

Detalles Metodológicos del Modelo GLM Bayesiano con Estructura AR(1)

José Mauricio Gómez Julián

Table of Contents

1	Marco Bayesiano para Modelado de Series Temporales con Estructura Autorregresiva	2
1.1	Especificación del Modelo Bayesiano con AR(1)	2
1.1.1	Estructura del Modelo Completo	2
1.1.2	Modelo Base de Referencia	3
1.1.3	Especificación de Priors	3
1.2	Validación Cruzada Temporal con Leave-Future-Out (LFO)	3
1.2.1	Implementación del Rolling-Origin con Ventana Deslizante	3
1.2.2	Criterios de Evaluación Predictiva	4
1.2.3	Criterio de Victoria por Fold	5
1.3	Inferencia Bayesiana con Hamiltonian Monte Carlo	5
1.3.1	Configuración del Muestreador NUTS	5
1.3.2	Diagnósticos de Convergencia	6
1.4	Determinación de Estabilidad Temporal y Soporte	6
1.4.1	Métrica de Soporte	6
1.4.2	Umbrales de Robustez	6
1.4.3	Interpretación del Soporte	6
1.5	Análisis Comparativo con Metodología ECM-MARS	7
1.5.1	Diferencias Fundamentales	7
1.5.2	Ventajas del Enfoque Bayesiano	7
1.5.3	Limitaciones Relativas	7
1.6	Implementación Técnica y Optimizaciones	8
1.6.1	Gestión de Memoria y Paralelización	8
1.6.2	Manejo de Casos Degenerados	8
1.7	Pseudocódigo del Pipeline Completo	8
1.8	Notas Técnicas y Consideraciones Especiales	9

1.8.1	Escalamiento y Estabilidad Numérica	9
1.8.2	Interpretación de Métricas Negativas	9
1.8.3	Extensiones Potenciales	9
1.9	Conclusiones Metodológicas	9

1 Marco Bayesiano para Modelado de Series Temporales con Estructura Autorregresiva

El presente documento detalla la metodología implementada para el análisis de relaciones causales entre variables de producción y circulación mediante un Modelo Lineal Generalizado Bayesiano (BGLM) con estructura autorregresiva de primer orden AR(1). Este enfoque difiere sustancialmente de la metodología ECM-MARS al adoptar un paradigma inferencial bayesiano completo, permitiendo cuantificar la incertidumbre predictiva de manera integral y comparar modelos mediante criterios de información predictiva.

1.1 Especificación del Modelo Bayesiano con AR(1)

1.1.1 Estructura del Modelo Completo

Sea Y_t la variable dependiente y X_t la variable independiente en el tiempo t . El modelo completo se especifica como:

$$Y_{t,s} = \alpha + \beta_0 \cdot t_s + \sum_{i=1}^L \beta_i X_{t-i,s} + \epsilon_t$$

donde:

- $Y_{t,s}$ es la variable dependiente estandarizada: $(Y_t - \mu_Y)/\sigma_Y$
- t_s es el índice temporal estandarizado: $(t - \mu_t)/\sigma_t$
- $X_{t-i,s}$ son los rezagos estandarizados de la variable independiente
- L es el número máximo de rezagos (por defecto $L = 3$)
- ϵ_t sigue un proceso AR(1): $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$, con $\eta_t \sim N(0, \sigma^2)$

La estandarización es crítica para la eficiencia del algoritmo NUTS¹ (No-U-Turn Sampler), garantizando que todos los parámetros operen en escalas comparables y evitando problemas de convergencia en el muestreador Hamiltoniano.

¹ NUTS (No-U-Turn Sampler) es una extensión adaptativa del algoritmo Hamiltonian Monte Carlo que ajusta automáticamente la longitud de las trayectorias para

1.1.2 Modelo Base de Referencia

Como benchmark de comparación, se estima un modelo base que excluye la información de la variable independiente:

$$Y_{t,s} = \alpha + \beta_0 \cdot t_s + \epsilon_t$$

Este modelo captura únicamente la tendencia temporal y la estructura autorregresiva intrínseca de Y , sirviendo como hipótesis nula efectiva contra la cual evaluar la contribución predictiva de los rezagos de X .

1.1.3 Especificación de Priors

Los priors se especifican en la escala estandarizada para mantener consistencia:

- **Coefficientes de regresión** (β_i): $\beta_i \sim N(0,1)$
- **Intercepto** (α): $\alpha \sim t_3(0,2.5)$ (Student-t con 3 grados de libertad)
- **Desviación estándar residual** (σ): $\sigma \sim \text{Exponencial}(1)$
- **Coefficiente AR(1)** (ϕ): Prior implícito uniforme en $(-1,1)$ para estacionariedad

La elección de priors débilmente informativos² refleja un balance entre regularización suave y flexibilidad para aprender de los datos. La distribución Student-t para el intercepto proporciona robustez ante valores atípicos, mientras que el prior exponencial para σ garantiza positividad con cola pesada.

