BORRADORES DE ECONOMÍA

Inflation Drivers in Colombia: A Hemisphere Neural Network Approach

By:
Franky Juliano Galeano-Ramírez
Nicolás Martínez-Cortés
Hernán Dario Perdomo-Sánchez
Marlon Salazar

No. 1370
October, 2026

Inflation Drivers in Colombia: A Hemisphere Neural Network Approach*

Franky Juliano Galeano-Ramírez** Nicolás Martínez-Cortés
Hernán Dario Perdomo-Sánchez Marlon Salazar

The opinions expressed in this document are solely those of the authors and do not necessarily reflect the views of Banco de la República or its Board of Directors.

Abstract

This paper develops a neural network model to capture nonlinear relationships between inflation and a broad set of macroeconomic variables that may provide insights into the factors behind deviations of inflation from its target in Colombia. The framework estimates implicit latent factors for distinct economic domains and uses them to construct a historical decomposition of the inflation gap. Our specification organizes the information set into five hemispheres covering short-run inflation expectations and inertia, long-run inflation expectations, real activity, the real exchange rate, and cost pressures from international commodity prices and domestic factors. The results suggest that real activity is not the only source of predictive information about inflationary pressures. Inflation expectations, inertia, and cost-related factors also account for a substantial share of observed inflation movements, with their relative contributions varying across inflationary episodes. Overall, the framework provides a flexible and interpretable empirical decomposition of inflation dynamics that can complement structural approaches.

Keywords: Inflation Dynamics; Inflation Decomposition; Neural Networks; Interpretable Machine Learning.

JEL classification: E31, E32, E37, C45, C53.

*We thank Carlos Huertas, Adolfo Cobo and Viviana Alfonso for their valuable comments and suggestions throughout the development of this research. We are also grateful to Philippe Goulet Coulombe for generously sharing the code used in the paper "*A Neural Phillips Curve and a Deep Output Gap*," which served as a valuable reference for our empirical implementation. We also thank José Pulido and the other participants of the Banco de la República Seminar for their helpful comments.

**All authors are Specialized Economists at Banco de la República, Bogotá, Colombia. E-mails: fgaleara@banrep.gov.co (F. Galeano-Ramírez); nmartico@banrep.gov.co (N. Martínez-Cortés); hperdosa@banrep.gov.co (H. Perdomo-Sánchez); and msalazsi@banrep.gov.co (M. Salazar).

Determinantes de la inflación en Colombia: aproximación mediante una red neuronal por hemisferios*

Franky Juliano Galeano-Ramírez** Nicolás Martínez-Cortés
Hernán Dario Perdomo-Sánchez Marlon Salazar

Las opiniones expresadas en este documento son exclusivamente responsabilidad de los autores y no reflejan necesariamente las opiniones del Banco de la República ni de su Junta Directiva.

Resumen

Este artículo desarrolla un modelo de redes neuronales para capturar relaciones no lineales entre la inflación y un amplio conjunto de variables macroeconómicas asociadas con sus desviaciones respecto a la meta en Colombia. El modelo estima factores latentes para distintas dimensiones económicas y permite construir una descomposición histórica de la brecha de inflación. La información se organiza en cinco hemisferios: expectativas e inercia de corto plazo, expectativas de largo plazo, actividad económica, tasa de cambio real y presiones de costos. Los resultados sugieren que la actividad económica no es la única fuente de información predictiva sobre las presiones inflacionarias. Las expectativas, la inercia y los costos también tienen contribuciones relevantes, cuya importancia varía entre episodios inflacionarios. En conjunto, el modelo ofrece una descomposición flexible e interpretable de la dinámica inflacionaria que puede complementar los enfoques estructurales.

Palabras clave: Dinámica de la inflación; Descomposición de la inflación; Redes Neuronales; Aprendizaje Automático Interpretable.
Clasificación JEL: E31, E32, E37, C45, C53.

*Agradecemos a Carlos Huertas, Adolfo Cobo y Viviana Alfonso por sus valiosos comentarios y sugerencias durante el desarrollo de esta investigación. Asimismo, agradecemos a Philippe Goulet Coulombe por compartir generosamente el código utilizado en el artículo "*A Neural Phillips Curve and a Deep Output Gap*", el cual sirvió como una valiosa referencia para nuestra implementación empírica. También agradecemos a José Pulido y a los demás participantes del Seminario del Banco de la República por sus útiles comentarios.

**Todos los autores son economistas especializados del Banco de la República, Bogotá, Colombia. Correos electrónicos: fgaleara@banrep.gov.co (F. Galeano-Ramírez); nmartico@banrep.gov.co (N. Martínez-Cortés); hperdosa@banrep.gov.co (H. Perdomo-Sánchez); and msalazsi@banrep.gov.co (M. Salazar).

1. Introduction

Conducting monetary policy under an inflation-targeting regime entails a fundamental observability problem: the structural variables that inform policy decisions, most notably the output gap (y_t) and inflation expectations (E_t), are latent constructs that cannot be directly observed in the data. In a small open economy such as Colombia, this challenge is compounded by the prevalence of supply shocks, substantial exchange rate volatility, and the persistence arising from institutional indexation mechanisms. In this context, conventional approaches to estimating marginal costs and inflation expectations, both of which are key determinants of inflation dynamics, may be insufficient. These approaches typically rely on linear frameworks to infer unobservable variables and impose smoothness assumptions that may fail to capture abrupt adjustments and nonlinear patterns in the data, thereby limiting their ability to accurately characterize the underlying inflation process.

This paper proposes an alternative empirical strategy to decompose deviations of inflation from its target in Colombia, based on Hemisphere Neural Networks (HNN) and, in particular, the factorized architecture (HNN-F) introduced by Goulet Coulombe (2025). The proposed approach combines the flexibility of deep learning with a structured representation that facilitates economic interpretation, distinguishing it from conventional machine learning implementations that operate as black boxes. Information is organized into thematic blocks, referred to as hemispheres, representing short-run expectations and inflation inertia, long-run expectations, real activity, the real exchange rate, and costs. This structure enables the supervised extraction of latent states from a rich and systematically organized set of macroeconomic indicators while reducing reliance on restrictive ex ante parametric assumptions.

The primary contribution of this paper is to adapt this framework to the institutional and macroeconomic characteristics of the Colombian economy. Relative to the original specification for the United States, we introduce two extensions. First, we incorporate a dedicated Real Exchange Rate hemisphere, informed by terms of trade indicators and risk premia, to capture nonlinearities in exchange rate pass-through. Second, we refine the Costs hemisphere by incorporating measures of domestic price pressures, including regulated fuel prices and agricultural input costs, as well as wage measures whose dynamics differ from those of international commodity prices. We also follow Goulet Coulombe (2025) in decomposing inflation expectations into two latent components: a short-run component associated with inertia and indexation, and a long-term component related to the credibility of the inflation

target. This decomposition enables a more granular characterization of inflation persistence.

The empirical results suggest substantial heterogeneity across inflationary episodes. The historical decomposition of the latent states points to important differences in the drivers of inflation dynamics across episodes. The model associates the 2006–2008 episode with a combination of strong activity indicators and elevated cost pressure contributions, whereas the 2014–2017 increase in predicted inflation exhibits a substantial predictive contribution from the real exchange rate hemisphere following the decline in oil prices.

For the post-pandemic period from 2021 to 2024, the framework registers a sequence of overlapping predictive contributions, beginning with a rebound in the contribution of the activity hemisphere and followed by rising domestic and external cost pressure indicators that coincided with stronger indexation patterns and elevated short-run expectations. In general, the decomposition of deviations of inflation from its target suggests that real activity is not the only factor that could help explain inflation dynamics. Factors related to inflation expectations and cost pressures also make substantial contributions and could be relevant for explaining deviations of inflation from its target.

The macroeconomic coherence of the estimated latent states is assessed by comparing them with independent benchmark indicators. The HNN's implicit real activity factor exhibits a close association with the output gap estimates produced by the Central Bank of Colombia (Banco de la República) (Pulido-Mahecha et al., 2024; Ramos-Veloza et al., 2026), while the contribution of the exchange rate hemisphere displays dynamics that are consistent with measures of real exchange rate misalignment (Salazar-Díaz et al., 2023). Taken together, these results suggest that the HNN framework could serve as a useful complementary diagnostic tool for monetary policy analysis.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the HNN architecture and its adaptation to the Colombian economy. Section 4 describes the data and the corresponding transformations. Section 5 presents the empirical results and the historical decomposition of inflation deviations. Section 6 examines the robustness of the results using alternative measures of the latent variables. Section 7 presents the pseudo-out-of-sample evaluation. Finally, Section 8 concludes.

2. Literature Review

Recent advances in inflation modeling have increasingly leveraged deep learning to revisit the Phillips Curve (PC) and the output gap, two constructs that are central to inflation targeting regimes but notoriously difficult to measure. A milestone in this literature is Goulet Coulombe (2025), who proposes the HNN, an architecture that imposes an interpretable separation of the latent drivers of inflation, such that the final layer corresponds to an approximation of the Neural Phillips Curve. In its U.S. application, the HNN combines flexible nonlinear function approximation with economic structure and attributes the 2021 inflation surge first to a loss of anchoring in inflation expectations and subsequently to an unusually tight real economy. These results suggest that standard Phillips Curves may understate the role of real activity when latent gaps are extracted from richer information sets (Goulet Coulombe, 2025). Subsequent work extends the hemispherical architecture to density forecasting, with dedicated mean and variance hemispheres and regularization mechanisms designed to limit overfitting. This extension suggests that the hemispherical approach can be applied beyond point forecasts and incorporated into other structured macroeconomic models (Coulombe et al., 2023; Goulet Coulombe, 2025).

Viewed against the broader macro–ML literature, the HNN sits at the intersection of three strands: (i) output-gap and PC estimation; (ii) deep learning for macroeconomic forecasting; and (iii) interpretable AI for policy use. On the second front, Coulombe et al. (2022) show that machine learning's predictive edge over classical econometrics in macro data arises chiefly from capturing nonlinearities, with standard factor regularization continuing to perform well for dimensionality reduction. Empirically, modern architectures have proved competitive in inflation prediction across settings: transformers sometimes outperform strong time–series baselines at specific frequencies and horizons (Tschuchnig et al., 2023), while models that aggregate disaggregated price forecasts using ML (e.g., random forests) can outperform aggregate-only approaches, particularly during volatile periods (Boaretto and Medeiros, 2023). These developments, though heterogeneous in method and scope, converge on the insight that data-rich, nonlinear methods can improve inflation forecasting without entirely sacrificing interpretability if their structure is aligned with economic theory—precisely the motivation behind hemispherical designs.

Turning to Colombia, the literature on the Phillips Curve and the determinants of inflation has a long empirical tradition that highlights heterogeneity across CPI components, the importance of supply shocks, and the persistence associated with indexation. Early evidence

documents a nonlinear and convex Phillips Curve for Colombia, with policy-relevant asymmetries in the sacrifice ratio (Gómez and Julio, 2000), followed by further evidence supporting nonlinearities in the Phillips Curve (Nigrinis-Ospina, 2004). More recently, a disaggregated, bottom-up Phillips Curve finds a relatively flat aggregate relationship but steeper sectoral trade-offs, with supply shocks affecting tradable and food prices and indexation playing an important role in nontradable and regulated prices (Lanau et al., 2018).

The Colombian policy modeling framework also incorporates this disaggregation: the semi-structural 4GM includes four Phillips Curves for tradables, nontradables, food, and regulated prices, and links their marginal costs to both the output gap and the real exchange rate gap, while incorporating external price pressures such as oil and imported food into the supply block (González et al., 2020). In parallel, research on output gap measurement in Colombia emphasizes the sensitivity of the estimated gap to the underlying methodology and to commodity price cycles, reinforcing the case for flexible, data-driven extraction within a framework guided by economic structure (Cobo-Serna, 2005; González et al., 2013).

