Series Temporales Bayesianas para la Proyección de Poblaciones Rurales

Bayesian Time Series Models for Rural Population Forecasting

Daniel Freire-Navarrete[1],
Oscar Llano-Moscuy[2] y Byron Morocho-Caisatoa[2]

[1]Universidad Nacional de Chimborazo
Av. Antonio José de Sucre kilómetro 1 1/2 vía a Guano, Riobamba, Ecuador
[2]Universidad Central del Ecuador
Av. América y Av. Universitaria, Quito, Ecuador
*ollano793@gmail.com

PALABRAS CLAVE:

Demografía, series temporales, estadística bayesiana

RESUMEN:

Estimar las variaciones en un horizonte de tiempo "h" para el número de personas en una población, ha sido un problema recurrente en los últimos años. El presente trabajo propone una solución al considerar la serie temporal como un proceso estocástico, evaluando su comportamiento con la información presente en su propia estructura mediante modelos del tipo autorregresivo, eliminando la necesidad de interacciones con variables exógenas de difícil recolección, especialmente en poblaciones rurales. Junto con el uso del enfoque bayesiano para la estimación de los parámetros como alternativa a la falta de ajuste, de esta manera las predicciones conseguidas poseerán una robustez considerable.

KEYWORDS:

Mathematics, demography, time series, Bayesian statistics

ABSTRACT:

Estimating variations over a time horizon "h" in the number of individuals within a population has been a recurrent challenge in recent years. This study proposes a solution by modeling the time series as a stochastic process and evaluating its behavior based on the information contained in its own structure through autoregressive models. This approach eliminates the need for interactions with exogenous variables, which are often difficult to collect—especially in rural populations. Furthermore, by employing a Bayesian framework for parameter estimation as an alternative to traditional model fitting, the resulting forecasts exhibit considerable robustness.

• Recibido: 2 de febrero de 2026 • Aceptado: 24 de abril de 2026 • Publicado en línea: 2 de octubre de 2026

1. INTRODUCCIÓN

La elaboración de proyecciones poblacionales ha sido un problema abordado mediante diferentes enfoques provenientes en su mayoría de las ciencias exactas, sin embargo, muchos de estos resultados no han sufrido cambios considerables desde su concepción. Un ejemplo sería el método de componentes demográficos por cohortes (MCDC), el cual desde los años 50 ha sido el procedimiento dominante para estimar – proyectar una población al estar basado en el principio de la ecuación compensadora, resaltando la necesidad de información externa con un nivel de confiabilidad elevado para conseguir resultados robustos [1].

Con el paso de los años, diferentes modificaciones a este modelo han surgido en razón de las características predominantes en determinadas poblaciones, mismas que se

Copyright:© 2026 por los autores
Creative Commons Atribución 4.0 CC-BY

enfocan principalmente en adaptaciones a la forma de estimación, pasando del tipo determinísticas a probabilísticas o también llamadas estocásticas, sin embargo, a pesar de estos avances los resultados alcanzados presentan problemáticas y limitaciones considerables, producto de estar sujetas por supuestos de compleja verificación, así como variables de difícil recolección; el lector podrá encontrar comparativas con múltiples enfoques en textos como [2].

Ante estos limitantes, nuevos enfoques han empezado a tomar relevancia gracias al desarrollo de los avances computacionales, un primer acercamiento lo presenta Blaconá Teresa Maria, en el cual consideran a cada punto en una serie temporal como una realización de un proceso estocástico, donde el modelamiento se realiza solo con la información de la estructura interna de la serie, eliminando la necesidad de incorporar información adicional [3]. Característica crucial para casos en los que acceder a datos sobre la mecánica demográfica para una población se convierte en un limitante significativo.

Estos métodos se contrastan de forma significativa con los enfoques tradicionales, al permitir realizar tanto estimaciones puntuales como intervalos de confianza para un valor futuro desconocido, sin la necesidad de grandes registros de información histórica. A pesar de ello, su eficiencia está fuertemente vinculada a una construcción teórica matemática, que en algunos casos limita su aplicabilidad reflejándose en una falta de ajuste. Este comportamiento ha sido explorado por múltiples autores donde se evidencia una evolución paulatina para estos modelos gracias a la incorporación de nuevos conceptos, permitiendo profundizar en modificaciones significativas [3].

Andrade & Barreto desarrollaron una de estas adaptaciones, analizando la robustez y las estimaciones obtenidas mediante un tipo de estos modelos descritos, los llamados autorregresivos. En donde sus propiedades se ven potenciadas al utilizar un enfoque bayesiano en la estimación de los parámetros, alterando el esquema frecuentista al considerar cada uno de estos una variable aleatoria con su propia forma y comportamiento, reflejando la información previa proveniente de estudios anteriores dando un protagonismo importante a el entorno de los datos sin la necesidad de una incorporación directa en forma de variable [4].

En el presente artículo se propone una modificación al enfoque tradicional mediante la reformulación de la verosimilitud utilizada en la representación de los datos, y la implementación de algoritmos de Cadenas de Markov Monte Carlo para el muestreo de la distribución posterior, considerando su naturaleza no cerrada. A través de estas adaptaciones, se busca mejorar la precisión de proyecciones en contextos donde la información es limitada, pero existe conocimiento relevante del entorno, como ocurre en poblaciones rurales.

2. MÉTODO DE INVESTIGACIÓN

La presente investigación se esquematiza bajo un diseño cuantitativo, elección fundamentada por el uso de los diferentes métodos estadísticos tanto del tipo descriptivo e inferencial en las etapas de estimación y validación del modelo propuesto. Este proceso será abordado mediante un enfoque bayesiano, ofreciendo una particular ventaja al incorporar el conocimiento de los patrones, tendencias y demás comportamientos presentes en el contexto propio de la serie temporal, mismos que son reflejados mediante distribuciones de probabilidad conocidas como previas y cuyo entendimiento necesita de un desarrollo particular [5].

Este enfoque al ser una extensión de la teoría frecuentista tradicional necesitará de una especificación de conceptos clave sin

abarcar una descripción profunda. El componente bayesiano será desarrollado de forma secuencial, empezando por la verosimilitud empleada para representar los datos, la justificación de las distribuciones previas y finalmente el método de muestreo de la distribución posterior encontrada con estos componentes, mediante el desarrollo descrito se espera una comprensión adecuada de los conceptos y resultados obtenidos por parte del lector.

2.1. Series temporales bajo la metodología "Box Jenkins"

Un primer acercamiento a su estudio formal fue presentado y analizado por el autor Temponi de Oliveira definiéndolas como un conjunto de observaciones para una variable aleatoria que se encuentra indexada en el tiempo, entendiéndola como la realización de un proceso estocástico [6]. Generalmente son denotas por $Z_t(\omega)$ donde ω representa su naturaleza aleatoria al ser un elemento del espacio de probabilidad (Ω, F, P) y el subíndice $(t \in N)$ simboliza el punto o periodo de tiempo específico. Para el desarrollo del presente artículo se simbolizará por $\{Z_1, Z_2, \ldots,\}$ o $\{Z_t, t \in N\}$ a la serie temporal en cuestión, siendo Z_t la observación de la serie en el instante *t*=1,2, ...

Bajo la metodología "Box Jenkins" se ajusta un modelo para explicar un valor Z_t, mediante un grupo de observaciones $Z=\{Z_{t-1}, Z_{t-2}, \ldots\}$ donde la$cov(Z_i, Z_j) \neq 0 \; \forall \, i \neq j$ al ser elementos de la propia serie. Al conjunto Z se lo conoce como rezagos, mismos que junto con errores no observables, variables exógenas, variables estacionales entre otras facilitan la representación [7]. El presente artículo hará uso de los modelos autorregresivos mismos que expresan a un valor de la serie de tiempo como una función lineal de sus valores pasados, cuya comprensión en profundidad puede encontrarse en el primer y tercer capítulo de [2].

2.2. Estimación frecuentista: Modelos AR(p)

Este modelo centra su construcción y posterior validación en 2 ejes principales, el supuesto que una colección de variables Z son originadas por un proceso estocástico del tipo estacionario en el sentido débil, lo cual aprovecha su estructura y propiedades particulares para una representación eficiente. El segundo es una serie de restricciones sobre el polinomio característico asociado a este tipo de modelos, imponiendo a sus raíces complejas un módulo mayor a uno asegurando la invertibilidad propia de los modelos lineales estacionarios y generando restricciones para los coeficientes respectivos, asegurando un comportamiento estable ante perturbaciones o shocks marcados en el periodo de tiempo considerado [3].