1.2 Validación Cruzada Temporal con Leave-Future-Out (LFO)

1.2.1 Implementación del Rolling-Origin con Ventana Deslizante

A diferencia de la validación cruzada tradicional que viola el principio de causalidad temporal, implementamos Leave-Future-Out (LFO) con ventana deslizante (*sliding window*). Este esquema respeta estrictamente el ordenamiento temporal y simula el contexto real de pronóstico prospectivo.

El procedimiento se estructura así:

1. **Inicialización:** El conjunto de entrenamiento inicial comprende el 70% de los datos o un mínimo de 90 observaciones, lo que sea mayor.

maximizar la eficiencia del muestreo sin requerir tuning manual del número de pasos leapfrog.

² Los priors débilmente informativos proporcionan regularización suave sin imponer creencias fuertes a priori. En el límite de datos abundantes, su influencia se desvanece, recuperando estimaciones similares a máxima verosimilitud.

2. **Horizonte de prueba:** Cada fold evalúa $h = 12$ meses hacia adelante, simulando pronósticos anuales.
3. **Paso entre folds:** El origen se desplaza 12 meses entre evaluaciones sucesivas.
4. **Ventana deslizante:** El tamaño del conjunto de entrenamiento permanece constante³, descartando observaciones antiguas al incorporar nuevas. Esto contrasta con la ventana expansiva donde el conjunto de entrenamiento crece monótonicamente.

La ventaja de la ventana deslizante radica en su sensibilidad a cambios de régimen y su capacidad de adaptación a dinámicas no estacionarias en sentido amplio. Mientras que una ventana expansiva asume parámetros constantes a lo largo de toda la historia, la ventana deslizante reconoce que las relaciones económicas pueden evolucionar, manteniendo solo la “memoria reciente” más relevante para el pronóstico.

1.2.2 Criterios de Evaluación Predictiva

Para cada fold k , comparamos el desempeño predictivo del modelo completo contra el modelo base mediante dos criterios complementarios:

1.2.2.1 Expected Log Predictive Density (ELPD)

El ELPD⁴ cuantifica la densidad predictiva logarítmica esperada para nuevas observaciones:

$$\text{ELPD}_k = \sum_{t \in \text{test}_k} \log \int p(y_t | \theta) p(\theta | \text{datos}_{\text{train},k}) d\theta$$

En la práctica, aproximamos esta integral mediante el estimador log-sum-exp sobre las muestras posteriores:

$$\widehat{\text{ELPD}}_k = \sum_{t \in \text{test}_k} \log \left(\frac{1}{S} \sum_{s=1}^S p(y_t | \theta^{(s)}) \right)$$

³ El tamaño constante del conjunto de entrenamiento en sliding window garantiza que todos los folds tengan poder estadístico comparable, evitando el sesgo hacia folds tardíos que ocurre con ventanas expansivas.

⁴ ELPD (Expected Log Predictive Density) es el criterio de evaluación predictiva fundamental en estadística bayesiana, generalizado por WAIC y LOO-CV. Mide la capacidad del modelo para asignar alta probabilidad a nuevas observaciones.

donde $\theta^{(s)}$ son las muestras posteriores y S es el número de iteraciones post-warmup.

La diferencia $\Delta\text{ELPD}_k = \text{ELPD}_{k,\text{full}} - \text{ELPD}_{k,\text{base}}$ mide la ganancia en capacidad predictiva. Un $\Delta\text{ELPD} > 0$ indica que el modelo completo asigna mayor probabilidad a los datos observados fuera de muestra.

1.2.2.2 Root Mean Square Error (RMSE) y Métricas Clásicas

Complementamos el ELPD con métricas tradicionales en la escala original de las variables:

- **RMSE:** $\sqrt{\frac{1}{n_{\text{test}}} \sum_t (y_t - \hat{y}_t)^2}$
- **MAE:** $\frac{1}{n_{\text{test}}} \sum_t |y_t - \hat{y}_t|$
- **sMAPE:** $\frac{100}{n_{\text{test}}} \sum_t \frac{2|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$
- **R²:** $1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}$ (protegido contra $SST \approx 0$)

donde \hat{y}_t es la media posterior de las predicciones, obtenida mediante `posterior_epred()` y transformada de vuelta a la escala original.

1.2.3 Criterio de Victoria por Fold

Un modelo completo “gana” en el fold k si y solo si:

$$\begin{cases} \Delta\text{ELPD}_k > 0 & \text{(mejor densidad predictiva)} \\ \Delta\text{RMSE}_k < 0 & \text{(menor error cuadrático)} \end{cases}$$

Este criterio dual exige superioridad tanto en términos probabilísticos (ELPD) como determinísticos (RMSE), evitando victorias espurias por mejoras en una sola dimensión.