Colombia's recent experience with machine learning for inflation forecasting and expectations provides useful guidance for the application of machine learning and deep learning models to the Colombian economy. On the forecasting side, Cárdenas-Cárdenas et al. (2023) deploy LSTM networks to predict headline inflation up to one year ahead, finding that models that augment univariate inflation with additional predictors perform particularly well at medium horizons. Complementing the use of deep neural networks, Roldán-Ferrín and Parra-Polanía (2025) show that random forests combined with Google Trends can improve monthly nowcasts of Colombian inflation, illustrating the potential gains from incorporating high-frequency external signals through regularized machine learning methods. Regarding expectations and Phillips Curve specification, recent evidence indicates that the measured relationship between inflation expectations and inflation depends on how expectations are constructed and observed, with implications for the representation of the expectations hemisphere in neural Phillips Curves (Sánchez-Jabba and Villabón-Hinestroza, 2024; Sánchez-Jabba et al., 2023). Taken together, these contributions suggest a practical empirical strategy for Colombia: assemble a broad, high-frequency panel that includes sectoral price indices, real activity indicators, exchange rate and commodity price gaps, survey and market-based expectations, and digital-trace proxies; then use a hemispherical architecture to (i) supervise the extraction of a Colombia-specific output gap from indicators of economic tightness and external variables, and (ii) distinguish expectation-driven from gap-related inflation pressures in a way that remains economically interpretable and useful for policy discussion.

In sum, the case for adapting the HNN to Colombia rests on three pillars. First, the hemispherical structure offers interpretability that is otherwise costly to obtain in deep models, mapping naturally to the Colombian PC's sectoral and expectations–gap decomposition (Goulet Coulombe, 2025; González et al., 2020). Second, Colombia-specific ML evidence shows that both deep and tree-based models already add value to inflation forecasting when fed richer signals, paving the way for HNN-style latent-state extraction (Cárdenas-Cárdenas et al., 2023; Roldán-Ferrín and Parra-Polanía, 2025). Third, the institutional and structural features emphasized by the Colombian literature—supply shocks, indexation, exchange-rate pass-through, and commodity cycles—are precisely the kinds of nonlinearities and interactions for which a supervised, interpretable deep architecture is well suited (Lanau et al., 2018; Gómez and Julio, 2000; González et al., 2013). A Colombian application should therefore emphasize sectoral inputs, robust reality checks to prevent overfitting, and careful construction of expectations series, with the ultimate goal of delivering both better forecasts and a transparent decomposition of inflation dynamics that is usable within the inflation-targeting framework.

3. Model

3.1. Architecture

This subsection outlines the structure of the HNN model. The model establishes a predictive relationship between the dependent variable y_t and a set of predictors organized into k hemispheres, denoted as $H_1, \ldots, H_k$. Each hemisphere comprises a block of variables designed to capture conceptually distinct sources of information.

$$y_t = f(H_{1,t}, \ldots, H_{k,t}), \tag{1}$$

The HNN architecture processes each hemisphere through a neural network that generates latent contributions $h_{t,j}$, representing nonlinear transformations of the predictors within hemisphere H_j. These contributions are combined additively in the final aggregation layer:

$$\hat{y}_t = \sum_{j=1}^{k} h_{t,j}. \tag{2}$$

In the baseline HNN proposed by Goulet Coulombe (2025), each hemisphere is modeled using a multilayer perceptron (MLP) — a supervised feedforward neural network composed of multiple layers: an input layer, one or more hidden layers, and an output layer. Each layer contains at least one neuron, referred to as a perceptron. The input layer contains n input variables $X = \{x_1, \ldots, x_n\}$ and the output layer contains m output variables $Y = \{y_1, \ldots, y_m\}$. An MLP with k hidden layers iteratively computes:

$$x^{(\ell+1)} = \sigma\big(W^{(\ell)} x^{(\ell)} + b^{(\ell)}\big), \qquad \ell = 0, \ldots, k-1, \tag{3}$$

This formulation describes how the network transforms the data layer by layer, where $x^{(0)}$ denotes the hemispheric inputs and $x^{(\ell+1)}$ represents the output of layer ℓ after applying weights, biases, and activation functions. The activation function σ (e.g., ReLU, sigmoid, tanh) introduces non-linearity, enabling the network to learn complex patterns. $W^{(\ell)}$ is the weight matrix for layer ℓ, with dimensions $[n_{\ell+1} \times n_\ell]$, where n_ℓ is the number of neurons in layer ℓ. This matrix determines how the inputs are linearly combined. $b^{(\ell)}$ is the bias vector for the layer ℓ, of dimension $n_{\ell+1}$, which allows the activation function to shift and thus increase flexibility. The parameter set

$$\theta = \{W^{(\ell)}, b^{(\ell)}\}_{\ell=0}^{k-1}$$

is updated during training by the algorithm and is estimated by minimizing the squared-error loss, which quantifies the difference between the predicted values $\hat{y}_t(\theta)$ and the true outcomes y_t:

$$\min_{\theta} \sum_{t} \left(y_t - \hat{y}_t(\theta)\right)^2. \tag{4}$$

Optimization is performed using gradient-based algorithms, such as stochastic gradient descent or Adam. Gradients are computed efficiently through the backpropagation algorithm, which enables the network to iteratively adjust its parameters and thereby learn complex mappings from inputs to outputs.

3.2. Factorized Hemisphere Neural Network (HNN-F)

To enhance interpretability, the factorized model (HNN-F) decomposes the effect of selected hemispheres ($h_{t,j}$) into a latent state (denoted as $g_{j,t}$) and a time-varying coefficient (de-

noted as $\gamma_{j,t}$). This factorization requires some assumptions on $g_{j,t}$ and $\gamma_{j,t}$[1]. Consider the restricted set where variable k is excluded from the set of predictors originally included in the hemisphere j, which is denoted as $\mathcal{H} \setminus k$. The factorization followed is $\gamma_{j,t} = h_{\gamma_j}(t)$ and $g_{j,t} = h_{g_j}(H_j \setminus t)$. For its implementation, an additional hemisphere containing only t is created. Then, in the final layer, rather than summing some contributions ($h_{t,j}$), some outputs of the last layer are multiplied together. That is, the output of the hemisphere that contains only t will be multiplied by that of $\mathcal{H} \setminus t$ and the product will be added to the rest of the sum that constitutes the HNN.

The final layer in HNN-F is therefore given by:

$$\hat{y}_t = \sum_{j=1}^{n^*} h_{\gamma_j}(t)\, h_{g_j}(H_j \setminus t) \; + \; \sum_{k \in J_0} h_k(t). \tag{5}$$

Hemispheres indexed by $j = 1, \ldots, n^*$ enter multiplicatively, while those in J_0 enter additively.

To address scale indeterminacy in the factorized terms, the coefficients $h_{\gamma_j}(t)$ are constrained to be nonnegative and normalized to have sample mean one. These assumptions follow the identification logic in the original formulation.

3.3. Application to Inflation Deviations in Colombia

In this section, we specify a neural network model of inflation deviations for the Colombian economy, following and extending the approach of Goulet Coulombe (2025). Specifically, we estimate the factors associated with deviations of inflation from its target. To guide the selection of relevant variables, we draw on the standard New Keynesian Phillips Curve, in which inflation dynamics are related to inflation expectations, marginal costs, typically represented by the output gap, cost pressures stemming from commodity price cycles, and an idiosyncratic disturbance term. We emphasize that this theoretical structure is used to organize the information set and assign economic content to each hemisphere, rather than to impose a structural identification assumption.

In addition, we incorporate the real exchange rate, which is particularly relevant in a small open economy (Gali and Monacelli, 2005). The dependent variable is defined as the devi-

[1]Following Goulet Coulombe (2025), we refer to the factorized model as HNN-F. For simplicity, throughout the remainder of the paper, we use HNN to refer to this model.

ation of annual inflation from the central bank's inflation target, $\hat{\pi}_t = \pi_t - \alpha$, so that the inflation gap is expressed directly relative to the target.[2] The resulting inflation dynamics are described by

$$\hat{\pi}_t = \sum_{i=1}^{n} \lambda_i \pi_{t-i} + \theta E_t(\pi_{t+i}) + \beta y_t + \gamma z_t + \varepsilon_t \tag{6}$$

where π_t denotes the annual inflation rate measured using the Consumer Price Index (CPI), α denotes the Banco de la República's inflation target, set at 3.0%, $E_t(\pi_{t+i})$ denotes inflation expectations, y_t denotes the output gap, z_t denotes the real exchange rate gap, and ε_t represents supply disturbances. The first term on the right-hand side captures the contribution of lagged inflation to current inflation dynamics. The flexibility of the HNN framework does not impose the parametric restrictions required to interpret the estimated coefficients as structural parameters or to identify structural shocks underlying deviations of inflation from its target. Instead, the model provides an empirical decomposition based on latent states estimated through the HNN architecture, allowing us to characterize the relative contributions of different groups of variables that could help explain deviations of inflation from its target.

Based on this framework, we estimate the factors associated with inflation deviations using the HNN architecture. Each term in Equation (6) is represented by a separate hemisphere. Following the approach of Goulet Coulombe (2025), the first hemisphere captures inflation persistence by jointly incorporating annual inflation rates for food, goods, services, and regulated prices, together with short-run inflation expectations derived from both survey-based and market-based measures. We additionally introduce a hemisphere dedicated to medium and long-term inflation expectations. The output gap hemisphere draws on a broad set of indicators covering economic activity, investment, consumption, labor market dynamics, sentiment, and sectoral activity, thereby accounting for heterogeneity across sectors. The exchange rate hemisphere comprises indicators closely related to exchange rate dynamics, including terms of trade and risk premia. Finally, the cost pressure hemisphere captures pressures arising from domestic gasoline and gas prices, agricultural input prices, labor costs, transportation costs, and producer price indicators, among others. Domestic gasoline and

[2] Although the official inflation target varied during the early part of the sample, these changes affect only a small fraction of the estimation period. We therefore use the current 3% target throughout the analysis, as it represents the long-run inflation objective consistent with price stability, and define the inflation gap accordingly.

gas prices are particularly relevant because they do not fully reflect international price movements due to the presence of subsidies. The resulting specification for estimated inflation deviations is given by:

$$\hat{\pi}_t = \underbrace{\beta_t \overbrace{EST_t}^{h_{t,EST}}}_{\mathcal{H}_{EST}} + \underbrace{\theta_t \overbrace{ELT_t}^{h_{t,ELT}}}_{\mathcal{H}_{ELT}} + \underbrace{\gamma_t \overbrace{g_t}^{h_{t,g}}}_{\mathcal{H}_g} + \underbrace{\delta_t \overbrace{RER_t}^{h_{t,RER}}}_{\mathcal{H}_{RER}} + \underbrace{\zeta_t \overbrace{c_t}^{h_{t,c}}}_{\mathcal{H}_c} \tag{7}$$

This is a five-factor model in which $\mathcal{H}_{EST}$, $\mathcal{H}_{ELT}$, $\mathcal{H}_g$, $\mathcal{H}_{RER}$, and $\mathcal{H}_c$ represent the contributions of short-run expectations, long-run expectations, real activity, the real exchange rate, and cost pressures, respectively. Each contribution can be decomposed into a time-varying coefficient, β_t, θ_t, γ_t, δ_t, or ζ_t, and a latent factor, $h_{t,EST}$, $h_{t,ELT}$, $h_{t,g}$, $h_{t,RER}$, or $h_{t,c}$. The time-varying coefficients are constrained to be positive to reflect the expected directional relationship between each factor and inflation dynamics. Both the latent factors and the time-varying coefficients are estimated flexibly within the HNN architecture, allowing the model to capture nonlinearities in the relationships between the underlying determinants and deviations of inflation from its target. Each latent factor summarizes the information contained in a set of observed variables assigned to its corresponding hemisphere, providing a parsimonious representation of their common information content.

The five hemispheres are modeled as separate components without direct connections between them. This structure limits interactions across groups of variables, but allows the contribution of each hemisphere to the deviation of inflation from its target to be examined separately. In this way, the HNN combines the flexibility of a nonlinear neural network with an economically interpretable structure that clarifies how different groups of variables relate to the inflation gap. The resulting five-hemisphere HNN architecture is illustrated in Figure 1.

3.4. Estimation and Tuning

Each hemisphere $\mathcal{H}$ consists of a fully connected feed-forward neural network configured with 3 hidden layers and 700 neurons per layer to capture the nonlinearities of the Colombian macroeconomic dataset. Consistent with the HNN framework, we estimate time-varying coefficients using dedicated trend hemispheres, each specified with 3 hidden layers and 100 neurons. This dual architecture allows for the flexible estimation of time-varying predictive sensitivities without over-parameterizing low-frequency components.