Matemáticamente, sea $\{\tilde{Z}_t, t \in N\}$ un proceso estacionario no centrado usado como elemento base para un modelo lineal autorregresivo de orden "p", donde el valor presente del proceso $\tilde{Z}_t$ puede ser expresado como una combinación lineal de sus "p" valores pasados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \ldots, \tilde{Z}_{t-p}$ y un ruido blanco débil $\{a_t; t \in N\}$ considerado a este como proceso i.i.d. $N(0, \sigma^2)$, escrito de la siguiente manera:

$$\tilde{Z}_t = \mu + \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \ldots + \phi_p \tilde{Z}_{t-p} + a_t \qquad (1)$$

Donde μ representa un valor constante, esta expresión se reescribe usando el operador de retardo, obteniendo:

$$\phi(B)\tilde{Z}_t = \mu + a_t \qquad (2)$$

Con $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$, acorde a la literatura revisada este polinomio es crucial para la estabilidad del modelo al asegurar condiciones de estacionariedad e invertibilidad, resultado de poseer raíces con un módulo mayor a uno garantizando la ausencia de raíces unitarias que afecten las estimaciones, además esta restricción limita el rango de los coeficientes del modelo a $|\phi_i| < 1$

al estar asociados de forma directa con el valor de las mismas.

El enfoque frecuentista ofrece varias técnicas para la estimación de coeficientes y demás parámetros requeridos, estas han evolucionado de aproximaciones basadas en la ecuación de Yule Walker expuesta por Temponi de Oliveira, pasando por el uso de mínimos cuadrados generalizados (MCG) como una alternativa a mínimos cuadrados ordinarios (MCO) al no cumplir los supuestos requeridos dada la dependencia temporal [4]. Estas han evolucionado centrándose finalmente en los métodos enfocados en la maximización de la función de verosimilitud y sus variantes, cuya característica es la necesidad mínima de supuestos teóricos fuertes [5].

Uno de los factores determinantes para la eficiencia de los métodos mencionados en especial de los basados en la verosimilitud, se concentra en la selección del número de términos autorregresivos a emplear, lo cual es realizado en su mayoría mediante el análisis de los retardos significativos en correlogramas de las funciones de autocorrelación y autocorrelación parcial para este tipo de modelo, elección complementada mediante el uso de criterios de información basada en la verosimilitud del modelo planteado y el uso de la distancia o información de Kullback – Leibler, permitiendo una elección robusta sin la necesidad de complejos pasos intermedios [6].

2.3. Enfoque bayesiano para el modelado de Series Temporales

Es posible gracias al teorema de Bayes, cuya representación en términos de variables aleatorias y proporcionalidad está dada por:

$$p(\theta|y) \propto p(y|\theta) * p(\theta) \tag{3}$$

donde $p(y|\theta)$ simboliza la función de verosimilitud de los datos representando el comportamiento observado o conocimiento actual. Por otro lado $p(\theta)$ es llamada distribución previa, encargada de cuantificar la información disponible personificando el conocimiento anterior y cambiando la interpretación usual de valor especificado a una variable aleatoria. Finalmente $p(\theta|y)$se denomina distribución posterior reflejando la actualización del conocimiento previo reduciendo la incertidumbre inherente del objeto de estudio [7].

Los elementos mencionados anteriormente son expuestos en detalle por Rangel y Karakani Masoumi donde se resaltan los siguientes conceptos:

- **Distribución previa:** Objeto análisis particular, debido a que su elección determina posteriores con una forma funcional simple o no. Existen diferentes tipos de previas como lo son aquellas informativas, no informativas, propias, impropias, conjugadas y no conjugadas, diferenciándose en cómo cuantifican el conocimiento previo mediante los valores que determinan su forma llamados hiperparámetros. Para el análisis de modelos lineales se prioriza el uso del tipo informativo al poseer numerosas técnicas de control y paso del conocimiento existente al espacio de los parámetros involucrados [13].

- **Verosimilitud:** Definida en función a la naturaleza de los datos y motiva en muchas ocasiones el uso de previas especificas dado el dominio de sus parámetros. Representa la probabilidad de obtener la muestra de una variable dado un valor para los parámetros de la misma, considerando que dicha variable sigue una determinada distribución de probabilidad. En el modelamiento de series temporales este elemento es definido en el planteamiento del ruido blanco débil, fijándola como una distribución normal con varianza constante y media igual a la sumatoria de los rezagos multiplicados por los coeficientes [13].

- **Distribución Posterior:** Por definición significa la probabilidad de θ para tomar algún valor dentro de su espacio paramétrico dado los datos y, representando una actualización de las creencias iniciales y personifica el pilar de la inferencia bayesiana al permitir caracterizar a los parámetros buscados mediante sus propiedades. Su estimación directa en muchos casos suele no ser posible debido a formas funcionales complejas al usar diferentes distribuciones previas, ante esta problemática para modelos lineales se propone el uso de métodos de muestreo de la distribución objetivo mediante técnicas de Cadenas de Markov Monte Carlo.

Finalmente, el modelo planteado para las series temporales debe tener la capacidad de realizar inferencias sobre los posibles valores desconocidos para una variable, lo cual es realizado mediante modificaciones a la posterior conseguida a partir de un conjunto de datos y siendo posible predecir el valor para un conjunto desconocido $\tilde{y}$ obteniendo así la llamada distribución predictiva posterior cuya expresión es:

$$p(\tilde{y}|y)=\int p(\tilde{y},\theta|y)d\theta=\int p(\tilde{y}|\theta)\,p(\theta|y)d\theta \qquad (4)$$

La cual es obtenida aplicando diferentes teoremas y el supuesto de independencia condicional de y dado $\tilde{y}$, expresando como esta distribución es un promedio de las predicciones condicionales sobre la distribución posterior de θ. Presenta complejas dificultades para su cálculo debido a su forma funcional, por lo cual se han desarrollado técnicas a fin de aproximar sus resultados [8].

2.4. Estimación bayesiana de modelos AR(p)

Parte sobre una definición apropiada de los elementos que conforman la distribución posterior, donde la literatura revisada propone en su mayoría dos tipos de verosimilitud (aproximada y exacta) para este tipo de modelos con dependencia temporal, distinguiéndose en como abordan la problemática de las últimas "p" observaciones. El presente artículo hará uso de la verosimilitud "aproximada", cuya estructura permite usar diferentes tipos de distribuciones previas y su desarrollo puede ser considerado como una extensión a los planteamientos determinados por el enfoque frecuentista cuyo desarrollo es expuesto por [3].

La función de verosimilitud aproximada parte de una serie de tiempo $\{Z_t, t=1,2,..,N\}$ la cual se asume generada por un modelo autorregresivo centrado en 0 de orden "p":

$$Z_t=\phi_1 Z_{t-1}+\ldots+\phi_p Z_{t-p}+a_t \qquad (5)$$

Donde $\phi=(\phi_1,\phi_2,\ldots,\phi_p)\in R^p$ y $\{a_t, t=1,\ldots,N\}$ es ruido blanco débil denominado como errores o innovaciones y es distribuido según $N(0,\tau^{-1})$ y $\tau=1/\sigma_a{}^2>0$, despejando la expresión anterior se obtiene:

$$a_t=Z_t-\phi_1 Z_{t-1}-\ldots-\phi_p Z_{t-p} \qquad (6)$$

Donde los errores seguirán distribuyéndose normalmente solo desplazando su centro, por lo cual es factible escribir la densidad conjunta para las innovaciones subsiguientes $\{a_t, t=p+1,\ldots,N\}$ como:

$$f(a_{p+1},\ldots,a_N|\phi,\tau)\propto \tau^{\frac{N-p}{2}}\exp\left\{\frac{-\tau}{2}\sum_{t=p+1}^{N}a_t^2\right\} \qquad (7)$$

