

# Detalles Metodológicos de Modelos de Estado-Espacio Bayesianos con Selección de Variables

José Mauricio Gómez Julián

## Table of Contents

|       |   |   |
|-------|---|---|
| 1     | Marco Teórico de Modelos de Estado-Espacio para Series Temporales ..... | 2 |
| 1.1   | Representación General Estado-Espacio .....                             | 2 |
| 1.1.1 | Formulación Matemática .....  | 2 |
| 1.1.2 | Ventajas del Framework Estado-Espacio .....                             | 3 |
| 1.2   | Componentes Estructurales Implementados .....                           | 3 |
| 1.2.1 | Nivel Local (Local Level - LL).....                                     | 3 |
| 1.2.2 | Tendencia Lineal Local (Local Linear Trend - LLT) .....                 | 3 |
| 1.2.3 | Componente Estacional (Opcional).....                                   | 4 |
| 1.3   | Selección Bayesiana de Variables con Spike-and-Slab .....               | 4 |
| 1.3.1 | Motivación Económica.....   | 4 |
| 1.3.2 | Prior Spike-and-Slab .....  | 4 |
| 1.3.3 | Calibración de Hiperparámetros.....                                     | 4 |
| 1.4   | Inferencia Bayesiana mediante MCMC.....                                 | 5 |
| 1.4.1 | Algoritmo de Muestreo.....  | 5 |
| 1.4.2 | Configuración MCMC .....  | 5 |
| 1.4.3 | Diagnósticos de Convergencia.....                                       | 5 |
| 1.5   | Validación Temporal con Leave-Future-Out .....                          | 5 |
| 1.5.1 | Esquema de Validación.....  | 5 |
| 1.5.2 | Procedimiento por Fold.....   | 5 |
| 1.6   | Métricas de Evaluación Predictiva .....                                 | 6 |
| 1.6.1 | Expected Log Predictive Density (ELPD) .....                            | 6 |
| 1.6.2 | Métricas de Error Puntual.....  | 6 |
| 1.6.3 | Métricas de Calibración .....   | 6 |
| 1.7   | Selección de Estructura Óptima .....                                    | 7 |

|        |  |    |
|--------|--|----|
| 1.7.1  | <b>Grid de Estructuras</b>                           | 7  |
| 1.7.2  | <b>Criterio de Selección</b>                         | 7  |
| 1.8    | <b>Criterio de Victoria y Estabilidad</b>            | 7  |
| 1.8.1  | <b>Victoria por Fold</b>                             | 7  |
| 1.8.2  | <b>Métricas de Estabilidad</b>                       | 7  |
| 1.9    | <b>Implementación Computacional y Optimizaciones</b> | 7  |
| 1.9.1  | <b>Arquitectura de Procesamiento</b>                 | 7  |
| 1.9.2  | <b>Gestión de Casos Especiales</b>                   | 8  |
| 1.9.3  | <b>Estructura de Datos de Salida</b>                 | 8  |
| 1.10   | <b>Análisis Comparativo con Metodologías Previas</b> | 8  |
| 1.10.1 | <b>Tabla Comparativa de Enfoques</b>                 | 8  |
| 1.10.2 | <b>Ventajas Distintivas de BSTS</b>                  | 9  |
| 1.10.3 | <b>Limitaciones Relativas</b>                        | 9  |
| 1.11   | <b>Extensiones y Desarrollos Futuros</b>             | 9  |
| 1.11.1 | <b>Mejoras Inmediatas</b>                            | 9  |
| 1.11.2 | <b>Extensiones Metodológicas</b>                     | 10 |
| 1.12   | <b>Pseudocódigo del Pipeline Completo</b>            | 10 |
| 1.13   | <b>Conclusiones Metodológicas</b>                    | 11 |

# 1 Marco Teórico de Modelos de Estado-Espacio para Series Temporales

Los modelos de estado-espacio constituyen un framework unificado para el análisis de series temporales que separa explícitamente la señal subyacente del ruido observacional. Esta metodología implementa inferencia bayesiana completa sobre componentes estructurales de series temporales (Bayesian Structural Time Series - BSTS), combinando la flexibilidad de la descomposición estructural con la selección automática de variables mediante priors spike-and-slab. El enfoque permite cuantificar la incertidumbre en todos los niveles del modelo mientras mantiene interpretabilidad económica de los componentes.