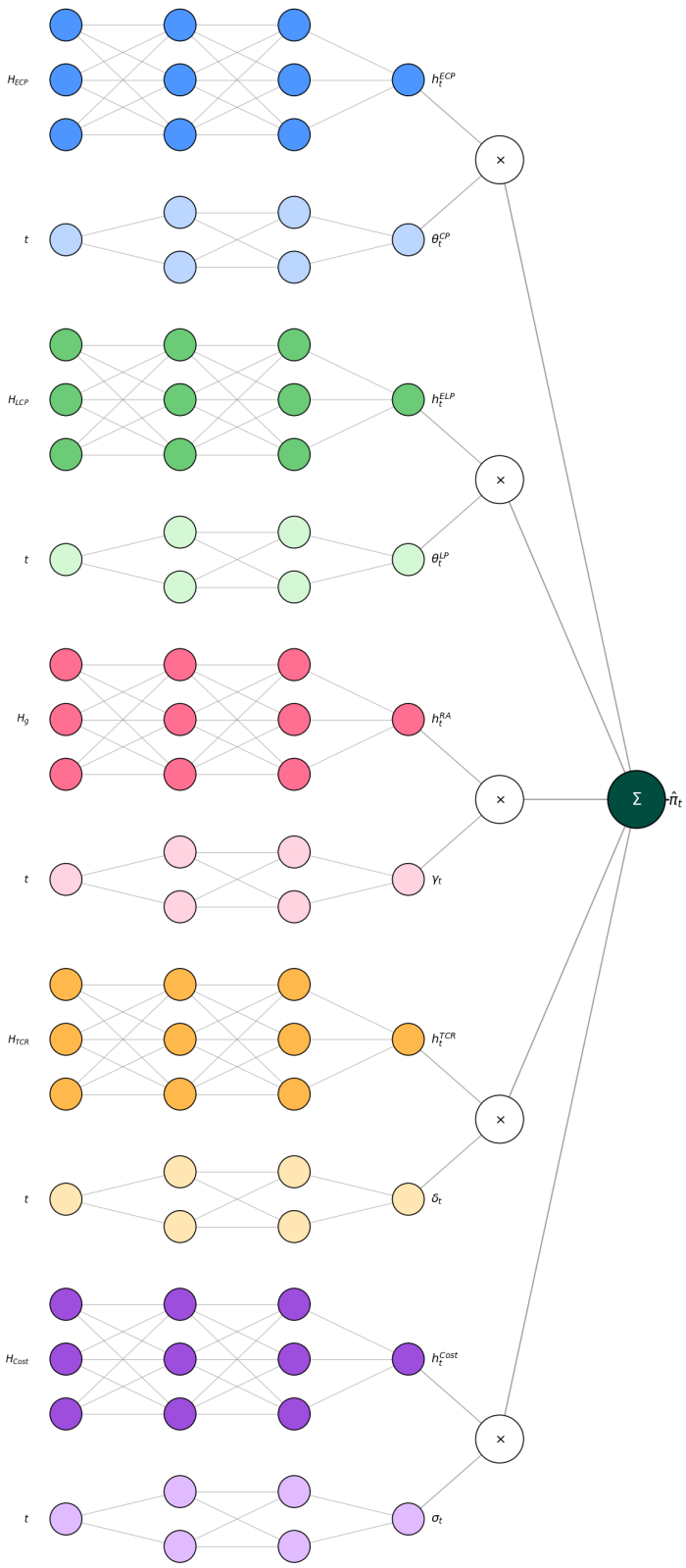

Figure 1: Hemisphere Neural Network's Architecture, with five hemispheres, four regressors per hemisphere, and (for each hemisphere) three hidden layers of three neurons each.

We estimate the model using the Adam optimizer with a learning rate of 0.01 within a bagging (bootstrap aggregating) framework of $B = 1{,}000$ replications. The sample is split at October 2022: the first 85% of the observations form the training set and the remaining 15% are held out as a test set. Within each replication, 80% of the training observations are used to fit the network and the remaining 20% serve as an out-of-bag validation set. Because the training observations are resampled at random within each replication, serial dependence is broken during fitting and the algorithm treats observations as independent.

To prevent overfitting, we employ early stopping with a maximum of 400 epochs and a patience window of 100 epochs, alongside a dropout rate of 0.2 across all layers. Inputs $X_{t,k} \in \mathcal{H}_j$ are normalized by dividing by $\sqrt{\text{card}(\mathcal{H}_j)}$ to ensure balanced a priori variance across components. The final estimates for the latent components $h_{t,j}$ are derived by averaging the outputs of the bootstrap models, with uncertainty quantified via the resulting empirical distribution.

3.5. Regularization and Overfitting Control

The HNN methodology embeds overfitting control as a design feature of the estimation itself. Overfitting refers to the inability of a model to generalize beyond the estimation sample: an estimator flexible enough that, in addition to the structural relationship linking the predictors to inflation, it also fits noise that is specific to the sample at hand. In terms of the bias-variance decomposition of prediction error, overfitting arises when the variance component dominates. The low bias of a highly flexible model is then achieved by absorbing sample-specific fluctuations that, by their nature, do not recur, so that a strong in-sample fit coexists with deteriorating out-of-sample performance. Keeping this variance component in check is therefore essential both for forecast accuracy and for the stability of the latent decompositions at the core of our analysis.

Following Goulet Coulombe (2025), these controls operate at three distinct stages of the estimation procedure: the architecture of the HNN model, the estimation process itself, and the regularization techniques applied during training. Each stage addresses a different source of overfitting, and together they improve the model's ability to generalize by reducing both the variance of the estimator and the extent to which it captures sample-specific noise. We discuss each stage in turn.

Architecture of the HNN model. The first layer of control is built into the network before any parameter is estimated, since the architecture itself restricts the space of functions the model

can represent. Predictors are partitioned into economically motivated hemispheres that are processed in isolation throughout the hidden layers and interact only in the final aggregation layer. Relative to a fully connected network, this separability removes by construction the vast set of cross-group interactions, which constitutes the dominant source of spurious fit in high-dimensional settings, thereby reducing the variance of the estimator while preserving the economic interpretability of its components (Hastie, 2017).

The identification restrictions play a complementary role: each contribution is factorized into the product of a latent state and a time-varying coefficient, where the latter is restricted to take non-negative values, via an absolute-value transformation applied before the final layer, and to evolve slowly as a function of time alone. These restrictions shrink the set of observationally equivalent decompositions, stabilizing the estimated gaps and coefficients across replications.

In addition, as noted above, the standardized inputs of hemisphere j are divided by $\sqrt{\text{card}(\mathcal{H}_j)}$, where $\text{card}(\mathcal{H}_j)$ denotes the number of predictors assigned to that hemisphere. This equalizes the a priori variance of every hemisphere's output under the ridge-type shrinkage induced by early stopping; without it, data-rich hemispheres would mechanically absorb variation, including spurious variation, at the expense of the others (Goulet Coulombe, 2025).

As described in Section 4, all predictors enter through lagged values and their Moving Average Rotations (MARX; Coulombe et al., 2021). Under the implicit ridge prior, this representation acts as a smoothness prior on the lag polynomials, shrinking the coefficients of neighboring lags toward one another and penalizing the erratic lag profiles that typically arise from fitting noise rather than persistent economic dynamics.

Estimation procedure. As mentioned above, the sample is first divided into a training set and a test set. The training set comprises all observations up to October 2022, corresponding to approximately 85% of the full sample, while the test set contains the remaining 15%. The test set is completely excluded from the estimation procedure and is never used to update the network parameters or to select the final model during training.

The model is then estimated 1,000 times using a random resampling scheme similar to bagging (Breiman, 1996). Importantly, this resampling procedure is applied exclusively to the 85% of observations that constitute the training set. In each replication, the training set is further divided into two subsets: 80% of the observations in the training set are used to train the network, while the remaining 20% are reserved as an out-of-bag validation set.

Thus, the 80% and 20% shares refer exclusively to the original 85% training sample and do not involve the 15% test set. The out-of-bag observations are never used to update the network weights and are reserved exclusively for evaluating the network during the estimation process. Each replication also starts from a different random initialization of the network weights, resulting in a collection of independently trained neural networks.

After completing the 1,000 replications, the estimates corresponding to the training sample are obtained by averaging the out-of-bag predictions available for each observation, that is, the predictions generated when the observation belongs to the 20% validation subset. Consequently, every fitted value is computed exclusively from models that were not trained on that particular observation. For the test period, the reported predictions are obtained by averaging the predictions from all 1,000 independently estimated networks, since none of these observations are used during the estimation procedure. Finally, uncertainty bands are constructed from the dispersion of predictions across the 1,000 replications.

This estimation procedure controls overfitting through two complementary channels. The first is the averaging itself. The noise that any individual network absorbs depends on the particular subsample it was trained on and on its random initial weights, and both ingredients change from one replication to the next; the structural signal, in contrast, is common to all replications. Averaging across the ensemble therefore cancels a large share of the idiosyncratic component while preserving the common one, directly reducing the variance of the final estimator. The second channel is the out-of-bag construction. Since a flexible model fits best the observations it was trained on, which is precisely where memorization concentrates, computing every fitted value only from networks that never saw that observation guarantees that the reported historical series reflect genuine predictive content rather than memorized noise. Any overfitting that survives at the level of an individual replication remains confined to in-sample fits that never enter the reported estimates or their uncertainty bands.

Regularization techniques. The final layer of overfitting control operates within each of the 1,000 replications, through regularization techniques that constrain the learning process. The primary mechanism is early stopping. During each replication, the network weights are updated using only the 80% estimation subset, while predictive performance is evaluated after every epoch on the remaining 20% validation (out-of-bag) subset.

Because gradient-based optimization learns the dominant and persistent patterns in the data before fitting idiosyncratic fluctuations, the number of training epochs effectively determines the complexity of the network. Training therefore stops once the validation error has failed

to improve for a predetermined number of epochs, and the parameter values corresponding to the lowest validation error are retained. In this way, estimation halts before the network begins fitting noise rather than the underlying signal.

A second source of regularization is dropout. During training, each hidden neuron is independently deactivated with probability 0.2 at every epoch. Because the set of active neurons changes across training epochs, the network cannot rely on specific combinations of neurons to memorize individual observations. Instead, it is encouraged to learn distributed representations that remain robust to these random perturbations, thereby improving its ability to generalize to unseen data (Srivastava et al., 2014).

Finally, optimization is performed using full-batch Adam, so that every weight update is computed from the exact gradient evaluated on the entire 80% estimation subset rather than on randomly selected mini-batches. This produces smooth and stable validation-error trajectories, allowing the early-stopping criterion to identify a well-defined optimum while ensuring that the variability across replications reflects the resampled data instead of stochastic optimization noise (Goulet Coulombe, 2025).

Taken together, these regularization strategies reduce the risk of overfitting by limiting model complexity, preventing the memorization of idiosyncratic observations, and stabilizing the optimization process. The effectiveness of this framework is assessed through the pseudo-out-of-sample forecasting exercise presented in Section 7, where the HNN is repeatedly re-estimated over rolling windows and its forecasting performance is compared with both univariate time-series models and alternative machine-learning approaches.

Since genuine overfitting deteriorates predictive performance on unseen observations, competitive out-of-sample accuracy provides empirical evidence that the HNN captures the systematic determinants of inflation deviations rather than fitting transitory noise.

3.6. Identification and Interpretation: What the Model Does and Does Not Deliver

The HNN provides a flexible and internally consistent empirical decomposition of the predictive content carried by different groups of variables, and shows how their contributions to inflation deviations vary across episodes. By estimating the latent states jointly with their mapping into inflation, the framework summarizes the information contained in a rich set of macroeconomic indicators into economically interpretable hemispheres while allowing for

nonlinear relationships between the underlying variables and inflation deviations. The model can therefore exploit the flexibility of neural networks while retaining a structured representation of the forces shaping inflation dynamics. The resulting decomposition should thus be interpreted as an empirical characterization of the predictive contributions of different information blocks rather than as a structural decomposition of inflation.

