Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(2): 3334–3380.
DOI: 10.3934/math.2025155
Received: 10 November 2024
Revised: 02 February 2025
Accepted: 14 February 2025
Published: 21 February 2025

Research article

Estimating the Consumer Price Index using the lognormal diffusion process with exogenous factors: The Colombian case

Antonio Barrera[1,2], Arnold de la Peña Cuao[3], Juan José Serrano-Pérez[3] and Francisco Torres-Ruiz[2,3,*]

[1] Department of Mathematical Analysis, Statistics and Operations Research and Applied Mathematics, University of Málaga, Bulevar Louis Pasteur 31, 29010, Málaga, Spain

[2] Institute of Mathematics of the University of Granada (IMAG), Calle Ventanilla 11, 18001, Granada, Spain

[3] Department of Statistics and Operations Research, University of Granada, Avenida de Fuente Nueva s/n, 18071, Granada, Spain

* **Correspondence:** Email: fdeasis@ugr.es; Tel: +34-958-241000, ext 20056.

Abstract: In this paper, a model based on the lognormal diffusion process with exogenous factors was considered, aiming to describe the dynamics of the basic Consumer Price Index (CPI) in Colombia. To this end, a Bayesian procedure was employed for the selection of factors using real data from different economic sources such as the Central Bank (*Banco de la República de Colombia*), among others. A model with five exogenous factors (economic variables) was obtained, carrying out maximum likelihood estimation for its parameters. Fitting and forecasting procedures under different conditions showed that the proposed model outperformed the predictions made by the Central Bank. In addition, potential future economic scenarios were analyzed for testing purposes. This model provided valuable insight into the main determinants of inflation in Colombia, reflecting the importance of the factors under the control of economic authorities, labor market dynamics, and external economic conditions. This new approach also gave rise to asking questions concerning impact analysis of economic shocks, and the search of possible scenarios for given CPI dynamics.

Keywords: lognormal diffusion process; exogenous factors; inference in diffusion processes; Bayesian model selection; economic forecasts
Mathematics Subject Classification: 60J60, 62M20

1. Introduction

The modeling of dynamic economic variables is a constantly evolving area of research, covering diverse economic topics such as business cycle theories, asset valuation, monetary policy, financial stability, energy demand, and natural disaster impact assessment, among others. The importance of this type of modeling stems from the need to understand and predict the behavior of key indicators, such as economic growth, inflation, unemployment, and financial asset prices. However, modeling these variables has presented great statistical and mathematical challenges due to the very nature of the data and its multiple characteristics, such as its nonlinearity, structural changes, sensitivity to external factors and complex interrelationships between economic factors.

In recent years, there has been a growing uptake of advanced techniques to address these complex economic challenges. These techniques include modified traditional econometric models, such as vector autoregressive (VAR) with error correction models and co-integration analysis (see, for instance, Chávez and Rodríguez [1]). Bayesian models have also acquired relevance by allowing the incorporation of a priori information and the quantification of parameter uncertainty [2, 3]. General equilibrium models have been proven useful for capturing interrelationships between economic sectors and agents (Montaud et al. [4]), while models such as Markov-switching have been helpful for the analysis of business cycles and regime shifts, as in the study carried out by Gomes et al. [5] on the Brazilian economy. In addition, hybrid approaches were developed in order to combine traditional econometric models with machine learning techniques, allowing complex and nonlinear relationships to be approximated (Cui et al. [6]).

Though these models have sought to accurately capture the complexity and dynamics of real markets and economies, they present significant challenges in terms of data requirements, computational cost and understandability. Bayesian models, for instance, allow for the incorporation of a priori information, but it can be difficult to specify and estimate. VAR and general equilibrium models, designed to capture interrelationships between economic variables, require large sample sizes and numerous parameters, which can lead to overfitting problems and difficulties in results interpretation. On the other hand, machine learning models, while powerful and flexible, often act as *black boxes*, making it difficult to understand how they make their predictions.

In this context, diffusion processes have emerged as a widespread alternative in the field of quantitative finance. They have been used to model variables such as asset prices, interest rates, exchange rates, valuation of financial derivatives, optimal portfolio management problems, and investment strategy selection. As an example, Wang et al. [7] investigated a time-consistent investment strategy under the mean-variance criterion for an investor saving for retirement. Zhu et al. [8] proposed an efficient simulation-based method to approximate the solution of the continuous-time dynamic portfolio optimization problem for multiple assets and state variables. Liu et al. [9] investigated variance swap pricing under a hybrid model combining the stochastic double-Heston volatility with the Cox-Ingersoll-Ross stochastic interest rate model. Moreover, Tao et al. [10] studied the valuation of geometric Asian options where the underlying asset follows a sub-fractional Brownian movement.

Different approaches, apart from diffusion processes, have also been considered. Such is the case of stochastic volatility models (i.e., Alakkari et al. [11]) or inhomogeneous Poisson processes (Liang et al., [12]), which allow modeling specific aspects of certain economic variables. Different models have incorporated more realistic features, such as time-varying parameters, jumps, and the interaction between different economic variables (Herzog [13], Mehrdoust, and Noorani [14]).

1.1. Consumer Price Index modeling in Colombia

The modeling of inflation in Colombia, as measured by the basic Consumer Price Index (CPI), is of crucial importance given the *Banco de la República de Colombia* (hereinafter referred as "Central Bank") constitutional responsibility to maintain low and stable inflation. This mandate led to the adoption of the inflation targeting scheme in 1999, in which the Central Bank commits itself to achieving and maintaining an annual inflation target defined by its Board of Directors. This institutional change underscores the fundamental importance of understanding and forecasting inflation dynamics for the effective conduct of monetary policy in the country.

Within the framework of inflation targeting, the Central Bank uses the intervention interest rate as its main monetary policy instrument. Decisions on this rate are taken based on a thorough analysis of the current economic situation and future projections, seeking to have an influence in aggregate demand and inflation expectations to guide them toward the established target. Nevertheless, the effectiveness of these decisions depends mostly on the accuracy of inflation forecasts and on the understanding of the monetary policy transmission mechanisms. It is worth highlighting that, despite the Central Bank's efforts, recent events such as the COVID-19 pandemic and geopolitical conflicts, including the Russia-Ukraine war and the Israel-Palestine conflict, have generated significant inflationary pressures, pushing inflation above the target. This situation further emphasizes the need for robust modeling tools that can capture and predict inflation dynamics in contexts of high volatility and uncertainty.

The relevance of accurate forecasts and the complexity of monetary policy are crucial issues in this context. Policy formulation by the Central Bank is much less straightforward than it appears, mainly because monetary policy often takes a long time to affect the economy. Monetary policymakers must therefore aim to have the right policy in place at the time in the future when their measures will start to take effect, which implies not only correctly predicting the effects of their policies, but also forecasting where the economy will be once the policy has had its impact. Studies on inflation in Colombia have covered various aspects, from the identification of its determinants to the evaluation of different forecasting methodologies (see, for instance, González et al. [15], Cárdenas-Cárdenas et al. [16]). This diversity of approaches reflects the complexity of the inflationary phenomenon and the need to address it from multiple perspectives so as to achieve a more complete comprehension and develop more effective policy tools.

Traditional econometric models, such as structural VAR models, regression models, input-output matrix (IPM), and time series analysis (i.e., Rodríguez [17]), have been mainly used to identify the factors that influence inflation in Colombia. These approaches have made it possible to capture the dynamic relationships between inflation and other macroeconomic variables, as well as to quantify their impact. Amongst the factors identified are inflation expectations, the exchange rate, international commodity prices, wage costs, and demand pressures. For example, González et al. [18] compared forecasts from VAR and ARX (autoregressive with an exogenous input) models, finding that the exchange rate and food prices are significant in explaining inflation. Jaramillo and Tovar [19] highlighted the role of the Value Added Tax (VAT) using an econometric estimation relating price level to the VAT and the Producer Price Index. In order to analyze the relationship between monetary policy and inflation, more sophisticated models have been employed to incorporate more realistic features of the economy. For instance, Aristizábal [20] explored the nonlinear relationship between money and inflation using artificial neural networks, while Villa et al. [21]) estimated an optimal Taylor-type monetary policy rule for Colombia using regime-switching models.

However, despite the diversity of models used, few studies have applied diffusion processes to model economic variables in Colombia. Diffusion processes, like geometric Brownian motion and the Ornstein-Uhlenbeck process, are widely used in finance to model asset prices and interest rates due to their ability to reflect uncertainty and the stochastic evolution of variables. Nonetheless, its application in modeling other macroeconomic variables such as the CPI has been rather limited. In this scenario, diffusion processes emerge as a promising alternative for modeling inflation dynamics in Colombia. These processes offer several significant advantages over traditional approaches, as they are based on stochastic differential equations, which allow capturing the continuous and stochastic nature of economic variables, as well as incorporating exogenous factors that influence their dynamics.

Not only do diffusion processes allow us to obtain point forecasts, but also confidence intervals and probability distributions for future inflation. An additional advantage to consider is that they allow analytical expressions for quantities of interest, such as the mean and variance of the process, to be obtained, which facilitates interpretation and inference. This is particularly useful in the economic context, where the interpretation of the results is crucial for policy decisions and the communication with economic agents. By having explicit expressions for the process steps and moments, sensitivity analysis can be performed and the impact of different scenarios on the variables of interest can be assessed.

This paper focuses on the use of a nonhomogeneous lognormal diffusion process to model the dynamics of the CPI. This is carried out by considering a particular case of the exogenous function that allows the incorporation of relevant external factors and proposing a strategy for the selection of such factors based on the selection of regressors in linear models. Its aim is to show how this approach can improve the modeling and forecasting of the CPI compared to the traditional models used by the Central Bank for its forecasts by properly capturing the external influences and the stochastic nature of the variable. This new approach is expected to provide a deeper understanding of inflation dynamics and to offer useful tools for monetary policy decision-making and inflation risk management in Colombia.

The paper is structured as follows: Section 2 presents the general version of the model, showing its main characteristics and particularizing it to the case of linear exogenous factors. For such particularization, the maximum likelihood (ML) inference is proposed. Special attention is paid to the estimation of the mean function, obtaining the probability density function (pdf) of its ML estimator. Another point of interest lies in the selection of the exogenous factors, which is carried out by using variable selection from the Bayesian point of view. Section 3 is dedicated to an application for modeling and analyzing the CPI in Colombia from data extracted from official sources. From a wide range of possible influential exogenous factors, the selection procedure determines a subset of five exogenous factors with which the model is estimated. An analysis of the predictive capacity of the model is carried out, making predictions up to eight posterior quarters, considering three cases: the first, using observed data until March 2022 (Section 3.1) and two latter updates, using data until June and September 2022 (Sections 3.2 and 3.3). Results are compared with the observed data of the CPI and with the official projections that the Central Bank made at the time. The results reveal that the model appropriately fits the real data and significantly improves the projections of Colombia's economic authorities. The model is also tested for predictions outside the sample time and under different economic scenarios in Section 3.4. The paper finishes with some conclusions and comments about future research lines.

2. An overview of the nonhomogeneous lognormal diffusion process

As it is well-known, the lognormal diffusion process, or geometric Brownian motion, has its origin in the randomization of the Malthusian-type differential equation $dx_t = \mu\, x_t\, dt$, where μ represents the growth rate of a certain population modeled by x_t. The randomization process consists of replacing μ with a random rate that is affected by a white noise, $\mu + \sigma\xi_t$, giving rise to the stochastic differential equation $dX_t = \mu\, X_t dt + \sigma\, X_t dW_t$, where W_t represents a Wiener process. The solution to this equation is the lognormal diffusion process, in its homogeneous version, which is a model widely used in finance, being used to represent the price of some financial assets following the fluctuations of the financial markets (Black-Scholes model).

The nonhomogeneous version of the process results from considering a variable growth rate, represented by a continuous function $h(t)$. The introduction of this function allows the consideration of external influences (generally called exogenous factors) that can modify the behavior of the endogenous variable modeled by the process. On the other hand, the fact that the mean function of the process is expressed in terms of $h(t)$ allows this process to be used for modeling a wide range of dynamic phenomena associated with growth curves. This is achieved with a convenient choice of said function. Recent works illustrate this comment, including particular growth curves such as the logistic, the Richards, Bertalanffy, Hubbert (see [22, 23] and references therein), as well as others recently introduced such as the hyperbolastic curves [24]. The recent work by Albano et al. [25] goes further and considers a generic growth curve that includes a wide range of curves of this type.

In this section we will focus our attention on the nonhomogeneous lognormal diffusion process with linear exogenous factors (Section 2.2), introduced from the general version of the process summarized in Section 2.1. Sections 2.3 and 2.4 are devoted to the inference of the model, whereas Section 2.5 focuses on the problem of choosing exogenous factors. Finally, Section 2.6 summarizes the computational process to carry out the inference in the model as well as to estimate the main characteristics of the process and approximate the probability density function of its mean function.

2.1. Formulation of the model

Next, we present a brief general description of the model (see [26] and references therein for details). Later, the attention will be focused on a particular case of the function $h(t)$, which will be used in the rest of the paper.

Let us consider $I = [t_0, +\infty)$, a real interval ($t_0 \geq 0$), $\Theta \subseteq \mathbb{R}^k$ an open set, and $h_{\boldsymbol{\theta}}(t)$ a continuous, bounded, and differentiable function on I and depending on $\boldsymbol{\theta} \in \Theta$. The nonhomogeneous lognormal diffusion process, $\{X(t); t \in I\}$, is the solution of the stochastic differential equation

$$dX(t) = h_{\boldsymbol{\theta}}(t)X(t)dt + \sigma X(t)dW(t), \qquad X(t_0) = X_0, \tag{2.1}$$

where $W(t)$ is a standard Wiener process, independent on $X_0 = X(t_0)$, $t \geq t_0$. More in detail, $X(t)$ is given by

$$X(t) = X_0 \exp\left(H_{\boldsymbol{\xi}}(t_0, t) + \sigma(W(t) - W(t_0))\right), \qquad t \geq t_0 \tag{2.2}$$

with

$$H_{\xi}(s,t) = \int_s^t h_{\theta}(u)du - \frac{\sigma^2}{2}(t-s), \qquad \xi = (\theta^T, \sigma^2)^T, \; s < t. \tag{2.3}$$

Regarding the distribution of the process, if the initial distribution X_0 is lognormal, $\Lambda_1\left[\mu_0; \sigma_0^2\right]$, or degenerate ($P[X_0 = x_0] = 1$), the finite-dimensional distributions of the process are lognormal. Concretely, $\forall n \in \mathbb{N}$ and $t_1 < \cdots < t_n$, vector $(X(t_1), \ldots, X(t_n))^T$ follows an n-dimensional lognormal distribution $\Lambda_n[\boldsymbol{\varepsilon}, \boldsymbol{\Sigma}]$, where the components of vector $\boldsymbol{\varepsilon}$ and matrix $\boldsymbol{\Sigma}$ are $\varepsilon_i = \mu_0 + H_{\xi}(t_0, t_i)$ and $\sigma_{ij} = \sigma_0^2 + \sigma^2(\min(t_i, t_j) - t_0)$, $i, j = 1, \ldots, n$, respectively. Consequently, the transition probability distribution is lognormal, that is, for $s < t$, the variable $X(t)|X(s) = y$ follows a lognormal distribution. Concretely,

$$X(t) \mid X(s) = y \rightsquigarrow \Lambda_1\left[\ln y + H_{\xi}(s,t), \sigma^2(t-s)\right], \;\; s < t. \tag{2.4}$$

From these distributions, the main characteristics of the process can be calculated. Among them, we can highlight those employed in practical applications with fitting and forecasting purposes, which can be expressed jointly in the form

$$G_{\xi}^{\lambda}(t|z,\tau) = M_{\xi}(t|z,\tau)^{\lambda_1} \exp\left(\lambda_2 \left(\lambda_3\, \sigma_0^2 + \sigma^2(t-\tau)\right)^{\lambda_4}\right), \tag{2.5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^T$ and $M_{\xi}(t|z,\tau) = \exp\left(z + H_{\xi}(\tau, t)\right)$. Table 1 includes some of these characteristics according to the values of $\boldsymbol{\lambda}$, τ, and z.

Table 1. Values used to obtain the n-th moment, the mode, and quantile functions from $G_{\xi}^{\lambda}(t|z,\tau)$. z_α is the α-quantile of a standard normal distribution.

Function	Expression	z	τ	$\boldsymbol{\lambda}$
n-th moment	$E[X(t)^n]$	μ_0	t_0	$(n, n^2/2, 1, 1)^T$
n-th conditional moment	$E[X(t)^n \vert X(s) = y]$	$\ln y$	s	$(n, n^2/2, 0, 1)^T$
mode	$Mo[X(t)]$	μ_0	t_0	$(1, -1, 1, 1)^T$
conditional mode	$Mo[X(t) \vert X(s) = y]$	$\ln y$	s	$(1, -1, 0, 1)^T$
α-quantile	$C_\alpha[X(t)]$	μ_0	t_0	$(1, z_\alpha, 1, 1/2)^T$
α-conditional quantile	$C_\alpha[X(t) \vert X(s) = y]$	$\ln y$	s	$(1, z_\alpha, 0, 1/2)^T$

2.2. The case of linear exogenous factors

Sometimes, the dynamic evolution of a variable can be affected by external influences represented by variables (exogenous factors) whose temporal behavior is assumed to be known. Related to the lognormal process, usually the way of introducing the external variables in the model is by means of a linear time function $h_{\theta}(t) = \theta_0 + \sum_{j=1}^{q} \theta_j F_j(t)$, with $\theta_j \in \mathbb{R}$ and F_j time-continuous functions, $j = 1, \ldots, q$. Examples of the use of this methodology can be seen in Tintner and Sengupta [27] or in Gutiérrez et al. [28] to model the evolution of the Gross Domestic Product in Spain.

By noting $\mathbf{u}(s,t) = \left(t - s, \int_s^t F_1(\tau)d\tau, \ldots, \int_s^t F_q(\tau)d\tau\right)^T$, expression (2.3) becomes $H_{\beta}(s,t) = \mathbf{u}(s,t)^T \boldsymbol{\beta}$, with $\beta_0 = \theta_0 - \dfrac{\sigma^2}{2}$ and $\beta_j = \theta_j$, $j = 1, \ldots, q$. It is noticed that all expressions in the previous subsection are valid, replacing $\boldsymbol{\xi}$ with $\boldsymbol{\varphi} = (\boldsymbol{\beta}^T, \sigma^2)^T$. In particular, the mean of the process is

$$m_{\boldsymbol{\eta},\boldsymbol{\varphi}}(t) = E[X_0] \exp\left(\mathbf{u}(t_0, t)^T \boldsymbol{\beta} + \frac{1}{2}\sigma^2(t - t_0)\right), \tag{2.6}$$

being $\boldsymbol{\eta}$ the parametric vector associated to X_0. On the other hand, the conditioned version is given by

$$m_{\boldsymbol{\varphi}}(t|s) = y \exp\left(\mathbf{u}(s, t)^T \boldsymbol{\beta} + \frac{1}{2}\sigma^2(t - s)\right) \tag{2.7}$$

where $X(s) = y$ is assumed.

Let us observe that, in both cases, a linear combination of the parameters involved appears in the exponential. This, together with the fact that the distributions that give rise to these means are lognormal, allows a study of these functions along the lines of Land [29], who carries out an in-depth study in the case of the distribution lognormal, or in the lines of Gutiérrez et al. [30, 31], who adapts the situation to the case of the lognormal diffusion process.

2.3. *Inference on the model*

Let us consider a discrete sampling of the process, based on d sample paths at time instants $t_{i1} = t_1 < t_{i2} < \ldots < t_{i,n_i}$, $i = 1, \ldots, d$. Denote by $\mathbf{X} = \left(\mathbf{X}_1^T | \cdots | \mathbf{X}_d^T\right)^T$, where $\mathbf{X}_i = (X(t_{i1}), \ldots, X(t_{i,n_i}))^T$, $i = 1, \ldots, d$. In addition, we assume the general case in which $X(t_1)$ is lognormally distributed $\Lambda_1[\mu_1, \sigma_1^2]$ (for the degenerate case $\sigma_1 = 0$ and $\mu_1 = \ln x_0$ must be considered, being x_0 the initial value).

In [26], and considering a generic function $h_{\boldsymbol{\theta}}(t)$, the authors carry out a detailed inferential study aimed at the ML estimation of the parameters of the process. The final result is a system of equations for which, in general, it is not possible to find solutions explicitly. For this reason, it is necessary to employ numerical methods or metaheuristic procedures, which in turn raise various problems such as the search for initial solutions or the delimitation of the parametric space. However, in the case in which the exogenous factors are linear, the system does have an explicit solution. Furthermore, it is possible to calculate the distribution of the estimators as well as that of certain characteristics associated with the process, such as the mean function.