Si dividimos las "N" observaciones disponibles de la serie temporal en los conjuntos $Z_p=\{Z_1,\ldots,Z_p\}$ agrupando las "p" primeras observaciones y $Z=\{Z_{p+1},\ldots,Z_N\}$ como el resto de las observaciones. Es factible relacionar los ruidos $\{a_{p+1},\ldots,a_N\}$ con las observaciones Z de la siguiente forma:

$$a_{p+1}=Z_{p+1}-\phi_1 Z_{p+2}-\phi_2 Z_{p+3}-\ldots-\phi_N Z_{p+N} \qquad (8)$$

despejando:

$$Z_{p+1}=\phi_1 Z_{p+2}+\phi_2 Z_{p+3}+\ldots+\phi_N Z_{p+N}+a_{p+1} \quad (9)$$

Como posee un Jacobiano unitario, es posible reescribir la densidad de probabilidad en función del conjunto de observaciones Z, asumiendo como fijas las "p" primeras observaciones Z_p al hacer uso correcto de los índices "t", resultando:

$$f(Z,\phi,\tau)\propto\tau^{\frac{N-p}{2}} \cdot\exp\left\{\frac{-\tau}{2}\sum_{t=p+1}^{N}\left(Z_t-\phi_1 Z_{t-1}-\ldots-\phi_p Z_{t-p}\right)^2\right\} \quad (10)$$

cuya escritura se facilita definiendo las siguientes matrices:

$$Z=\begin{bmatrix}Z_{p+1}\\Z_{p+2}\\\vdots\\Z_N\end{bmatrix}, X=\begin{bmatrix}Z_1 & \cdots & Z_p\\Z_2 & \cdots & Z_{p+1}\\\vdots & \cdots & \vdots\\Z_{N-p} & \cdots & Z_N\end{bmatrix}, \Phi=\begin{bmatrix}\phi_1\\\phi_2\\\vdots\\\phi_p\end{bmatrix} \quad (11)$$

concluyendo con la siguiente expresión para la función de verosimilitud aproximada asociada a los parámetros (ϕ,τ):

$$f(Z|\phi,\tau)=L(\phi,\tau|Z) \propto\tau^{\frac{N-p}{2}}\cdot\exp\left\{\frac{-\tau}{2}\left[(Z-X\Phi)'(Z-X\Phi)\right]\right\} \quad (12)$$

La denominación como verosimilitud "aproximada" se debe al hecho de asumir fijas las "p" primeras observaciones Z_p como lo refleja el desarrollo presentado. Sin embargo, esta función representa sólo la densidad de probabilidad conjunta de las observaciones Z no considera la densidad de probabilidad conjunta de las "p" primeras observaciones más que solo para cálculos específicos. Por lo tanto, la función $L(\phi,\tau\vee Z)$, se encuentra condicionada únicamente al conjunto Z.

La cual será un elemento central para el desarrollo de la distribución posterior, donde los parámetros $\phi\, y\, \tau$ poseen sus propias distribuciones de probabilidad que se ajusten al dominio establecido en el desarrollo expuesto. En la literatura revisada como una alternativa se presenta el uso de distribuciones del tipo no informativo como la previa de "Jeffrey" resultando en una posterior normal gamma permitiendo calcular las expresiones correspondientes al poseer una expresión cerrada. Otra opción es el uso de previas informativas mediante esquemas semiconjugados o conjugados obteniendo en muchas ocasiones expresiones no cerradas [9].

2.5. Simulación de la Distribución Posterior

Para estimar los parámetros deseados es necesario el uso de la distribución posterior, misma que presenta desafíos y limitaciones, dado que en muchos casos posee una forma funcional compleja. Ante esta problemática una de las alternativas más usadas es recurrir a métodos de simulación basados en Cadenas de Markov Monte Carlo (MCMC) que convergen a la posterior deseada mediante "n" iteraciones en múltiples cadenas, limitando la efectividad a los recursos computacionales y el esquema de salto o muestreo del espacio de parámetros [10].

Uno de los métodos más usados para estos algoritmos es "Metropolis - Hastings" que se caracteriza por generar una cadena de muestras dependientes que se distribuyen conforme la posterior buscada mediante diferentes elementos involucrados, disminuyendo su efectividad en modelos de alta dimensionalidad. Esta limitante en su eficiencia está fuertemente vinculada con la elección de la distribución propuesta y una mala elección puede ocasionar baja tasa de aceptación, pasos cortos y problemas de convergencia provocando un muestreo ineficiente del espacio parametral [11].

En respuesta a esta limitación se presenta el algoritmo Hamiltoniano Monte Carlo (HMC) que incorpora información del gradiente de la log-verosimilitud y realiza la simulación deseada basándose en dinámicas físicas, confluyendo en trayectorias más largas y coherentes, evitando estancamientos en regiones de alta densidad y explorando más eficientemente el espacio parametral incluso

en alta dimensión. Con estas propiedades, la precisión y eficiencia computacional se mejoran significativamente lo cual es provechoso para el modelo planteado en los siguientes párrafos [12].

3. DESARROLLO Y APLICACIÓN DEL MODELO BAYESIANO

El presente capítulo expondrá el desarrollo y aplicación del modelo propuesto con base en los conceptos definidos en párrafos anteriores, primero se expondrá el modelo propuesto y todos sus elementos, se describirá la selección de hiperparámetros para las distribuciones previas y finalmente se expondrá los resultados presentados en la fase de estimación y validación. Para tales fines, es necesario realizar adaptaciones a la función de verosimilitud incorporando nuevos elementos que justifiquen su pertinencia dentro del esquema sugerido para las previas, y la correcta elección del tipo de método MCMC para el muestreo de la simulación posterior encontrada, lo cual será realizado mediante el paquete "bayesforecast" disponible para el lenguaje R.

3.1. Mejora de la distribución predictiva y predictiva posterior

El primer elemento por incorporar será la función de verosimilitud, adaptando la incorporación del intercepto respectivo y la parametrización de la distribución de las innovaciones en términos de la dispersión. Ahora dada una serie temporal $\{Z_t, t=1,2,..,N\}$ se asume generada por el siguiente modelo autorregresivo de orden "p":

$$Z_t=\phi_0+\phi_1 Z_{t-1}+\ldots+\phi_p Z_{t-p}+a_t \tag{13}$$

Donde $\phi=(\phi_1,\phi_2,\ldots,\phi_p)\in R^p$, $\phi_0\in R$ y $\{a_t, t=1,\ldots,N\}$ corresponde al ruido blanco débil distribuido por $N(0,\sigma^2)$, permitiendo reescribir la verosimilitud planteada con base en las observaciones Z_p y Z de la siguiente manera:

$$f(Z|\phi_0,\phi,\sigma^2)=(2\pi\sigma^2)^{\frac{N-p}{2}} \cdot\exp\left\{\frac{-1}{2\sigma^4}\sum_{t=p+1}^{N}(Z_t-\phi_0-\phi_1 Z_{t-1}-\ldots-\phi_p Z_{t-p})^2\right\} \tag{14}$$

por lo tanto:

$$L(\phi_0,\phi,\sigma^2|Z)\propto(\sigma^2)^{\frac{(N-p)}{2}} *\exp\left\{\frac{-1}{2\sigma^2}-(N-p-1)\phi_0\right\}+ *\exp\left\{\sum_{t=p+1}^{N}(Z_t-\phi_0-\phi_1 Z_{t-1}-\ldots-\phi_p Z_{t-p})^2\right\} \tag{15}$$

reescribiendo los elementos de la sumatoria con las matrices definidas anteriormente:

$$L(\phi_0,\phi,\sigma^2|Z)\propto(\sigma^2)^{\frac{(N-p)}{2}} *\exp\left\{\frac{-1}{2\sigma^2}-(N-p-1)\phi_0\right\}+ *\exp\left[(Z-X\Phi)^T(Z-X\Phi)\right] \tag{16}$$

Siendo la función de verosimilitud aproximada construida a partir de la densidad conjunta de los innovaciones $\{a_t, t=1,\ldots,N\}$ pero parametrizada en función de la varianza y considerando un intercepto en el proceso autorregresivo al no estar centrado. Los parámetros establecidos serán las variables aleatorias para incorporar en el modelamiento a través de las distribuciones previas, por tanto, se recuerda la estructura del teorema de Bayes.