## 1.1 Representación General Estado-Espacio

### 1.1.1 Formulación Matemática

Un modelo de estado-espacio se define mediante dos ecuaciones fundamentales:

### Ecuación de Observación:

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

### Ecuación de Estado (Transición):

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

donde: -  $y_t$  es la observación en tiempo  $t$  -  $\alpha_t$  es el vector de estado latente de dimensión  $m$  -  $x_t$  es el vector de regresores exógenos de dimensión  $k$  -  $\beta$  son los coeficientes de regresión -  $Z_t$  conecta el estado con las observaciones -  $T_t$  es la matriz de transición del estado -  $R_t$  y  $Q_t$  definen la estructura de varianza del proceso de estado -  $\epsilon_t$  es el error de observación

#### 1.1.2 Ventajas del Framework Estado-Espacio

1. **Descomposición estructural:** Separa tendencia, estacionalidad y componentes irregulares
2. **Manejo natural de datos faltantes:** El filtro de Kalman interpola automáticamente
3. **Incertidumbre completa:** Distribuciones posteriores para estados y pronósticos
4. **Flexibilidad modular:** Componentes agregables según necesidad

## 1.2 Componentes Estructurales Implementados

### 1.2.1 Nivel Local (Local Level - LL)

El modelo más simple incluye solo un nivel estocástico:

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

donde  $\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$  representa innovaciones en el nivel. Este modelo es apropiado para series con nivel variable pero sin tendencia sistemática.

### 1.2.2 Tendencia Lineal Local (Local Linear Trend - LLT)

Extiende el modelo LL agregando una tendencia estocástica:

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

donde: -  $\mu_t$  es el nivel en tiempo  $t$  -  $\delta_t$  es la pendiente (tasa de cambio) en tiempo  $t$  -  $\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$  son innovaciones del nivel -  $\eta_{\delta,t} \sim N(0, \sigma_\delta^2)$  son innovaciones de la pendiente

### 1.2.3 Componente Estacional (Opcional)

Para series con patrones estacionales:

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

donde  $S$  es el período estacional (e.g., 12 para datos mensuales) y  $\eta_{\gamma,t} \sim N(0, \sigma_{\gamma}^2)$ .

## 1.3 Selección Bayesiana de Variables con Spike-and-Slab

### 1.3.1 Motivación Económica

En contextos de alta dimensionalidad con múltiples rezagos potenciales, la selección de variables es crucial para: - Evitar sobreajuste - Mantener interpretabilidad - Identificar predictores genuinos

### 1.3.2 Prior Spike-and-Slab

Para cada coeficiente  $\beta_j$  definimos un prior jerárquico:

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

donde: -  $\gamma_j \in \{0,1\}$  es el indicador de inclusión -  $\tau^2 v_j$  es la varianza del slab (componente no-nulo) -  $\delta_0$  es una masa puntual en cero

El prior sobre los indicadores es:

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

con  $\pi_j$  calibrado según el tamaño esperado del modelo.

### 1.3.3 Calibración de Hiperparámetros

Los hiperparámetros se configuran mediante:

1. **Tamaño esperado del modelo:**  $E[\sum \gamma_j] = \max(1, \min(5, k))$  donde  $k$  es el número de predictores
2. **Peso de información prior:**  $w = 0.01$  (prior débilmente informativo)
3. **Shrinkage diagonal:**  $\kappa = 0.5$  para regularización moderada

Esta configuración balancea flexibilidad con parsimonia, permitiendo que los datos dominen la inferencia mientras se mantiene regularización suave.

## 1.4 Inferencia Bayesiana mediante MCMC

### 1.4.1 Algoritmo de Muestreo

La inferencia se realiza mediante un esquema híbrido de Gibbs sampling:

1. **Estados latentes** ( $\alpha_{1:T}$ ): Forward Filtering Backward Sampling (FFBS)
2. **Varianzas** ( $\sigma_\epsilon^2, \sigma_\mu^2, \sigma_\delta^2$ ): Gibbs con priors inverse-gamma conjugados
3. **Coefficientes e indicadores** ( $\beta, \gamma$ ): Stochastic Search Variable Selection (SSVS)

### 1.4.2 Configuración MCMC

- **Iteraciones totales:** 2000
- **Burn-in:** 500 (25%)
- **Thinning:** No aplicado (todas las muestras post-burn)
- **Cadenas:** 1 (paralelización a nivel de pares)
- **Semilla:** Fija para reproducibilidad