Next we are going to focus on the linear case, adapting the general situation posed in [26]. To do this, we consider vector $\mathbf{V} = \left[\mathbf{V}_0^T | \mathbf{V}_1^T | \cdots | \mathbf{V}_d^T\right]^T = \left[\mathbf{V}_0^T | \mathbf{V}_{(1)}^T\right]^T$, built from $\mathbf{X}$ by means of the following change of variables:

$$V_{0i} = X_{i1}, \ i = 1, \ldots, d,$$
$$V_{ij} = (\Delta_i^{j+1,j})^{-1/2} \ln \frac{X_{i,j+1}}{X_{ij}}, \ i = 1, \ldots, d; j = 1, \ldots, n_i - 1,$$

where $\Delta_i^{j+1,j} = t_{i,j+1} - t_{ij}$.

From (2.4), and taking into account this change of variables, the density of $\mathbf{V}$ becomes

$$\frac{\exp\left(-\frac{1}{2\sigma_1^2}(\ln \mathbf{v}_0 - \mu_1 \mathbf{1}_d)^T (\ln \mathbf{v}_0 - \mu_1 \mathbf{1}_d)\right)}{\prod_{i=1}^{d} v_{0i} \left(2\pi\sigma_1^2\right)^{\frac{d}{2}}} \frac{\exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{v}_{(1)} - \mathbb{U}^T \boldsymbol{\beta}\right)^T \left(\mathbf{v}_{(1)} - \mathbb{U}^T \boldsymbol{\beta}\right)\right)}{(2\pi\sigma^2)^{\frac{n}{2}}} \tag{2.8}$$

with $\ln \mathbf{v}_0 = (\ln v_{01}, \ldots, \ln v_{0d})^T$ and $n = \sum_{i=1}^{d}(n_i - 1)$. Here, $\mathbf{1}_d$ represents the d-dimensional vector whose components are all equal to one, while $\mathbb{U} = [\mathbf{U}_1|\cdots|\mathbf{U}_d]$ is a $(q+1)\times n$ block-column matrix whose i-th block is given by the $(q+1)\times(n_i - 1)$ matrix.

$$\mathbf{U}_i = \left[(\Delta_i^{2,1})^{-1/2}\,\mathbf{u}(t_{i1}, t_{i2})\,\Big|\ldots\Big|\,(\Delta_i^{n_i,n_{i-1}})^{-1/2}\,\mathbf{u}(t_{i,n_{i-1}}, t_{i,n_i})\right].$$

From (2.8), it is deduced that:

- $\mathbf{V}_0$ and $\mathbf{V}_{(1)}$ are independents,
- the distribution of $\mathbf{V}_0$ is lognormal $\Lambda_d\left[\mu_1\mathbf{1}_d; \sigma_1^2\mathbf{I}_d\right]$,
- $\mathbf{V}_{(1)}$ is distributed as an n-variate normal distribution $N_n\left[\mathbb{U}^T\boldsymbol{\beta}; \sigma^2\mathbf{I}_n\right]$,

where $\mathbf{I}_d$ and $\mathbf{I}_n$ are the identity matrices of order d and n, respectively.

Now, suppose that $\boldsymbol{\eta} = (\mu_1, \sigma_1^2)^T$ and $\boldsymbol{\varphi}$ are functionally independents. Then, for a fixed value $\mathbf{v}$ of the sample, (2.8) can be seen as the likelihood function, from which we obtain the ML estimators of the parameters. Concretely,

$$\widehat{\mu}_1 = \frac{1}{d}\sum_{i=1}^{d}\ln v_{0i} \text{ and } \widehat{\sigma}_1^2 = \frac{1}{d}\sum_{i=1}^{d}(\ln v_{0i} - \widehat{\mu}_1)^2, \tag{2.9}$$

while

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\mathbb{U}\mathbb{U}^T)^{-1}\mathbb{U}\mathbf{v}_{(1)},\\ \widehat{\sigma}^2 &= \frac{1}{n-1}\mathbf{v}_{(1)}^T[\mathbf{I}_{n-1} - \mathbb{U}^T(\mathbb{U}\mathbb{U}^T)^{-1}\mathbb{U}]\mathbf{v}_{(1)}.\end{aligned} \tag{2.10}$$

Regarding the distributions of the estimators, we have to note:

- $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\varphi}}$ are independents,
- $\widehat{\mu}_1 \rightsquigarrow N_1[\mu_1, \sigma_1^2/d]$, and $\dfrac{d\widehat{\sigma}_1^2}{\sigma_1^2} \rightsquigarrow \chi_{d-2}^2$, both are independents,
- The distribution of $\widehat{\boldsymbol{\beta}}$ is normal $N_{q+1}\left[\boldsymbol{\beta}, \sigma^2(\mathbb{U}\mathbb{U}^T)^{-1}\right]$, and it is independent of $\dfrac{(n-1)\widehat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-q-2}^2$.

2.4. Estimation of the mean of the process and its pdf

From the estimators previously obtained, the estimation of the parametric functions (2.5) is immediate by virtue of Zehna's theorem. In particular, focusing again on the case of a linear exogenous factor, and for the mean functions given by (2.6) and (2.7), it is verified that $\widehat{m}_{\eta,\varphi}(t) = m_{\widehat{\eta},\widehat{\varphi}}(t)$ and $\widehat{m}_{\varphi}(t|s) = m_{\widehat{\varphi}}(t|s)$.

In economic series, it is common to make predictions about the variable of interest based on prior knowledge of the behavior of the series, which implies the use of $\widehat{m}_{\varphi}(t|s)$. For this reason, we focus on describing some aspects concerning this function.

Regarding its distribution, it is possible to obtain its pdf. Indeed, taking into account the distributions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$, and their independence, the pdf of $\mathbf{u}(s,t)^T\widehat{\boldsymbol{\beta}} + \frac{1}{2}\widehat{\sigma}^2(t-s)$ is calculated from the convolution product of their respective densities. A simple change of variable leads to the pdf corresponding to $\widehat{m}_{\varphi}(t|s)$, resulting in

$$f_{\widehat{m}_{\varphi}(t|s)}(m)=\frac{\exp\left(-\frac{\left(\ln\left(\frac{m}{y}\right)-\mathbf{u}(s,t)^T\widehat{\boldsymbol{\beta}}\right)^2}{2\widehat{\sigma}^2A^{\mathbb{U}}_{s,t}}\right)}{m\widehat{\sigma}\sqrt{2\pi A^{\mathbb{U}}_{s,t}}}\sum_{l=0}^{\infty}\frac{\left(\frac{n-q-2}{2}\right)_l}{l!}\left(\frac{\widehat{\sigma}(t-s)}{(n-1)\sqrt{2A^{\mathbb{U}}_{s,t}}}\right)^l H_l\left(\frac{\ln\left(\frac{m}{y}\right)-\mathbf{u}(s,t)^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}\sqrt{2A^{\mathbb{U}}_{s,t}}}\right),m>0 \tag{2.11}$$

where $A^{\mathbb{U}}_{s,t}=\mathbf{u}(s,t)^T(\mathbb{U}\mathbb{U}^T)^{-1}\mathbf{u}(s,t)$,

$$(\alpha)_j=\frac{\Gamma(\alpha+j)}{\Gamma(\alpha)}=\begin{cases}\alpha(\alpha+1)\cdots(\alpha+j-1) & j\geq 1\\ 1 & j=0\,,\end{cases}$$

and H_l is the Hermite polynomial $H_l(x)=l!\sum_{u=0}^{[l/2]}\frac{(-1)^u}{u!(l-2u)!}(2x)^{l-2u}$, $[x]$ being the integer part of x.

2.5. Choice of the exogenous factors

One of the main motivations for considering the linear form of the function $h_{\boldsymbol{\theta}}(t)$ is being able to include in the model external influences to the endogenous variable. However, in practice, there may be various sources of external influence, which motivates having to address the problem of selecting among all possible candidates. To this end, we propose the following strategy, based on the selection of regressors in linear models.

Considering a partition $t_0<t_1<\cdots<t_N$, the Euler-Maruyama scheme applied to the stochastic differential equation (2.1) provides

$$X_j-X_{j-1}=h_{\boldsymbol{\theta}}(t_{j-1})X_{j-1}\Delta_j+\sigma X_{j-1}\Delta W_j,\qquad j=1,\ldots,N,$$

where $X_j=X(t_j)$, $\Delta_j=t_j-t_{j-1}$ and $\Delta W_j=W_{t_j}-W_{t_{j-1}}$.

From this expression, we have

$$\frac{X_j-X_{j-1}}{X_{j-1}\Delta_j^{1/2}}=h_{\boldsymbol{\theta}}(t_{j-1})\Delta_j^{1/2}+\frac{\sigma}{\Delta_j^{1/2}}\Delta W_j,\qquad j=1,\ldots,N,$$

which we can express as $y_j=h_{\boldsymbol{\theta}}(t_{j-1})\Delta_j^{1/2}+u_j$ where

$$y_j=\frac{X_j-X_{j-1}}{X_{j-1}\Delta_j^{1/2}}\text{ and }u_j=\frac{\sigma\Delta W_j}{\Delta_j^{1/2}},$$

with u_j being independent variables distributed according to a normal with zero mean and variance σ^2. Consequently, we have a linear regression model in which the explanatory variables are determined by the exogenous factors F_j (affected by Δ_j at each time instant t_j). The selection of factors can be done by several procedures as:

- Classical procedures, based on sequential *stepwise* methods, using selection criteria such as the partial correlation coefficient or the Akaike statistic.

- Bayesian procedures for variable selection in normal regression models using intrinsic priors for model parameters and the uniform prior for models (see Moreno et al. [32–34] for details). If h exogenous factors are considered, the optimal model will be the one having the highest posterior probability in the class of all possible sub-models $\mathfrak{M}$, $M_0 \in \mathfrak{M}$ being the regression model that does not include any exogenous factor. In this way, if we note by $\mathbf{y}_{N\times 1}$ the data vector of the dependent variable, and by $\mathcal{F}_i$ the $N \times (q_i + 1)$ design matrix for a model M_i with q_i exogenous factors, the posterior probability of model M_j is given by

$$P(M_j|\mathbf{y}, \mathcal{F}_j) = \frac{B_{j0}(\mathbf{y}, \mathcal{F}_j)}{1 + \displaystyle\sum_{M_i \in \mathfrak{M}-M_0} B_{i0}(\mathbf{y}, \mathcal{F}_i)}, \qquad M_j \in \mathfrak{M}, \tag{2.12}$$

where

$$B_{i0}(\mathbf{y}, \mathcal{F}_i) = \frac{2}{\pi}(q_i + 2)^{q_i/2} \int_0^{\pi/2} \frac{\sin^{q_i}\varphi \, (N + (q_i + 2) \, \sin^2\varphi)^{(N-q_i-1)/2}}{(N\mathcal{B}_{i0} + (q_i + 2) \, \sin^2\varphi)^{(N-1)/2}} \, d\varphi$$

is the Bayes factor for comparing the nested models M_0 versus M_i, which depends on $\mathcal{B}_{i0} = \mathrm{RSS}_i/\mathrm{RSS}_0$, the ratio of the residual sum of squares.

2.6. Computing process

Assume that we have observed d sample paths of the endogenous variable, $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,n_i})^T$, $i = 1, \ldots, d$, at time instants $t_{i1} = t_1 < t_{i2} < \ldots < t_{i,n_i}$, and that $F_1, \ldots, F_h$ are exogenous factors that can affect the evolution of said variable, whose values at time instants $t_{i1}, \ldots, t_{i,n_i}$, $i = 1, \ldots, d$, are also known. Next, we detail the algorithm that should be used in practice to select the exogenous factors to include in the model (steps 1 to 6); to estimate the lognormal diffusion process with exogenous factors (steps 7 to 13); to determine its most important characteristics (step 14), and to approximate the pdf of its mean function (step 15):

1) Determine $\mathcal{T} = \bigcup_{i=1}^{d} \{t_{i1}, \ldots, t_{i,n_i}\}$. Let $\mathcal{T} = \{t_1, \ldots, t_N\}$ with $t_1 < t_2 < \ldots < t_N$.

2) For $r = 1, \ldots, N$, compute $\overline{x}_r = \dfrac{\sum_{i=1}^{d} \sum_{j=1}^{n_i} x_{ij} \, \mathbf{1}_{t_r}(t_{ij})}{\sum_{i=1}^{d} \sum_{j=1}^{n_i} \mathbf{1}_{t_r}(t_{ij})}$, where $\mathbf{1}_{t_r}$ is the indicator function of the subset $\{t_r\}$ of the set $\{t_{ij}\}_{i=1,\ldots,d; j=1,\ldots,n_i}$.

3) Calculate $Y = (y_2, \ldots, y_N)^T$, where $y_r = \dfrac{\overline{x}_r - \overline{x}_{r-1}}{\overline{x}_{r-1}(t_r - t_{r-1})^{1/2}}$, $r = 2, \ldots, N$.

4) Consider the 2^h regression models M_i defined by

$$Y = \mathbf{\Delta}^{1/2} \mathcal{F}_i \, \boldsymbol{\theta}_i + U,$$

where $\mathbf{\Delta} = diag((t_2 - t_1), \ldots, (t_N - t_{N-1}))$, $\boldsymbol{\theta}_i = (\theta_0, \ldots, \theta_{q_i})^T$, $\mathcal{F}_i = (\mathbf{1}_{N-1}|\mathbf{F}_{l_1}| \ldots |\mathbf{F}_{l_{q_i}})$ with $\mathbf{F}_{l_d} = (F_{l_d}(t_1), \ldots, F_{l_d}(t_{N-1}))^T$, $d = 1, \ldots, q_i$, and $U \rightsquigarrow N_{N-1}[0, \sigma^2 \mathbf{I}_{N-1}]$; M_0 being the regression model that does not include any exogenous factor, that is, $Y = \mathbf{\Delta}^{1/2} \mathcal{F}_0 \, \boldsymbol{\theta}_0 + U$, where $\mathcal{F}_0 = \mathbf{1}_{N-1}$ and $\boldsymbol{\theta}_0 = \theta_0$.

5) For $i = 0, \ldots, 2^h - 1$, compute the residual sum of squares $RSS_i = Y^T(\mathbf{I}_{N-1} - \mathbf{H}_i)Y$ of the regression model M_i, where $\mathbf{H}_i = \mathbf{\Delta}^{1/2}\mathcal{F}_i\left(\mathcal{F}_i^T\mathbf{\Delta}\mathcal{F}_i\right)^{-1}\mathcal{F}_i^T\mathbf{\Delta}^{1/2}$.
6) For $i = 1, \ldots, 2^h - 1$, use (2.12) to calculate $P(M_i|\mathbf{y}, \mathcal{F}_i)$ from $\mathcal{B}_{i0} = \mathrm{RSS}_i/\mathrm{RSS}_0$ and select those exogenous factors that provide the highest posterior probability, which we will simply denote $F_1, \ldots, F_q$.
7) If $X(t_1) \rightsquigarrow \Lambda_1[\mu_1, \sigma_1^2]$, determine $\widehat{\mu}_1 = \dfrac{1}{d}\displaystyle\sum_{i=1}^{d} \ln x_{i1}$ and $\widehat{\sigma}_1^2 = \dfrac{1}{d}\displaystyle\sum_{i=1}^{d}(\ln x_{i1} - \widehat{\mu}_1)^2$. If $X(t_1) = x_0$, consider $\widehat{\mu}_1 = \ln x_{i1}$.
8) Calculate the vector $\mathbf{v}_{(1)} = \left[\mathbf{v}_1^T|\cdots|\mathbf{v}_d^T\right]^T$ where

$$\mathbf{v}_i^T = \left[(t_{i2} - t_{i1})^{-1/2} \ \ln\frac{x_{i2}}{x_{i1}}\,\Big|\ldots\Big|\,(t_{i,n_i} - t_{i,n_{i-1}})^{-1/2} \ \ln\frac{x_{i,n_i}}{x_{i,n_{i-1}}}\right], \quad i = 1, \ldots, d.$$

9) For $k = 1, \ldots, q$, compute the cubic spline $S_k(t)$ that interpolates the knots $(t_1, F_k(t_1))$, …, $(t_N, F_k(t_N))$, preserving the shape of the data for each exogenous factor.
10) For $i = 1, \ldots, d$ and $j = 1, \ldots, n_i$, approximate $\mathbf{u}(t_{ij}, t_{ij-1})$ by

$$\widetilde{\mathbf{u}}(t_{ij}, t_{ij-1}) = \left(t_{ij} - t_{i,j-1}, \int_{t_{i,j-1}}^{t_{ij}} S_1(\tau)d\tau, \ldots, \int_{t_{i\,j-1}}^{t_{ij}} S_q(\tau)d\tau\right)^T.$$

11) Approximate the block-column matrix $\mathbb{U}$ by $\widetilde{\mathbb{U}} = [\widetilde{\mathbf{U}}_1|\cdots|\widetilde{\mathbf{U}}_d]$ where

$$\widetilde{\mathbf{U}}_i = \left[(t_{i2} - t_{i1})^{-1/2}\,\widetilde{\mathbf{u}}(t_{i1}, t_{i2})\,\Big|\ldots\Big|\,(t_{i,n_i} - t_{i,n_{i-1}})^{-1/2}\,\widetilde{\mathbf{u}}(t_{i,n_{i-1}}, t_{i,n_i})\right].$$

12) Calculate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_q)^T = (\widetilde{\mathbb{U}}\widetilde{\mathbb{U}}^T)^{-1}\widetilde{\mathbb{U}}\,\mathbf{v}_{(1)}$ and $\widehat{\sigma}^2 = \dfrac{1}{n-1}\,\mathbf{v}_{(1)}^T\left[\mathbf{I}_{n-1} - \widetilde{\mathbb{U}}^T(\widetilde{\mathbb{U}}\widetilde{\mathbb{U}}^T)^{-1}\widetilde{\mathbb{U}}\right]\mathbf{v}_{(1)}$.
13) Compute $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_0, \ldots, \widehat{\theta}_q)^T$ where $\widehat{\theta}_0 = \widehat{\beta}_0 + \widehat{\sigma}^2/2$ and $\widehat{\theta}_j = \widehat{\beta}_j$, $j = 1, \ldots, q$.
14) Obtain the estimates of the n-th moment, the mode, and quantile functions for distributions $X(t)$ or $X(t) \mid X(s) = y$, with $s < t$, from $G_{\widehat{\xi}}^{\lambda}(t|z, \tau)$ in (2.5) according to the values of λ, τ, and z in Table 1, approximating $H_{\widehat{\xi}}(\tau, t)$ by

$$\widetilde{H}_{\widehat{\xi}}(\tau, t) = \widetilde{\mathbf{u}}(\tau, t)^T\,\widehat{\theta} + (\widehat{\sigma}^2/2)\,(t - \tau) = \widetilde{\mathbf{u}}(\tau, t)^T\,\widehat{\beta},$$

where

$$\widetilde{\mathbf{u}}(s, t) = \left(t - s, \int_s^t S_1(\tau)d\tau, \ldots, \int_s^t S_q(\tau)d\tau\right)^T,$$

and replacing t_0 by t_1 and μ_0 by $\widehat{\mu}_1$ in Table 1.
15) Approximate the values of the pdf of the mean function, $f_{m_{\widehat{\varphi}(t|s)}}(m)$, $m > 0$, considering a finite number of terms in the series expansion given by (2.11), and replacing $\mathbb{U}$ by $\widetilde{\mathbb{U}}$ and $\mathbf{u}(s, t)$ by $\widetilde{\mathbf{u}}(s, t)$.

3. Application

In this section we apply the proposed model for analyzing the behavior of the basic CPI in Colombia. For this purpose, we consider quarterly data between March 2003 and September 2024 for this index and for a set of macroeconomic indicators (see Table 2) that have been selected a priori as possible exogenous factors that may affect the CPI behavior. Prior to this analysis, a thorough correlation analysis was conducted to identify and exclude variables exhibiting high degrees of correlation, thereby mitigating risks associated with collinearity or multicollinearity. Due to the diversity of the multiple variables under study, the data had to be taken from different sources: *Banco de la República de Colombia* (showed here as "Central Bank"); *Departamento Administrativo Nacional de Estadística (DANE)*; and *Superintendencia Financiera de Colombia*.

The selection of this period is not arbitrary; it corresponds to the beginning of the implementation of the inflation targeting regime in Colombia, a key strategy within the macroeconomic stabilization policy adopted by the Central Bank. This approach, which prioritizes the control of inflation through tools such as interest rates, establishes a structured framework for analyzing the dynamics of the CPI in a more consistent context and aligned with the country's current economic policy strategy. The 2003–2024 range represents a homogeneous period in terms of macroeconomic strategy, enabling a more precise analysis of the effects of exogenous factors on inflation within the inflation-targeting regime. This approach ensures that the data used is consistent with the current economic and political context, and that the model's results are relevant for present-day decision-making.

The use of absolute and relative variables (r) is justified by the requirement of modeling the long range structural relations and also the short range dynamics of the inflation in Colombia. Such strategy allowed us to consider nonstationary economic series, showing different velocities of the transmissions of the economic effects and, finally, showing the real economic politics of the country, where all the above variables might be relevant for the inflationary dynamics.