Adaptando a la nomenclatura utilizada la expresión descrita en la sección anterior se escribe como:

$$p(\theta|Z)=\frac{p(Z\vee\theta)p(\theta)}{p(Z)} \tag{17}$$

donde $p(\theta|Z)$ es la distribución posterior buscada, sin embargo, debido a que $p(Z)$ es una constante de normalización, la expresión anterior puede ser reescrita de la siguiente manera:

$$p(\theta|Z)\propto p(Z|\theta)*p(\theta) \tag{18}$$

estableciendo a $p(\theta)$ como representación de la distribución previa para el vector de parámetros en la verosimilitud $p(Z|\theta)$, en el presente artículo dicho vector está formado por:

$$\theta=\left\{\phi_0,\phi_1,\ldots,\phi_n,\sigma^2\right\}=\left\{\phi_0,\phi,\sigma^2\right\} \tag{19}$$

En función de un muestreo óptimo de la distribución posterior, se propone una relación independiente entre estos elementos, permitiendo establecer la siguiente expresión para hallar la distribución posterior con base en un modelo jerárquico:

$$p(\theta|Z) \propto p(Z|\theta)*p(\phi_0)*p(\phi)*p(\sigma^2) \tag{20}$$

donde $\phi=(\phi_1,\phi_2,\ldots,\phi_p)\in R^p, \phi_0 \in R \; y \; \sigma^2 \in R^+$, por ende, las distribuciones previas deben cumplir con el requisito de coincidir con el dominio del parámetro buscado, proponiendo el uso de las siguientes distribuciones con base en lo expuesto por Andrade & Barreto y Temponi de Oliveira [19, 7].

- $\phi_0 \;\; t-student(v,\mu_o,\sigma_0)$,
- $\Phi \;\; Np(\mu_p,\Sigma_p) con \Sigma_p diagonal$,
- $\sigma^2 \;\; Gamma(\alpha,\beta)$,

Estableciendo a $v,\mu_o,\sigma_0,\mu_p,\Sigma_p,\alpha,\beta$ como hiperparámetros a ser definidos o elicitados, cuyos valores provendrán del conocimiento de un experto del dominio y deberá ser traducido al espacio del problema mediante procesos que controlen los posibles sesgos agregados [23].

Con esta estructura es posible encontrar la forma de la distribución posterior, como un paso previo se instaura las expresiones correspondientes al kernel o núcleo de las distribuciones previas, para utilizar correctamente los operadores de proporcionalidad. Las expresiones serán:

1) $$P(\phi_0|\mu_0,\sigma_0,v) \propto \left(1+\frac{1}{v}\left[\frac{\phi_0-\mu_0}{\sigma_0}\right]^2\right)^{\frac{-v+1}{2}},$$

2) $$P(\Phi|\mu_p,\Sigma_p) \propto \exp\left\{\frac{-1}{2}(\Phi-\mu_p)^T\Sigma_p^{-1}(\Phi-\mu_p)\right\},$$

3) $$P(\sigma^2|\alpha,\beta) \propto (\sigma^2)^{\alpha-1}*\exp\{-\beta\sigma^2\}.$$

Reemplazando en la expresión adaptada a la nomenclatura utilizada se tiene:

$$p(\theta|Z) \propto p(Z|\theta)*p(\phi_0)*p(\phi)*p(\sigma^2) \tag{21}$$

$$\begin{gathered} p(\theta|Z) \propto L\left(Z|\phi_0,\Phi,\sigma^2\right)*p(\phi_0|\mu_0,\sigma_0,v) \\ *p(\Phi|\mu_p,\Sigma_p)*p(\sigma^2|\alpha,\beta) \\ p(\theta|Z) \propto \left(\sigma^2\right)^{\frac{N-p}{2}} \\ *\exp\left\{\frac{-1}{2\sigma^4}-(N-p+1)\phi_0+(Z-X\Phi)^T(Z-X\Phi)\right\} \\ *\left(\sigma^2\right)^{\alpha-1}*\exp\{-\beta\sigma^2\} \\ *\exp\left\{\frac{-1}{2}\Phi^T\Sigma_p^{-1}\Phi\right\}*\left(1+\frac{1}{v}\left[\frac{\phi_0-\mu_0}{\sigma_0}\right]^2\right)^{\frac{-v+1}{2}}. \end{gathered} \tag{22}$$

Reagrupando se obtiene la siguiente expresión correspondiente a la distribución posterior:

$$\begin{gathered} p(\theta|Z) \propto \left(\sigma^2\right)^{\frac{N-p}{2}+\alpha-1}*\exp\left\{\frac{-1}{2\left(\sigma^2\right)^2}-\beta\sigma^2\right\} \\ *\exp\left\{(Z-X\Phi)^T(Z-X\Phi)-\frac{-1}{2}\Phi^T\Sigma_p^{-1}\Phi\right\} \\ *\left(1+\frac{1}{v}\left[\frac{\phi_0-\mu_0}{\sigma_0}\right]^2\right)^{\frac{-v+1}{2}}*\exp\{-n\phi_0\}. \end{gathered} \tag{23}$$

Dado que no posee una expresión cerrada, se procederá con su muestreo mediante técnicas MCMC. Esta estructura también limita el cálculo de la distribución predictiva posterior, misma que permitirá realizar proyecciones $Z_{N+1},Z_{N+2},\ldots,Z_{N+h}$ para un horizonte planteado (h). Obligando a una adaptación de la expresión vista en la sección anterior, denotando a los conjuntos Z_h que representa los "h" valores de la serie a ser previstos y Z_r las últimas "r" observaciones:

$$Z_h=(Z_{N+1},Z_{N+2},\ldots,Z_{N+h})', \tag{24}$$

$$Z_r=(Z_{N-(r-1)},Z_{N-(r-2)},\ldots,Z_N)'. \tag{25}$$

Sea $P(\theta|Z)$ la densidad posterior calculada con base en los parámetros propuestos, la densidad predictiva posterior será:

$$P(Z_h \vee Z_r) = \int_{\phi_0} \int_{\Phi} \int_{\sigma^2} P(Z_h|\theta) * P(\theta|Z_r)\, d\phi_0\, d\Phi\, d\sigma^2 \quad (26)$$

Donde $\int_{\Phi} \ldots d\Phi$ representa las "p" integrales siguientes $\int_{\phi_1} \ldots \int_{\phi_p} d\phi_1 \ldots d\phi_p$. Esta expresión resultante al tener nuevamente una expresión cerrada será estimada por medio de la técnica seleccionada.

3.2. Preparación de los datos

El modelo desarrollado será aplicado en datos temporales sobre la población cotizante beneficiaria perteneciente a un seguro social, el cual se encuentra especializado en la protección de la población rural. Esta serie temporal tiene su punto de origen en el 31 de enero del año 2021 y su punto final en el 28 de febrero del año 2025 poseyendo una frecuencia mensual con 50 observaciones. Esta extensión permite la partición en dos conjuntos de datos con eje en el 29 de febrero del 2024, los datos antes de esta fecha permitirán la estimación de los parámetros buscados y posteriores a este punto temporal realizarán el papel de control o prueba a las proyecciones encontradas, haciendo un símil al aprendizaje de máquina usado ampliamente.

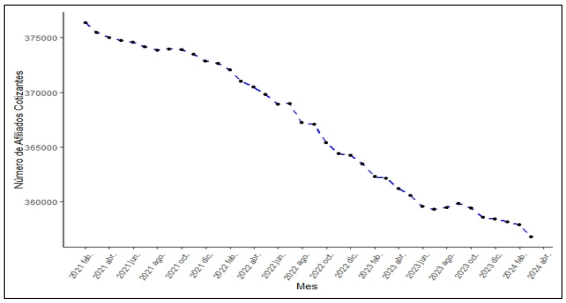

Figura 1. Beneficiarios Cotizantes de enero 2021 a febrero 2024.

En la figura 1, se representa el segmento de tiempo elegido para estimar los parámetros definidos en la verosimilitud de los datos. Sin embargo, la serie evidencia una tendencia decreciente con ligeros picos de aumento originados en su mayoría por causas no aleatorias, lo que sugiere la presencia de no estacionariedad tanto en su componente de tendencia como en posibles patrones estacionales.

Frente a esta situación, la literatura revisada de Alonso & Semaán respalda la aplicación de la prueba de HEGY como una herramienta para identificar la presencia de raíces unitarias estacionales y no estacionales, diferenciándose de otros contrastes tradicionales al proporcionar la respectiva diferenciación a aplicarse sobre la serie dado el tipo raíz presente. En el presente estudio, la prueba se aplicó sobre la serie original con el objetivo de establecer la transformación más adecuada [20].

