. **Model fitting:**
 - **Base:** Only structural components without regressors
 - **Full:** Structural components + regressors with spike-and-slab
3. **Probabilistic prediction:**
 - Generate complete posterior predictive distribution
 - Extract mean, standard deviation, and quantiles
4. **Evaluation:**
 - Calculate ELPD, point metrics, and calibration

1.6 Predictive Evaluation Metrics

1.6.1 Expected Log Predictive Density (ELPD)

For each future observation y_t^* :

$$\text{ELPD}_t = \log \int p(y_t^* | \theta) p(\theta | \mathcal{D}_{\text{train}}) d\theta$$

Approximated via MCMC samples as:

$$\widehat{\text{ELPD}}_t = \log \left(\frac{1}{S} \sum_{s=1}^S p(y_t^* | \theta^{(s)}) \right)$$

where we assume normality: $p(y_t^* | \theta^{(s)}) = N(y_t^*; \mu_t^{(s)}, \sigma_t^{(s)})$.

1.6.2 Point Error Metrics

Using the posterior mean as point prediction:

- **RMSE:** $\sqrt{\frac{1}{h} \sum_{t=1}^h (y_t - \bar{y}_t)^2}$
- **MAE:** $\frac{1}{h} \sum_{t=1}^h |y_t - \bar{y}_t|$

where $\bar{y}_t = \frac{1}{S} \sum_{s=1}^S y_t^{(s)}$ is the posterior mean.

1.6.3 Calibration Metrics

We evaluate the quality of quantified uncertainty:

- **80% Coverage:** Proportion of observations within 80% credible interval
- **95% Coverage:** Proportion of observations within 95% credible interval
- **PIT (Probability Integral Transform):** $u_t = F_{Y_t}(y_t)$ where F_{Y_t} is the predictive CDF

If the model is well-calibrated, PIT values follow a uniform distribution on $[0, 1]$.

1.7 Optimal Structure Selection

1.7.1 Structure Grid

We systematically evaluate:

| Model | Level | Trend | Seasonality | Free Parameters |
|-------|------------|------------|-------------|--|
| LL | Stochastic | No | Optional | $\sigma_\mu^2, \sigma_\epsilon^2$ |
| | c | | | |
| LLT | Stochastic | Stochastic | Optional | $\sigma_\mu^2, \sigma_\delta^2, \sigma_\epsilon^2$ |

| Model | Level | Trend | Seasonality | Free Parameters |
|-------|-------|-------|-------------|-----------------|
| | c | c | | |

1.7.2 Selection Criterion

The optimal structure is selected through:

1. **Primary:** Highest average ΔELPD across folds
2. **Secondary:** Highest support (proportion of winning folds)
3. **Tertiary:** Highest ΔRMSE (error reduction)

This hierarchy prioritizes probabilistic predictive capacity over point fit.

1.8 Victory and Stability Criterion

1.8.1 Victory per Fold

A full model “wins” in fold f if:

$$\begin{cases} \Delta\text{ELPD}_f > 0 & \text{(better predictive density)} \\ \Delta\text{RMSE}_f > 0 & \text{(lower error, i.e., } \text{RMSE}_{\text{base}} - \text{RMSE}_{\text{full}} > 0) \end{cases}$$

Note: Unlike GLM-AR(1), here $\Delta\text{RMSE} = \text{RMSE}_{\text{base}} - \text{RMSE}_{\text{full}}$, so positive values indicate improvement.

1.8.2 Stability Metrics

- **Support:** $\frac{\text{wins}}{\text{folds}} = \text{proportion of victorious folds}$
- **Strict threshold:** support ≥ 0.70 with minimum 5 folds
- **Moderate threshold:** support ≥ 0.60 with minimum 5 folds

1.9 Computational Implementation and Optimizations

1.9.1 Processing Architecture

1. **Parallelization by pairs:** The 84 pairs are processed sequentially
2. **MCMC efficiency:** Single chain per model (speed vs diagnostics trade-off)
3. **Matrix caching:** Reuse of decompositions in Kalman filter

1.9.2 Special Case Management

The code includes robust handling of:

- **Zero variance:** Skip fold if $\text{sd}(y) \approx 0$
- **Convergence failures:** Try-catch with informative messages
- **Singular matrices:** Automatic regularization in Kalman filter
- **Degenerate predictions:** Fallback to quantile-based intervals

1.9.3 Output Data Structure

Each pair generates:

```
list(  
  best_summary = tibble(      # Best model summary  
    spec, folds, wins, support,  
    dELPD_mean, dRMSE_mean, ...  
  ),  
  best_results = tibble(      # Details per fold  
    fold, n_train, n_test,  
    ELPD_base, ELPD_full, dELPD,  
    RMSE_base, RMSE_full, dRMSE,  
    cover80, cover95, pit, ...  
  ),  
  all_summaries = tibble()    # ALL tested structures  
)
```

1.10 Comparative Analysis with Previous Methodologies

1.10.1 Comparative Approach Table

| Aspect | ECM-MARS | GLM-AR(1) | BSTS |
|----------------------------|---------------------|---------------------|------------------------------|
| Paradigm | Hybrid frequentist | Parametric Bayesian | Structural Bayesian |
| Prerequisites | I(1), cointegration | None | None |
| Non-linearity | MARS (splines) | No | No (but flexible components) |
| Variable selection | Manual (pre-tests) | No | Automatic (spike-slab) |
| Temporal components | ECM only | Global AR(1) | Level, trend, seasonality |
| Uncertainty | Implicit bootstrap | Complete posterior | Complete posterior + states |
| Interpretability | High (economic) | Medium | High (decomposition) |
| Computational cost | Medium | High | Very high |