In the following sections we will see that the CPI in Colombia can be appropriately modeled by a nonhomogeneous lognormal diffusion process with $h_\theta(t)$ a linear time function of a subset of the exogenous factors included in Table 2.

Table 2. Macroeconomic indicators candidates to be exogenous factors.

Oil price (Brent ref.), X_1	RMR (representative market rate), X_2
Houses final consumption expenditure, X_3	Manufacturing real production, X_4
Political intervention rate, X_5	Unemployment rate, X_6
Occupation rate, X_7	Oil production, X_8
Internal demand, X_9	Gross capital formation, X_{10}
Gross fixed capital formation, X_{11}	Exports, X_{12}
Federal funds effective interest rate, X_{13}	Coffee production (r), X_{14}
Retail sales, X_{15}	Government final consumption expenditure (r), X_{16}
Overall participation rate, X_{17}	Consumption interest rate (r), X_{18}

The steps required to use the model are summarized as follows:

- **Specification of the model. Selection of exogenous factors.** The specification of the model makes reference to the selection of the exogenous factors defining the linear function $h_\theta(t)$. It will be carried out by means of the Bayesian method showed in Section 2.5. This step must always be done before the estimation of the model. Indeed, economic conditions may change in each time interval, so exogenous factors may vary from one time interval to another. Otherwise, note that the procedure guarantees the selection of those variables with more influence on the endogenous variable. In this way, redundant information is avoided, as well as variables with little influence over CPI or with information contained in other factors. This is intended to respect the principle of parsimony, according to which the simplest model that leads to similar conclusions should be chosen.

- **ML estimation of the model.** Once the factors have been chosen, the estimation of the parameters can start. In order to do this, expressions (2.9) (in the case of having multiple sample paths of the process, when the initial distribution is not degenerate) and (2.10) (parameters of the transition distributions of the process) must be employed. In the particular case of this application study, since only one sample path is available, the initial distribution is then degenerate and only the parameters of the transition distributions must be estimated. Once the estimates of the parameters have been calculated, it is possible to obtain those corresponding to some characteristics of the model, such as those indicated in Table 1 (and whose functional form is given in (2.5)). In the specific case of the mean function, it can also be obtained an estimate of the density function of its estimator, with the expression given in (2.11).
- **Use the model to make predictions.** Perhaps the main virtue that can be asked of a model such as the one proposed in this article is to be able to make accurate forecasts on the variable of interest, in this case the CPI in Colombia. In practice, monetary authorities make predictions about the economic variables attending to multiple scenarios, by considering projections for future times. In particular, the Central Bank establishes, in each report on monetary policy, the prediction of the next eight quarters, updating those made in the previous report. In the application study, this approach has been simulated. Thus:

  – First, the model has been estimated according to available data until March 2022. This time was chosen as a reference due to an observed change of tendency in the CPI evolution, followed by an acceleration of the inflation. Once the model has been estimated, projections until March 2024 have been calculated.
  – Afterward, two updates have been made; one considering data until June 2022, and the other taking data until September 2022. The projections have been updated until June 2024 and September 2024, respectively.

  It is noteworthy that, for each estimated model, a previous study for the selection of the exogenous factors has been carried out, resulting in the same chosen factors for each case. On the other hand, the values considered for the exogenous variables used to make the predictions were those actually observed in each quarter. This validates whether the model is capable of reliably predicting CPI values within a given economic framework defined by the values of the exogenous factors.

Finally, trying to make future projections, it seems obvious that different scenarios must be taken into account. Such scenarios depend on different economical situations that could appear at future times. Therefore, to finish the application study, some scenarios with variations of certain exogenous factors are considered. With this, the ability of the model to predict the behavior of the CPI under different variations of the factors which might represent fluctuations in the markets or internal policies is tested.

3.1. Observed data until March 2022

According to the steps described previously, from the data on the quarterly relative variation of the CPI, and those of the exogenous factors in Table 2 from March 2003 to March 2022, a Bayes selection of the factors was carried out by using (2.12) to evaluate the posterior probabilities of the $2^{18} = 262144$ plausible regression models to explain the relative variation of the endogenous variable. It is noteworthy that although the number of models is high, the implementation carried out allows obtaining the best subset of exogenous factors in a fairly short time, specifically 40.27 seconds.

Table 3 contains the four best models. For each one, the posterior probability and the ratio of the square sum of the residuals of the model and the null one (that without exogenous factors) are shown.

Table 3. Four best subsets of factors using the Bayes selection procedure (data from March 2003 to March 2022).

M_j	$P(M_j \vert \mathbf{y}, \mathbf{F}_j)$	$\mathcal{B}_{j,0}$
$\{X_2, X_5, X_7, X_{14}, X_{18}\}$	0.001033	0.6202
$\{X_1, X_6, X_{14}, X_{18}\}$	0.001006	0.6446
$\{X_2, X_6, X_{14}, X_{15}, X_{18}\}$	0.000841	0.6240
$\{X_2, X_3, X_6, X_{14}, X_{18}\}$	0.000755	0.6260

According to Table 3, the optimal subset of exogenous factors to explain the relative variation of CPI are "RMR (representative market rate)" (X_2), "Political intervention rate" (X_5), "Occupation rate" (X_7), "Coffee production (r)" (X_{14}), and "Consumption interest rate (r)" (X_{18}), with a posterior probability equal to 0.0001033. The estimated optimal regression model obtained is

$$\widehat{Y} = -0.071417 + 0.000285\,X_2 + 0.000753\,X_5 + 0.000421\,X_7 + 0.011104\,X_{14} + 0.033064\,X_{18},$$

with an R-squared (percent of variance explained) of $100(1 - \mathcal{B}_{opt,0})\% = 38\%$, and an adjusted R-squared of 33.57%.

The set of selected exogenous factors includes tools of the monetary policies (intervention rate), an indicator of the labor market (occupation rate), and factors showing the external and internal economical conditions, like the representative market rate (RMR) and the consumption interest rate. We emphasize the selection of the coffee production, given its historical and economical role in the history of Colombia. Therefore, the selection captures the interaction between monetary policy, dynamics of the labor market, and external factors, emphasizing the key role of the exporting agricultural sector in the configuration of the national economic framework.

With these results, and in order to fit the quarterly CPI data, we will consider the nonhomogeneous lognormal diffusion process $X(t)$ given by (2.2) with $h_{\boldsymbol{\theta}}(t) = \theta_0 + \sum_{j=1}^{5} \theta_j F_j(t)$, with $F_1 = X_2$, $F_2 = X_5$, $F_3 = X_7$, $F_4 = X_{14}$, and $F_5 = X_{18}$ the exogenous factors selected from Table 3.

According to Section 2.3, the ML estimation of this model requires evaluating the integral of the factors between every two correlative time instants. To do this, a piecewise cubic Hermite interpolation polynomial that preserves the shape of the data for each factor is determined. Later, the integral of said polynomial is evaluated.

The ML estimate of φ is given by $\widehat{\varphi} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_5, \widehat{\sigma}^2)^T$, being $\hat{\beta}_0 = -8.652283 \times 10^{-2}$, $\hat{\beta}_1 = 3.237701 \times 10^{-6}$, $\hat{\beta}_2 = 7.960460 \times 10^{-4}$, $\hat{\beta}_3 = 6.126848 \times 10^{-4}$, $\hat{\beta}_4 = 3.891693 \times 10^{-3}$, $\hat{\beta}_5 = 4.324282 \times 10^{-2}$, and $\widehat{\sigma}^2 = 2.568703 \times 10^{-5}$. Finally, the ML estimate of θ is $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_5)^T$, where $\hat{\theta}_0 = -8.650998 \times 10^{-2}$ and $\hat{\theta}_j = \hat{\beta}_j$, $j = 1, \ldots, 5$.

The estimation of the parameters of the model allows us to obtain the one corresponding to different characteristics of the model. Specifically, in this section we show, among other characteristics, the estimated mean and α-quantile ($\alpha = 0.025, 0.975$) functions associated with the estimated probability distribution of the diffusion process, given its initial value. Figure 1 displays the estimated values of these parametric functions. The estimated conditional mean captures the various observed trends and fluctuations in the CPI, successfully reflecting the slower and faster periods of growth over the time period considered.

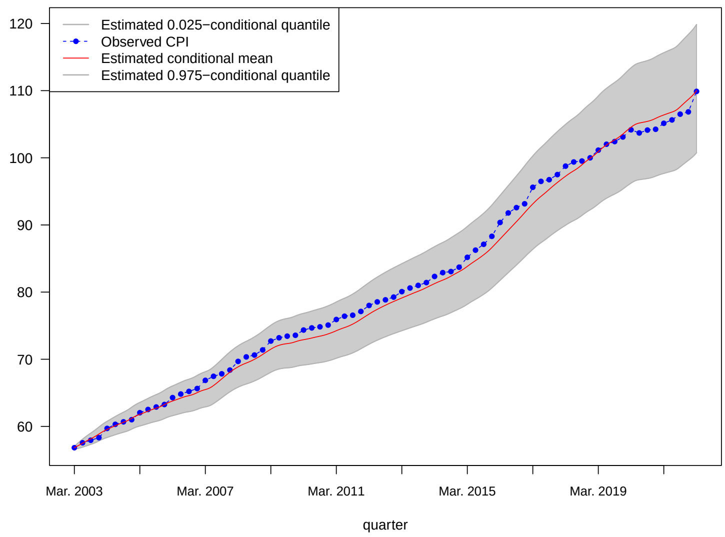


Figure 1. Estimated mean and α-quantile ($\alpha = 0.025, 0.975$) functions together with the observed CPI data (data from March 2003 to March 2022).

To assess the goodness of fit, one can use the Relative Absolute Error (RAE) of the mean according to the expression

$$\text{RAE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|m_i - \widehat{m}_\varphi(t_1|t_0)|}{m_i},$$

where m_i is the value of the observed mean at time t_i, and $\widehat{m}_\varphi(t_i|t_0)$ is the value of the estimated mean function of the process at the same time instant. In this case, the value obtained of the RAE is 0.011854.

In addition, we have approached the pdf of the ML estimator of the mean function conditioned on the first time instant, $f_{\widehat{m}_\varphi(t|t_0)}$. To this end, Eq (2.11) has been used considering the first 20 terms in the series expansion that appears in said expression. All calculated pdfs are bell-shaped and approximately symmetrical. It is noticed that the standard deviation of $\widehat{m}_\varphi(t|t_0)$ grows significantly over the quarters, as can be seen in Figure 2. Inset figure shows the pdfs for some quarters between December 2003 and March 2022. Table 4 summarizes some characteristics calculated from them.

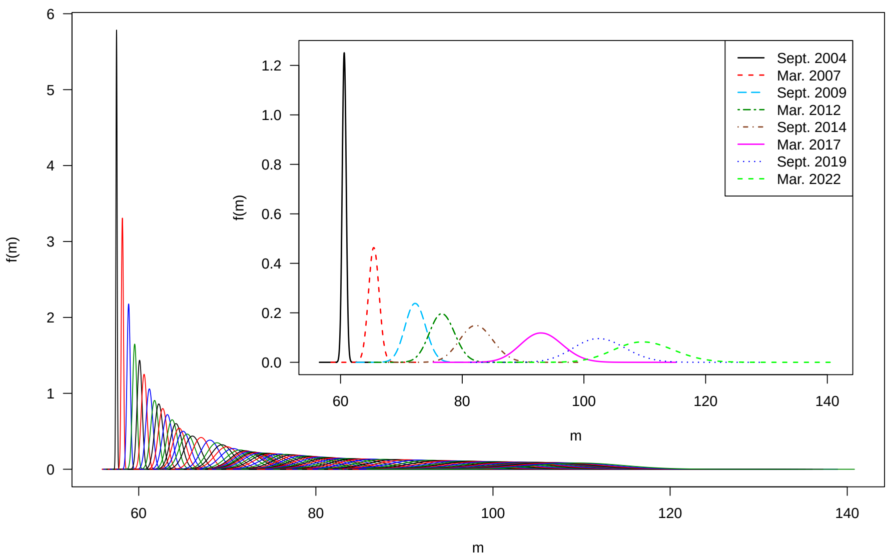


Figure 2. Quarterly pdfs of the ML estimator of the conditional mean of CPI given its initial value. Inset: densities for some quarters (data from March 2003 to March 2022).

Table 4. For some quarters, main characteristics of the pdf of the ML estimator of the conditional mean of CPI given its initial value (data from March 2003 to March 2022).

	Sept. 2004	Mar. 2007	Sept. 2009	Mar. 2012
0.025-quantile	59.978026	63.790800	69.068131	72.772410
mean	61.257831	66.229241	73.089864	77.702973
median	60.600066	65.456969	72.279846	76.654827
mode	60.598376	65.448012	72.246171	76.607481
0.975-quantile	61.186609	67.045504	75.415793	80.421508
standard deviation	0.730479	1.153506	1.862250	2.286174
	Sept. 2014	Mar. 2017	Sept. 2019	Mar. 2022
0.025-quantile	77.188523	86.625207	94.648334	100.677583
mean	83.467394	94.488564	104.242556	111.784800
median	82.276553	93.004139	102.502187	109.783977
mode	82.193457	92.884384	104.807747	109.574761
0.975-quantile	87.255919	99.239365	110.212547	118.758235
standard deviation	2.935494	3.687006	4.525536	5.259546

Finally, regarding the predictions, it can be seen (Table 5) that those made by the model for the next eight quarters (providing almost identical values for the three characteristics under consideration) lead to values that are very similar to the real observed values, and notably improve the projections made by the Central Bank.

Table 5. CPI forecasts for the last eight quarters using the estimated conditional probability distribution of the diffusion process vs. forecasts of the Central Bank and real observed values (data from March 2003 to March 2022).

		June 2022	Sept. 2022	Dec. 2022	Mar. 2023
Model	$\widehat{E}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	111.77	114.24	117.33	120.54
Predictions	$\widehat{Me}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	111.76	114.24	117.33	120.53
	$\widehat{Mo}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	111.76	114.23	117.32	120.52
	$\widehat{C}_{0.025}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	110.66	112.65	115.33	118.16
	$\widehat{C}_{0.975}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	112.88	115.86	119.36	122.95
Forecast of Central Bank		111.43	112.42	113.19	115.75
Observed values		112.04	114.47	117.01	121.44
		June 2023	Sept. 2023	Dec. 2023	Mar. 2024
Model	$\widehat{E}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	123.15	125.35	127.76	130.04
Predictions	$\widehat{Me}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	123.14	125.34	127.75	130.03
	$\widehat{Mo}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	123.12	125.32	127.73	130.00
	$\widehat{C}_{0.025}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	120.44	122.33	124.44	126.43
	$\widehat{C}_{0.975}[X(t)\|X(\text{Mar. } 2022) = 109.89]$	125.91	128.43	131.15	133.74
Forecast of Central Bank		116.85	117.89	118.60	120.63
Observed values		123.82	125.36	126.86	129.66

It is especially noteworthy the fact that the official projections get worse as the prediction instant advances in time, while the predictions made by the model remain at values close to the real observed values, which indicates the good performance of the model in the long-term scenario. Figure 3 provides a fairly clear visualization of this comment. In this graph, the predicted CPI values for each quarter are represented together with the actual observed values and the projections made by the Central Bank. The vertical sections in the shaded area represent the α-quantile intervals ($\alpha = 0.025, 0.975$) calculated for the distributions conditioned on the value of the process observed in March 2022.

On the other hand, Figure 4 shows the pdfs of the ML estimator of the conditional mean for the last eight quarters given the observed CPI value in March 2022. It can be seen that all density functions are bell-shaped and approximately symmetrical. From these densities, the most common quantiles, which provide appropriate forecasts of the CPI, are summarized in Table 6. Finally, Figure 5 shows the forecasts provided by the obtained means, together with the observed values and the projections made by the Central Bank. The interpretation of the shaded area is identical as that of Figure 3.

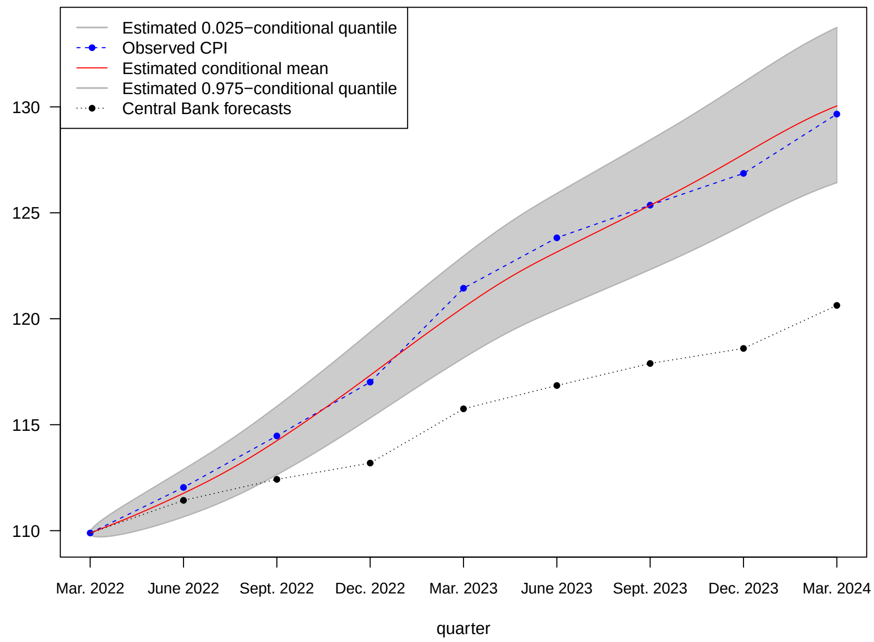


Figure 3. CPI forecasts for the last eight quarters provided by the estimated conditional mean of the model fitted given the value of CPI in March 2022 (data from March 2003 to March 2022).

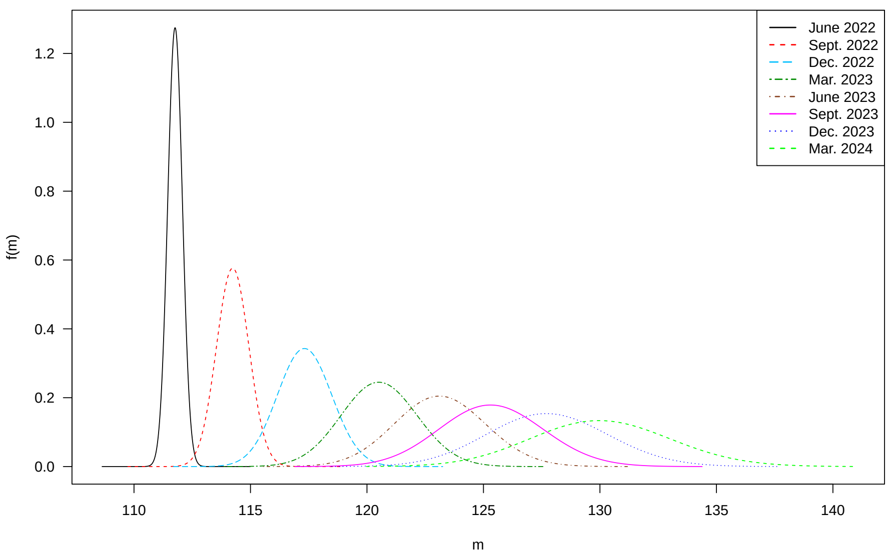


Figure 4. For the last eight quarters, pdfs of the ML estimator of the conditional mean function given the value in March 2022 (data from March 2003 to March 2022).

Table 6. CPI forecasts for the last eight quarters using the pdf of the ML estimator of the conditional mean function vs. forecasts of the Central Bank and real observed values (data from March 2003 to March 2022).