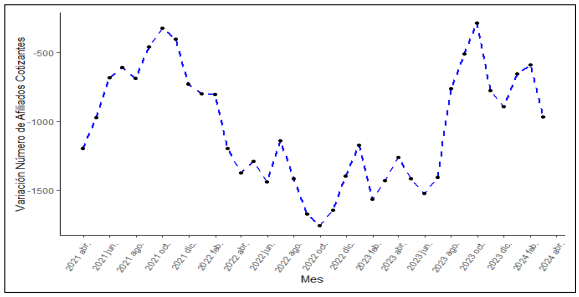

Figura 2. Serie de tiempo diferenciada y suavizada.

El resultado de la prueba de HEGY evidenció la presencia de una raíz unitaria no estacional del tipo bimensual, por lo que se aplicó la transformación correspondiente para inducir estacionariedad en la serie; de forma conjunta se implementó un proceso de suavización exponencial con un valor α de 0.5 destinado a mitigar el efecto de picos elevados asociados a componentes no aleatorios y evitar inconvenientes al momento de simular la distribución posterior, obteniendo una serie más estable y adecuada para la estimación y modelación, lo cual no afecta a la verosimilitud de la variable transformada conforme al teorema de factorización [13].

Tabla 1. Pruebas estadísticas para verificar la estacionariedad.

Prueba	Estadístico	V. Critico (95%)	Decisión
Dickey Fuller Aumentado (ADF)	-2.7236	-2.93	Se Rechaza H0
Phillips Perron (PP)	-7.61	-1.7834	Se Rechaza H0
Supremum ADF (SADF)	0	1.14	No Se Rechaza H0
KPSS	0.1321	0.463	No Se Rechaza H0

Finalmente, en la tabla 1 se presentan los resultados de las diferentes pruebas aplicadas a la serie de tiempo representada en la figura 2. Con las cuales descartamos la presencia de raíces unitarias no estacionales (ADF y PP), no se observa la presencia de un comportamiento explosivo (SADF) y existe un comportamiento estacionario alrededor del valor central (KPSS). Los resultados permiten utilizar la serie de tiempo para la estimación de los parámetros buscados con el modelo planteado en párrafos anteriores.

3.3. Estimación de Parámetros

Previo a la estimación un paso fundamental es la elección adecuada para el número de términos autorregresivos, misma que ha de considerar las distribuciones previas escogidas. El presente artículo adaptará un enfoque sencillo para esta etapa, expuesto por Correa & Barrera-Causil concentrándose en un análisis gráfico de la distribución de la variable objetivo dada las transformaciones realizadas y el uso adecuado de las propiedades del modelo considerado [21]. Con esto se controla la influencia que pueden tener valores extremos en el espacio de parámetros al respetar la estructura interna del comportamiento histórico de la serie temporal para diferentes valores de “p”.

Mediante la figura 3 es viable analizar la distribución de la variable objetivo haciendo a un lado la dependencia temporal de forma momentánea, permitiendo visualizar el comportamiento general, donde cabe resaltar la presencia de una simetría alrededor de un valor central propio de una variable distribuida normalmente. Dicho valor representa un decrecimiento constante próximo a 1.000, por tanto, el uso de un intercepto en conjunto con la estimación realizada con los valores pasados facilitará ajustar las proyecciones realizadas a este valor referencial.

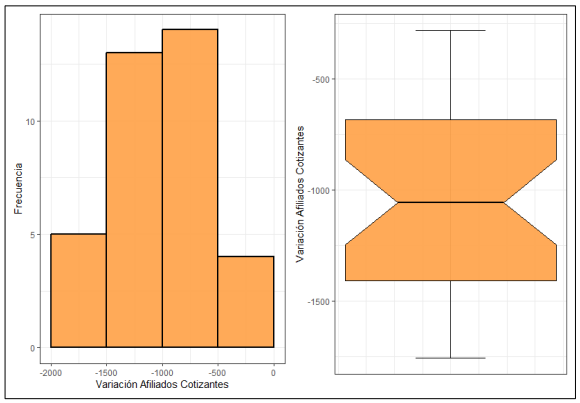

Figura 3. Distribución de la variable objetivo.

En tal contexto, al intercepto se lo suele considerar como aquel elemento que captura la parte del valor esperado y no es explicada por los rezagos. Es incorporada mediante una distribución t-Student “generalizada” con un valor medio de -175, grados de libertad igual a 25 y desviación estándar de 15. Con esta configuración se desea representar un decrecimiento constante con una variación elevada proporcionando flexibilidad para representar adecuadamente los cambios estructurales en la serie.

Los coeficientes autorregresivos serán incorporados mediante el uso de una distribución normal multivariante con media cero y matriz de covarianza diagonal igual a 0.25, representando la independencia entre los mismos, facilitando la interpretación de sus efectos sobre el valor actual de la serie. De igual forma, se impone un truncamiento al dominio de la distribución a -1 y 1 buscando la estabilidad del modelo al respetar los limitantes establecidos [14]. Finalmente, la varianza del error seguirá una distribución Gamma con media igual 500 y dispersión 50 adaptándose a shocks externos sin comprometer la estabilidad del sistema.

Esta elección aprovecha las propiedades asociadas a los modelos autorregresivos y permite una simulación con menor coste computacional para la distribución posterior mediante métodos MCMC con diferentes valores de p. Se realizan comparaciones del tipo Leave-One-Out Cross Validation (LOO-CV) bayesiano para cada valor p, algoritmo que utiliza el logaritmo de la densidad predictiva para cada punto evaluado eliminando la necesidad de reentrenar el modelo "n" veces, con esto se valora su capacidad de predicción fuera de la muestra [22].

Tabla 2. Comparación LOO de valores planteados

Métrica	p = 1	p = 2	p = 3
elpd_loo	-277.0 ± 4.7	-277.3 ± 4.5	-276.7 ± 4.4
p_loo	0.7 ± 0.2	1.2 ± 0.3	1.6 ± 0.4
looic	554.0 ± 9.4	554.7 ± 9.1	553.2 ± 8.8
K (Pareto)	Todos < 0.7	Todos < 0.7	Todos < 0.7

Al simular una distribución posterior inicial con 10.000 iteraciones y definir un periodo de calentamiento de 5.000 con 4 cadenas independientes para valores de p comprendidos entre 1 y 3 como lo sugiere la literatura revisada, se obtiene como resultado general la no existencia de observaciones influyentes ni mal ajustadas (K Pareto) para ninguno de los valores considerados. Sin embargo, las métricas consideradas "elpd_loo" y "looic" presentan un desempeño predictivo similar. El empate es resuelto mediante la métrica "p_loo" (número efectivo de parámetros del modelo) dictaminando una elección preferente por un modelo simple con solo un término autorregresivo haciendo uso del principio de parsimonia al no haber razones técnicas que justifiquen una complejidad elevada.

Con la elección de "p=1" se procede con la estimación final de los parámetros $\phi_1, \phi_0 \, y \, \sigma^2$, para esto es necesario redefinir los valores de los hiperparámetros planteados en la fase inicial. Lo cual será realizado mediante un estudio de simulación, tomando las combinaciones de puntos comprendidos en intervalos cuyos valores centrales son los valores descritos en párrafos anteriores. Estas combinaciones serán consideradas como valores de entrada para la respectiva función de distribución previa, adaptando el esquema expuesto por Bockting, se toma estos resultados como estimaciones para los valores de la serie temporal comparando sus resultados con los reales mediante una métrica de evaluación [6].

La evaluación se realizó con un total de 10.000 combinaciones seleccionadas al azar de los intervalos construidos y su rendimiento fue evaluado mediante la métrica de la raíz cuadrada del error medio (RMSE en inglés). Resultando en valores comprendidos entre 558.3 y 3726.1, escogiendo aquella composición que minimiza la métrica mencionada por los siguientes valores:

- $\phi_0 \;\; t-student\left(v=19\,;\mu_o=-123{,}2323\,;\sigma_0=17{,}27273\right)$,
- $\phi_1 \;\; N\left(\mu_p=0\,;\sigma^2=0{,}3661616\right)$, al considerar "p = 1"
- $\sigma^2 \;\; Gamma\left(\alpha=5727{,}273\,;\beta=8{,}434343\right)$.