This distinction is important for interpreting the estimated contributions. The HNN does not impose the restrictions required to identify structural shocks or causal effects, and the latent states should not be interpreted as causal factors. Several variables included in the different hemispheres are jointly determined with inflation: inflation expectations may respond to realized inflation, nominal wages and the statutory minimum wage are partly indexed to past inflation, and real activity and the real exchange rate respond to monetary policy, which itself responds to inflation developments. Consequently, the estimated contributions reflect predictive co-movements among correlated information blocks rather than partial equilibrium responses to individual variables. This limitation does not undermine the empirical exercise, but instead defines the appropriate interpretation of the estimated decomposition. The flexibility of the HNN allows these relationships to be represented without imposing a specific functional form, making it possible to capture nonlinearities that may be difficult to accommodate in more restrictive approaches.

The economic coherence and generalizability of the estimated decomposition are assessed through the robustness and out-of-sample exercises presented in Sections 6 and 7. The comparisons with independent measures of real activity, inflation expectations, and real exchange rate misalignment provide evidence on the coherence of the estimated latent states, while the pseudo out-of-sample exercises assess whether the estimated relationships generalize to observations that were not used for estimation. The strong performance across these exercises suggests that the HNN captures systematic information in the data rather than simply memorizing the estimation sample. These results should be interpreted in light of the broader literature on machine learning and deep learning, which emphasizes that the predictive relationships estimated by flexible learning algorithms do not, by themselves, provide causal interpretation (Mullainathan and Spiess, 2017; Zhao and Hastie, 2021; Farrell et al., 2021). This limitation is intrinsic to the predictive nature of these methods rather than specific to the HNN architecture.

Although the robustness and out-of-sample exercises do not establish exogeneity or causal identification, they provide evidence that the estimated latent states and their contributions contain economically relevant and generalizable information. The HNN is therefore best

viewed as a complementary empirical tool that exploits the flexibility of neural networks to capture nonlinear relationships and organize a rich information set into interpretable latent states, while structural models remain necessary when the objective is to identify structural shocks or causal effects.

4. Data

The baseline model described in the previous section uses monthly data and is estimated over January 2005–October 2025. The dependent variable considered in the main analysis, as previously mentioned, is the deviation of headline inflation (year-over-year percentage change in the Consumer Price Index) from the central bank target (3%), thus:

$$\pi_t - \pi^T = \left(\frac{CPI_t}{CPI_{t-12}} - 1 \right) - 3\% \tag{8}$$

Each hemisphere is fed a set of economic indicators. The details of the composition of each hemisphere are in Table 1 and the complete list of variables and mnemonics is available in Appendix A. Predictors are made stationary using year-over-year (YoY) percentage changes or absolute differences with respect to the same period of the previous year. Following Goulet Coulombe (2025), for each variable $X_{t,k}$ we include four lags together with moving averages of orders 2 and 4. This transformation, known as the Moving Average Rotation of X (MARX), alters the implicit prior of certain machine learning algorithms applied to time series data without requiring them to be recoded.

Table 1: Defining $\mathcal{H}'s$

H	**Content**
EST	Inflation expectations from surveys and market-based (Break-Even Inflation for 1 and 2 years), π_t components, t
ELT	Inflation expectations from market-based measures (BEI and FBEI to horizons greater than 3 years), t
g_t	Labor market, industrial production, national accounts and economic sectorial indicators, t
RER_t	Real exchange rate indicators, risk premia, terms of trade index, t
c_t	PPI, commodity prices, labor costs, freight rates, t

First, the short-run hemisphere includes one- and two-year-ahead inflation expectations, taken from Banco de la República's monthly analyst survey and from market-based measures

(break-even inflation). We also include several CPI sub-indices, since consumers may pay attention to individual prices to a degree that differs from their weight in the headline consumption basket. This hemisphere is further motivated by the persistence of Colombian inflation: items with a large weight in the basket, such as rents, are indexed to past inflation.

Second, we add a long-term inflation expectations hemisphere, which includes BEI for longer horizons (from three to eight years) and Forward Break-Even Inflation (five years to three years). We do not include survey measures owing to data availability (the central bank has collected five-year-ahead expectations only since 2019), and we exclude longer horizons, since they are unlikely to be relevant for short-run inflation dynamics.

The economic activity hemisphere includes indicators of aggregate demand, labor market dynamics, and sectoral heterogeneity. We consider the leading and contemporaneous indicators that central bank staff monitor in the nowcasting process (Eduardo and Gonzalo, 2018; Galeano-Ramírez et al., 2021; Cote-Barón et al., 2023).

The next hemisphere gathers various indicators of inflationary pressures originating in the exchange rate. We include direct measures in nominal and real terms, as well as the terms of trade, which capture foreign currency inflows and outflows from exports — mostly commodities. Over the sample period, the exchange rate is highly correlated with oil price fluctuations, as discussed in the following section. We also add risk premium measures, which co-move with the fiscal outlook and with risk aversion in international financial markets.

Finally, the last hemisphere gathers indicators aimed at capturing inflationary pressures not covered by the other hemispheres — those that the New Keynesian Phillips Curve absorbs in the error term and labels cost-push. In particular, we consider international prices of agricultural goods and inputs (fertilizers), freight rates (domestic and international), nominal wages, and producer price indices (PPI).

5. Empirical Results

5.1. Baseline Results

The estimation results yield the latent factors associated with predicted deviations of inflation from its target. The corresponding time-varying coefficient dynamics are presented in Figures B.2 and B.3 in Appendix B.

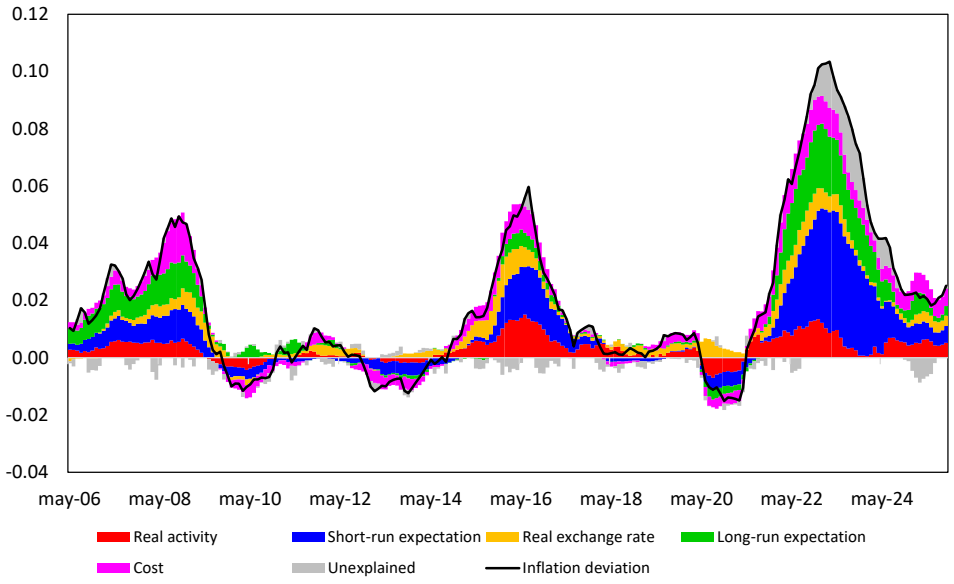

Figure 2: Determinants of inflation deviations from the target estimated through an HNN model

Additionally, the model allows for a historical decomposition of inflation deviations from the target based on the contribution of each hemisphere. Figure 2 presents the contribution of each latent group, estimated using the HNN, to the deviation of inflation from its target. The results point to three episodes characterized by sizable and sustained inflationary deviations above the target: 2006-–2008, 2014—2017, and 2021—2024. Outside these episodes, inflation remains broadly aligned with the target, particularly during 2009—2014 and 2018—2020, when deviations are modest and short-lived, indicating a relatively stable inflation environment.

5.2. Determinants of Inflation Between 2006 and 2008

Between 2006 and 2008, Colombia experienced a pronounced acceleration in inflation, which intensified in 2007. The literature identifies several stylized facts that explain inflation dynamics during this period. First, adverse supply shocks played a central role. In 2007, a drought associated with the El Niño phenomenon, followed by episodes of excessive rainfall, significantly affected agricultural production and increased the prices of unprocessed food (Abril-Salcedo et al., 2016, 2020; Bejarano-Salcedo et al., 2020; Romero and Naranjo-Saldarriaga, 2024). Second, global commodity prices rose sharply, particularly for oil and food commodities used in biofuel production. This increase was driven by strong global de-

mand from China and India, as well as rising geopolitical tensions. Higher commodity prices raised production costs and translated into increases in domestic regulated prices, including gasoline and public transportation (Banco de la República, 2006, 2007, 2008; Vargas-Herrera et al., 2009).

These supply pressures materialized in a context of strong aggregate demand. Favorable terms of trade, accommodative credit conditions, and robust external demand, particularly due to increased exports to Venezuela, reinforced inflationary pressures in the economy. From late 2007 onward, the interaction between persistent demand pressures and intensifying cost pressures, amplified by domestic supply disruptions, was critical in driving the rise in inflation. In this context, the literature has identified an increase in inflation expectations, which affected wage and price formation throughout the inflationary period (Banco de la República, 2006, 2007, 2008; Vargas-Herrera et al., 2009).

Our decomposition of inflation above the target is consistent with the literature. The model allocates a growing share of the inflation gap to cost pressures, driven by indicators of the agricultural disruptions caused by the El Niño phenomenon and by higher global commodity prices; within this block, the weight falls primarily on the food and regulated-price segments. The activity hemisphere contributes positively over the same window, and both blocks precede a rise in the contributions of short-run and long-run inflation expectations. This sequencing accords with the backward-looking expectation formation documented by Vargas (2016), under which past inflation outcomes strongly influence current expectations. The real exchange rate hemisphere also gains weight during the episode. Banco de la República tightened monetary policy over this phase, and the contributions of real activity and inflation expectations decline thereafter, alongside the convergence of inflation toward the long-run target by 2009.

5.3. Determinants of Inflation Between 2014 and 2017

Beginning in 2014, Colombia experienced a pronounced external shock associated with the collapse in international oil prices, which triggered a sharp depreciation of both the nominal and real exchange rates and, in turn, led to a deterioration in the terms of trade. This exchange rate adjustment intensified pass-through to domestic prices, exerting upward pressure on inflation, particularly in the imported goods and services components. These external forces were further compounded by domestic cost pressures, including a temporary increase in the value-added tax and a strengthening of domestic demand, which together amplified the upward impulse to headline inflation (Banco de la República, 2014, 2015, 2016).

Concurrently, the El Niño episode around 2015—2016 generated a severe agricultural supply shock that drove a large spike in food prices and further pushed inflation above target (Abril-Salcedo et al., 2016, 2020; Bejarano-Salcedo et al., 2020; Romero and Naranjo-Saldarriaga, 2024). These transitory shocks also affected inflation expectations at the onset, reinforcing price pressures. As weather conditions normalized and agricultural production recovered toward the end of 2017, the temporary supply pressures eased and contributed to the deceleration of food price inflation, facilitating a gradual return of overall inflation toward the target (Banco de la República, 2014, 2015, 2016).

The decomposition in Figure 2 aligns with the documented historical narrative. The HNN assigns a positive contribution to the real exchange rate hemisphere over the depreciation episode, together with a rising weight on the cost hemisphere, which overlaps chronologically with the El Niño shock; both blocks precede an increase in the estimated contribution of short-run inflation expectations. Real activity also enters with a positive weight, which moderates after 2017 as the central bank adopts a contractionary stance. Over the same window, the contributions of inflation expectations decline and inflation converges back toward the target.