		June 2022	Sept. 2022	Dec. 2022	Mar. 2023
Model	$E[\widehat{m}(t\vert\text{Mar. 2022})]$	112.14	114.60	117.67	120.87
Predictions	$Me[\widehat{m}(t\vert\text{Mar. 2022})]$	111.76	114.24	117.32	120.52
	$Mo[\widehat{m}(t\vert\text{Mar. 2022})$	111.76	114.23	117.31	120.50
	$C_{0.025}[\widehat{m}(t\vert\text{Mar. 2022})]$	111.15	112.89	115.06	117.37
	$C_{0.975}[\widehat{m}(t\vert\text{Mar. 2022})]$	112.36	115.58	119.58	123.69
Forecast of Central Bank		111.43	112.42	113.19	115.75
Observed values		112.04	114.47	117.01	121.44
		June 2023	Sept. 2023	Dec. 2023	Mar. 2024
Model	$E[\widehat{m}(t\vert\text{Mar. 2022})]$	123.52	125.75	128.16	130.43
Predictions	$Me[\widehat{m}(t\vert\text{Mar. 2022})]$	123.13	125.33	127.74	130.02
	$Mo[\widehat{m}(t\vert\text{March 2022})$	123.10	125.29	127.69	129.95
	$C_{0.025}[\widehat{m}(t\vert\text{Mar. 2022})]$	119.37	121.04	122.76	124.30
	$C_{0.975}[\widehat{m}(t\vert\text{Mar. 2022})]$	126.93	129.68	132.81	135.88
Forecast of Central Bank		116.85	117.89	118.60	120.63
Observed values		123.82	125.36	126.86	129.66

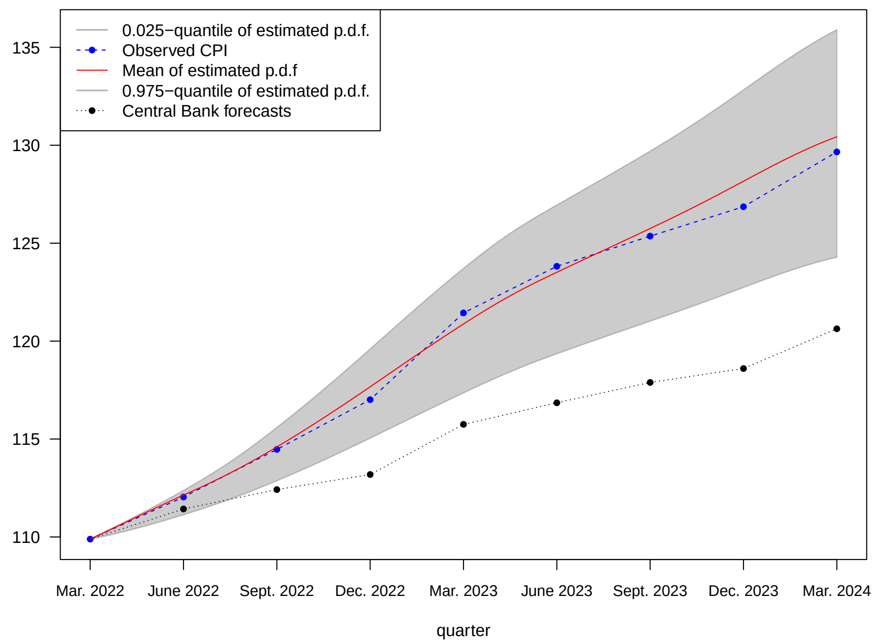


Figure 5. CPI forecasts for the last eight quarters provided by the mean of the ML estimator of the conditional mean function given the value of CPI in March 2022 (data from March 2003 to March 2022).

3.2. First update: Data until June 2022

In this case, data from March 2003 to June 2022 is considered. The selection of factors led to the same subset as in the previous case, even showing greater posterior probability difference with the second selected model than in the previous case (see Table 7).

Table 7. Four best subsets of factors using the Bayes selection procedure (data from March 2003 to June 2022).

M_j	$P(M_j \vert \mathbf{y}, \mathbf{F}_j)$	$\mathcal{B}_{j,0}$
$\{X_2, X_5, X_7, X_{14}, X_{18}\}$	0.001220	0.6113
$\{X_2, X_6, X_{14}, X_{15}, X_{18}\}$	0.000975	0.6154
$\{X_2, X_5, X_{14}, X_{17}, X_{18}\}$	0.000905	0.6167
$\{X_2, X_3, X_6, X_{14}, X_{18}\}$	0.000881	0.6172

The estimated optimal regression model now becomes

$$\hat{Y} = -0.079523 + 0.000031\, X_2 + 0.000695\, X_5 + 0.000426\, X_7 \ + 0.010918\, X_{14} + 0.040885\, X_{18},$$

with R-squared of 39% and an adjusted R-squared equal to 0.346.

The ML estimates were $\hat{\theta}_0 = -9.063434 \times 10^{-2}, \hat{\theta}_1 = 3.290187 \times 10^{-6}, \hat{\theta}_2 = 7.761643 \times 10^{-4}, \hat{\theta}_3 = 6.201280 \times 10^{-4}, \hat{\theta}_4 = 3.516982 \times 10^{-3}, \hat{\theta}_5 = 4.733693 \times 10^{-2}$, and $\hat{\sigma}^2 = 2.523352 \times 10^{-5}$. Figure 6 shows the estimated mean and α-quantile parametric functions together with the observed values of the CPI. Again, a good fit to the observed data can be observed, with the estimated mean capturing the oscillations observed in the CPI in the time interval considered. The value of the RAE in this case is equal to 0.0124.

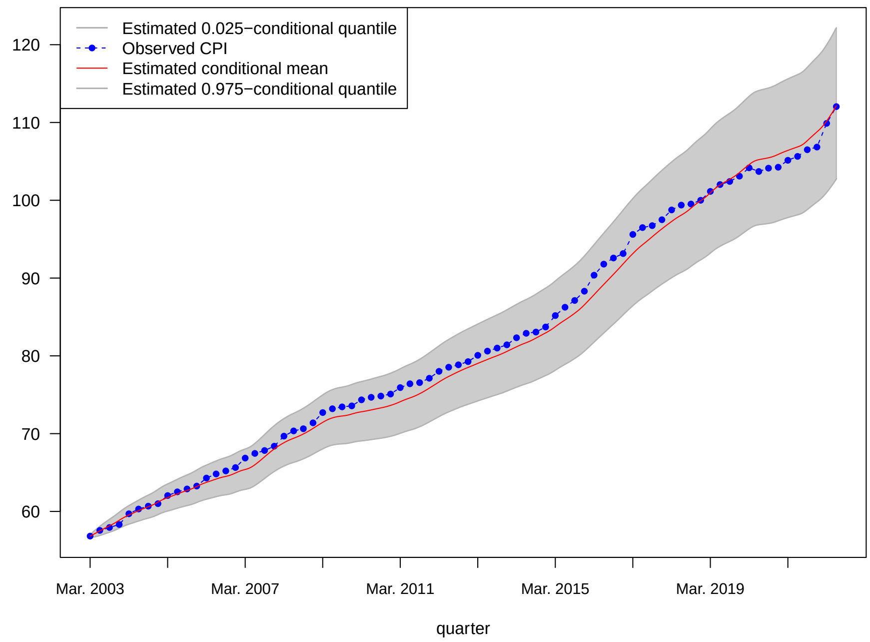


Figure 6. Estimated mean and α-quantile ($\alpha = 0.025, 0.975$) functions together with the observed CPI in the first update case (data from March 2003 to June 2022).

As in the previous case, the model performs very well (see results in Table 8) given predictions closer to the finally observed data than those made by the Central Bank. This behavior remains the same across all the prediction intervals, being even more noticeable in the long-term. Figures 7 and 8 show, respectively, predicted CPI values together with projections made by the Central Bank and the forecast from the obtained mean, all together with the central 95% confidence band. On the other hand, the pdfs of the ML estimator of the conditional mean function for each of the last eight quarters, given the observed CPI in June 2022, are also shown in Figure 9, with the central tendency characteristics and some relevant quantiles summarized in Table 9.

Table 8. CPI forecasts for the last eight quarters using the estimated conditional probability distribution of the diffusion process vs. forecasts of the Central Bank and real observed values (data from March 2003 to June 2022).

		Sept. 2022	Dec. 2022	Mar. 2023	June 2023
Model	$\widehat{E}[X(t)\|X(\text{June 2022}) = 112.04]$	114.62	117.81	121.11	123.75
Predictions	$\widehat{Me}[X(t)\|X(\text{June 2022}) = 112.04]$	114.61	117.81	121.11	123.75
	$\widehat{Mo}[X(t)\|X(\text{June 2022}) = 112.04]$	114.61	117.81	121.10	123.74
	$\widehat{C}_{0.025}[X(t)\|X(\text{June 2022}) = 112.04]$	113.49	116.18	119.06	121.33
	$\widehat{C}_{0.975}[X(t)\|X(\text{June 2022}) = 112.04]$	115.75	119.46	123.19	126.21
Forecast of Central Bank		113.80	114.91	117.91	119.62
Observed values		114.47	117.01	121.44	123.82
		Sept. 2023	Dec. 2023	Mar. 2024	June 2024
Model	$\widehat{E}[X(t)\|X(\text{June 2022}) = 112.04]$	125.94	128.32	130.46	132.27
Predictions	$\widehat{Me}[X(t)\|X(\text{June 2022}) = 112.04]$	125.93	128.31	130.45	132.25
	$\widehat{Mo}[X(t)\|X(\text{June 2022}) = 112.04]$	125.91	128.29	130.43	132.23
	$\widehat{C}_{0.025}[X(t)\|X(\text{June 2022}) = 112.04]$	123.19	125.25	127.09	128.62
	$\widehat{C}_{0.975}[X(t)\|X(\text{June 2022}) = 112.04]$	128.73	131.44	133.89	135.99
Forecast of Central Bank		121.01	121.87	123.89	124.89
Observed values		125.36	126.86	129.66	131.26

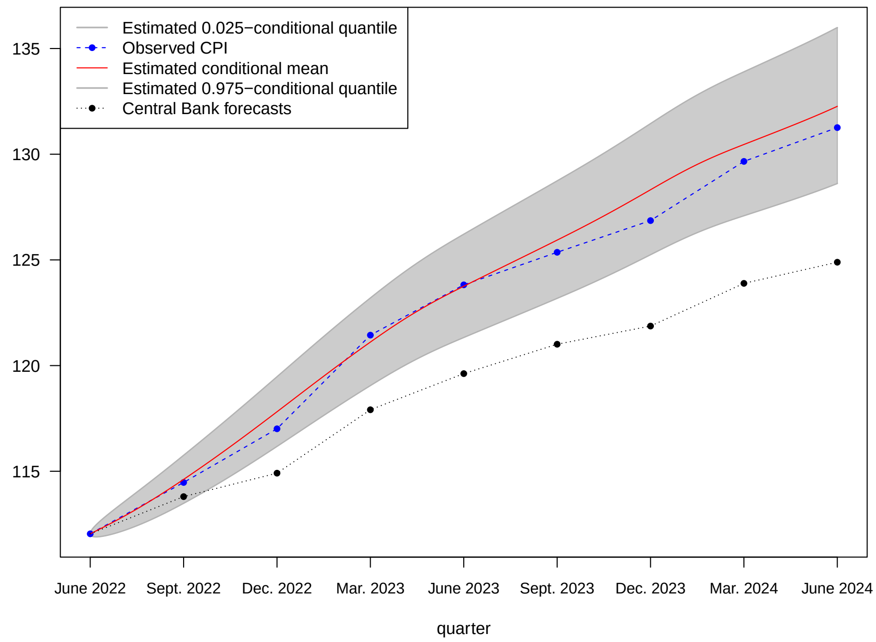


Figure 7. CPI forecasts for the last eight quarters provided by the estimated conditional mean of the model fitted given the value of CPI in June 2022 (data from March 2003 to June 2022).

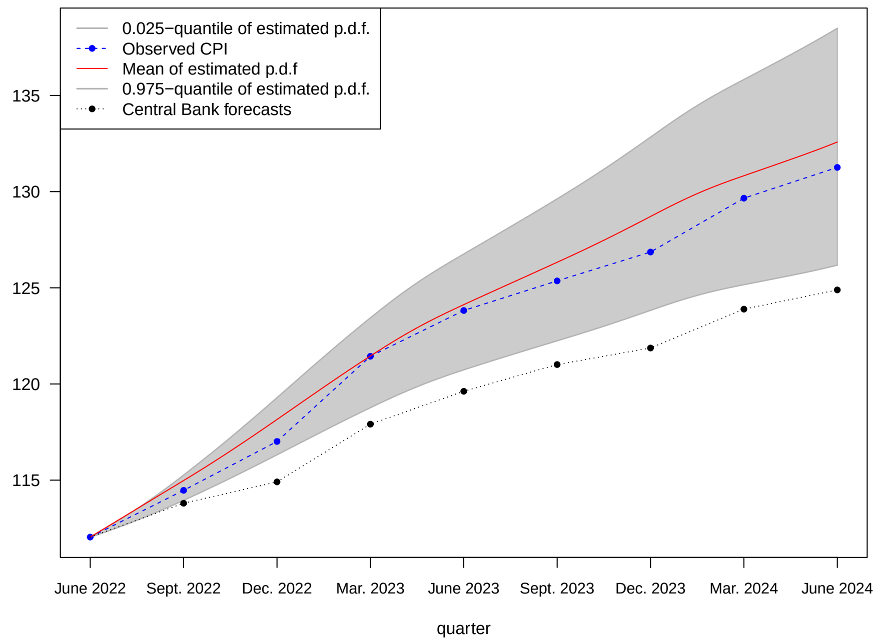


Figure 8. CPI forecasts for the last eight quarters provided by the mean of the ML estimator of the conditional mean function given the value of CPI in June 2022 (data from March 2003 to June 2022).

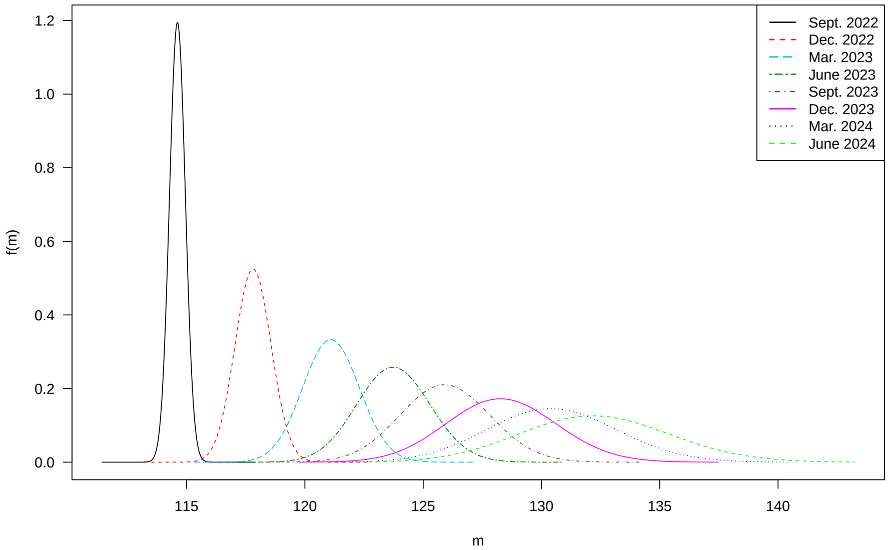


Figure 9. For the last eight quarters, pdfs of the ML estimator of the conditional mean function given the value in June 2022 (data from March 2003 to June 2022).

Table 9. CPI forecasts for the last eight quarters using the pdf of the ML estimator of the conditional mean function vs. forecasts of the Central Bank and real observed values (data from March 2003 to June 2022).

		Sept. 2022	Dec. 2022	Mar. 2023	June 2023
Model	$E[\widehat{m}(t\|\text{June 2022})]$	114.98	118.15	121.45	124.12
Predictions	$Me[\widehat{m}(t\|\text{June 2022})]$	114.61	117.81	121.10	123.74
	$Mo[\widehat{m}(t\|\text{June 2022})$	114.61	117.80	121.09	123.72
	$C_{0.025}[\widehat{m}(t\|\text{June 2022})]$	113.96	116.33	118.77	120.74
	$C_{0.975}[\widehat{m}(t\|\text{June 2022})]$	115.25	119.28	123.43	126.75
Forecast of Central Bank		113.80	114.91	117.91	119.62
Observed values		114.47	117.01	121.44	123.82
		Sept. 2023	Dec. 2023	Mar. 2024	June 2024
Model	$E[\widehat{m}(t\|\text{June 2022})]$	126.33	128.71	130.83	132.58
Predictions	$Me[\widehat{m}(t\|\text{June 2022})]$	125.92	128.30	130.44	132.24
	$Mo[\widehat{m}(t\|\text{June 2022})$	125.89	128.26	130.38	132.16
	$C_{0.025}[\widehat{m}(t\|\text{June 2022})]$	122.25	123.83	125.17	126.19
	$C_{0.975}[\widehat{m}(t\|\text{June 2022})]$	129.62	132.83	135.82	138.49
Forecast of Central Bank		121.01	121.87	123.89	124.89
Observed values		125.36	126.86	129.66	131.26

3.3. Second update: Data until September 2022

In the second updated case, with observed data until September 2022, the selected factors were again the same as in the two previous cases, as it is shown in Table 10.

Table 10. Four best subsets of factors using the Bayes selection procedure (data from March 2003 to September 2022).

M_j	$P(M_j \mid \mathbf{y}, \mathbf{F}_j)$	$\mathcal{B}_{j,0}$
$\{X_2, X_5, X_7, X_{14}, X_{18}\}$	0.001531	0.5808
$\{X_2, X_5, X_{14}, X_{17}, X_{18}\}$	0.001131	0.5860
$\{X_2, X_6, X_{14}, X_{15}, X_{18}\}$	0.001064	0.5870
$\{X_2, X_6, X_8, X_{13}, X_{14}, X_{18}\}$	0.001036	0.5664

The estimated optimal regression model, with an R-squared of 42% (adjusted R-squared equal to 0.378), becomes

$$\hat{Y} = -0.083114 + 0.000322\, X_2 + 0.000693\, X_5 + 0.000433\, X_7 + 0.010987\, X_{14} + 0.043761\, X_{18}.$$

The ML estimates were $\hat{\theta}_0 = -8.950815 \times 10^{-2}, \hat{\theta}_1 = 3.257744 \times 10^{-6}, \hat{\theta}_2 = 7.750878 \times 10^{-4}, \hat{\theta}_3 = 6.173011 \times 10^{-4}, \hat{\theta}_4 = 3.540570 \times 10^{-3}, \hat{\theta}_5 = 4.642698 \times 10^{-2}$, and $\hat{\sigma}^2 = 2.491456 \times 10^{-5}$. Parametric estimated mean and α-quantiles functions are shown in Figure 10. The same comments made in the previous sections can be made. The RAE in this case is equal to 0.0121.

In this case, the predictions made by the model remain closer to the observed data and still outperform the Central Bank's projections, which mostly even fall outside the confidence band for the model, as it can be seen in Table 11 and Figure 11. The summary of central tendency characteristics and some relevant quantiles are shown in Table 12, related with Figures 12 and 13.

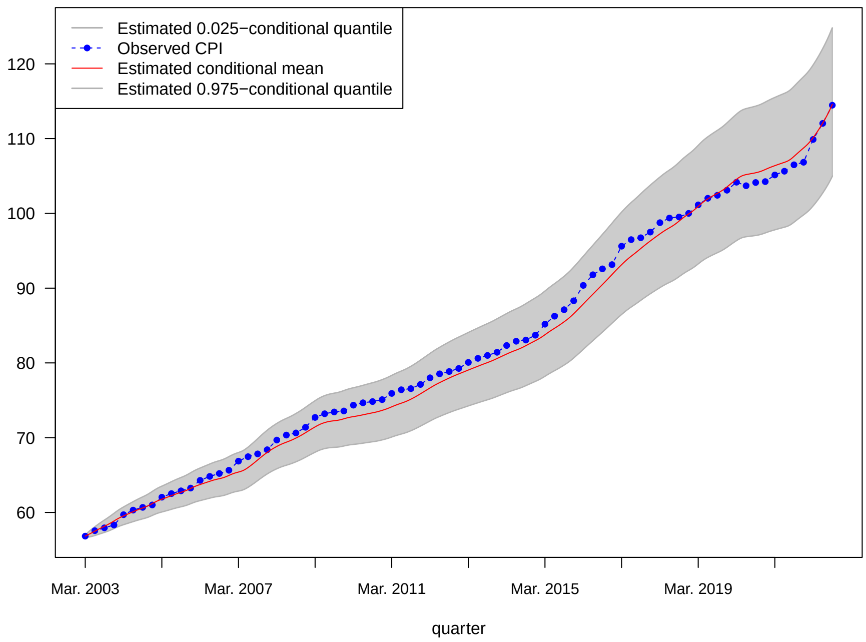


Figure 10. Estimated mean and α-quantile ($\alpha = 0.025, 0.975$) functions together with the observed CPI in the second update case (data from March 2003 to September 2022).

Table 11. CPI forecasts for the last eight quarters using the estimated conditional probability distribution of the diffusion process vs. forecasts of the Central Bank and real observed values (data from March 2003 to September 2022).