Con estos valores se procedió a la simulación de la distribución posterior, fijando un total de 900.000 iteraciones y un periodo de calentamiento 500.000 con 5 cadenas, obteniendo los siguientes resultados:

Tabla 3. Resultados de la Estimación de Parámetros

Parámetro	Promedio	se	5%	95%
ϕ_0	-126,6599	0,0211	-156,8499	-97,1302
σ^2	677,4417	0,0103	662,7933	692,2028
ϕ_1	0,7767	0,0001	0,6204	0,9291

Las estimaciones presentadas en la tabla 3 evidencian una significancia notable de los parámetros muestreados, reflejada en errores estándar reducidos y en intervalos de credibilidad que no incluyen el valor cero. En particular, el parámetro ϕ_1 muestra una actualización relevante comparado con la información previa, trasladando su valor central de 0 a una estimación positiva de 0.7767 marcando la presencia de una autocorrelación fuerte en la serie y valida la inclusión de una estructura autorregresiva de

orden uno.

Por su parte los parámetros ϕ_0 y σ^2 presentan un ajuste de sus valores promedio consistentes con las creencias iniciales, manteniendo coherencia con la información previa, pero refinando su estimación a partir de los datos observados. Estos comportamientos reflejan la capacidad del modelo para actualizar de manera diferenciada las creencias iniciales, al identificar patrones estructurales de la serie diferenciándolos del ruido.

De esta manera, el modelo se considera informativo en la medida en que los datos aportan evidencia sustancial para la estimación de los parámetros, más allá de un simple ajuste mecánico. Esto se refleja en la precisión de las estimaciones y en la capacidad de modificar de forma significativa las creencias iniciales cuando la información empírica lo requiere.

Por otro lado, la coherencia del modelo se refleja en la integración estable entre los valores representados por las distribuciones previas y la información presente en los datos. No se presentan conflictos en la inferencia, como estimaciones inestables o intervalos contradictorios, sino que se observa una actualización lógica de los parámetros. El modelo ajusta con mayor intensidad aquellos parámetros donde los datos lo demandan y mantiene estabilidad donde la información previa resulta adecuada.

La convergencia del algoritmo utilizado para el muestreo de la distribución posterior se refleja en las siguientes figuras.

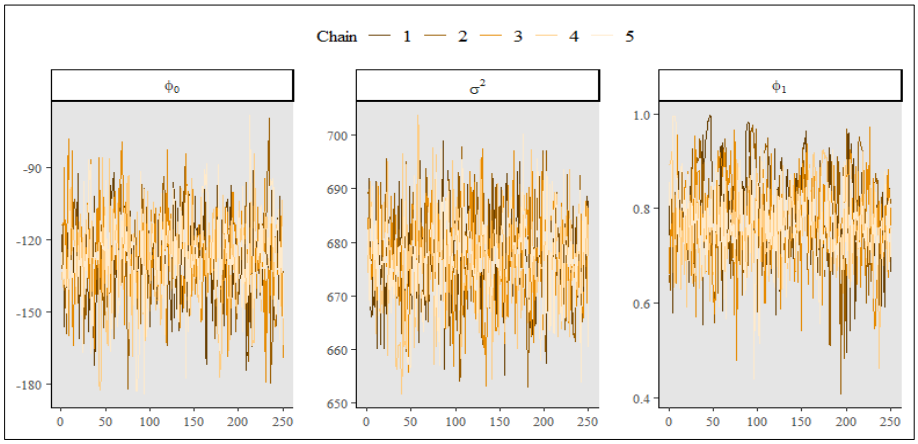

Figura 4. Gráfico de traza para las últimas 250 observaciones.

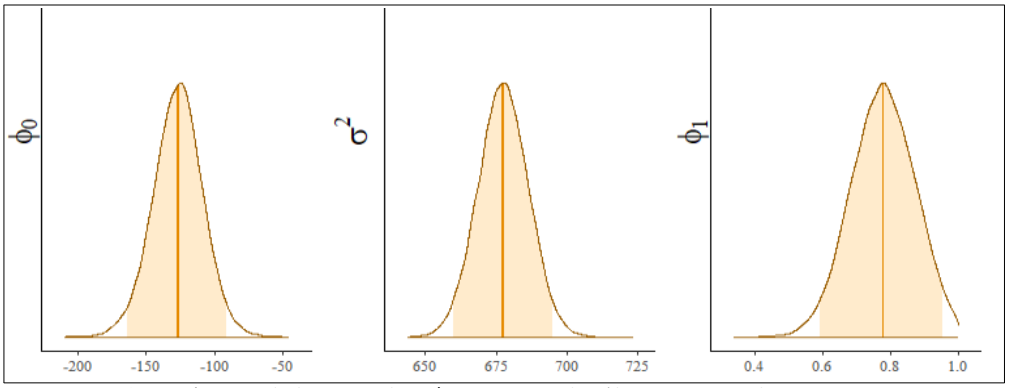

Figura 5. Densidad marginal parámetros para las últimas 10000 observaciones.

Por motivos de representación, las anteriores figuras fueron realizadas con un número determinado de observaciones. Sin embargo, a pesar del limitante mencionado el gráfico de traza muestra un comportamiento estacionario en todas las cadenas utilizadas lo cual junto con el bajo valor para el error estándar posterior (SE) evidencian una baja dispersión de la distribución posterior. Comportamiento reforzado por la densidad marginal de cada parámetro, pues su forma se aproxima de forma considerable a una distribución normal demostrando que las cadenas no están atrapadas en regiones no representativas.

3.4. Validación de errores

Si bien la convergencia y las estimaciones obtenidas son adecuadas, la validación de los supuestos asociados a los errores se define como una fase fundamental tanto para el enfoque frecuentista como para el tradicional. Para este propósito se presentan los siguientes gráficos.

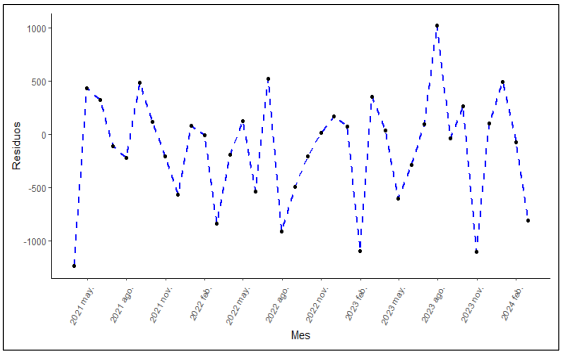

Figura 6. Serie de tiempo para los errores obtenidos.

Por lo descrito en párrafos anteriores, el error obtenido en la estimación de los puntos temporales debe ser ruido blanco débil distribuido acorde a una distribución normal de media 0 y varianza σ^2. Mediante la figura 6 se observa un comportamiento aleatorio en los errores sin una estructura o patrón. Esto es complementado por la figura 7, donde el histograma y qqplot correspondiente muestran un comportamiento casi normal y los correlogramas respectivos marcan una ausencia de una autocorrelación significativa.

Tabla 4. Resultados de pruebas estadísticas aplicadas sobre los residuos.

Prueba	Estadístico	P-Valor / V. Critico (95%)	Decisión
Box Ljung	10,912	0,5365	No Se Rechaza H0
Jarque Bera	0,854	0,6525	No Se Rechaza H0

Finalmente, los comportamientos resaltados deben ser comprobados mediante los resultados de las pruebas estadísticas con el respectivo P-valor o valor crítico. Las decisiones correspondientes reflejan que la serie de tiempo es ruido blanco distribuido normalmente, dando por válido los supuestos centrales en los que está cimentado el modelo bayesiano desarrollado.

4. RESULTADOS

El ajuste obtenido por el modelo propuesto presenta un ajuste razonable para la proyección de la población planteada como se demuestra a continuación.

En términos de las variaciones bimensuales suavizadas, los valores reales para el periodo de tiempo analizado se adaptan de manera satisfactoria al intervalo de credibilidad (al 90%) para los valores simulados. Aplicando la teoría asociada a la prueba de HEGY por Alonso & Semaán, es factible regresar las observaciones a su valor original, y junto con el uso de la distribución predictiva posterior se posibilita hallar las proyecciones del segundo grupo [20]. El cual se encuentra reservado para evaluar la capacidad predictiva del modelo, cuyos resultados se encuentran en la tabla 9 y figura 9.

Al evaluar la predicción con el valor real mediante las métricas de error más comunes y entendibles por usuarios no afines al enfoque bayesiano se obtienen los valores de RMSE: 3197.836 y MAE: 2807.667.