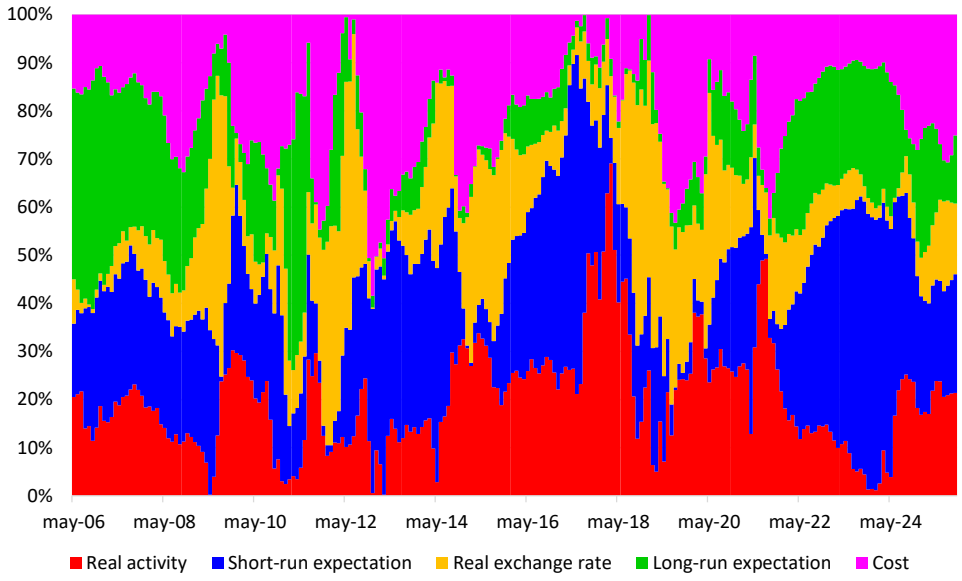

Figure 3: Absolute contribution of the determinants of inflation deviations from the target estimated through an HNN model

5.4. Determinants of Inflation Between 2021 and 2024

The convergence of inflation toward the target was abruptly interrupted by the COVID-19 pandemic in 2020. The sharp contraction in economic activity caused inflation to fall below the central bank's target range. In response, the central bank reduced its policy interest rate as part of an expansionary monetary policy aimed at mitigating downward price pressures stemming from the collapse in real activity and the persistence of significant excess productive capacity, conditions inherent to the pandemic shock (Banco de la República, 2020a,b).

In the aftermath of the pandemic, the economy experienced a series of overlapping and multifaceted disturbances. A robust rebound in economic activity ensued, driven by accumulated household savings during lockdowns and reinforced by expansionary fiscal and monetary measures. This resurgence was reflected in a sharp recovery of domestic consumption, which unfolded in an environment of accommodative monetary policy. Moreover, favorable international financial conditions, improved external demand prospects, and rising terms of trade further bolstered the recovery (Banco de la República, 2021a,b).

During the initial phase following the pandemic, inflationary pressures resurfaced, primarily driven by increases in the prices of food and regulated goods. These were largely attributable to rising international commodity prices and higher production costs stemming from elevated international freight rates and supply-side constraints, particularly in global shipping and the availability of raw materials. In addition, idiosyncratic shocks, such as roadblocks associated with the *Paro Nacional*, exacerbated pressures on food prices (Banco de la República, 2021b; Gáfaro-González et al., 2022). Adverse weather conditions and logistical disruptions arising from social protests in 2021 further impaired domestic food production, intensifying inflationary pressures.

External factors also contributed significantly. Bottlenecks in global value chains led to higher costs for intermediate inputs, thereby feeding into domestic inflation. These effects were intensified by the escalation of the Russia–Ukraine conflict, which drove up the prices of agricultural inputs, notably fertilizers, and consequently affected primary food prices. These external shocks related to rising costs unfolded in a context of a positive output gap, which contributed to demand-driven inflation, alongside a notable depreciation of the real exchange rate (Banco de la República, 2022).

By 2023, inflationary dynamics became increasingly shaped by inertia. Inflation exhibited high persistence, with backward-looking indexation mechanisms playing a key role in sus-

taining price increases in sectors such as rents and transport. Additional pressures emerged from substantial increases in the minimum wage and from adjustments in fuel prices aimed at reducing the deficit in the Fuel Price Stabilization Fund (FEPC). Following this period, inflation gradually began to decline as a result of the central bank's contractionary monetary policy, a trend that continued throughout 2024 (Banco de la República, 2023, 2024).

The HNN estimation indicates that short-run expectations and inflation inertia enter the decomposition with negative contributions, in line with the disinflation observed over that window. This occurs alongside a further depreciation of the Colombian peso that accompanies a renewed decline in international oil prices. In the recovery phase, the cost hemisphere gains weight, echoing the mechanisms documented in the literature: the increase is driven by domestic and external indicators of production costs, and it falls primarily on the food segment. Short-run expectations and inertia then turn positive and come to account for a substantial share of the estimated inflation gap. Long-run expectations move in the same direction, as does the real exchange rate hemisphere, mirroring the sustained depreciation trend in the data.

Subsequently, in 2022, the decomposition shifts toward a higher relative importance of cost-related variables. The onset of the conflict in Ukraine coincided with a global surge in commodity prices, particularly for energy, agricultural products, and fertilizers. This external pattern occurred alongside persistent global supply chain bottlenecks. At the same time, the weight of the real exchange rate intensified as the Colombian peso depreciated, alongside tighter external financial conditions and monetary tightening in advanced economies, which the model links to imported inflation. Within the HNN framework, the interaction among demand indicators, domestic and external cost pressures, and exchange rate depreciation accounts for the persistent estimated inflation overshoot, which peaks in 2022.

The joint variation of these cost and demand indicators was associated with a significant rise in both short- and long-term expectations during this period, which is reflected in their increased weight in the decomposition. This pattern highlights the predictive importance of expectations in accounting for deviations of predicted inflation from its target within the neural network. In response, the central bank adopted a decisive contractionary monetary policy. The objective of this tightening cycle was to moderate real activity growth toward a sustainable level and to support the re-anchoring of inflation expectations. Following this policy shift, the decomposition then records a decline in the contributions of real activity and inflation expectations, displaying a pattern consistent with the policy's objectives.

5.5. Extension: labor and non-labor costs

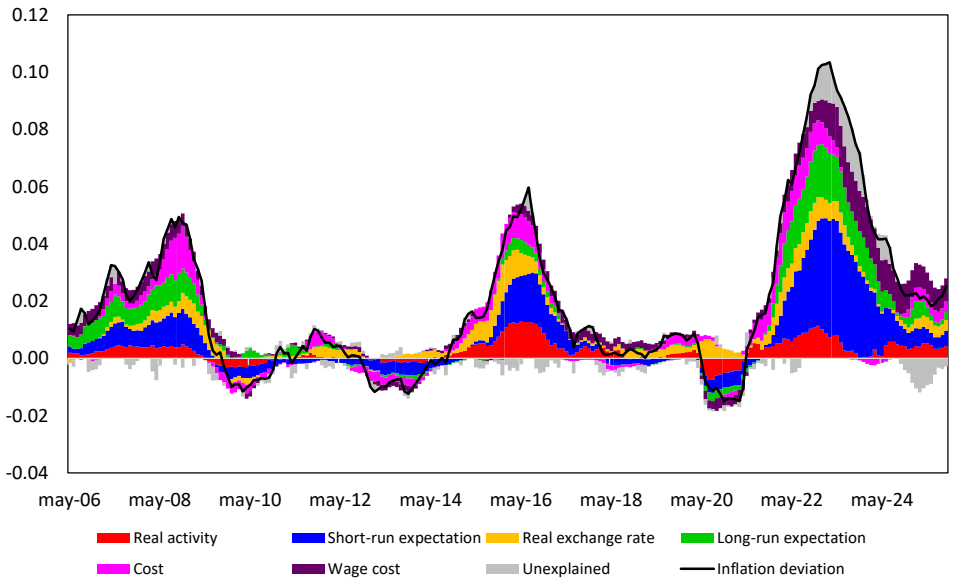

Figure 4: Determinants of inflation deviations from the target estimated through an HNN model with a specific hemisphere of wage costs

This section illustrates how the model can be read under an alternative hemisphere structure. Here, the cost hemisphere is split into two: a labor cost hemisphere and a non-labor cost hemisphere. All other features of the specification remain unchanged. This extension is particularly relevant in light of the changes experienced by Colombia in recent years, regarding the magnitude of adjustments to the statutory minimum wage (SMW). These adjustments have also been transmitted to some extent to the broader wage structure of the economy. From a modeling perspective, this extension provides a robustness check for the results obtained from the baseline model (more comprehensive robustness exercises are presented in Section 6).

Colombian legislation, together with strands of New Keynesian economic theory, identifies headline inflation and productivity gains[3] as key determinants of wage adjustments, particularly those of the statutory minimum wage, with the objective of preserving macroeconomic stability. On the one hand, incorporating these factors seeks to prevent the erosion of workers' purchasing power while ensuring that productivity improvements are shared with employees. On the other hand, it aims to avoid generating upward inflationary pressures

[3]Law 278 of 1996.

and adverse effects on production and employment. In the short run, the statutory minimum wage is interpreted as a production cost that affects firms throughout the supply chain, whereas in the medium run its effects may propagate through price indexation mechanisms if the initial inflationary pressures become persistent. The literature also discusses potential offsetting effects of minimum wage increases on prices, output, and employment through stronger aggregate demand.

Figure 5 illustrates the annual growth rate of the nominal statutory minimum wage, including the transportation allowance, relative to the sum of headline inflation and productivity gains recorded in the previous year. The figure reveals a marked increase in these excess wage adjustments following the COVID-19 pandemic. On average, the series approximately doubled during the 2021–2025 period compared with 2006–2020.

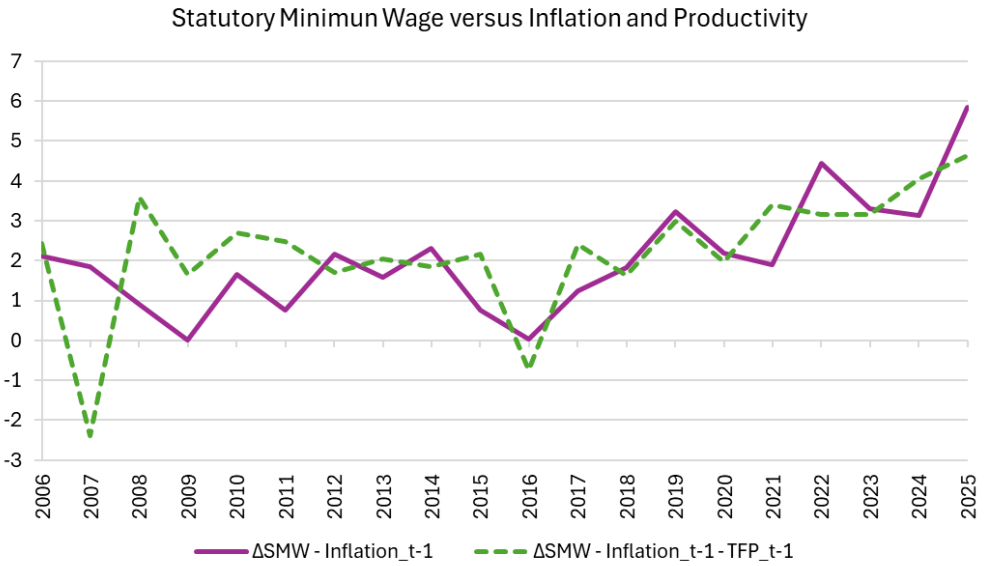

Figure 5: Increase in the nominal Statutory Minimum Wage each year compared to total inflation and Total Factor Productivity of the previous year.

These wage dynamics are reflected in the model's decomposition. The estimated time-varying coefficient (Appendix C) indicates a positive relationship between labor costs and inflation, suggesting the presence of upward inflationary pressures following wage increases. When computing the contribution of this hemisphere, the model identifies these pressures as particularly relevant from 2022 onward, consistent with the shift in the magnitude of statutory minimum wage adjustments and their spillover to the rest of the wage distribution in the economy, as described above. Similarly, the cost hemisphere that excludes wage cost variables, which primarily captures the prices of intermediate inputs, is also associated with

a positive estimated time-varying coefficient and reflects the evolution of international prices.

6. Comparison and Robustness

In this section, we use the latent factors from the HNN model, which summarize groups of variables associated with different inflation determinants, to examine the estimated coefficients and latent components that jointly account for the contributions described in the previous section. These latent factors can be interpreted as implicit representations of inflationary pressures arising from relevant economic domains, such as real activity, the real exchange rate, and inflation expectations. We then examine these implicit factors and compare them with measures obtained using alternative methodologies.