		Dec. 2022	Mar. 2023	June 2023	Sept. 2023
Model	$\widehat{E}[X(t)\|X(\text{Sept. }2022) = 114.47]$	117.64	120.90	123.52	125.70
Predictions	$\widehat{Me}[X(t)\|X(\text{Sept. }2022) = 114.47]$	117.63	120.90	123.52	125.69
	$\widehat{Mo}[X(t)\|X(\text{Sept. }2022) = 114.47]$	117.63	120.89	123.51	125.68
	$\widehat{C}_{0.025}[X(t)\|X(\text{Sept. }2022) = 114.47]$	116.49	119.24	121.44	123.26
	$\widehat{C}_{0.975}[X(t)\|X(\text{Sept. }2022) = 114.47]$	118.79	122.58	125.63	128.18
Forecast of Central Bank		116.04	119.40	121.33	123.07
Observed values		117.01	121.44	123.82	125.36
		Dec. 2023	Mar. 2024	June 2024	Sept. 2024
Model	$\widehat{E}[X(t)\|X(\text{Sept. }2022) = 114.47]$	128.07	130.21	132.09	134.12
Predictions	$\widehat{Me}[X(t)\|X(\text{Sept. }2022) = 114.47]$	128.07	130.20	132.07	134.11
	$\widehat{Mo}[X(t)\|X(\text{Sept. }2022) = 114.47]$	128.05	130.18	132.05	134.08
	$\widehat{C}_{0.025}[X(t)\|X(\text{Sept. }2022) = 114.47]$	125.30	127.12	128.70	130.45
	$\widehat{C}_{0.975}[X(t)\|X(\text{Sept. }2022) = 114.47]$	130.90	133.36	135.54	137.87
Forecast of Central Bank		123.90	126.30	127.37	128.35
Observed values		126.86	129.66	131.26	132.24

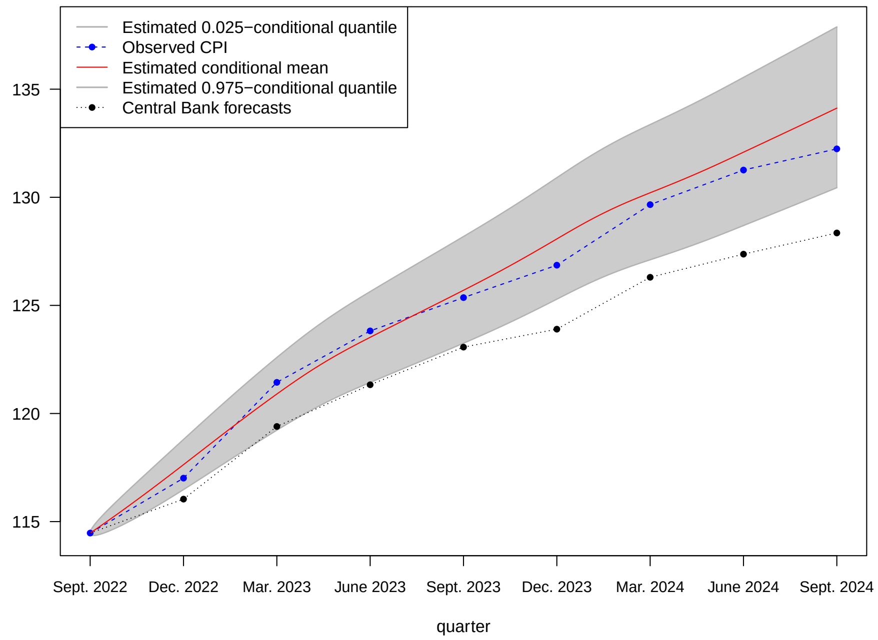


Figure 11. CPI forecasts for the last eight quarters provided by the estimated conditional mean of the model fitted given the value of CPI in September 2022 (data from March 2003 to September 2022).

Table 12. CPI forecasts for the last eight quarters using the pdf of the ML estimator of the conditional mean function vs. forecasts of the Central Bank and real observed values (data from March 2003 to September 2022).

		Dec. 2022	Mar. 2023	June 2023	Sept. 2023
Model	$E[\widehat{m}(t\|\text{Sept. }2022)]$	117.98	121.25	123.89	126.08
Predictions	$Me[\widehat{m}(t\|\text{Sept. }2022)]$	117.63	120.89	123.51	125.69
	$Mo[\widehat{m}(t\|\text{Sept. }2022)$	117.63	120.89	123.50	125.67
	$C_{0.025}[\widehat{m}(t\|\text{Sept. }2022)]$	116.91	119.40	121.30	122.69
	$C_{0.975}[\widehat{m}(t\|\text{Sept. }2022)]$	118.34	122.38	125.72	128.70
Forecast of Central Bank		116.04	119.40	121.33	123.07
Observed values		117.01	121.44	123.82	125.36
		Dec. 2023	Mar. 2024	June 2024	Sept. 2024
Model	$E[\widehat{m}(t\|\text{Sept. }2022)]$	128.45	130.57	132.39	134.33
Predictions	$Me[\widehat{m}(t\|\text{Sept. }2022)]$	128.06	130.19	132.06	134.10
	$Mo[\widehat{m}(t\|\text{Sept. }2022)$	128.03	130.14	132.00	134.01
	$C_{0.025}[\widehat{m}(t\|\text{Sept. }2022)]$	124.20	125.47	126.49	127.70
	$C_{0.975}[\widehat{m}(t\|\text{Sept. }2022)]$	131.95	134.99	137.79	140.76
Forecast of Central Bank		123.90	126.30	127.37	128.35
Observed values		126.86	129.66	131.26	132.24

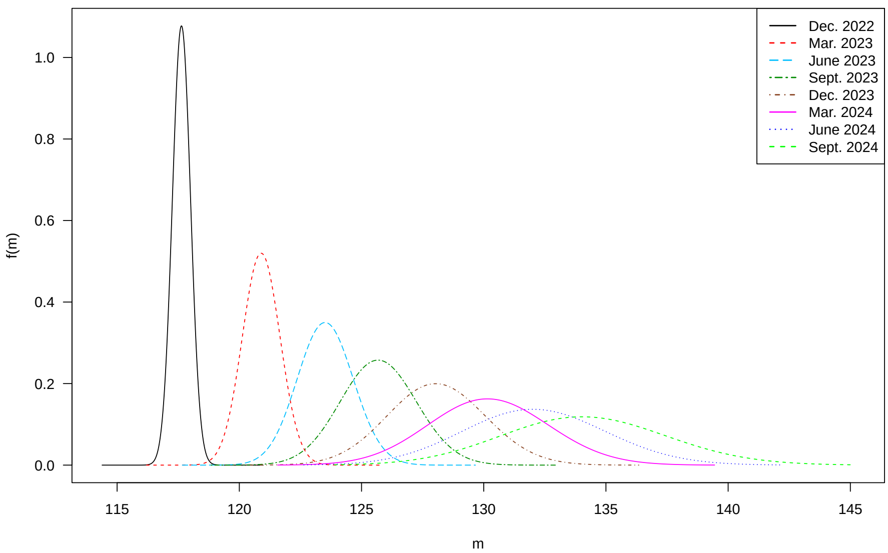


Figure 12. For the last eight quarters, pdfs of the ML estimator of the conditional mean function given the value in September 2022 (data from March 2003 to September 2022).

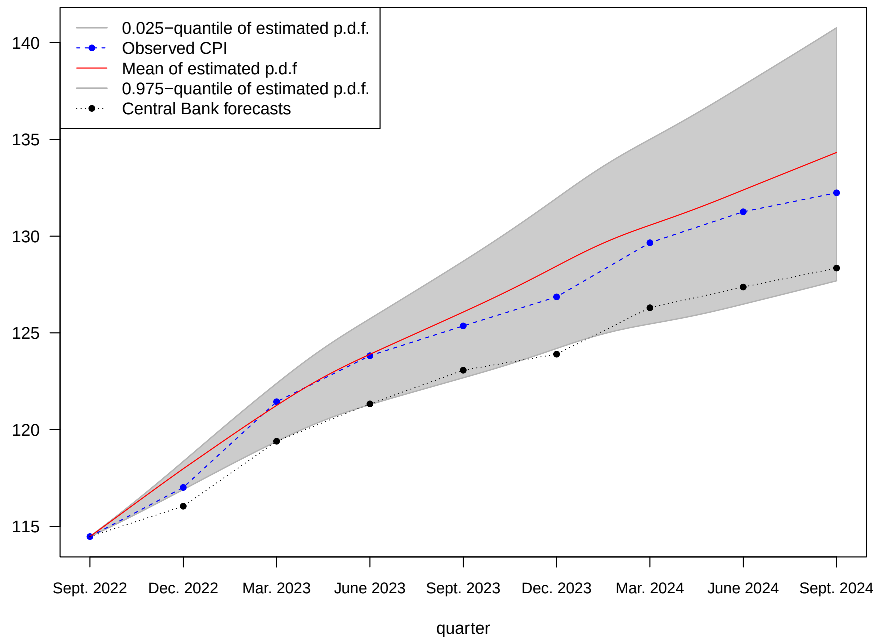


Figure 13. CPI forecasts for the last eight quarters provided by the mean of the ML estimator of the conditional mean function given the value of CPI in September 2022 (data from March 2003 to September 2022).

3.4. Analysis of potential future scenarios

In this section, in order to test the influence of the exogenous factors considered in the CPI forecasts, different future economic scenarios have been designed, specifically between December 2024 and September 2025. These scenarios have been drawn up on the basis of possible variations in the selected exogenous factors up to September 2024. Trying to show the current macroeconomic reality, these scenarios can be useful resources for economic policies.

In this context, the Central Bank, in its October 2024 monetary policy report, highlights that the Colombian economy is in a recovery phase, gradually moving toward its potential productivity capacity. As of December 2023, the Central Bank has made gradual interest rate reductions as a key measure to stimulate economic growth, to alleviate inflationary pressures, and to bring inflation to the 3.0% annual target. This approach aims to balance output recovery and price stability, ensuring the consolidation of the disinflation process within a framework of strong and sustainable growth.

However, these policies may involve internal and external risks, such as currency devaluation, or changes in commodity prices (e.g., coffee or oil). Some challenges may also arise, some related to the sustainability of public finances, which could limit monetary easing and affect the return to the macroeconomic balance expected by 2025.

Each scenario evaluates the individual or combined impact of some of the selected factors on the CPI. The scenarios are:

- **Scenario 1** (dynamism of labor market): The individual influence of the occupation rate on the CPI values is analyzed. For this purpose, it is considered that the occupation rate oscillates between its last observed value and said value increased (scenario 1A) or decreased (scenario

1B) on a quarterly basis by 1%.

- **Scenario 2** (monetary policy): The combined impact of the political intervention and consumer interest rates on CPI values is studied. In this case, it is assumed that both rates vary between their last observed values and those resulting from an increase (Scenario 2A) or a decrease (Scenario 2B) of these last observed values on a quarterly basis. Concretely, the political intervention rate varies by 0.75 units quarterly, while the consumer interest rate is modified in its relative form so that in absolute terms it varies by 1.4 units quarterly.
- **Scenario 3** (external risk): The combined impact of the representative market rate and coffee production on CPI values is evaluated. To this end, it is assumed that both the representative market rate and coffee production fluctuate between their last observed values and those resulting from an increase (Scenario 3A) or a decrease (Scenario 3B) of these last observed values on a quarterly basis. The considered variation for the representative market rate is 3%, while coffee production is modified in its relative form so that in absolute terms it varies by 5% quarterly.

By considering data until September 2024, the selection of factors has led to the same five variables as in all previous cases, with posterior probability 0.00474588. The estimation of the model has been carried out, resulting in $\hat{\sigma}^2 = 2.570359 \times 10^{-5}$ and

$$h_{\widehat{\theta}}(t) = \hat{\theta}_0 + \hat{\theta}_1 X_2 + \hat{\theta}_2 X_5 + \hat{\theta}_3 X_7 + \hat{\theta}_4 X_{14} + \hat{\theta}_5 X_{18},$$

where $\hat{\theta}_0 = -9.211354 \times 10^{-2}$, $\hat{\theta}_1 = 2.868627 \times 10^{-6}$, $\hat{\theta}_2 = 6.652347 \times 10^{-4}$, $\hat{\theta}_3 = 5.724394 \times 10^{-4}$, $\hat{\theta}_4 = 2.800971 \times 10^{-3}$, and $\hat{\theta}_5 = 5.396948 \times 10^{-2}$.

Figure 14 shows the estimated mean and α-quantile parametric functions ($\alpha = 0.025, 0.975$) together with the observed values of the CPI (the RAE in this case is equal to 0.0131).

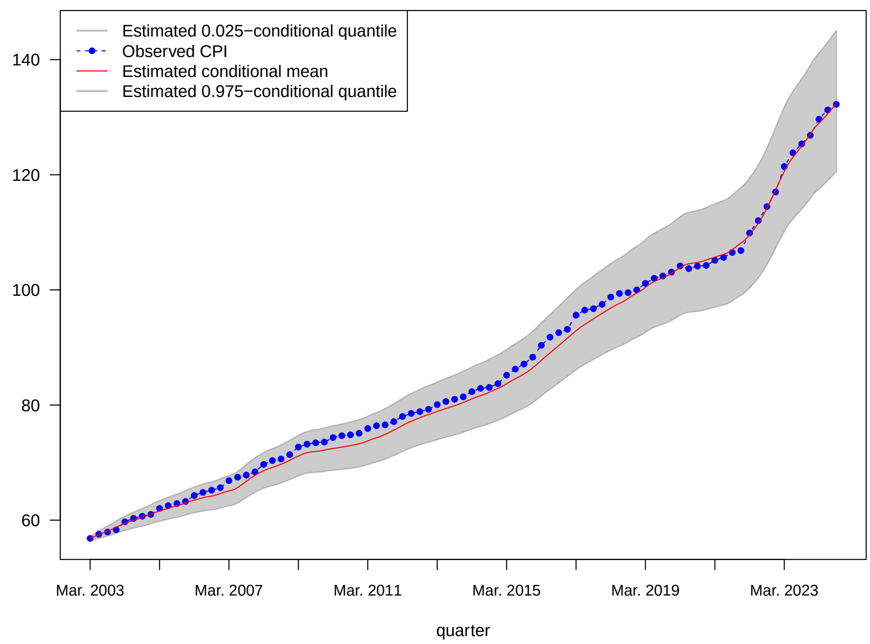


Figure 14. Estimated mean and α-quantile ($\alpha = 0.025, 0.975$) functions together with the observed CPI (data from March 2003 to September 2024).

Next, CPI projections in each quarter have been calculated, according to the methodology used in the previous sections. More specifically, in each scenario, a set of values - equally spaced - is considered for each factor that increases or decreases its value (the rest of the factors are kept constant), ranging from the lower to the higher values resulting from the proposed variation, and the CPI forecasts are calculated for all combinations of the values of the factors.

The following sections summarize the results obtained under each scenario.

3.4.1. Scenario 1. Dynamism of the labor market.

This scenario analyzes the impact of variations in the occupation rate on the CPI, keeping other factors constant. The analysis is particularly relevant in the current context of post-pandemic economic recovery, where the behavior of the labor market is key to understanding inflationary pressures. Figures 15 and 16 show the evolution of the predicted CPI in scenarios 1A (increase in occupation rate) and 1B (reduction in occupation rate), respectively. As expected, the magnitude of the CPI forecasts grows on a quarterly basis in both scenarios. Next, there is a more detailed description of the CPI performance.

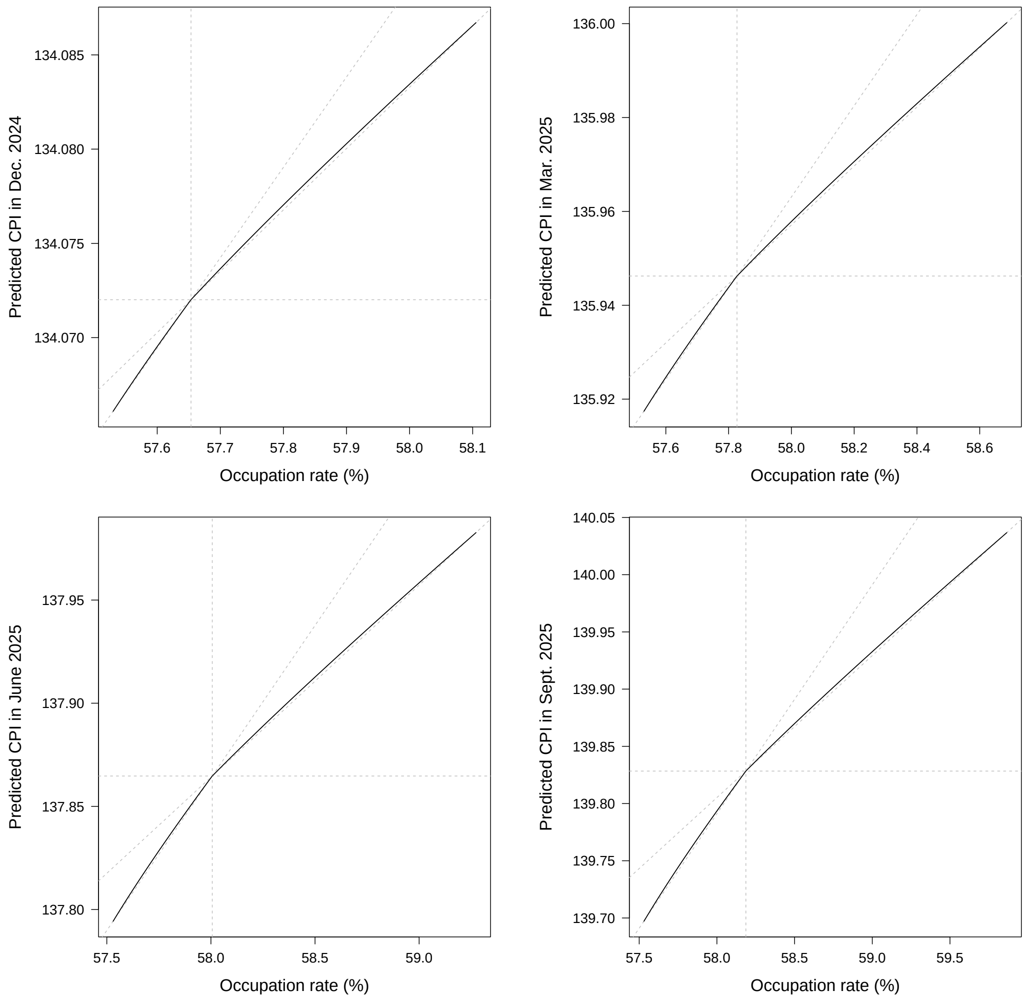


Figure 15. Quarterly evolution of the predicted CPI according to scenario 1A.

In scenario 1A and in all quarters considered, the CPI increases with the occupancy rate and an almost piecewise linear behavior of the CPI predictions is observed, different below and above a value of the occupancy rate, so that the CPI slows down its growth above this value although it grows faster and faster with each quarter.

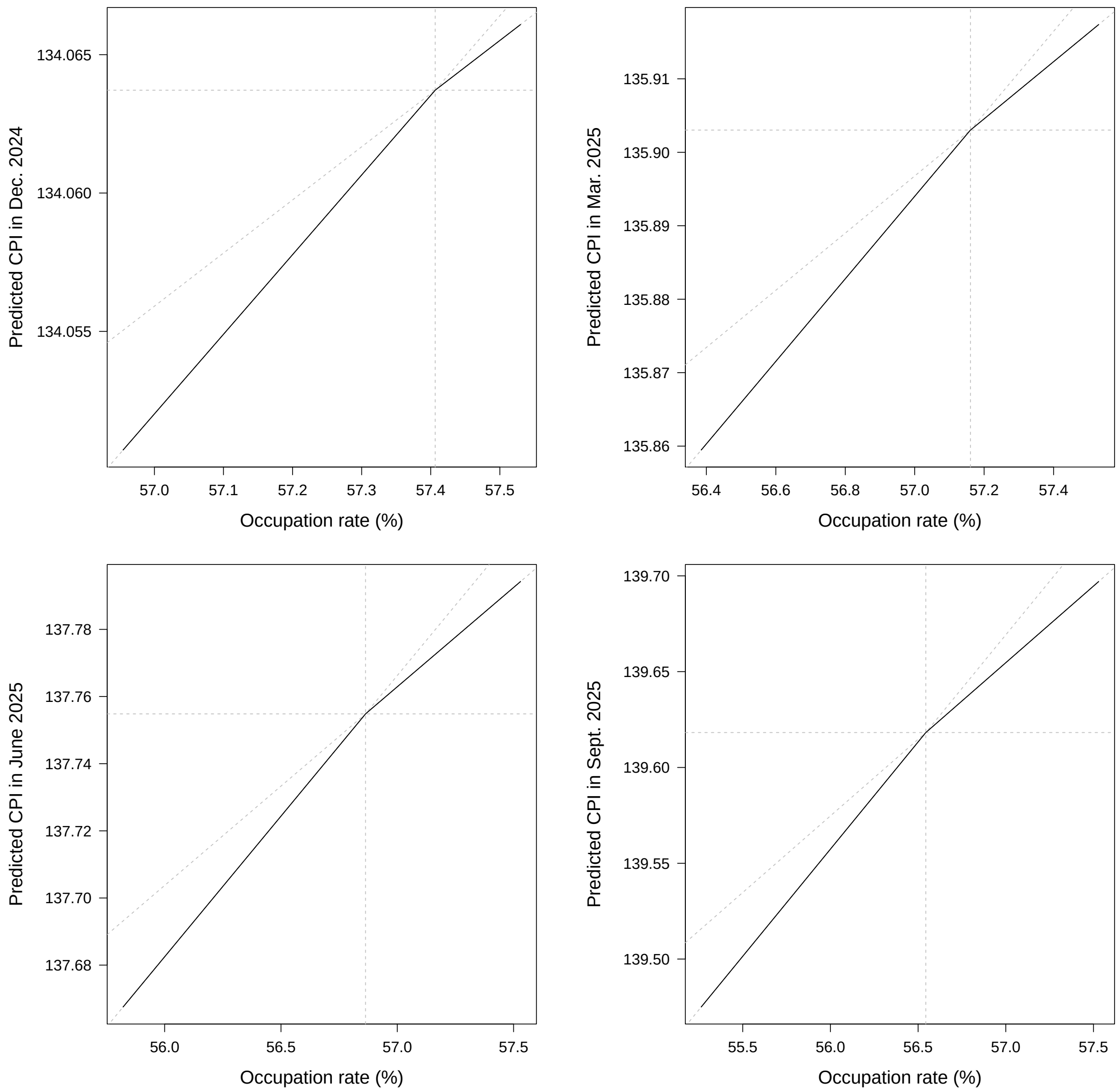


Figure 16. Quarterly evolution of the predicted CPI according to scenario 1B.