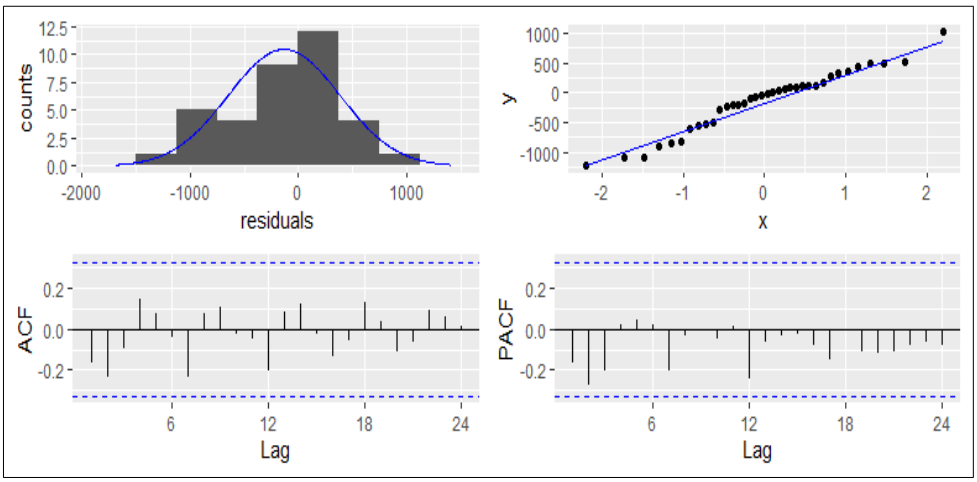

Figura 7. Distribución de los errores.

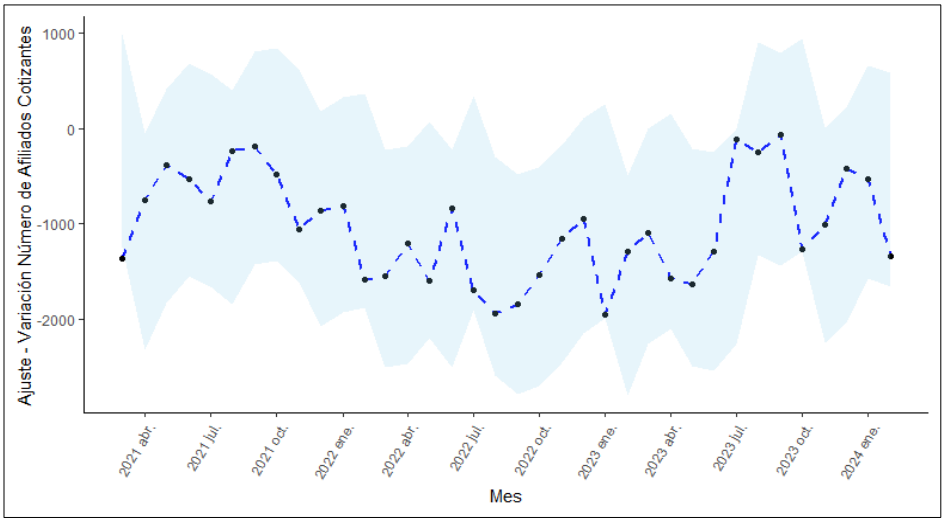

Figura 8. Ajuste de las predicciones realizadas.

Tabla 5. Proyecciones realizadas para un periodo h=12

			Valores Truncados para las Proyecciones			
Año	**Mes**	**Variación Predicha**	**L. Inferior Proyección**	**Proyección**	**L. Superior Proyección**	**Valor Real**
2024	Marzo	-1.158,36	355.602	356.726	357.852	355.290
2024	Abril	-941,7049	354.677	355.852	357.002	354.051
2024	Mayo	-787,6527	352.701	355.064	357.380	352.725
2024	Junio	-645,5906	350.866	354.419	357.918	352.394
2024	Julio	-550,781	349.106	353.868	358.534	353.218
2024	Agosto	-479,4369	347.424	353.388	359.218	355.071
2024	Septiembre	-423,1642	345.788	352.965	359.969	354.922
2024	Octubre	-368,947	344.210	352.596	360.756	355.193
2024	Noviembre	-332,8064	342.649	352.263	361.560	356.002
2024	Diciembre	-328,294	341.143	351.935	362.373	356.647
2025	Enero	-282,6483	339.662	351.652	363.237	357.102
2025	Febrero	-252,5481	338.216	351.400	364.102	356.703

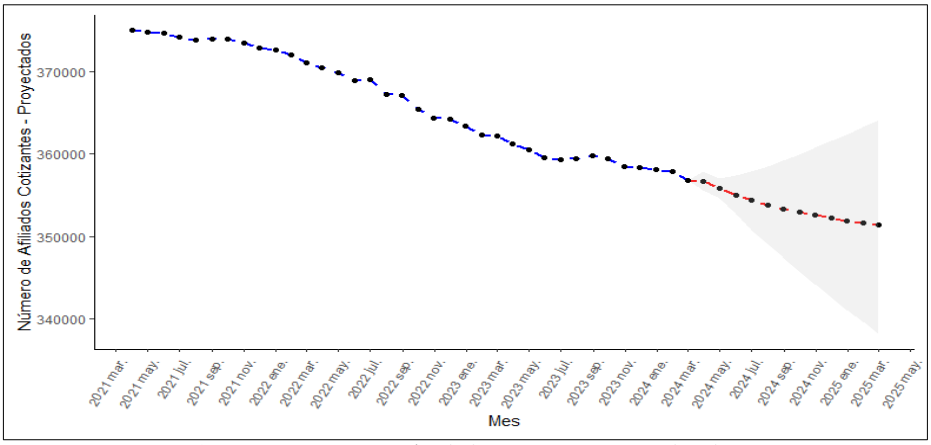

Figura 9. Representación de las proyecciones realizadas.

Lo cual a criterio del experto del dominio de los datos corresponde a un ajuste razonable dado el contexto social y político que atravesaba Ecuador en el periodo de tiempo usado como prueba.

Comparativa con el enfoque frecuentista

La comparación con el procedimiento tradicional (Metodologías Box Jenkins) permitirá el contraste del desempeño conseguido con la estimación bayesiana propuesta, para esto se adaptará una versión simplificada de la metodología expuesta por Huertas & Donado la cual se concentra en la aplicación de modelos Auto-ARIMA [24].

Al realizar el ajuste automático sobre la serie utilizada en la estimación bayesiana, se determinó como mejor frecuentista a un modelo $SARIMA(1,0,0)x(1,0,0)$, cuya validación fundamental de supuestos se resume en la siguiente tabla:

Tabla 6. Resultados de pruebas estadísticas

Prueba	Estadístico	P-Valor / V. Critico (95%)	Decisión
Box Ljung	5,6744	0,3392	No Se Rechaza H0
Jarque Bera	1,1077	0,5747	No Se Rechaza H0

Dado el no rechazo de las hipótesis nulas, se puede afirmar que los residuos del modelo presentan independencia y se distribuyen aleatoriamente conforme una ley normal. Sin embargo, la estructura seleccionada presenta una complejidad elevada al necesitar de una combinación de términos autorregresivos no estacionales y estacionales de orden 1, lo cual no es suficiente para alcanzar una predicción elevada conforme lo demuestra la siguiente tabla:

Tabla 7. Proyecciones frecuentistas vs bayesianas para un periodo h=12.

Año	Mes	Proyección Frecuentista	Proyección Bayesiana	Valor Real
2024	Marzo	356.839	356.726	355.290
2024	Abril	355.829	355.852	354.051
2024	Mayo	354.920	355.064	352.725
2024	Junio	353.956	354.419	352.394
2024	Julio	352.672	353.868	353.218
2024	Agosto	351.263	353.388	355.071
2024	Septiembre	349.744	352.965	354.922
2024	Octubre	348.470	352.596	355.193
2024	Noviembre	347.254	352.263	356.002
2024	Diciembre	345.922	351.935	356.647
2025	Enero	344.557	351.652	357.102
2025	Febrero	343.378	351.400	356.703

Al transformar los valores de las variaciones predichas, la diferencia entre los ajustes es notoria a partir del mes de agosto del año 2024, donde la brecha entre los enfoques supera 2.000 individuos y continúa alargándose en años posteriores. El comportamiento descrito es la causa de las métricas obtenidas para el enfoque frecuentista con valores de 7223,282 y 5723,419 para RSME y MAE respectivamente.