It is important to clarify that the implicit factors recovered from the HNN are not direct estimates of an output gap, of real exchange rate misalignment, or of expectation de-anchoring. They are supervised summaries of each information block, extracted so as to predict the inflation gap, and their scale and sign are only identified up to the normalizations described in Section 3. The comparison therefore asks whether these summaries co-move with the economic dynamics identified by dedicated benchmarks, and should not be read as a validation of the latent factors as structural objects.[4]

As benchmarks, we use Banco de la República's official estimates of the output gap and the real exchange rate gap, produced by its quarterly monetary policy models: PATACON (González et al., 2011) and 4GM (Gonzalez et al., 2020; Ramos-Veloza et al., 2026). In both models, these latent variables are estimated within a dynamic stochastic general equilibrium framework. In addition to official estimates, the economic literature has a large set of models that use other specifications or methodologies.

Banco de la República staff estimate potential output and the output gap using a production function approach, univariate and multivariate filters, structural models under alternative expectation formation schemes (rational and adaptive), and a Bayesian Structural Vector Autoregression (BSVAR). These methodologies are then evaluated against predictive criteria and combined through a clustering approach that groups models according to their real-time performance (Pulido-Mahecha et al., 2024).

We compare the implicit real activity latent factor with the output gap (lagged two months)

[4]For comparison purposes, both the implicit latent factors and the official measures are standardized.

estimated by the Central Bank of Colombia obtained from the 4GM model (official) and the outcome of combining different methodologies. Furthermore, following the methodology of Goulet Coulombe (2025), we conduct a counterfactual exercise where the real activity hemisphere includes only this official output gap.

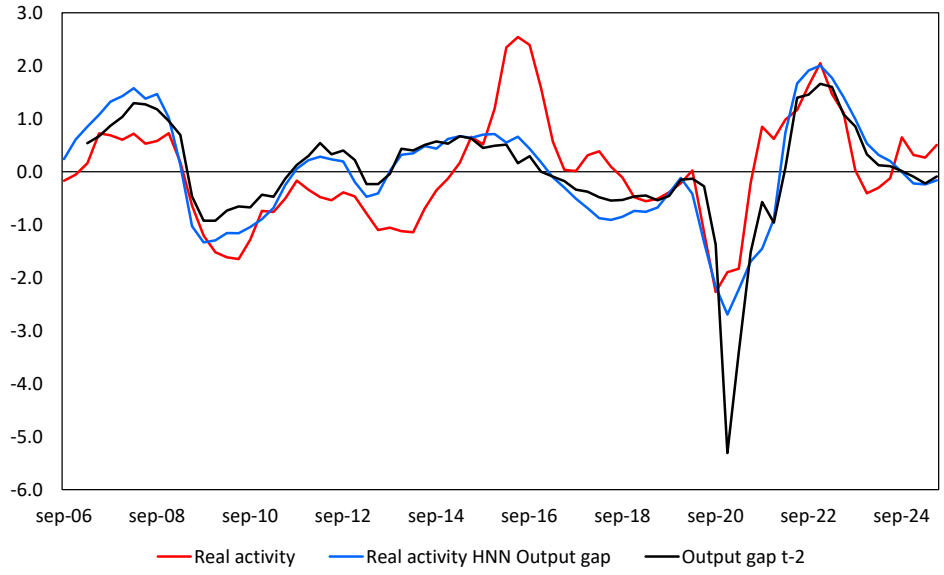

Figure 6: Comparison of the real activity latent factor obtained through an HNN model and a specific output gap measure

The results, illustrated in Figure 6, show that the implicit factor derived from the full set of activity variables exhibits a significant increase during the 2006–2008 inflationary period. This pattern is consistent with the estimates obtained when using only the output gap in the hemisphere, and it aligns closely with the official output gap series itself. This co-movement indicates that the HNN's activity factor tracks patterns consistent with demand-side pressures during this boom.

The implicit measure of real activity factor shows a further increase in the period from 2014 to 2017, albeit somewhat more moderate than the earlier episode. Both the comprehensive HNN factor and the output gap-based measure indicate that around 2016, economic growth could have contributed to explaining inflation through heightened demand pressures. Subsequently, the measures display a highly synchronized trajectory, capturing the severe economic contraction during the pandemic and the subsequent robust recovery linked to post-pandemic inflationary pressures. Finally, a modest positive contribution from economic

activity is observed in the most recent period for the implicit factor that aggregates all real activity variables.

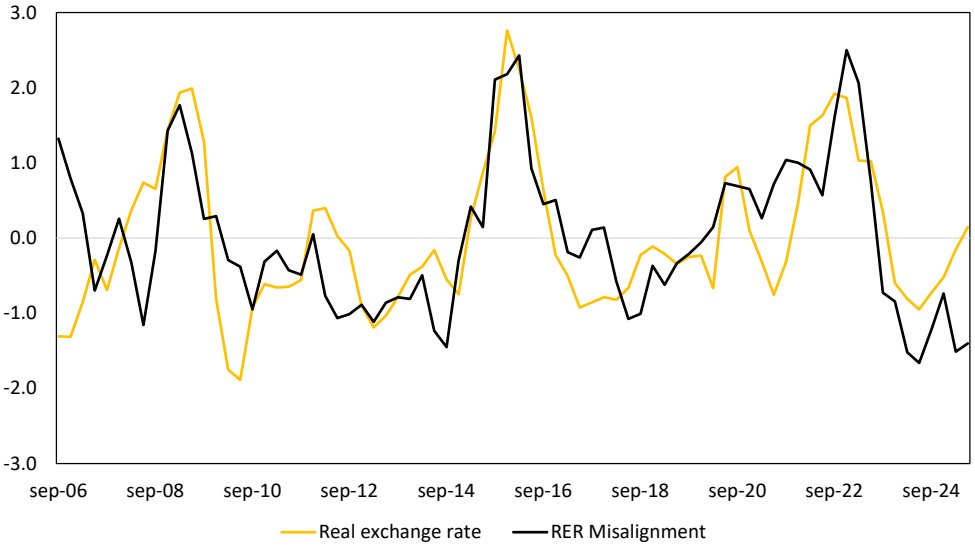

Figure 7: Comparison between the real exchange rate latent factor obtained through an HNN model and a specific estimate of real exchange rate misalignment

The estimated real exchange rate (RER) latent variable closely mirrors the measure of RER misalignment obtained with the Thousand BEERs methodology (Salazar-Díaz et al., 2023). As shown in Figure 7, the model's implicit factor for RER-induced inflationary pressure lines up closely with periods in which the observed RER stands above its equilibrium level (a positive misalignment, indicating depreciation). This relationship is particularly evident during the three main episodes of above-target inflation: 2006–2008, 2014–2017, and the post-pandemic period from 2021 to 2024. In each of these instances, the increase in the positive RER misalignment is matched by a concurrent rise in the HNN's implicit RER contribution, indicating that the model's exchange-rate factor co-moves closely with the misalignment measure, although both series are constructed from overlapping information and the comparison is therefore not an independent test.

Finally, we compare the inflation expectations latent factors estimated by the model, which summarize the behavior of short- and long-run expectations, with measures of inflation expectations de-anchoring based on the Gumbel and Clayton specifications reported in Muñoz-Martínez (2026). These indicators quantify the transmission of movements in short-run ex-

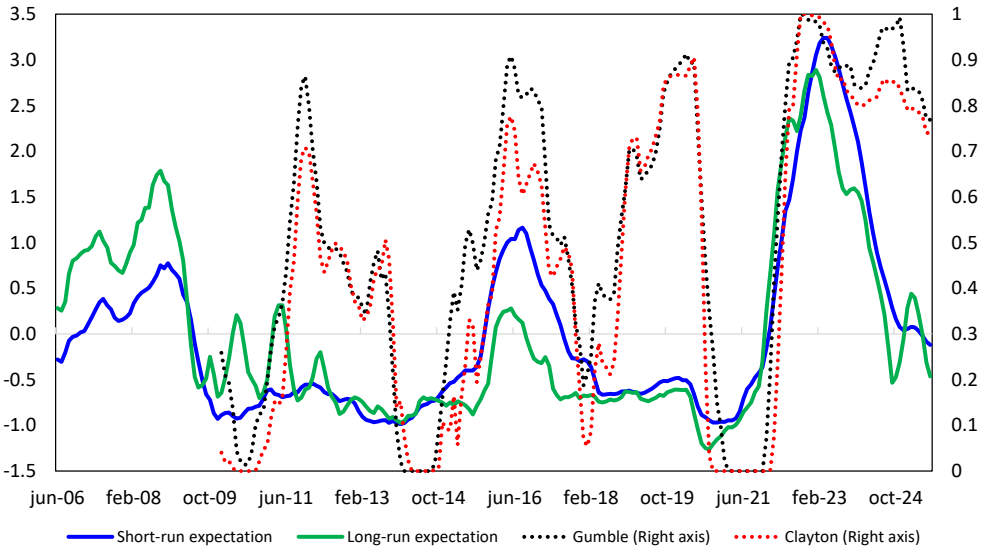

Figure 8: Comparison of the short- and long-run inflation expectations latent factor obtained through an HNN model with a measure of inflation expectations de-anchoring

pectations to long-run expectations and range between 0 and 1, where 0 corresponds to fully anchored expectations and 1 to the opposite case. Strictly speaking, expectations are unanchored when long-run expectations differ from the inflation target, that is, when agents believe that inflation will remain permanently away from it and adjust their wage and price decisions accordingly. The indicators used here, however, do not measure the level of long-run expectations, but rather the extent to which short-run dynamics contaminate long-run beliefs. It is important to emphasize that our estimated latent expectation variables do not constitute direct measures of de-anchoring. Instead, they capture the influence of expectations on deviations of inflation from its target through a nonlinear neural network framework.

The comparison is illustrated in Figure 8. We observe periods in which the latent factors that summarize short- and long-run expectations move closely together. In particular, there is a noticeable increase in the expectation-related latent factor during the periods 2015—2016 and 2022—2023, which is consistent with episodes identified as de-anchoring in the benchmark measures. This pattern suggests that episodes of de-anchoring in inflation expectations exhibit a close statistical association with deviations of inflation from its target. The model reproduces this co-movement, which we read as a consistency check on the ex-

pectations hemispheres rather than as independent evidence on the direction of causality between expectations and inflation.

7. Pseudo-Out-of-Sample Forecast Evaluation

We perform a pseudo-out-of-sample forecasting exercise to evaluate whether the HNN generalizes the relationships learned during estimation rather than memorizing the training sample. Although forecasting is not the primary objective of the proposed framework, predictive performance on previously unseen observations provides a natural empirical test of the regularization strategy described above and, consequently, of the model's ability to avoid overfitting.

The primary objective of the HNN is to recover economically interpretable latent components that explain the historical deviations of inflation from its target through the contemporaneous relationship defined in Equation (7), rather than to maximize forecasting accuracy. Consequently, the pseudo-out-of-sample exercise should not be interpreted as an attempt to establish the HNN as a forecasting model. Instead, it serves as an external validation of the estimated nonlinear structure that underlies the historical decomposition.

The same mapping between the explanatory variables and inflation deviations is used both to recover the historical determinants of inflation and to generate the pseudo-out-of-sample forecasts. Therefore, if the estimated relationships reflect genuine economic regularities rather than sample-specific noise, they should remain informative when applied to observations that were not used during estimation.

All models are evaluated under identical conditions: the same rolling windows, forecast origins, horizons, and target variable. The exercise is "pseudo" out-of-sample because the evaluation period is observed historically and, for models with exogenous predictors, forecasts are conditional on the realized values of the explanatory variables at the target date, that is, $\hat{y}_{t+h} = \hat{f}(x_{t+h})$, while the model parameters remain fixed at the values estimated using the information available up to time t. Consequently, the exercise evaluates the stability and predictive content of the estimated structural relationships rather than the ability to produce unconditional real-time forecasts.

Competitive forecasting performance in this setting therefore provides empirical evidence that the HNN captures persistent and economically meaningful relationships instead of memorizing transitory fluctuations in the estimation sample. Since the historical decomposition

presented in the previous section is generated from exactly the same estimated structure, satisfactory pseudo-out-of-sample performance increases confidence that the identified determinants of inflation deviations reflect robust economic mechanisms rather than artifacts of overfitting.