### 1.4.3 Diagnósticos de Convergencia

Aunque no se reportan explícitamente por eficiencia computacional, los diagnósticos estándar incluyen: - Inspección visual de trazas - Autocorrelación de las cadenas - Effective sample size para parámetros clave - Probabilidades de inclusión estables para  $\gamma_j$

## 1.5 Validación Temporal con Leave-Future-Out

### 1.5.1 Esquema de Validación

Implementamos Leave-Future-Out (LFO) con origen móvil:

1. **Conjunto inicial:** 80% de los datos o mínimo 30 observaciones
2. **Horizonte de pronóstico:**  $h = 6$  meses
3. **Paso entre orígenes:** 6 meses
4. **Tipo de ventana:** Expansiva (acumula historia)

Este esquema más conservador (horizonte y paso de 6 meses vs 12 en ECM-MARS) permite mayor resolución temporal en la evaluación de estabilidad.

### 1.5.2 Procedimiento por Fold

Para cada fold  $f$ :

1. **Preparación de datos:**
  - Dividir en train/test según el split
  - Escalar regresores por fold usando estadísticas del train

## 2. Ajuste de modelos:

- **Base:** Solo componentes estructurales sin regresores
- **Full:** Componentes estructurales + regresores con spike-and-slab

## 3. Predicción probabilística:

- Generar distribución predictiva posterior completa
- Extraer media, desviación estándar y cuantiles

## 4. Evaluación:

- Calcular ELPD, métricas puntuales y calibración

# 1.6 Métricas de Evaluación Predictiva

## 1.6.1 Expected Log Predictive Density (ELPD)

Para cada observación futura  $y_t^*$ :

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

Aproximado mediante muestras MCMC como:

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

donde asumimos normalidad:  $p(y_t^* | \theta^{(s)}) = N(y_t^*; \mu_t^{(s)}, \sigma_t^{(s)})$ .

## 1.6.2 Métricas de Error Puntual

Usando la media posterior como predicción puntual:

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

donde  $\bar{y}_t = \frac{1}{S} \sum_{s=1}^S y_t^{(s)}$  es la media posterior.

## 1.6.3 Métricas de Calibración

Evaluamos la calidad de la incertidumbre cuantificada:

- **Coverage 80%:** Proporción de observaciones en intervalo de credibilidad 80%
- **Coverage 95%:** Proporción de observaciones en intervalo de credibilidad 95%
- **PIT (Probability Integral Transform):**  $u_t = F_{Y_t}(y_t)$  donde  $F_{Y_t}$  es la CDF predictiva

Si el modelo está bien calibrado, los valores PIT siguen una distribución uniforme en  $[0,1]$ .

## 1.7 Selección de Estructura Óptima

### 1.7.1 Grid de Estructuras

Evaluamos sistemáticamente:

| Modelo | Nivel       | Tendencia   | Estacionalidad | Parámetros libres  |
|--------|-------------|-------------|----------------|--|
| LL     | Estocástico | No          | Opcional       | $\sigma_{\mu}^2, \sigma_{\epsilon}^2$                    |
| LLT    | Estocástico | Estocástica | Opcional       | $\sigma_{\mu}^2, \sigma_{\delta}^2, \sigma_{\epsilon}^2$ |

### 1.7.2 Criterio de Selección

La estructura óptima se selecciona mediante:

1. **Primario:** Mayor  $\Delta\text{ELPD}$  promedio across folds
2. **Secundario:** Mayor support (proporción de folds ganadores)
3. **Terciario:** Mayor  $\Delta\text{RMSE}$  (reducción de error)

Esta jerarquía prioriza capacidad predictiva probabilística sobre ajuste puntual.

## 1.8 Criterio de Victoria y Estabilidad

### 1.8.1 Victoria por Fold

Un modelo completo “gana” en el fold  $f$  si:

$$\begin{cases} \Delta\text{ELPD}_f > 0 & \text{(mejor densidad predictiva)} \\ \Delta\text{RMSE}_f > 0 & \text{(menor error, i.e., } \text{RMSE}_{\text{base}} - \text{RMSE}_{\text{full}} > 0) \end{cases}$$

Nota: A diferencia del GLM-AR(1), aquí  $\Delta\text{RMSE} = \text{RMSE}_{\text{base}} - \text{RMSE}_{\text{full}}$ , por lo que valores positivos indican mejora.