5. CONCLUSIÓN

Las proyecciones obtenidas con el modelo y enfoque propuesto evidencian un desempeño superior frente a metodologías tradicionales, las cuales, aunque cumplen los supuestos teóricos con la serie analizada, presentan métricas inferiores. Con el enfoque bayesiano se ha integrado de forma adecuada el conocimiento previo y controlado las limitaciones asociadas a la disponibilidad y naturaleza de los datos, resultado relevante al provenir de poblaciones rurales de difícil acceso. En este contexto, la especificación con un solo término autorregresivo resulta suficiente para cumplir condiciones de estabilidad y representar de manera consistente la dinámica observada, consolidando un ajuste robusto y metodológicamente sólido.

Los resultados obtenidos abren líneas de investigación relevantes para el contexto demográfico, permitiendo la inclusión de un enfoque no tradicional en proyecciones donde la escasez de registros administrativos completos y la heterogeneidad propia de las poblaciones rurales representan obstáculos para una adecuada estimación. El modelo bayesiano propuesto ha incorporado conocimiento previo de forma armoniosa, demostrando al enfoque propuesto como una herramienta adaptable y escalable.

Futuros trabajos podrían explorar la extensión del modelo hacia estructuras de orden superior, así como la incorporación de efectos estacionales mediante componentes SARIMA bayesianos. La implementación de algoritmos MCMC más eficientes, como variantes adaptativas del HMC, podrían reducir el costo computacional en aplicaciones con horizontes de proyección más extensos, potenciando la utilidad práctica del modelo en entornos con recursos limitados.

REFERENCIAS

[1] INEC. Estimaciones y proyecciones de la población del Ecuador. Revisión 2024. Quito: Instituto Nacional de Estadística y Censos; 2024.

[2] Mendoza-Ponce AV, Corona-Núñez RO, Galicia-Sarmiento L, García-Guerrero VM. Proyecciones poblacionales y económicas bajo diferentes escenarios para México. Papeles de Población (2019) 25(99):9-43. doi: 10.22185/24487147.2019.99.02

[3] Blaconá MT, CCG, CCG. Estudios de series temporales demográficas mediante modelos ARIMA de Box-Jenkins: una aplicación a series de la ciudad de Rosario, Argentina. Santiago de Chile: CEPAL; 1991. https://hdl.handle.net/11362/9245

[4] Andrade Guallpa CA, FUCM, PCLF, ACDP. Pronóstico de migración humana del Ecuador utilizando modelos de series temporales. Pro Sciences: Revista de Producción, Ciencias e Investigación (2023) 7(1):19-45.

[5] Andrade MG, Barreto GdA. Estimação e análise de robustez na abordagem bayesiana de modelos AR(p): estudo de caso em previsão de vazão. São Carlos: Instituto de Ciências Matemáticas e de Computação; 1999. p. 10-15.

[6] Temponi de Oliveira JR. Uso do amostrador de Gibbs e Metropolis-Hastings em análise bayesiana de modelos AR(p) [dissertação de mestrado]. São Carlos: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo; 1998. p. 20-30.

[7] Ordorica Mellado M. Pronóstico de las defunciones por medio de los modelos autorregresivos integrados de promedios móviles. Estudios Demográficos y Urbanos (2004) 19(3):545-578. https://www.scielo.org.mx/scielo.php?pid=S1405-74252004000400009&script=sci_arttext

[8] Rincón L. Una introducción a las series de tiempo. Ciudad de México: Universidad Nacional Autónoma de México, Facultad de Ciencias; 2024.

[9] Pérez JL. Una presentación de los mínimos cuadrados generalizados, y en particular, para funciones vectoriales. Revista Investigación Operacional (2011) 32(2):67-89.

[10] Gómez-Mejía A. Modelo de probabilidad máxima. Libre Empresa (2020) 17(1):121-138.

[11] López Montesinos A. Estudio del AIC y BIC en la selección de modelos de vida con datos censurados. Guanajuato: Centro de Investigación en Matemáticas, A.C.; 2011. p. 15-20.

[12] Casas Fernández A. Introducción al enfoque de la estadística bayesiana: aplicaciones a seguros. Cádiz: Universidad de Cádiz; 2023. https://rodin.uca.es/bitstream/handle/10498/32930/TFG_Antonio_Casas_Fernandez_Definitivo.pdf

[13] Rangel EN. ¿Qué es la estadística bayesiana? Journal of Basic Sciences (2015) 1(1):2-20.

[14] Stan Development Team. Stan user's guide. 2024. https://mc-stan.org/docs/2_36/stan-users-guide/posterior-prediction.html

[15] Wang YM, Park T. A brief tour of Bayesian sampling methods. IntechOpen; 2020. p. 8-10.

[16] Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. The American Statistician (1995) 49(4):327-335. doi: 10.1080/00031305.1995.10476177
[17] Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv; 2018.
[18] Correa JC, Barrera-Causil CJ. Introducción a la estadística bayesiana: notas de clase. Medellín: Instituto Tecnológico Metropolitano (ITM); 2018.
[19] García-Guerrero VM, Ordorica Mellado M. Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. Estudios Demográficos y Urbanos (2012) 27(2):409-448.
[20] Karakani Masoumi H, van Niekerk J, van Staden P. Bayesian analysis of AR(1) model. arXiv; 2016.
[21] Alonso JC, Semaán P. Prueba de HEGY en R: una guía. Apuntes de Economía (2010) (23):18-23.
[22] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing (2017) 27(5):1413-1432. doi: 10.1007/s11222-016-9696-4
[23] Bockting F, Radev ST, Bürkner PC. Simulation-based prior knowledge elicitation for parametric Bayesian models. arXiv preprint arXiv:2308.11672v2; 2024. https://arxiv.org/abs/2308.11672
[24] Huertas N, Donado S. Evaluación crítica de la función auto.arima de R en la identificación de procesos de series de tiempo. Proyecto de grado II; 2018.

ACERCA DE LOS AUTORES

Daniel Hernán Freire Navarrete es Ingeniero en Estadística por la Universidad Central del Ecuador (UCE) y Master en Matemática Aplicada. Actualmente se desempeña como Analista Medio de Metodologías de Registros Administrativos en el INEC, contribuyendo al diseño, estandarización y fortalecimiento metodológico para el aprovechamiento estadístico de fuentes administrativas. Sus principales áreas de análisis abarcan: metodologías y estadística de registros administrativos; evaluación y aseguramiento de la calidad de datos; modelos estadísticos y bayesianos. A lo largo de su trayectoria ha colaborado con instituciones académicas y públicas, así como con el sector privado, integrando tareas de depuración, consolidación, análisis y modelamiento de bases de datos, y traduciéndolas en productos de comunicación estadística.

Oscar Stalin Llano Moscuy es Ingeniero en Estadística por la Universidad Central del Ecuador y se desempeña como analista en el Banco del Instituto Ecuatoriano de Seguridad Social. Cuenta con formación y experiencia en modelado bayesiano, aprendizaje estadístico y análisis de series temporales, aplicados a problemas de clasificación, predicción y automatización de procesos en contextos institucionales. Su trabajo se orienta al análisis y tratamiento de grandes volúmenes de datos, así como al desarrollo de modelos cuantitativos que apoyan la toma de decisiones técnicas y estratégicas. Sus líneas de interés incluyen probabilidad, simulación, minería de datos y ciencia de datos aplicada, con énfasis en el diseño de estudios, la generación de indicadores y la elaboración de nuevas metodologías analíticas orientadas a la mejora continua de procesos y a la producción de evidencia estadística confiable.

Byron Morocho Caisatoa es egresado de la carrera de Ingeniería Estadística por la Universidad Central del Ecuador. Posee una sólida formación en análisis estadístico, modelamiento de datos y métodos cuantitativos, orientados a la toma de decisiones basada en evidencia. Actualmente se desempeña como Data Analyst / Research Assistant, donde desarrolla estudios enfocados en la evaluación del desempeño de medios publicitarios, análisis de competencia e identificación de dinámicas de mercado, a partir del tratamiento y análisis de grandes volúmenes de datos. Su experiencia

profesional incluye la gestión, depuración y validación de bases de datos, así como la construcción de indicadores y reportes analíticos para la planificación estratégica.