Based on the above, we perform the pseudo-out-of-sample exercise using a rolling window scheme. A rolling window evaluation consists of estimating the model on a sample of fixed length, producing forecasts from the end of that sample, and then shifting the entire estimation window forward to repeat the procedure, so that the model is re-estimated many times and evaluated at many different points of the recent history rather than at a single arbitrary date. In our implementation, at each forecast origin every model is estimated using only the most recent 180 monthly observations, and the resulting structure, held fixed, produces forecasts 1, 3, 6, and 12 months ahead. Given the computational cost of re-estimating the full ensemble at every forecast origin, the HNN is estimated with 50 replications in this exercise, following the forecasting protocol of Goulet Coulombe (2025), in which a reduced ensemble is sufficient to obtain stable point forecasts.

The window then advances three months and the procedure is repeated, yielding 15 forecast origins between May 2021 and October 2025, a period that covers both the inflationary surge and the subsequent disinflation. The HNN is compared against four benchmarks: a Naive forecast that carries the last observed value forward, an ARIMA whose specification is re-selected by information criteria within each window, and two flexible learners, a Random Forest (RF) and a fully connected deep neural network (DNN), which receive exactly the same predictors as the HNN but without the hemisphere structure.

Figure 9 summarizes the results by reporting, for each horizon, the RMSE of every model relative to that of the ARIMA, so that values below one indicate that a model outperforms the univariate benchmark. At the one-month horizon (upper-left panel), the HNN performs essentially on par with the ARIMA, with a relative RMSE of 1.06, while all remaining models are dominated by the univariate alternatives.

This is the expected pattern: at very short horizons, inflation deviations are largely explained by their own persistence, so that univariate benchmarks are difficult to outperform, in line with a well-documented regularity of the inflation forecasting literature (Atkeson et al., 2001; Stock and Watson, 2008). Even the Naive forecast loses to the ARIMA at this horizon (1.41), reflecting the rapid month-to-month movements of the evaluation period, while the RF (1.91) illustrates the cost of feeding a high-dimensional, unstructured predictor set into a flexible

learner when the signal-to-persistence ratio is low; the DNN (9.46) takes that cost to the extreme.

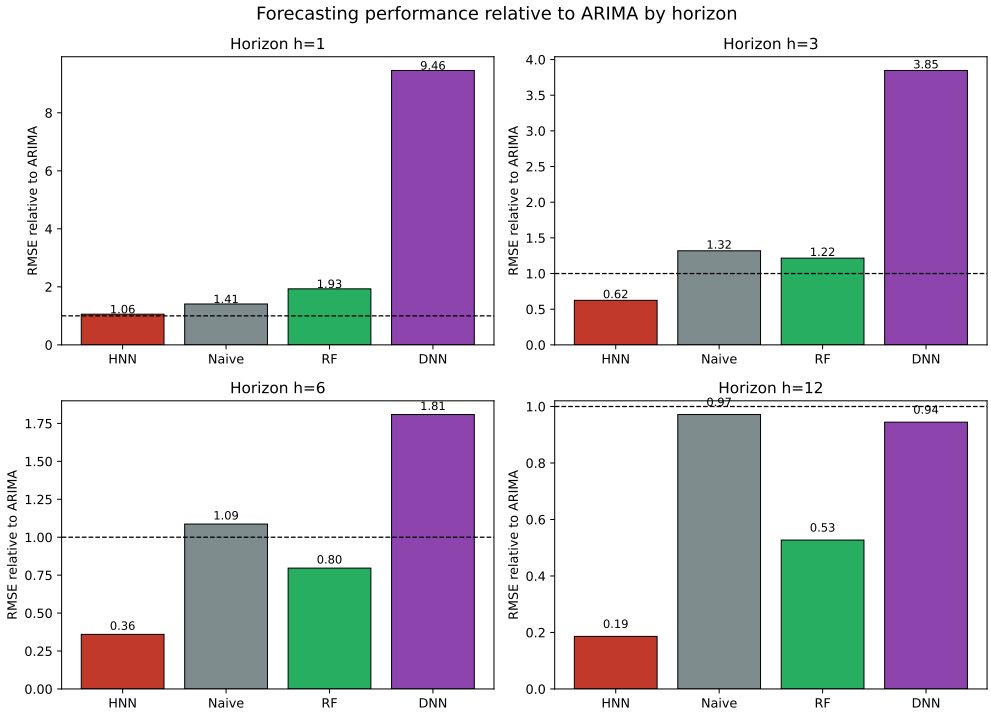

Figure 9: Each panel reports the root mean squared forecast error (RMSE) at horizon h, relative to an ARIMA benchmark selected by AIC, i.e., $\text{RMSE}_i/\text{RMSE}_{\text{ARIMA}}$ for each methodology i. Bars below the dashed line (= 1) indicate that the model outperforms ARIMA. Forecasts are generated from a rolling window of 180 monthly observations, with all models re-estimated every 3 months (15 forecast origins spanning May 2021–October 2025). HNN is the Hemisphere Neural Network; RF is a Random Forest and DNN a fully connected deep neural network, both using the same predictors as HNN without the hemisphere structure; Naive is the last observed value carried forward (random walk)

The ranking changes qualitatively as the horizon lengthens. At three months (upper-right panel), the HNN is the only model that outperforms the ARIMA, with a relative RMSE of 0.62, a first indication that the conditional information embedded in the hemispheres begins to dominate pure persistence once the forecast target moves beyond the very short run. At six months (lower-left panel), the relative RMSE of the HNN falls to 0.36, roughly one third of the univariate error, and the RF also crosses below parity (0.80), suggesting that the predictor set as a whole carries genuine medium-horizon information; the Naive forecast, by contrast, converges toward the ARIMA (1.09), as both effectively extrapolate the last observed conditions.

The twelve-month panel (lower-right) delivers the sharpest results. First, the HNN attains

a relative RMSE of 0.19: its forecast error is roughly one fifth of the univariate benchmark precisely at the horizon most relevant for monetary policy. Second, the Naive and ARIMA forecasts become nearly indistinguishable (0.96), consistent with the automatic order selection favoring highly persistent specifications whose long-horizon forecasts collapse toward the last observed value, so that both univariate benchmarks extrapolate the recent past and miss the turning points of the surge and the subsequent disinflation. Third, every model that exploits the predictor set outperforms the univariate benchmarks at this horizon (the RF at 0.53 and the DNN at 0.94); what separates the HNN (0.19) from the DNN is therefore not the information set, which is identical, but the hemisphere architecture that organizes it.

Three features of Figure 9 speak directly to the concern that the flexibility of the network relative to the sample size may translate into overfitting. First, the relative performance of the HNN improves monotonically with the horizon (1.06, 0.62, 0.36, 0.19). This is the opposite of the signature of overfitting: memorized noise carries no information about future observations, and its uselessness becomes more binding as the horizon lengthens, so a model that owed its in-sample fit to memorization would lose ground at long horizons rather than gain it.

The increasing advantage of the HNN indicates instead that the estimated structure captures systematic determinants of inflation deviations whose predictive content materializes precisely where pure persistence fades. Second, the restricted HNN outperforms the unrestricted DNN at every horizon despite sharing its information set and capacity, the empirical test anticipated in Section 3: the hemisphere restrictions act as effective regularization, curbing variance far more than they add bias.

Third, this performance is sustained across the fifteen recursive re-estimations of the model, spanning the entire recent inflation cycle, including both the inflation surge and the subsequent disinflation, periods in which extrapolation is particularly challenging. The consistency of the results across these episodes suggests that the estimated nonlinear structure captures persistent economic relationships rather than sample-specific noise. Since this same structure underlies the historical decomposition of inflation deviations, its ability to generalize outside the estimation sample strengthens confidence in the economic interpretation of the identified determinants.

8. Conclusion

This paper has proposed an HNN framework to decompose the deviation of inflation from its target in Colombia, adapting the architecture introduced by Goulet Coulombe (2025) to the institutional and structural characteristics of a small, commodity-dependent open economy. The approach combines the interpretability of structural macroeconomic modeling with the flexibility of deep learning, organizing predictors into thematic blocks – short-run expectations and inertia, long-run expectations, real activity, real exchange rate, and domestic and international cost pressures – and allowing the model to extract latent predictive components without imposing parametric laws of motion.

Several findings emerge from the analysis. First, the results suggest that real activity is not the only source of predictive information about the inflation gap in Colombia. In particular, we observe relevant contributions from inflation expectations and cost-related factors. This pattern is consistent with earlier disaggregated evidence for Colombia (Lanau et al., 2018) and with broader international evidence documenting a weakening relationship between measures of real activity and inflation. Second, the historical decomposition reveals that predicted inflation deviations are driven largely by inflation inertia, cost pressure indicators, and inflation expectations, both in the short run and in the long run.

Third, the framework provides a historically coherent narrative across three major inflationary episodes. In the estimated decomposition, the 2006–2008 acceleration reflects a combination of positive activity contributions, cost-push indicators, and rising global commodity prices. The 2014–2017 episode is characterized by an elevated contribution from the real exchange rate hemisphere following the decline in oil prices, alongside an increased contribution from the cost hemisphere during the chronological El Niño agricultural disruption. The post-pandemic surge of 2021–2024 is depicted as a sequence of overlapping predictive contributions—a rebound in the activity hemisphere's contribution, domestic and external cost-push indicators, the statistical co-movement with the Russia–Ukraine conflict, and, by 2023, a transition toward high predicted inflation persistence that historically coincided with indexation patterns in sectors such as rents and transport. Following the central bank's contractionary policy adjustment in 2022–2023, the decomposition records a subsequent decline in the contributions of real activity and inflation expectations, showing a statistical pattern consistent with a gradual convergence of the predicted inflation path toward its target.

Fourth, the external validity exercises suggest that the estimated latent states may generalize beyond the information used to estimate the model. The implicit real activity factor closely

tracks the output gap estimates produced by the Central Bank of Colombia's 4GM model, while the real exchange rate factor aligns with measures of real exchange rate misalignment derived from the Thousand BEERs methodology. The expectation-related latent variables are also consistent with copula-based measures of expectation de-anchoring, particularly during the 2015–2016 and 2022–2023 episodes. These high correlations suggest that the HNN latent decomposition, although derived from a predictive and associative architecture, captures co-movements that are closely aligned with established macroeconomic benchmarks rather than representing purely statistical artifacts. However, these results do not imply that the latent factors measure the same structural objects as the benchmark measures. Moreover, the model performs well in the pseudo-out-of-sample evaluation, and its relative performance across alternative benchmarks is consistent with the contributions of the different groups of inflation related variables. The stability of these results across observations not used for estimation provides additional evidence that the estimated decomposition is not merely driven by overfitting to the estimation sample. Taken together, these findings suggest that the latent factors capture economically relevant information about the inflation gap and can provide a useful complementary perspective to structural approaches.

From a methodological standpoint, this paper contributes to the growing literature on interpretable machine learning methods for macroeconomic analysis by extending their application to a developing economy. The HNN framework occupies a useful intermediate position between fully parametric structural models and black-box forecasting algorithms: it preserves the conceptual representation of the New Keynesian Phillips Curve while relaxing the linearity and stationarity restrictions that characterize conventional estimation approaches. For monetary policy institutions, this represents a practical advantage. The estimated decomposition can be visualized and discussed using the familiar language of contributions and gaps, while the HNN can simultaneously exploit the rich information contained in a high-dimensional panel and accommodate potentially nonlinear relationships among the fundamental factors associated with deviations of inflation from its target.

The results also offer useful insights for inflation-targeting regimes in small open economies with strong commodity linkages. The high weight of cost-push and indexation indicators in our decomposition is consistent with an inflation-convergence process in Colombia that is more protracted than conventional linear estimates imply, particularly when target deviations coincide with highly persistent inertial patterns. In this context, the early tightening cycle undertaken by Banco de la República in 2022 coincided, through the lens of the model, with a subsequent moderation in the predictive contributions of the activity and expectations

hemispheres before a stronger inertial component became embedded in the decomposition.

Two caveats delimit these conclusions. First, the hemispheres are not mutually exogenous: expectations, wages, real activity and the exchange rate are jointly determined with inflation and with one another, and the network does not correct for this simultaneity. The decomposition should therefore be read as an accounting of predictive content across correlated blocks of information, not as a set of structural elasticities, and in particular the limited weight assigned to real activity does not identify the slope of the Colombian Phillips Curve. Second, the latent states are identified only up to the normalizations imposed by the factorization, so their comparison with official gap estimates speaks to the internal coherence and timing of the decomposition rather than to measurement equivalence.