Table 13 displays the intervals over which the occupancy rate and the resulting CPI forecasts vary as well as precise data on the growth of CPI when the occupancy rate grows. For each quarter, the occupancy rate increases its value in the interval I_{X_7} and the resulting CPI predictions appear in the interval I_{CPI}, so that when the occupancy rate grows between the lower end of the interval I_{X_7} and x_7, the CPI increases approximately linearly until reaching the value $CPI(x_7)$ with a slope of s_1 and, when the occupancy rate grows between x_7 and the upper end of the interval I_{X_7}, the CPI also increases approximately linearly until reaching the value given by the upper end of the interval I_{CPI} with a slope of $s_2 < s_1$.

For illustrative purposes that could be extended to other cases, the following in-depth analysis could be considered: in December 2024, the occupancy rate increases between 57.53 and 58.1053 and the resulting CPI predictions are between 134.0661 and 134.0867, so that when the occupancy rate grows between 57.53 and 57.6538, the CPI increases approximately linearly until reaching the value 134.072 with a slope of 0.0479; and, when the occupancy rate grows between 57.6538 and 58.1053, the CPI also increases approximately linearly until reaching the value 134.0867 with a slope of 0.0325.

Table 13. Data on the growth of CPI when the occupancy rate grows (scenario 1A).

Quarter	I_{X_7}	I_{CPI}	x_7	$CPI(x_7)$	s_1	s_2
Dec. 2024	[57.53, 58.1053]	[134.0661, 134.0867]	57.6538	134.0720	0.0479	0.0325
Mar. 2025	[57.53, 58.6864]	[135.9174, 136.0002]	57.8267	135.9462	0.0972	0.0628
June 2025	[57.53, 59.2732]	[137.7943, 137.9827]	58.0065	137.8647	0.1479	0.0931
Sept. 2025	[57.53, 59.8659]	[139.6970, 140.0368]	58.1864	139.8284	0.2000	0.1241

On the other hand, in scenario 1B and in all quarters considered, the CPI decreases when the occupancy rate does so, and an almost piecewise linear behavior of the CPI predictions is observed, different before and after a value of the occupancy rate, so that the CPI accelerates its decrease below this value, although it decreases faster and faster with each quarter (please note that in Figure 16 and those related with scenarios of type B, the focus is studying how the CPI decreases when the exogenous factors decrease. Thus, the plots must be interpreted by looking at the axes corresponding to the factors from right to left).

Table 14 shows the intervals over which the occupancy rate and the resulting CPI forecasts vary as well as accurate data on the decrease of CPI when the occupancy rate decreases. This table should be interpreted the same way as Table 13, but in decreasing terms. For example, in December 2024, the occupancy rate decreases between 57.53 and 56.9547 and the resulting CPI forecasts are between 134.0661 and 134.0507, so that when the occupancy rate decreases between 57.53 and 57.4062, the CPI decreases approximately linearly until reaching the value 134.072 with a slope of 0.0192. When the occupancy rate decreases between 57.4062 and 56.9547, the CPI also decreases approximately linearly until reaching the value 134.0507 with a slope of 0.0288.

Table 14. Data on the decrease of CPI when the occupancy rate decreases (scenario 1B).

Quarter	I_{X_7}	I_{CPI}	x_7	$CPI(x_7)$	s_1	s_2
Dec. 2024	[56.9547, 57.53]	[134.0507, 134.0661]	57.4062	134.0720	0.0192	0.0288
Mar. 2025	[56.3851, 57.53]	[135.8595, 135.9174]	57.1606	135.9030	0.0389	0.0562
June 2025	[55.8213, 57.53]	[137.6675, 137.7943]	56.8636	137.7548	0.0591	0.0838
Sept. 2025	[55.2630, 57.53]	[139.4750, 139.6970]	56.5439	139.6182	0.0799	0.1241

Figure 17 allows us to compare the magnitude and range of variation of the CPI projections in scenarios 1A and 1B. It is noticed that, for each quarter, the magnitude and variation range of CPI increase more when the occupation rate increases (scenario 1A). Moreover, between quarters, both the magnitude and variation range of CPI increases in both scenarios, with a higher increase when the occupation rate augments.

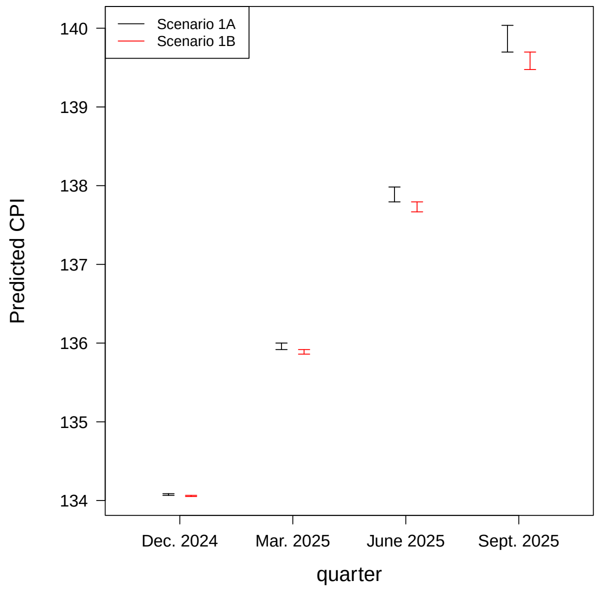


Figure 17. Comparison between CPI predictions under scenarios 1A and 1B.

For an economic point of view, these results demonstrate a direct relationship between employment rate and CPI, highlighting the key influence of labor dynamics on price evolution. A strengthened labor market tends to drive inflation through increased aggregate demand, while weakening employment reduces inflationary pressures through contracted consumption and disposable income. In the context of economic recovery, employment growth would not generate significant inflationary pressures as the economy recovers its productive capacity. This suggests that employment growth can coexist with price stability, provided that labor market expansion remains aligned with economic reactivation and convergence toward macroeconomic equilibrium.

3.4.2. Scenario 2. Monetary policy.

This scenario examines the joint effect of political intervention and consumption interest rates on the CPI, a fundamental interaction to understand the effectiveness of monetary policy. The analysis is especially relevant given the current context of gradual rate cuts made by the Central Bank, seeking to balance economic recovery with price stability.

Let's first consider scenario 2A. Figures 18 and 19 reveal the quarterly evolution of the predicted CPI in this case.

The perspective plots in Figure 18 show that, in all quarters considered, the CPI increases almost piecewise linearly with increasing values of the political intervention and consumption interest rates. Furthermore, regardless of the value of the consumption interest rate, the CPI shows a different linear behavior below and above a high value of the political intervention rate, so that the CPI accelerates its growth above this value, growing faster and faster with each quarter.

The intervals over which the factors and the resulting CPI forecasts vary are shown in Table 15, while Table 16 provides information about the growth of CPI when the political intervention rate grows.

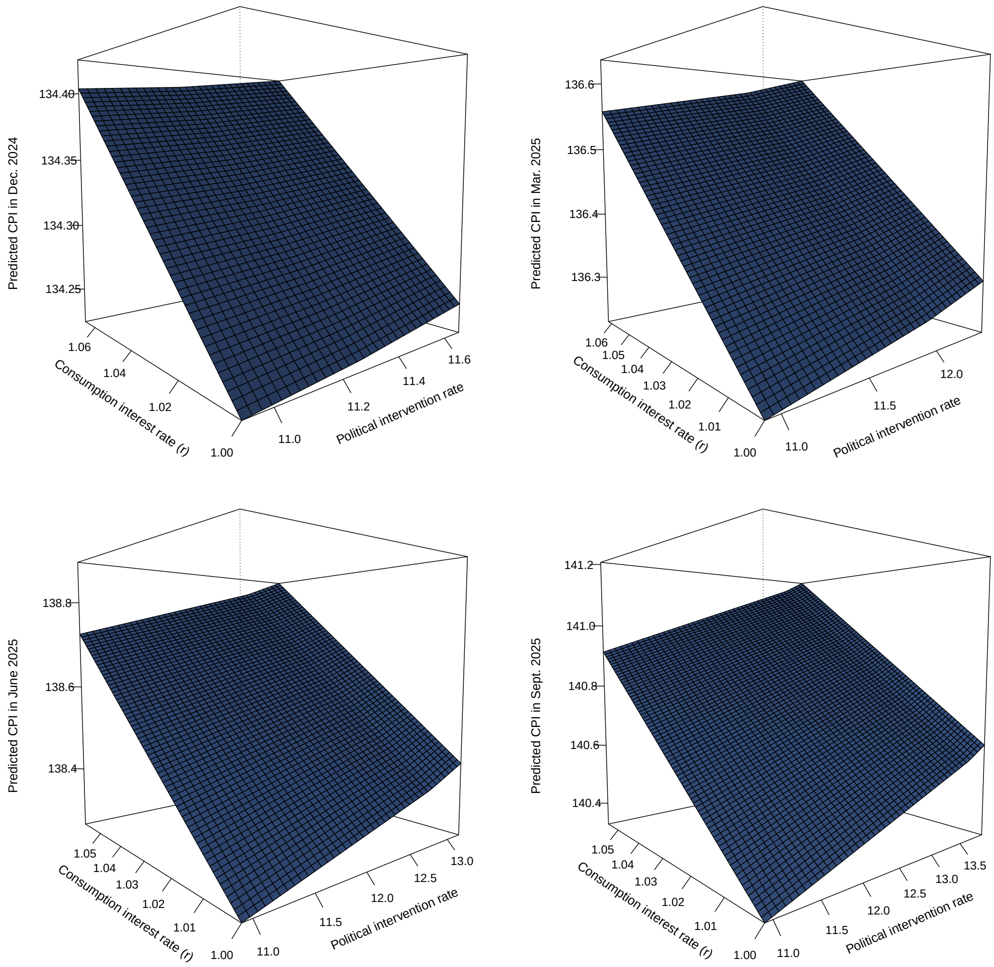


Figure 18. CPI quarterly evolution perspective plots in scenario 2A.

Table 15. Ranges of variation of factors and the resulting CPI predictions in scenario 2A.

Quarter	I_{X_5}	$I_{X_{18}}$	I_{CPI}
Dec. 2024	[10.92, 11.67]	[1.0, 1.0660]	[134.2241, 134.4250]
Mar. 2025	[10.92, 12.42]	[1.0, 1.0620]	[136.2286, 136.6360]
June 2025	[10.92, 13.17]	[1.0, 1.0583]	[138.2624, 138.8940]
Sept. 2025	[10.92, 13.92]	[1.0, 1.0551]	[140.3277, 141.2065]

Table 16. Data on the growth of CPI forecasts when the political intervention rate grows, regardless of the value of the consumption interest rate (scenario 2A).

Quarter	I_{X_5}	x_5	s_1	s_2
Dec. 2024	[10.92, 11.67]	11.2450	0.0223	0.0332
Mar. 2025	[10.92, 12.42]	11.9486	0.0457	0.0655
June 2025	[10.92, 13.17]	12.7662	0.0695	0.0978
Sept. 2025	[10.92, 13.92]	13.6473	0.0939	0.1309

Table 16 should be interpreted in a similar way to Table 13. For example, in June 2025, regardless of the value of the consumption interest rate, when the political intervention rate grows between 10.92 and 12.7662, the CPI increases approximately linearly with a slope of 0.0695 and, when the political intervention rate grows between 12.7662 and 13.17, the CPI also increases approximately linearly but with a slope of 0.0978.

Heat maps in Figure 19 show that in December 2024, high CPI values correspond to high values of the consumption interest rate, intermediate values of the CPI are obtained for intermediate values of the consumption interest rate, and low values of the CPI are associated with low values of the consumption interest rate. This suggests an unbalanced relationship between the two factors in predicting CPI. Nevertheless, as the quarters progress, this pattern evolves so that in September 2025, high CPI values correspond to high values of both rates, low CPI values are obtained for low values of both rates, and intermediate CPI values are associated either with low values of one rate and high values of the other, or with intermediate values of both rates. This indicates a quarterly evolution to a more balanced relationship between the two factors in predicting CPI.

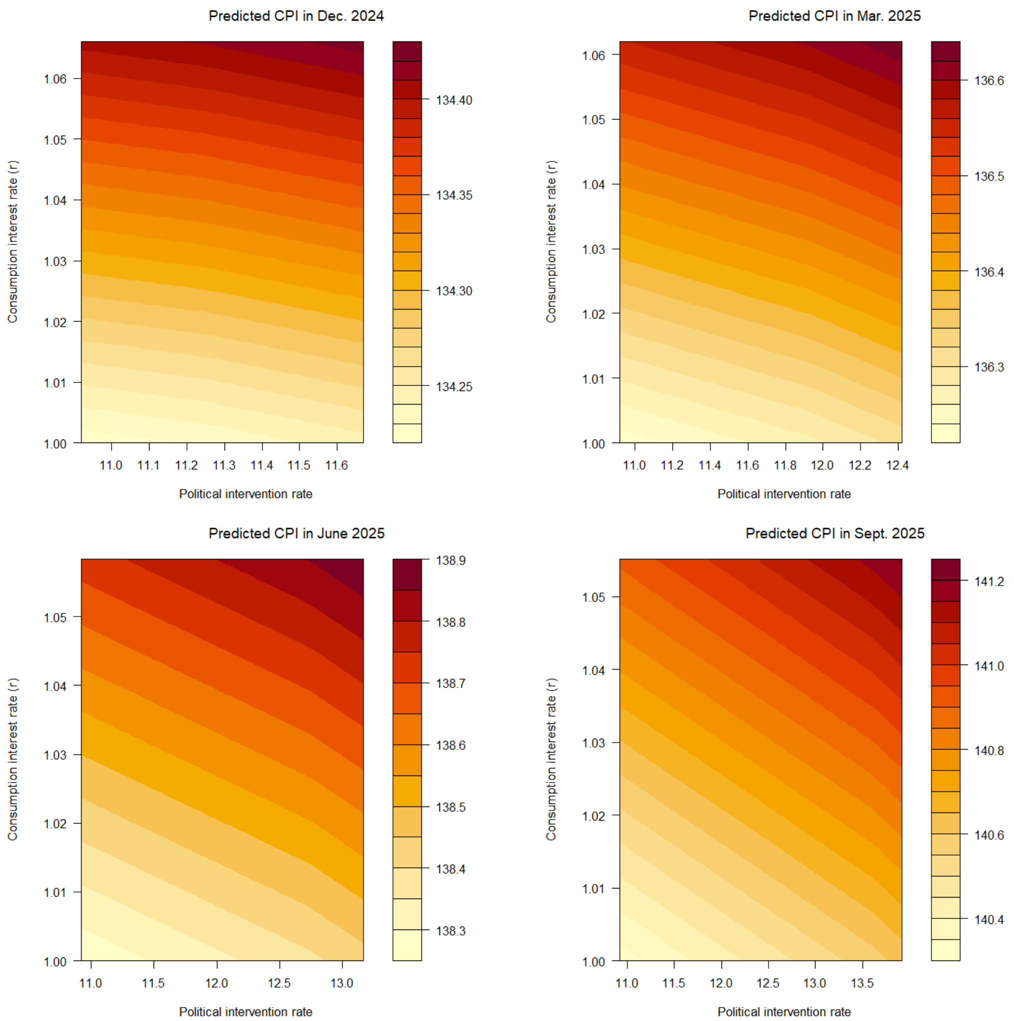


Figure 19. CPI quarterly evolution heat maps in scenario 2A.

On the other side, it can be seen that the ranges of heat values increase with the quarters. This supports the observation made before about the growth rates s_1 and s_2 in Table 16.

Similarly, for scenario 2B, Figure 20 depicts that the CPI decreases almost piecewise linearly with the political intervention and consumption interest rates.

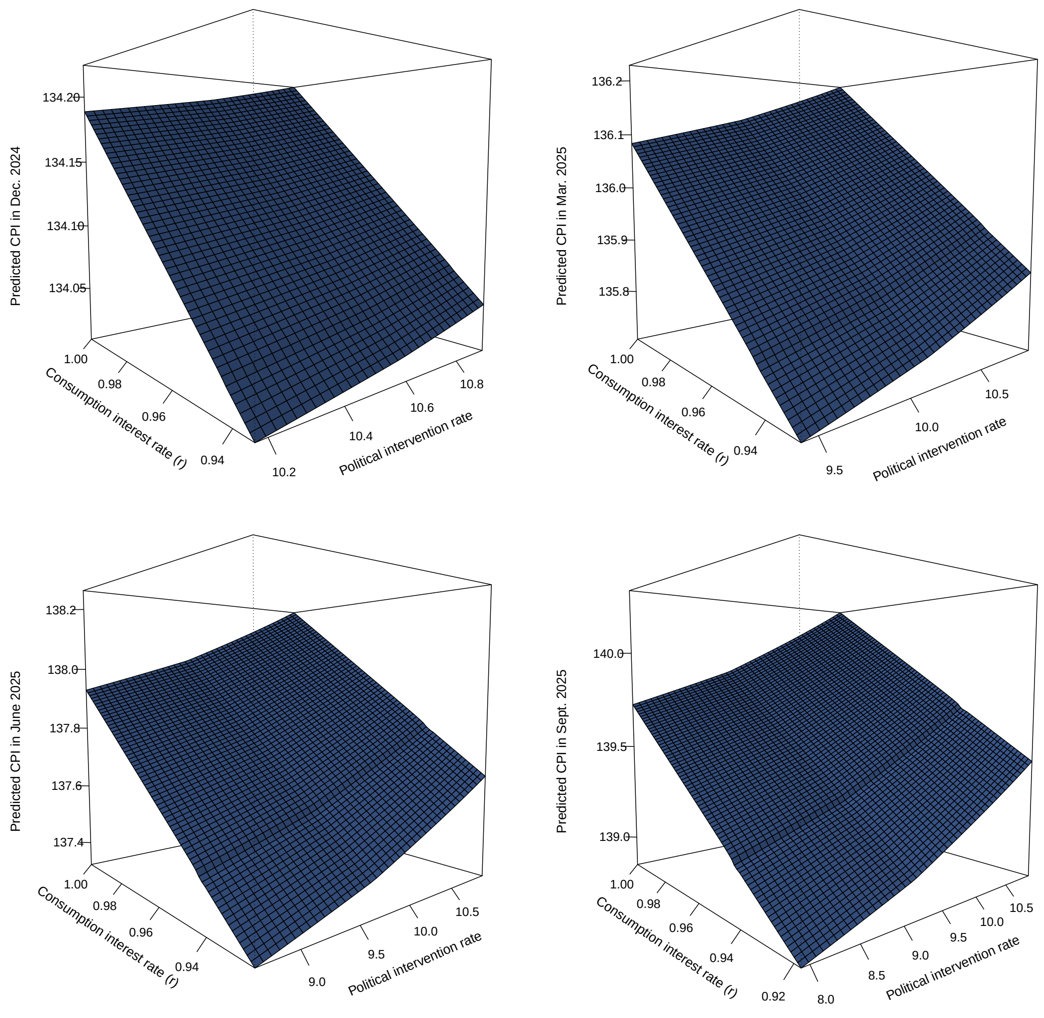


Figure 20. Quarterly evolution of the predicted CPI according to the scenario 2B.

Furthermore:

- Regardless of the value of the consumption interest rate, the CPI shows a different linear behavior above and below a certain value of the political intervention rate, so that the CPI slows down its decrease below this value although it decreases faster and faster with each quarter.
  Table 17 displays the intervals over which the factors and the CPI forecasts vary, whereas Table 18 provides precise information about the decrease of CPI forecasts when the political intervention rate decreases. Table 18 should be interpreted in a similar way to Table 16, but in decreasing terms. For example, in June 2025, regardless of the value of the consumption interest rate, when the political intervention rate decreases between 10.92 and 9.5931, the CPI decreases approximately linearly with a slope of 0.1712 and, when the political intervention rate decreases

between 9.5931 and 8.67, the CPI also decreases approximately linearly but with a slope of 0.1143.