1.3 Inferencia Bayesiana con Hamiltonian Monte Carlo

1.3.1 Configuración del Muestreador NUTS

La inferencia se realiza mediante el algoritmo NUTS⁵ implementado en Stan a través de `brms` con backend `cmdstanr`. Los parámetros de muestreo se calibran para garantizar convergencia robusta:

- **Cadenas:** 4 cadenas independientes

⁵ El algoritmo NUTS elimina la necesidad de ajustar manualmente el número de pasos en HMC mediante un criterio de parada basado en “U-turns” en el espacio de parámetros, donde la trayectoria comienza a retroceder hacia su origen.

- **Iteraciones totales:** 1500 por cadena
- **Warmup:** 750 iteraciones (50% para adaptación)
- **Adapt delta:** 0.95 (reduce pasos divergentes)
- **Max treedepth:** 12 (evita truncamiento prematuro del árbol)

La paralelización se implementa a nivel de cadenas (`parallel_chains = 4`), no dentro de cada cadena, para mantener la eficiencia del muestreador sin sobresuscripción del CPU.

1.3.2 Diagnósticos de Convergencia

Aunque no se reportan explícitamente en el código de producción por eficiencia, los diagnósticos estándar incluyen:

- **\hat{R} (Rhat):** Debe ser < 1.01 para todos los parámetros
- **ESS (Effective Sample Size):** Bulk-ESS y Tail-ESS > 400 por cadena
- **Divergencias:** Idealmente 0; pocas divergencias ($< 1\%$) pueden ser tolerables
- **BFMI (Bayesian Fraction of Missing Information):** > 0.2 indica buen comportamiento

1.4 Determinación de Estabilidad Temporal y Soporte

1.4.1 Métrica de Soporte

Definimos el soporte como:

$$\text{support} = \frac{\text{folds_pass}}{\text{folds}}$$

donde `folds_pass` cuenta los folds donde el modelo completo satisface el criterio dual de victoria. Esta métrica cuantifica la consistencia temporal del poder predictivo.

1.4.2 Umbrales de Robustez

Establecemos dos niveles de confianza:

- **Umbral estricto:** $\text{support} \geq 0.70$ con mínimo 5 folds válidos
- **Umbral moderado:** $\text{support} \geq 0.60$ con mínimo 5 folds válidos

El requisito de un mínimo absoluto de folds previene falsos positivos por denominadores pequeños (e.g., $1/1 = 100\%$ support no constituye evidencia robusta).

1.4.3 Interpretación del Soporte

Un soporte alto indica que la relación predictiva persiste a través de diferentes regímenes temporales. Valores de soporte:

- **0.80-1.00:** Relación muy estable, robusta a cambios de régimen
- **0.60-0.79:** Relación moderadamente estable, sensible a algunas épocas
- **0.40-0.59:** Relación inestable, dependiente del contexto temporal
- **< 0.40:** Relación espuria o altamente específica a períodos particulares

1.5 Análisis Comparativo con Metodología ECM-MARS

1.5.1 Diferencias Fundamentales

Aspecto	ECM-MARS	GLM Bayesiano AR(1)
Paradigma	Frecuentista con tests secuenciales	Bayesiano integral
Cointegración	Requisito explícito (Engle-Granger/Johansen)	Implícita en estructura AR(1)
No-linealidad	MARS con splines adaptativos	Lineal en parámetros
Incertidumbre	Errores estándar HAC	Distribución posterior completa
Comparación	Tests de significancia	ELPD + métricas predictivas
Complejidad	Alta (múltiples pre-tests)	Moderada (modelo directo)

1.5.2 Ventajas del Enfoque Bayesiano

1. **Cuantificación integral de incertidumbre:** Las distribuciones posteriores capturan toda la incertidumbre paramétrica, propagándola naturalmente a las predicciones.
2. **Comparación directa de modelos:** El ELPD proporciona una métrica unificada para comparar modelos no anidados sin ajustes por multiplicidad.
3. **Robustez a violaciones de supuestos:** Los priors regularizadores y la estructura AR(1) acomodan desviaciones moderadas de los supuestos clásicos.
4. **Interpretabilidad:** Los intervalos de credibilidad tienen interpretación probabilística directa, a diferencia de los intervalos de confianza frecuentistas.

1.5.3 Limitaciones Relativas

1. **Costo computacional:** El muestreo MCMC es órdenes de magnitud más lento que la estimación por MCO/MARS.

2. **Sensibilidad a priors:** En muestras pequeñas, la elección de priors puede influir sustancialmente en los resultados.
3. **Linealidad:** La ausencia de términos no lineales (splines) puede limitar la flexibilidad para capturar relaciones complejas.