### 1.8.2 Métricas de Estabilidad

- **Support:**  $\frac{\text{wins}}{\text{folds}}$  = proporción de folds victoriosos
- **Umbral estricto:** support  $\geq 0.70$  con mínimo 5 folds
- **Umbral moderado:** support  $\geq 0.60$  con mínimo 5 folds

## 1.9 Implementación Computacional y Optimizaciones

### 1.9.1 Arquitectura de Procesamiento

1. **Paralelización por pares:** Los 84 pares se procesan secuencialmente
2. **Eficiencia MCMC:** Una sola cadena por modelo (trade-off velocidad vs diagnósticos)
3. **Caching de matrices:** Reutilización de descomposiciones en Kalman filter

## 1.9.2 Gestión de Casos Especiales

El código incluye manejo robusto de:

- **Varianza nula:** Skip del fold si  $sd(y) \approx 0$
- **Fallas de convergencia:** Try-catch con mensajes informativos
- **Matrices singulares:** Regularización automática en Kalman filter
- **Predicciones degeneradas:** Fallback a intervalos basados en cuantiles

## 1.9.3 Estructura de Datos de Salida

Cada par genera:

```
list(  
  best_summary = tibble(      # Resumen del mejor modelo  
    spec, folds, wins, support,  
    dELPD_mean, dRMSE_mean, ...  
  ),  
  best_results = tibble(      # Detalles por fold  
    fold, n_train, n_test,  
    ELPD_base, ELPD_full, dELPD,  
    RMSE_base, RMSE_full, dRMSE,  
    cover80, cover95, pit, ...  
  ),  
  all_summaries = tibble()    # Todas las estructuras probadas  
)
```

## 1.10 Análisis Comparativo con Metodologías Previas

### 1.10.1 Tabla Comparativa de Enfoques

| Aspecto                       | ECM-MARS             | GLM-AR(1)             | BSTS                             |
|-------------------------------|----------------------|-----------------------|----------------------------------|
| <b>Paradigma</b>              | Frecuentista híbrido | Bayesiano paramétrico | Bayesiano estructural            |
| <b>Pre-requisitos</b>         | I(1), cointegración  | Ninguno               | Ninguno                          |
| <b>No-linealidad</b>          | MARS (splines)       | No                    | No (pero component es flexibles) |
| <b>Selección variables</b>    | Manual (pre-tests)   | No                    | Automática (spike-slab)          |
| <b>Componentes temporales</b> | ECM únicamente       | AR(1) global          | Nivel, tendencia, estacionalidad |
| <b>Incertidumbre</b>          | Bootstrap implícito  | Posterior completa    | Posterior completa +             |



| Aspecto                    | ECM-MARS         | GLM-AR(1) | BSTS                                |
|----------------------------|------------------|-----------|-------------------------------------|
| <b>Interpretabilidad</b>   | Alta (económica) | Media     | estados<br>Alta<br>(descomposición) |
| <b>Costo computacional</b> | Medio            | Alto      | Muy alto                            |

### 1.10.2 Ventajas Distintivas de BSTS

1. **Descomposición interpretable:** Separa señal de ruido con componentes económicamente significativos
2. **Selección automática:** Spike-and-slab identifica predictores relevantes sin pre-especificación
3. **Manejo de incertidumbre estructural:** Cuantifica incertidumbre en componentes no observables
4. **Robustez a especificación:** No requiere supuestos sobre orden de integración o cointegración
5. **Pronósticos por intervalo:** Intervalos de predicción naturales y bien calibrados

### 1.10.3 Limitaciones Relativas

1. **Costo computacional elevado:** FFBS + SSVS es intensivo, especialmente con muchos regresores
2. **Linealidad en predictores:** No captura interacciones no lineales como MARS
3. **Diagnósticos complejos:** Convergencia más difícil de verificar que MCO
4. **Sensibilidad a priors en muestras pequeñas:** Especialmente para varianzas de componentes

## 1.11 Extensiones y Desarrollos Futuros

### 1.11.1 Mejoras Inmediatas

1. **Componentes adicionales:**
  - **Efectos calendario:** Días hábiles, feriados móviles
  - **Cambios de nivel:** Detección automática de quiebres estructurales
  - **Regresores dinámicos:** Coeficientes variantes en el tiempo
2. **Priors informativos:**
  - Incorporar información económica en priors de inclusión