1.10.2 Distinctive Advantages of BSTS

1. **Interpretable decomposition:** Separates signal from noise with economically meaningful components
2. **Automatic selection:** Spike-and-slab identifies relevant predictors without pre-specification
3. **Structural uncertainty handling:** Quantifies uncertainty in unobservable components
4. **Robustness to specification:** Does not require assumptions about integration order or cointegration
5. **Interval forecasts:** Natural and well-calibrated prediction intervals

1.10.3 Relative Limitations

1. **High computational cost:** FFBS + SSVS is intensive, especially with many regressors
2. **Linearity in predictors:** Does not capture non-linear interactions like MARS
3. **Complex diagnostics:** Convergence harder to verify than OLS
4. **Sensitivity to priors in small samples:** Especially for component variances

1.11 Extensions and Future Developments

1.11.1 Immediate Improvements

1. **Additional components:**
 - **Calendar effects:** Business days, moving holidays
 - **Level shifts:** Automatic detection of structural breaks
 - **Dynamic regressors:** Time-varying coefficients
2. **Informative priors:**
 - Incorporate economic information in inclusion priors
 - Hierarchical priors for groups of related variables
3. **Enhanced validation:**
 - Stochastic cross-validation (Pareto smoothed importance sampling)
 - Backtesting with multiple horizons

1.11.2 Methodological Extensions

1. **Non-linearity:**
 - Gaussian processes for non-linear components
 - Bayesian neural networks for interaction capture
2. **Hierarchical models:**

- Partial pooling between related pairs
- Shared latent factors

3. Causality:

- Intervention models with causal inference
- Dynamic causal graphs

1.12 Complete Pipeline Pseudocode

INPUT: DataFrame with time series, variable specification
 OUTPUT: Ranking of pairs by predictive capacity and stability

```
# PREPARATION
1. LOAD data and clean names
2. IDENTIFY 6 circulation variables, 7 production
3. GENERATE 84 pairs (2 × 6 × 7)

# PROCESSING PER PAIR
FOR each pair (Y, X):

  4. BUILD matrix with Y, X, lags(X, 1:6)
  5. REMOVE observations with NA

  # STRUCTURE TUNING
  FOR each structure in {LL, LLT}:

    # LEAVE-FUTURE-OUT
    6. GENERATE LFO splits (init=80%, h=6, step=6)

    FOR each fold:
      7. SPLIT train/test
      8. SCALE regressors with train stats

    # MODELS
    9. FIT base_model:
      - state_spec = structure without regressors
      - MCMC with 2000 iter, 500 burn-in

    10. FIT full_model:
      - state_spec = structure
      - prior = spike_slab(X_lags, expected_size=5)
      - MCMC with 2000 iter, 500 burn-in

  # PREDICTION
  11. GENERATE predictive distributions (h=6)
  12. EXTRACT mean, sd, quantiles

# EVALUATION
13. CALCULATE:
```

```

- ELPD_base, ELPD_full → ΔELPD
- RMSE_base, RMSE_full → ΔRMSE
- Coverage 80%, 95%
- PIT values

14. DETERMINE victory:  $(\Delta\text{ELPD} > 0) \wedge (\Delta\text{RMSE} > 0)$ 

15. SUMMARIZE structure:
  - support = wins/folds
  - metric averages

16. SELECT best structure by ΔELPD
17. SAVE pair results

# RANKING AND EXPORT
18. ORDER pairs by (support DESC, ΔELPD DESC, ΔRMSE DESC)
19. FILTER winners by thresholds (0.70, 0.60)
20. EXPORT CSVs and visualizations
21. GENERATE plots for last fold (without re-fitting)

```

1.13 Methodological Conclusions

The BSTS methodology represents the most comprehensive and flexible approach of the three implemented, combining the rigor of Bayesian inference with the interpretability of structural decomposition. Unlike ECM-MARS which requires prior validation of strict econometric assumptions, and GLM-AR(1) which models only global serial dependence, BSTS decomposes the series into interpretable components while automatically selecting relevant predictors.

Empirical results suggest that BSTS identifies robust predictive relationships with high precision, albeit at substantially higher computational cost. The ability to quantify uncertainty at all levels—from structural components to predictions—makes it particularly valuable for applications where risk assessment is critical.

The implementation with a 6-month horizon (vs 12 in other methodologies) provides more granular evaluation of temporal stability, revealing that many relationships appearing stable at long horizons show variability at shorter scales. This suggests the importance of considering multiple horizons in predictive model evaluation.

The BSTS framework is especially appropriate when: - Component interpretation is a priority - There is uncertainty about which predictors to include - Well-calibrated prediction intervals are required - Data exhibit multiple sources of variation (trend, seasonality, regressors)

The combination of automatic variable selection with structural modeling positions BSTS as a bridge between classical econometric methods and modern machine learning techniques, maintaining interpretability while embracing the inherent complexity in economic data.