1.6 Implementación Técnica y Optimizaciones

1.6.1 Gestión de Memoria y Paralelización

- **Paralelización por pares:** Cada par $(X \rightarrow Y)$ se procesa secuencialmente, pero las cadenas MCMC dentro de cada modelo se paralelizan.
- **Liberación de memoria:** Los objetos de Stan se liberan implícitamente tras cada fold mediante el recolector de basura de R.
- **Backend cmdstanr:** Más eficiente que rstan para modelos grandes, con mejor gestión de memoria y diagnósticos.

1.6.2 Manejo de Casos Degenerados

El código incluye salvaguardas para:

- **Varianza casi nula:** Si $\sigma_Y \approx 0$ en el conjunto de entrenamiento, el fold se omite.
- **Errores de convergencia:** Los modelos que fallan en converger retornan NULL y se excluyen del análisis.
- **Log-verosimilitudes indefinidas:** Se verifican valores finitos antes de calcular ELPD.

1.7 Pseudocódigo del Pipeline Completo

PARA cada par (X, Y) en $\{\text{producción} \times \text{circulación}\} \times \{2 \text{ direcciones}\}$:

1. Construir matriz con Y , X , y $\text{lags}(X, 1:L)$
2. Eliminar filas con valores faltantes

PARA cada fold en $\text{rolling_splits}(\text{sliding_window})$:

3. Dividir en train/test
4. Estandarizar variables usando estadísticos del train
5. Ajustar modelo_base: $Y_s \sim 1 + t_s + \text{AR}(1)$
6. Ajustar modelo_full: $Y_s \sim 1 + t_s + X_{\text{lags}} + \text{AR}(1)$
7. Calcular ELPD para ambos modelos en test
8. Generar predicciones puntuales (posterior_epred)
9. Transformar predicciones a escala original
10. Computar métricas (RMSE, MAE, sMAPE, R^2)
11. Determinar victoria: $(\Delta \text{ELPD} > 0) \wedge (\Delta \text{RMSE} < 0)$

12. Calcular $\text{support} = \text{wins} / \text{total_folds}$
13. Promediar métricas across folds

14. Ranking por (support DESC, Δ ELPD DESC, Δ RMSE ASC)
15. Filtrar ganadores por umbrales de support
16. Exportar resultados y visualizaciones

1.8 Notas Técnicas y Consideraciones Especiales

1.8.1 Escalamiento y Estabilidad Numérica

El escalamiento por fold (no global) es crucial porque:

1. Las estadísticas de normalización deben calcularse solo con datos del train para evitar *data leakage*⁶
2. La escala de las variables puede cambiar sustancialmente entre ventanas temporales
3. NUTS requiere que todos los parámetros operen en escalas $\mathcal{O}(1)$ para eficiencia

1.8.2 Interpretación de Métricas Negativas

Es posible obtener $R^2 < 0$ en evaluación fuera de muestra cuando el modelo predice peor que la media muestral. Esto no indica error de cálculo sino genuino mal desempeño predictivo. De manera similar, ratios RMSE > 1 indican que el modelo completo empeora las predicciones respecto al baseline.

1.8.3 Extensiones Potenciales

El framework actual admite varias extensiones naturales:

- **Heterocedasticidad:** Modelar σ_t como función del tiempo o covariables
- **Mezcla de expertos:** Múltiples componentes AR con pesos variables
- **No-linealidad:** Splines o funciones base mediante $s(\cdot)$ en brms
- **Efectos aleatorios:** Estructura jerárquica para múltiples series
- **Pronóstico multi-paso:** Extender más allá de one-step-ahead

1.9 Conclusiones Metodológicas

La metodología GLM Bayesiana con AR(1) ofrece un marco robusto y principiado para evaluar relaciones predictivas entre variables económicas. A diferencia del enfoque ECM-MARS que requiere validación secuencial de múltiples supuestos

⁶ Data leakage ocurre cuando información del conjunto de prueba contamina el proceso de entrenamiento, inflando artificialmente las métricas de desempeño. El escalamiento con estadísticas globales constituiría una forma sutil pero perniciosa de leakage.

(estacionariedad, cointegración, velocidad de ajuste), el modelo bayesiano integra toda la incertidumbre en un framework unificado.

Los resultados sugieren que las relaciones más robustas (support > 0.70) corresponden típicamente a pares donde la teoría económica predice causalidad fuerte, mientras que relaciones con support moderado (0.60-0.70) pueden reflejar canales de transmisión indirectos o dependientes del régimen económico vigente.

La evaluación dual mediante ELPD y RMSE garantiza que los modelos seleccionados no solo ajusten bien en términos de error cuadrático sino que también capturen adecuadamente la estructura probabilística de los datos. Esta doble exigencia filtra efectivamente relaciones espurias que podrían parecer prometedoras bajo un solo criterio.