- Priors jerárquicos para grupos de variables relacionadas

### 3. Validación mejorada:

- Cross-validation estocástica (Pareto smoothed importance sampling)
- Backtesting con múltiples horizontes

## 1.11.2 Extensiones Metodológicas

### 1. No-linealidad:

- Gaussian processes para componentes no lineales
- Redes neuronales bayesianas para captura de interacciones

### 2. Modelos jerárquicos:

- Pooling parcial entre pares relacionados
- Factores latentes compartidos

### 3. Causalidad:

- Modelos de intervención con inferencia causal
- Grafos causales dinámicos

## 1.12 Pseudocódigo del Pipeline Completo

ENTRADA: DataFrame con series temporales, especificación de variables

SALIDA: Ranking de pares por capacidad predictiva y estabilidad

# PREPARACIÓN

1. CARGAR datos y limpiar nombres
2. IDENTIFICAR 6 variables circulación, 7 producción
3. GENERAR 84 pares ( $2 \times 6 \times 7$ )

# PROCESAMIENTO POR PAR

PARA cada par (Y, X):

4. CONSTRUIR matriz con Y, X, lags(X, 1:6)
5. ELIMINAR observaciones con NA

# TUNING DE ESTRUCTURA

PARA cada estructura en {LL, LLT}:

# LEAVE-FUTURE-OUT

6. GENERAR splits LFO (init=80%, h=6, step=6)

PARA cada fold:

7. DIVIDIR train/test
8. ESCALAR regresores con stats del train

# MODELOS

9. AJUSTAR modelo\_base:
  - state\_spec = estructura sin regresores
  - MCMC con 2000 iter, 500 burn-in

```

10. AJUSTAR modelo_full:
    - state_spec = estructura
    - prior = spike_slab(X_lags, expected_size=5)
    - MCMC con 2000 iter, 500 burn-in

# PREDICCIÓN
11. GENERAR distribuciones predictivas (h=6)
12. EXTRAER media, sd, cuantiles

# EVALUACIÓN
13. CALCULAR:
    - ELPD_base, ELPD_full → ΔELPD
    - RMSE_base, RMSE_full → ΔRMSE
    - Coverage 80%, 95%
    - PIT values

14. DETERMINAR victoria: (ΔELPD > 0) ∧ (ΔRMSE > 0)

15. RESUMIR estructura:
    - support = wins/folds
    - promedios de métricas

16. SELECCIONAR mejor estructura por ΔELPD
17. GUARDAR resultados del par

# RANKING Y EXPORT
18. ORDENAR pares por (support DESC, ΔELPD DESC, ΔRMSE DESC)
19. FILTRAR ganadores por umbrales (0.70, 0.60)
20. EXPORTAR CSVs y visualizaciones
21. GENERAR gráficos del último fold (sin re-ajustar)

```

## 1.13 Conclusiones Metodológicas

La metodología BSTS representa el enfoque más comprehensivo y flexible de los tres implementados, combinando la rigurosidad de la inferencia bayesiana con la interpretabilidad de la descomposición estructural. A diferencia del ECM-MARS que requiere validación previa de supuestos econométricos estrictos, y del GLM-AR(1) que modela solo dependencia serial global, BSTS descompone la serie en componentes interpretables mientras selecciona automáticamente predictores relevantes.

Los resultados empíricos sugieren que BSTS identifica relaciones predictivas robustas con alta precisión, aunque a un costo computacional sustancialmente mayor. La capacidad de cuantificar incertidumbre en todos los niveles —desde componentes estructurales hasta predicciones— lo hace particularmente valioso para aplicaciones donde la evaluación de riesgo es crítica.

La implementación con horizonte de 6 meses (vs 12 en otras metodologías) proporciona evaluación más granular de la estabilidad temporal, revelando que muchas relaciones que parecen estables en horizontes largos muestran variabilidad en escalas más cortas. Esto sugiere la importancia de considerar múltiples horizontes en la evaluación de modelos predictivos.

El framework BSTS es especialmente apropiado cuando: - La interpretación de componentes es prioritaria - Existe incertidumbre sobre qué predictores incluir - Se requieren intervalos de predicción bien calibrados - Los datos exhiben múltiples fuentes de variación (tendencia, estacionalidad, regresores)

La combinación de selección automática de variables con modelado estructural posiciona a BSTS como un puente entre los métodos econométricos clásicos y las técnicas modernas de machine learning, manteniendo interpretabilidad mientras abraza la complejidad inherente en los datos económicos.

---