- Regardless of the value of the political intervention rate, the CPI decreases almost linearly until a certain value of the rate is reached. Subsequently, the decrease slows down until another value of the rate is reached (very close to the previous value). Finally, from this last value of the rate, the CPI decreases linearly again with almost the same speed as at the beginning. This behavior is observed in all quarters, becoming more noticeable in the last quarters. The numerical values can be found in Table 19.

Table 17. Ranges of variation of factors and the resulting CPI predictions in scenario 2B.

Quarter	I_{X_5}	$I_{X_{18}}$	I_{CPI}
Dec. 2024	[10.17, 10.92]	[0.9340, 1.00]	[134.0087, 134.2241]
Mar. 2025	[9.42, 10.92]	[0.9293, 1.00]	[135.7057, 136.2286]
June 2025	[8.67, 10.92]	[0.9239, 1.00]	[137.3219, 138.2624]
Sept. 2025	[7.92, 10.92]	[0.9176, 1.00]	[138.8463, 140.3277]

Table 18. Data on the decrease of CPI forecasts when the political intervention rate decreases, regardless of the value of the consumption interest rate (scenario 2B).

Quarter	I_{X_5}	x_5	s_1	s_2
Dec. 2024	[10.17, 10.92]	10.5950	0.0559	0.0402
Mar. 2025	[9.42, 10.92]	10.1057	0.1130	0.0771
June 2025	[8.67, 10.92]	9.5931	0.1712	0.1143
Sept. 2025	[7.92, 10.92]	9.1473	0.2333	0.1534

Table 19. Data on the decrease of CPI forecasts when the consumption interest rate decreases, regardless of the value of the political intervention rate (scenario 2B).

Quarter	$I_{X_{18}}$	$x_{18,1}$	$x_{18,2}$	s_1	s_2	s_3
Dec. 2024	[0.9340, 1.00]	0.9422	0.9420	2.7348	1.8184	2.7126
Mar. 2025	[0.9293, 1.00]	0.9422	0.9414	5.3804	3.6638	5.2920
June 2025	[0.9239, 1.00]	0.9422	0.9408	8.0821	5.5690	7.8819
Sept. 2025	[0.9176, 1.00]	0.9422	0.9401	10.8678	7.5115	10.5083

As in the previous case, heat maps are shown in Figure 21. In this case, the same reading can be made as in scenario 2A, although now the relationship between the factors in predicting CPI in the first quarters is slightly more balanced.

The magnitude and range of variation of the CPI projections in scenarios 2A and 2B are compared in Figure 22. In this case, we can conclude that, for each quarter, the magnitude of CPI increases more in scenario 2A, while the range of variation of the CPI is slightly larger in scenario 2B and the first quarters, and much larger in scenario 2B and the last quarters. In addition, between quarters, both the magnitude and the range of variation of the CPI increase in both scenarios, with a somewhat more pronounced increase in scenario 2B.

Finally, from an economic perspective, the results reveal a direct relationship between interest rates and the CPI, reflecting the impact of negative supply shocks following the pandemic and the invasion of Ukraine. Central Banks initially responded with aggressive rate hikes, later transitioning to gradual reductions as positive supply shocks emerged and inflation moderated. The model captures this dynamic, showing how the CPI and interest rates adjust according to changing economic conditions.

Moreover, the sharp slope changes in Scenario 2B suggest that lower interest rates could trigger positive demand and supply shocks, stimulating growth and slowing inflation. This decline in the CPI reduces price indexation and lowers inflation expectations, further reinforcing the disinflationary trend quarter by quarter.

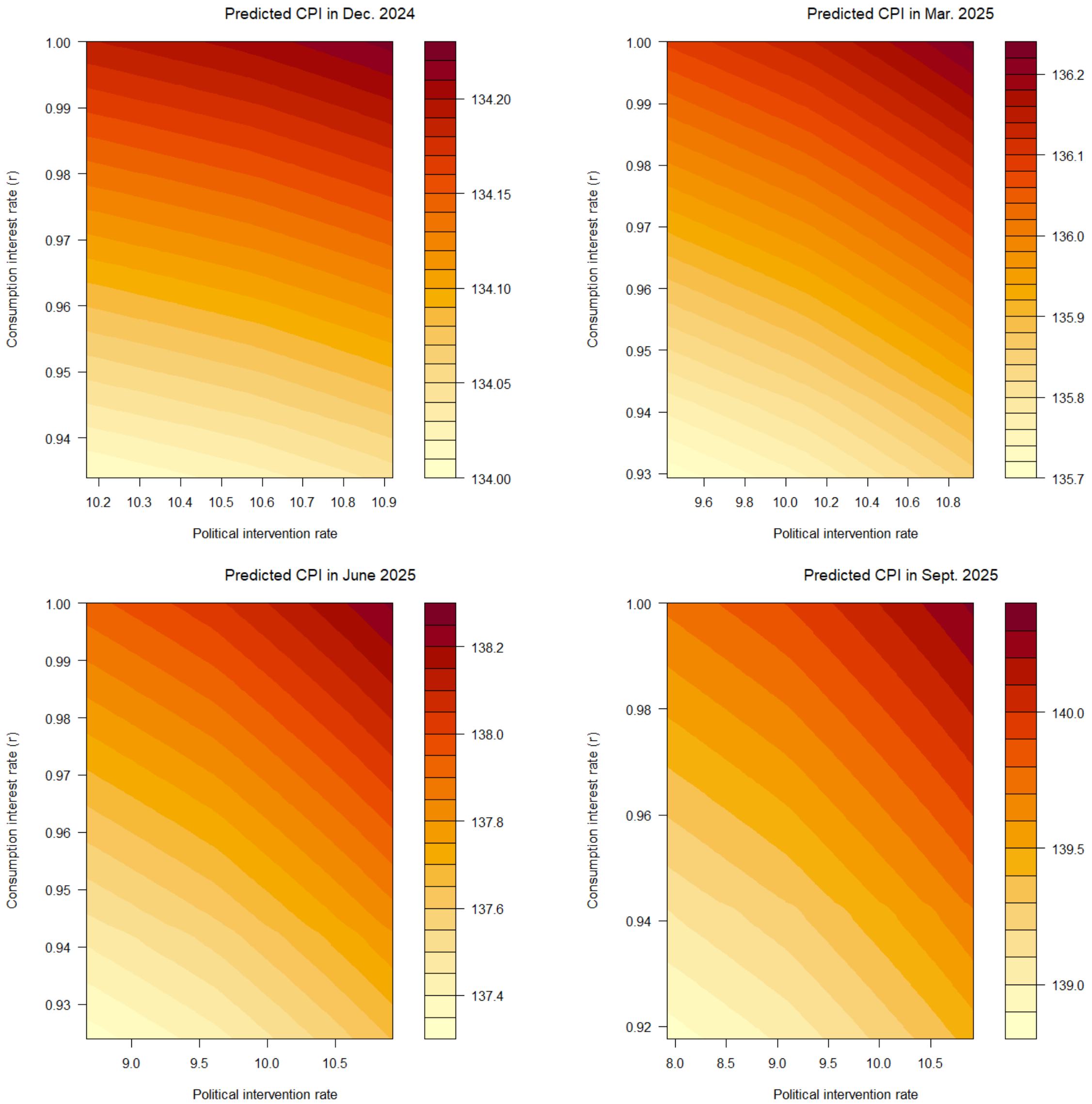


Figure 21. Quarterly evolution of the predicted CPI according to the scenario 2B.

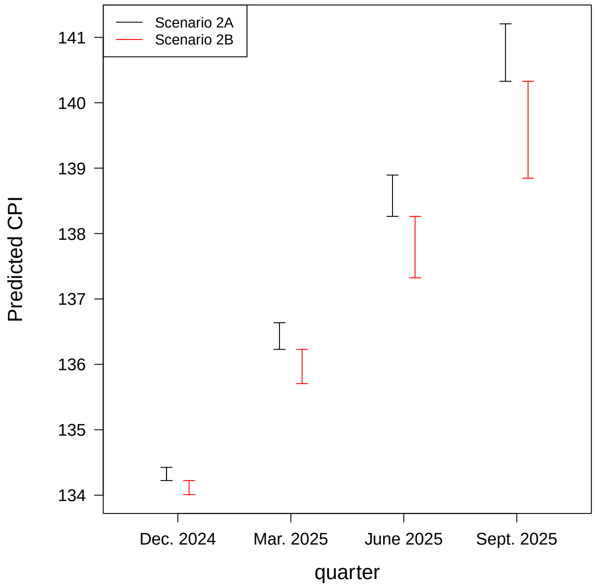


Figure 22. Comparison between CPI predictions under scenarios 2A and 2B.

3.4.3. Scenario 3. External risk.

This scenario evaluates the vulnerability of the Colombian economy to external shocks by simultaneously varying the representative market rate and coffee production, keeping all other factors constant. The relevance of this analysis lies in the open nature of the Colombian economy and the historical importance of the coffee sector in its productive structure.

The comments that can be made in this case are along the same lines as those made in scenario 2, and, for the sake of brevity, the predicted CPI behavior is described in a bit less detail.

Figures 23 and 24 show the quarterly evolution of the predicted CPI when the representative market rate and coffee production are simultaneously increased (scenario 3A).

The perspective plots in Figure 23 reveal that the predicted CPI increases almost piecewise linearly with increasing values of the representative market rate and coffee production. In addition, regardless of the value of the coffee production, the CPI shows a different linear behavior below and above a value of the RMR around a quarterly increasing point of the interval in which this rate varies, so that the CPI slows down its growth above this value, although it grows faster and faster with each quarter.

Heat maps in Figure 24 suggest an unbalanced relationship between the representative market rate and coffee production in predicting the CPI so that, in all quarters, high CPI values correspond to high values of the representative market rate and similarly with the respective intermediate and lower CPI values.

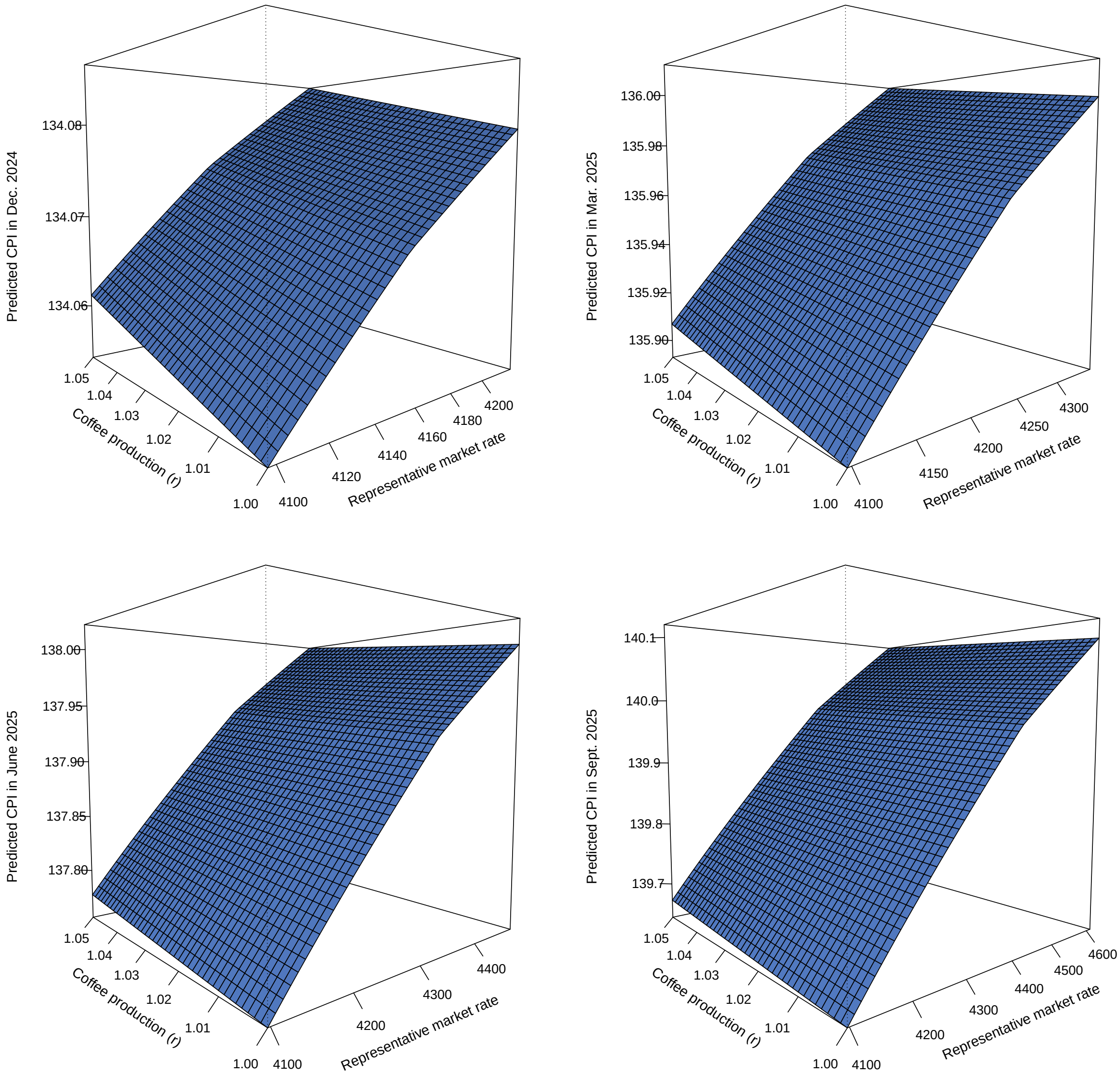


Figure 23. CPI quarterly evolution perspective plots in scenario 3A.

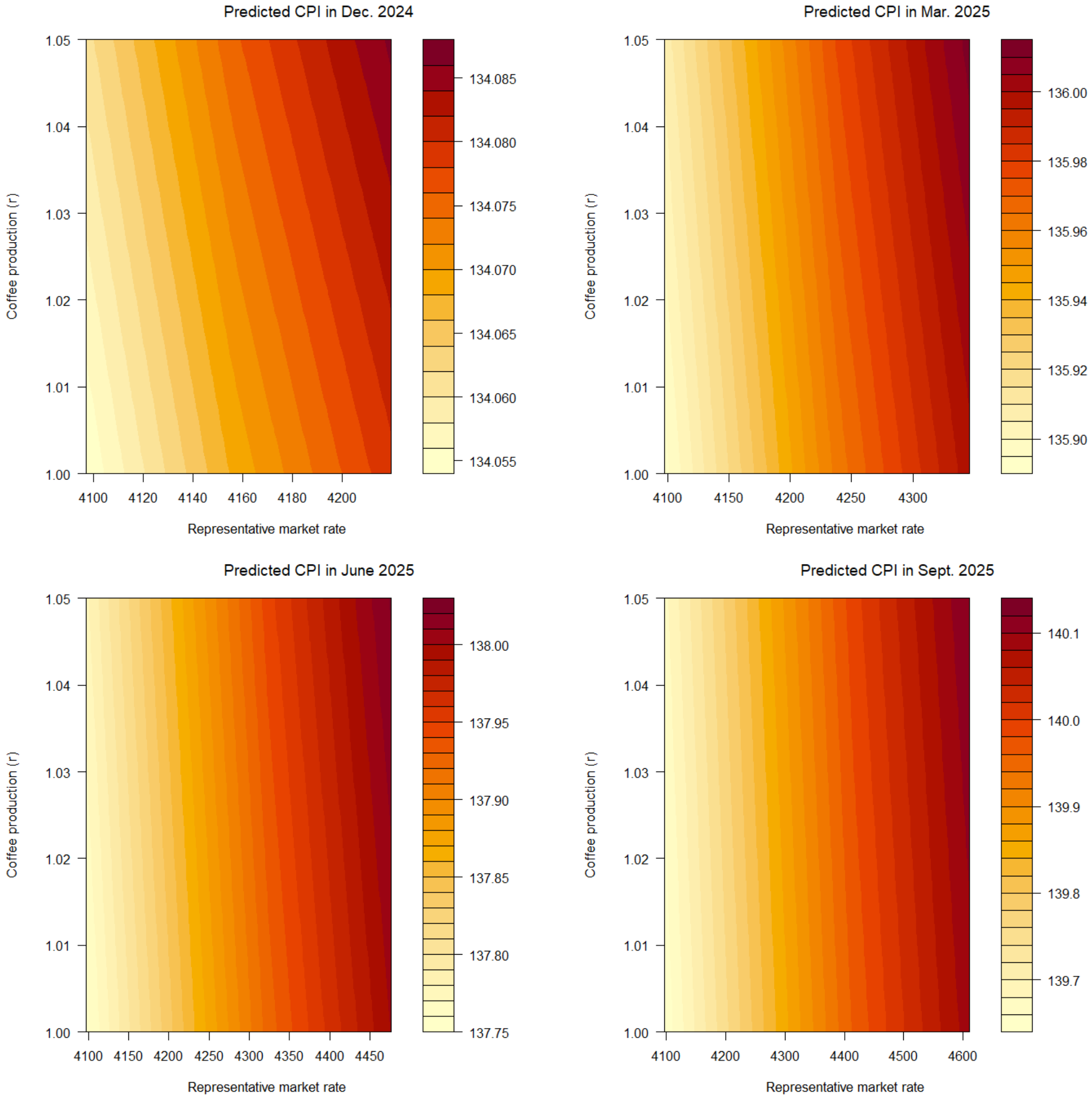


Figure 24. CPI quarterly evolution heat maps in scenario 3A.

On the other hand, Figure 25 shows that the predicted CPI decreases quasi-linearly when the representative market rate and coffee production are decreased (scenario 3B). Furthermore, except in September 2025, the CPI shows a different linear behavior below and above a value of the RMR that is moving further and further away from the last observed value, so that the CPI accelerates its decrease below this value, although it grows faster and faster with each quarter.

Heat maps in Figure 26 again indicate an unbalanced relationship between the representative market rate and coffee production in predicting the CPI, similar to that described in scenario 3A.

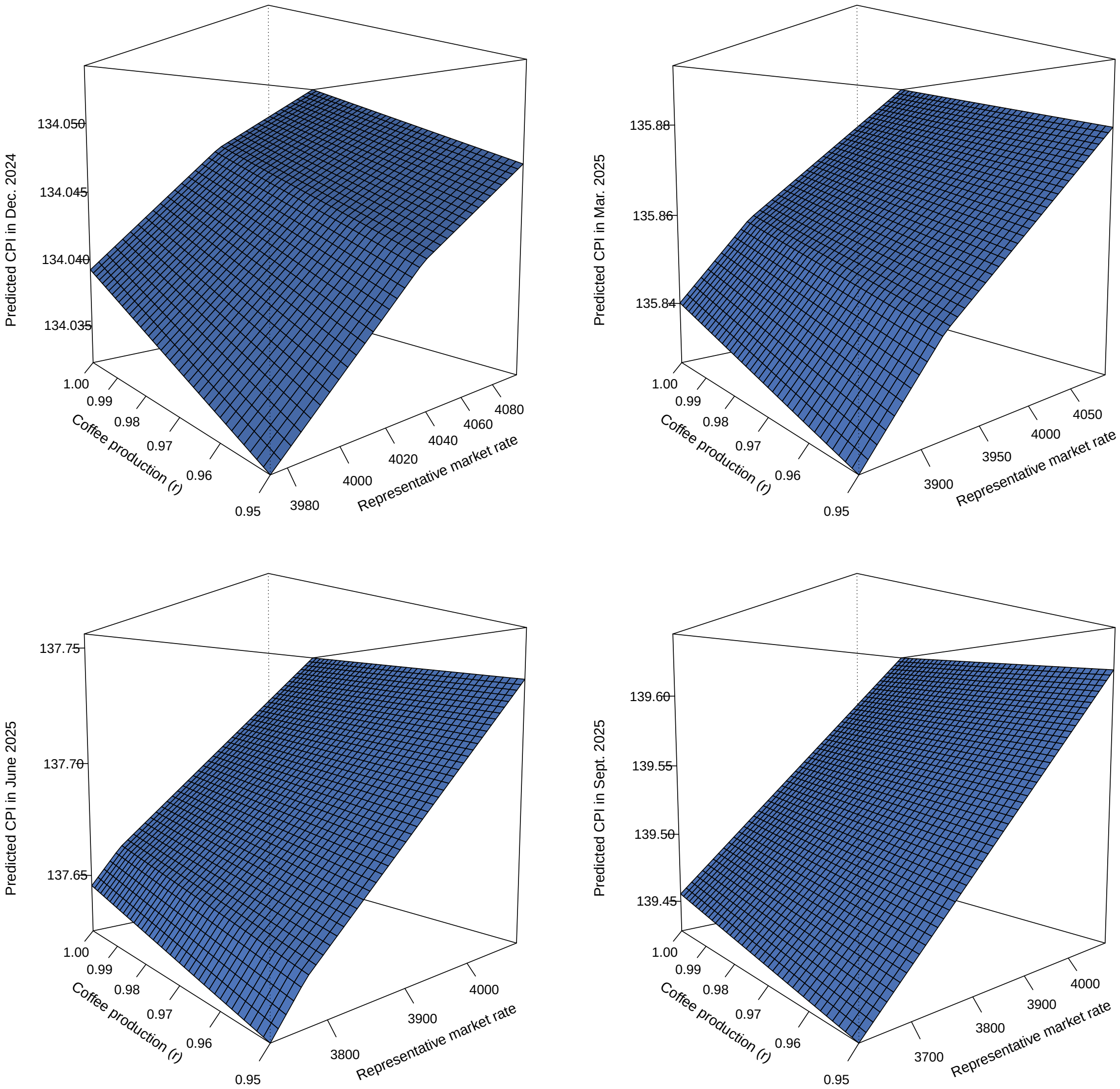


Figure 25. CPI quarterly evolution perspective plots in scenario 3B.

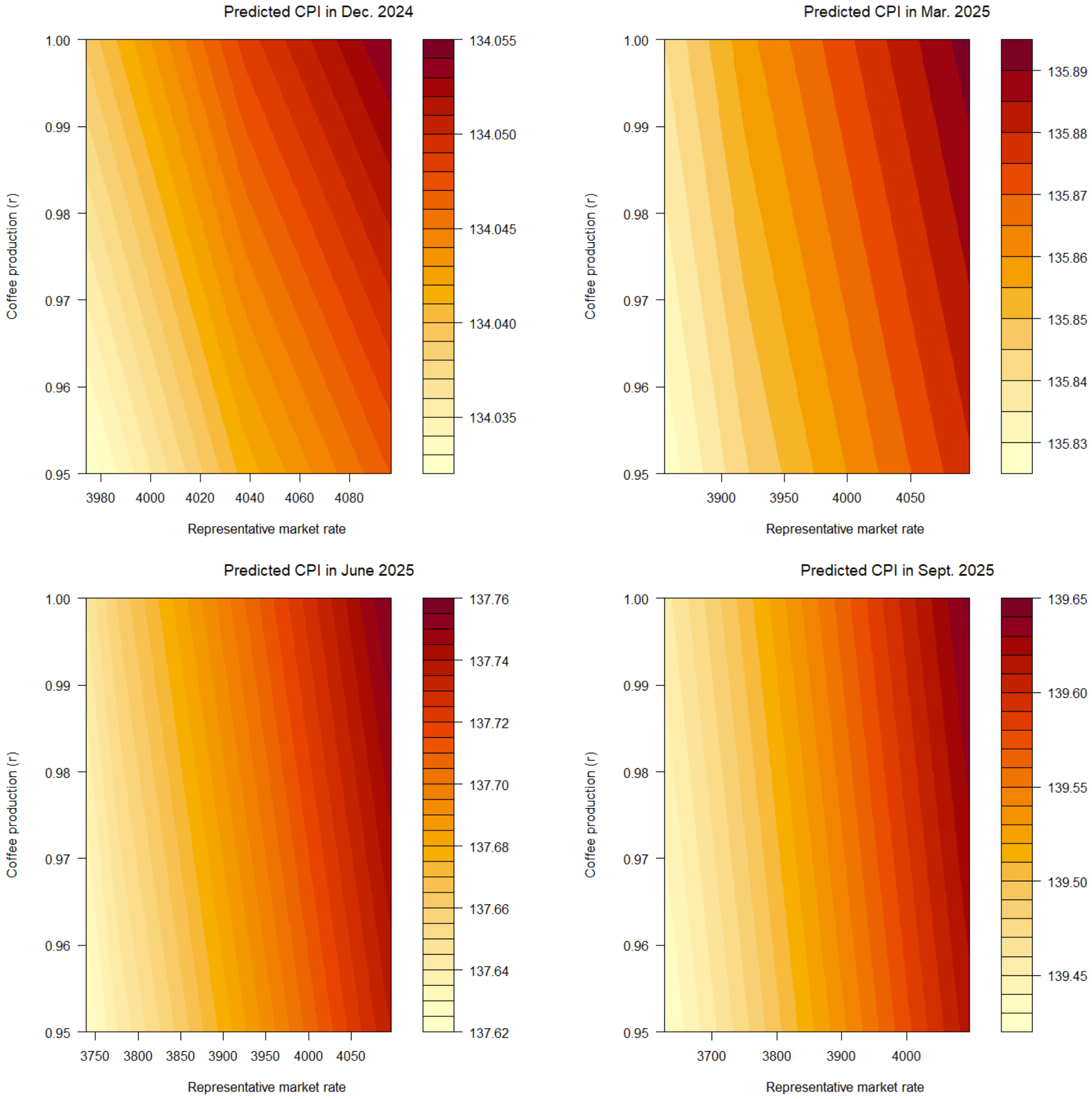


Figure 26. CPI quarterly evolution heat maps in scenario 3B.

Figure 27 allows us to compare the magnitude and range of variation of the CPI projections in scenarios 3A and 3B. In this case, we can conclude that, for each quarter, the magnitude and the range of variation of the CPI are greater in scenario 3A than in 3B. In addition, between quarters, both the magnitude and the range of variation of the CPI increase in both scenarios.

Finally, among scenarios 1A, 2A, and 3A, the one that can increase the predicted CPI the most is scenario 2A, followed far behind by scenarios 3A and 1A (see Figure 28(a)). Similarly, among scenarios 1B, 2B and 3B, the one that can reduce the predicted CPI the most is scenario 2B, followed far behind by scenarios 3B and 1B (see Figure 28 (b)).

The results of the model identify the consumption interest rate and monetary policy as the primary factors influencing the CPI, followed by the external sector and the labor market. In Scenario 3A, a depreciation of the Colombian peso boosts exports and coffee production, raising the CPI more than in Scenario 3B, where an appreciation supports disinflation. However, higher import and foreign debt costs moderate this impact. This suggests that persistent inflationary pressures driven by exchange rate fluctuations could hinder the disinflation strategy of monetary policy and delay inflation's convergence to its target.

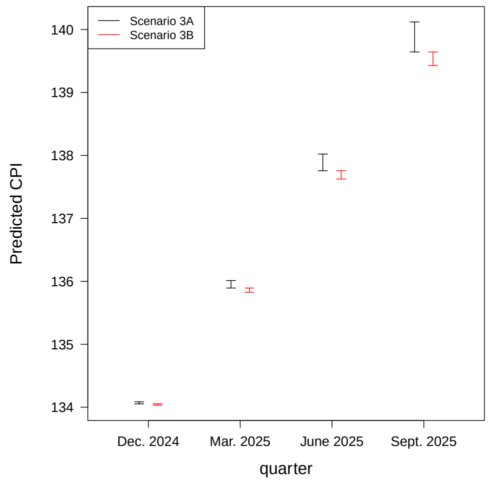
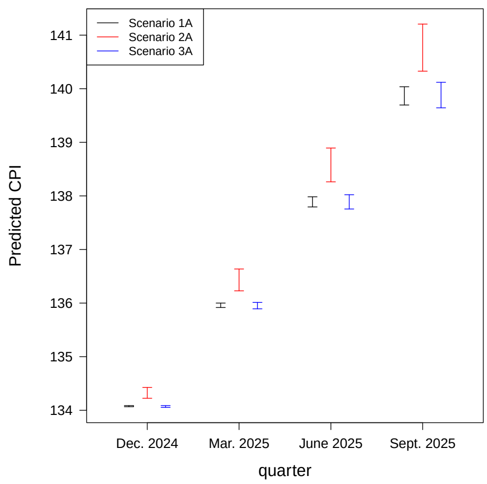
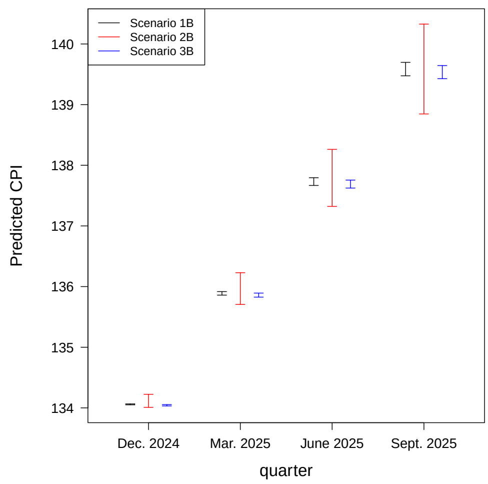


Figure 27. Comparison of CPI predictions under scenarios 3A and 3B.



(a)



(b)

Figure 28. Comparison of CPI predictions under scenarios type A (a) and type B (b).

4. Conclusions

In this work, a stochastic model to describe and forecast the dynamics of the CPI in Colombia is proposed. This model is based on a lognormal diffusion process with linear exogenous factors, which is able to take into account the random influence of other external economic variables. This process has been widely studied in the literature about stochastic processes. Furthermore, many of its properties are well-known and explicitly obtained, including those related with the ML estimation of its parameters. These make the process very treatable, as it could be seen in Section 2.

By considering recent datasets obtained from different official sources such as the *Banco de la República de Colombia* ("Central Bank"), *Departamento Administrativo Nacional de Estadística* and *Superintendencia Financiera de Colombia*, an in-depth modeling study has been carried out, revealing important findings.

Based on a broad set of economic indicators, a procedure is proposed for selecting the exogenous factors that influence the variable of interest. To this end, the use of variable selection methods in linear models from a Bayesian perspective has been proposed, based on the methodology derived from intrinsic a priori distributions. The regression models considered have been obtained by considering the Euler-Maruyama scheme applied to the stochastic differential equation that defines the lognormal diffusion process with exogenous factors. Absolute and relative variables were included, allowing us to model long-range structural relations as well as short-range dynamics. Results showed five key variables that influence stationary dynamics: RMR, political intervention rate, occupation rate, coffee production (relative), and consumption interest rate. This selection provides valuable insight into the main determinants of inflation in Colombia, reflecting the importance of factors under the control of economic authorities, labor market dynamics, and external economic conditions.

In regards to the predictive capability of the model, two options for predictions were considered: one based on the characteristics obtained from the transition probability distribution of the process, such as mean, median, or mode, and the other based on the estimation of the distribution of the ML estimators of the conditional mean functions of the process. In this case, the lognormal diffusion makes available the pdf of the ML estimators of the conditional mean, which may become important to consider predictions based on prior knowledge, as it is customary in economic studies. The flexibility of the model to provide different forecasting options offers policymakers a wide range of tools to assess inflation projections under different scenarios.

The proposed model has demonstrated remarkable predictive ability, consistently outperforming the forecasts of the Central Bank across different time horizons. The predictions obtained using the model outperform those made by the Colombian authorities in short, medium, and long-term horizons. In particular, considering data until March 2022, it accurately captured the inflationary upturn in mid-2022, which was not anticipated by official projections. The predictions maintained their accuracy across the two times the Central Bank updated its forecasts (June and September 2022), highlighting the model's robustness and its usefulness for economic planning.

Furthermore, the proposed approach considered in this work not only offers point estimates but also confidence intervals and probability distributions, allowing for a more precise quantification of uncertainty in CPI projections. Therefore, this study highlights the critical importance of incorporating exogenous factors into inflation forecasting models, particularly for emerging economies like Colombian, which are susceptible to diverse internal and external shocks. By integrating these factors, the model demonstrates an enhanced capacity to capture the nuanced fluctuations and evolving dynamics of inflation. This approach represents a significant advancement in the analytical toolkit available to central banks and economic policymakers for inflation analysis and forecasting. Moreover, the model facilitates comprehensive impact analysis of economic shocks, enabling a more detailed evaluation of how variations in these exogenous factors influence CPI projections, validating the capabilities of the model for reliable prediction in the context of a particular economic framework. This feature provides valuable insights for the calibration of economic policies and decision-making, giving the authorities the predictions depending on different economic scenarios. These characteristics of the model reveal also interesting future research lines concerning the study of inverse problems related with the possibility of determining those economic scenarios responsible for certain observed CPI dynamics.

Author contributions

A. De la Peña Cuao, A. Barrera, J. J. Serrano-Pérez and F. Torres-Ruiz: conceptualization, investigation, writing-original draft, review and editing; J. J. Serrano-Pérez: software development; F. Torres-Ruiz: supervision, funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research is partially supported by PID2020-1187879GB-100 and CEX2020-001105-M grants, funded by MCIN/AEI/10.13039/501100011033 (Spain).

Conflict of interest

Francisco Torres-Ruiz is a Guest Editor of special issue "Stochastic modeling and forecasting in dynamic systems" for AIMS Mathematics. Francisco Torres-Ruiz was not involved in the editorial review and the decision to publish this article. All author declare no conflict of interest in this paper.

References

1. P. Chávez, G. Rodríguez, Time changing effects of external shocks on macroeconomic fluctuations in Perú: Empirical application using regime-switching VAR models with stochastic volatility, *Rev. World Econ.*, **159** (2023), 505–544. https://doi.org/10.1007/s10290-022-00474-1
2. P. Clavijo-Cortés, Is unemployment hysteretic or structural? A Bayesian model selection approach, *Empir. Econ.*, **65** (2023), 2837–2866. https://doi.org/10.1007/s00181-023-02433-7
3. Z. Xie, Research on Bayesian financial panel data model based on BP neural network, In: *2023 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, 2023, 645–648. https://doi.org/10.1109/ICEDCS60513.2023.00124
4. J. M. Montaud, J. Dávalos, N. Pécastaing, Socioeconomic risks of extreme El Niño event-related road damages in Perú, *Environ. Model Assess.*, **27** (2022), 831–851. https://doi.org/10.1007/s10666-022-09830-9
5. F. A. R. Gomes, L. C. M. Melo, G. P. Soave, Flexible markov-switching models with evolving regime-specific parameters: An application to Brazilian business cycles, *Appl. Econ.,* **56** (2024), 1705–1722. https://doi.org/10.1080/00036846.2024.2305621

6. Q. Cui, S. Rong, F. Zhang, X. Wang, Exploring and predicting China's Consumer Price Index with its influence factors via big data analysis, *J. Intell. Fuzzy Syst.*, **46** (2024), 891–901. https://doi.org/10.3233/JIFS-234102
7. P. Wang, Y. Shen, L. Zhang, Y. Kang, Equilibrium investment strategy for a DC pension plan with learning about stock return predictability, *Insur. Math. Econ.*, **100** (2021), 384–407. https://doi.org/10.1016/j.insmatheco.2021.07.001
8. Y. Zhu, M. Escobar-Anel, M. Davison, A polynomial-affine approximation for dynamic portfolio choice, *Comput. Econ.*, **62** (2023), 1177–1213. https://doi.org/10.1007/s10614-022-10297-9
9. S. Liu, Y. Yang, H. Zhang, Y. Wu, Variance swap pricing under Markov-modulated jump-diffusion model, *Discrete Dyn. Nat. Soc.*, **2021** (2021), 9814605. https://doi.org/10.1155/2021/9814605
10. L. Tao, Y. Lai, Y. Ji, X. Tao, Asian option pricing under sub-fractional Vasicek model, *Quant. Finan. Econ.*, **7** (2023), 403–419. https://doi.org/10.3934/QFE.2023020
11. K. Alakkari, S. Yadav, P. Mishra, Measuring economic uncertainty in Syria: An approach to the stochastic volatility model, *Indian J. Econ. Dev.*, **18** (2022), 281–291. https://doi.org/10.35716/IJED/21242
12. D. Liang, B. Ewing, E. Cardella, L. Song, Probabilistic modeling of small business recovery after a hurricane: A case study of 2017 Hurricane Harvey, *Nat. Hazards Rev.*, **24** (2023), 05022012. https://doi.org/10.1061/(asce)nh.1527-6996.0000602
13. B. Herzog, Modeling inflation dynamics with fractional Brownian motions and Lévy processes, In: *Linear and non-linear financial econometrics-theory and practice*, 2020. https://doi.org/10.5772/intechopen.92292
14. F. Mehrdoust, I. Noorani, Implied higher order moments in the Heston model: A case study of S&P500 index, *Decisions Econ. Finan.*, **46** (2023), 477–504. https://doi.org/10.1007/s10203-023-00396-z
15. A. González, L. Mahadeva, J. D. Prada, D. Rodríguez, Policy analysis tool applied to Colombian needs: Patacon model description, *Borradores de Economía, num.656,* 2011, Banco de la República de Colombia. https://doi.org/10.32468/be.656
16. J. A. Cárdenas-Cárdenas, D. J. Cristiano-Botia, N. Martínez-Cortés, Colombian inflation forecast using long short-term memory approach, *Borradores de Economía, num.1241*, 2023, Banco de la República de Colombia. https://doi.org/10.32468/be.1241
17. H.Y. Rodríguez Pinzón, Estudio del fenómeno de inflación importada vía precios del petróleo y su aplicación al caso colombiano mediante el uso de modelos VAR para el periodo 2000–2009, *Estudios Gerenciales*, **27** (2011), 79–97. https://doi.org/10.1016/S0123-5923(11)70182-6
18. E. González Molano, L. F. Melo Velandia, A. G. Olarte, Pronósticos directos de la inflación colombiana, *Borradores de Economía, num.458*, 2007, Banco de la República de Colombia. https://doi.org/10.32468/be.458
19. C. R. Jaramillo Herrera, J. Tovar, Incidencia del impuesto al valor agregado sobre los precios en Colombia, *Documentos CEDE 2830*, 2007.
20. M. C. Aristizábal, Evaluación asimétrica de una red neuronal: Aplicación al caso de la inflación en Colombia, *Lecturas de Economía*, **65** (2009), 73–116. https://doi.org/10.17533/udea.le.n65a2641

21. E. Villa, M. A. Misas, A. F. Giraldo, Inflation targeting and an optimal Taylor rule for an open economy: Evidence for Colombia 1990–2011, *Lat. Am. J. Econ.*, **51** (2014), 41–83.

22. I. Luz-Sant'Ana, P. Román-Román, F. Torres-Ruiz, Modeling oil production and its peak by means of a stochastic diffusion process based on the Hubbert curve, *Energy*, **133** (2017), 455–470. http://doi.org/10.1016/j.energy.2017.05.125

23. A. Di Crescenzo, P. Paraggio, F. Torres-Ruiz, A Bertalanffy-Richards growth model perturbed by a time-dependent pattern, statistical analysis and applications, *Commun. Nonlinear Sci.*, **139** (2024), 108258. https://doi.org/10.1016/j.cnsns.2024.108258

24. A. Barrera, P. Román-Román, F. Torres-Ruiz, Hyperbolastic models from a stochastic differential equation point of view, *Mathematics*, **9** (2021), 1835. http://doi.org/10.3390/math9161835

25. G. Albano, V. Giorno, V., P. Román-Román, F. Torres-Ruiz, Study of a general growth model, *Commun. Nonlinear Sci. Numer. Simul.*, **107** (2022), 106100. http://doi.org/10.1016/j.cnsns.2021.106100

26. P. Román-Román, J. J. Serrano-Pérez, F. Torres-Ruiz, Some notes about inference for the lognormal diffusion process with exogenous factors, *Mathematics*, **6** (2018), 85. http://doi.org/10.3390/math6050085

27. G. Tintner, J. K. Sengupta, *Stochastic economics: Stochastic processes, control, and programming*, New York: Academic Press, 1972. https://doi.org/10.1016/C2013-0-11604-2

28. R. Gutiérrez, P. Román, D. Romero, F. Torres, Forecasting for the univariate lognormal diffusion process with exogenous factors, *Cybernet. Syst.*, **34** (2003), 709–724. http://doi.org/10.1080/716100279

29. C. E. Land, *Hypothesis tests and interval estimates,* In: *Lognormal distributions: Theory and applications*, New York: Marcel Dekker, 1988, 87–112. https://doi.org/10.1201/9780203748664

30. R. Gutiérrez, P. Román, F. Torres, Inference on some parametric functions in the univariate lognormal diffusion process with exogenous factors, *Test*, **10** (2001), 357–373. http://doi.org/10.1007/bf02595702

31. R. Gutiérrez, N. Rico, P. Román, F. Torres, Approximate and generalized confidence bands for the mean and mode functions of the lognormal diffusion process, *Comput. Stat. Data An.*, **51** (2007), 4038–4053. http://doi.org/10.1016/j.csda.2006.12.027

32. F. J. Girón, M. L. Martínez, E. Moreno, F. Torres, Objective testing procedures in linear models: Calibration of the p-values, *Scand. J. Stat.*, **33** (2006), 765–784. http://doi.org/10.1111/j.1467-9469.2006.00514.x

33. E. Moreno, F. J. Vázquez-Polo, M. A. Negrín, *Bayesian cost-effectiveness analysis of medical treatments*, Boca Raton: Chapman and Hall/CRC, 2019. https://doi.org/10.1201/9781315188850

34. E. Moreno, J. J. Serrano-Pérez, F. Torres-Ruiz, Consistency of Bayes factors for linear models, *Rev. Real Acad. Cienc. Exactas Fis. Nat. Ser. A-Mat.*, **119** (2025), 20. http://doi.org/10.1007/s13398-024-01685-x