Several avenues for future research emerge from the present analysis. First, addressing the simultaneity issues discussed in Section 3.6, for instance by embedding an identified structural block or by instrumenting the expectations and wage hemispheres, would strengthen the interpretation of the decomposition and provide a firmer basis for causal inference. Second, incorporating disaggregated price indices, such as tradables, non-tradables, food, and regulated goods, as separate hemispheres, in the spirit of the sectoral Phillips Curve literature, would allow for a more granular decomposition of the predictive patterns underlying aggregate inflation.

Appendix

A. Variable Description

Table A-1: Description of Variables by Hemisphere

Hemisphere	Variable Name	Source
EST (Short-run Expectations)	Inflation Expectations (1-year ahead)	Banco de la República's monthly survey
	CPI for food	DANE
	CPI for goods (excluding food and regulated items)	DANE
	CPI for services (excluding food and regulated items)	DANE
	CPI for regulated items	DANE
	Break-Even Inflation 1 year	Banco de la República
	Break-Even Inflation 2 years	Banco de la República
ELT (Long-run Expectations)	Break-Even Inflation 3 years	Banco de la República
	Break-Even Inflation 5 years	Banco de la República
	Break-Even Inflation 6 years	Banco de la República
	Break-Even Inflation 7 years	Banco de la República
	Break-Even Inflation 8 years	Banco de la República
	Forward Break-Even Inflation 5y - 3y	Banco de la República
Real Activity	Monthly Economic Activity Indicator:	
	- Mining and Quarrying	DANE
	- Manufacturing Industries	DANE
	- Construction	DANE
	- Utilities	DANE
	- Wholesale and Retail Trade	DANE
	- Information and Communication	DANE
	- Real Estate Activities	DANE
	- Professional Services	DANE
	Industrial Production Index	DANE
	Total National Employment	DANE
	Labor Force Participation Rate	DANE
	Employment-to-Population Ratio	DANE
	Unemployment Rate	DANE

Continued on next page

Hemisphere	Variable Name	Source
	Real Bank Transfers (ACH)	ACH Colombia and Banco de la República
	Real Mortgage Loan Disbursements	DANE, calculations by Banco de la República
	Real Consumer Loan Disbursements (in pesos)	Financial Superintendency of Colombia
	Real Manufacturing Production	DANE
	Cattle and Swine Slaughter	DANE
	Real Imports of Durable Consumer Goods (in pesos)	DANE
	Real Imports of Capital Goods (in pesos)	DANE
	Regulated Energy Demand	XM
Real Exchange rate	Real exchange Rate Index - adjusted with Producer Prices (PPI)	Banco de la República
	Real exchange Rate Index - adjusted with Consumer Prices (CPI)	Banco de la República
	Terms of trade index - International trade	Banco de la República
	Terms of trade index - Adjusted by Producer prices (IPP)	Banco de la República
	Nominal exchange rate - monthly average	Banco de la República
	CDS five years	Bloomberg
	EMBI	Bloomberg
Commodity prices and costs	International Fertilizer prices	Bloomberg
	International Metals prices	Bloomberg
	International palm oil prices	Bloomberg
	International soybean prices	Bloomberg
	International soybean oil prices	Bloomberg
	International corn prices	Bloomberg
	International sorghum prices	Bloomberg
	International wheat prices	Bloomberg
	International cotton prices	Bloomberg
	International phosphate prices	Bloomberg
	International DAP prices	Bloomberg
	International TSP prices	Bloomberg
	International Urea prices	Bloomberg
	International potassium chloride prices	Bloomberg
	Implicit import prices	Banco de la República
	International freight rates (container)	Bloomberg
	Relation of CIF and FOB imports	DANE, calculations by Banco de la República
	Industry nominal wages	DANE
	Retail nominal wages	DANE

Continued on next page

Hemisphere	Variable Name	Source
	Construction nominal wages	DANE
	Housing nominal wages	DANE
	Producer Prices Index - Produced and consumed	Banco de la República
	Producer Prices Index - Imported	Banco de la República
	Producer Prices Index - Exported	Banco de la República

B. Main contributions, latent factors and coefficients

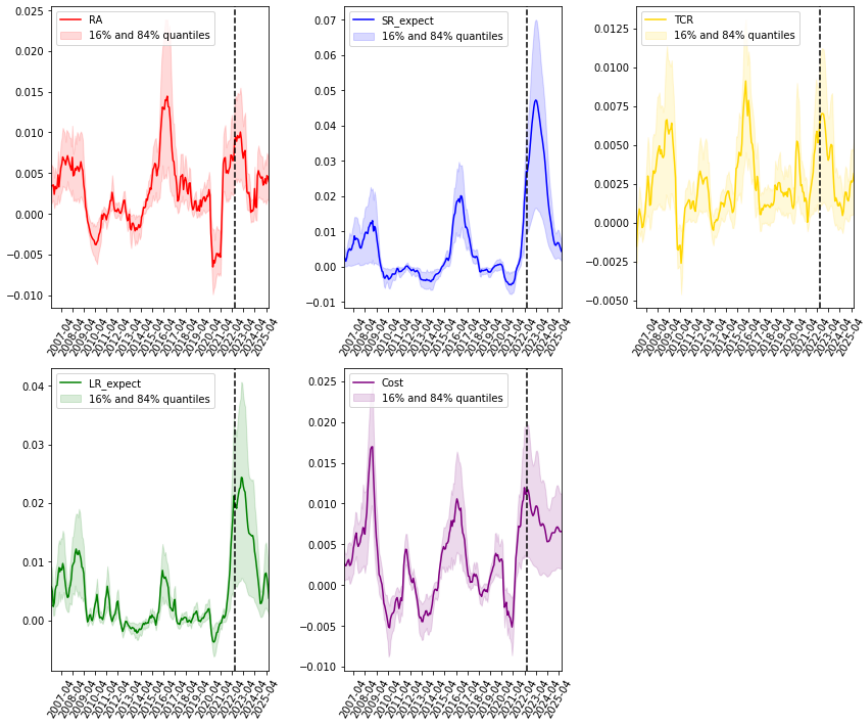

Figure B.1: Contribution plot

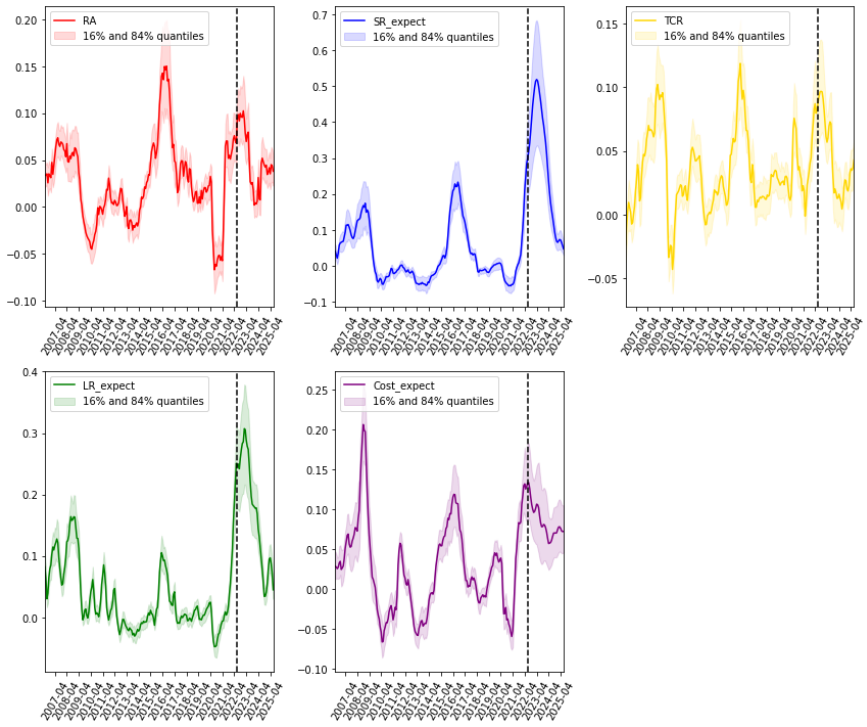

Figure B.2: Latent factors plot

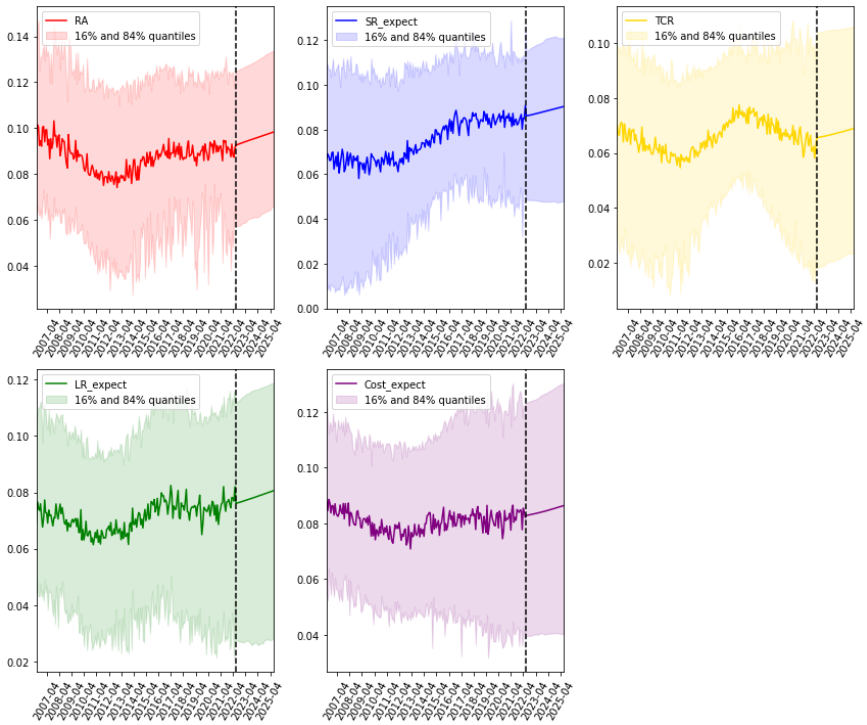

Figure B.3: Coefficients plot

C. Labor and Non-labor Costs extension: Main contributions, latent factors and coefficients

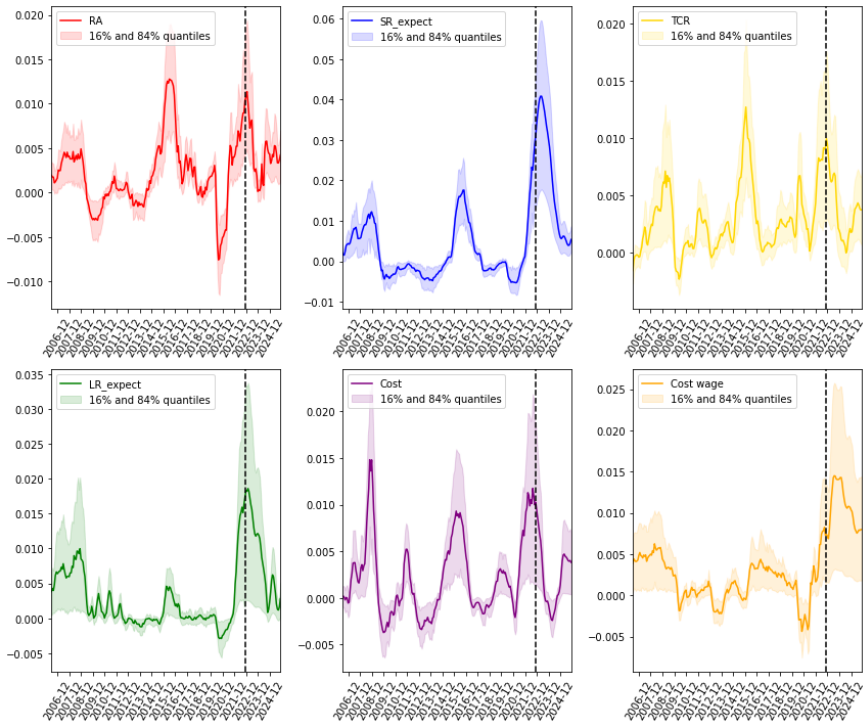

Figure C.1: Labor and Non-labor Costs extension: Contribution plot

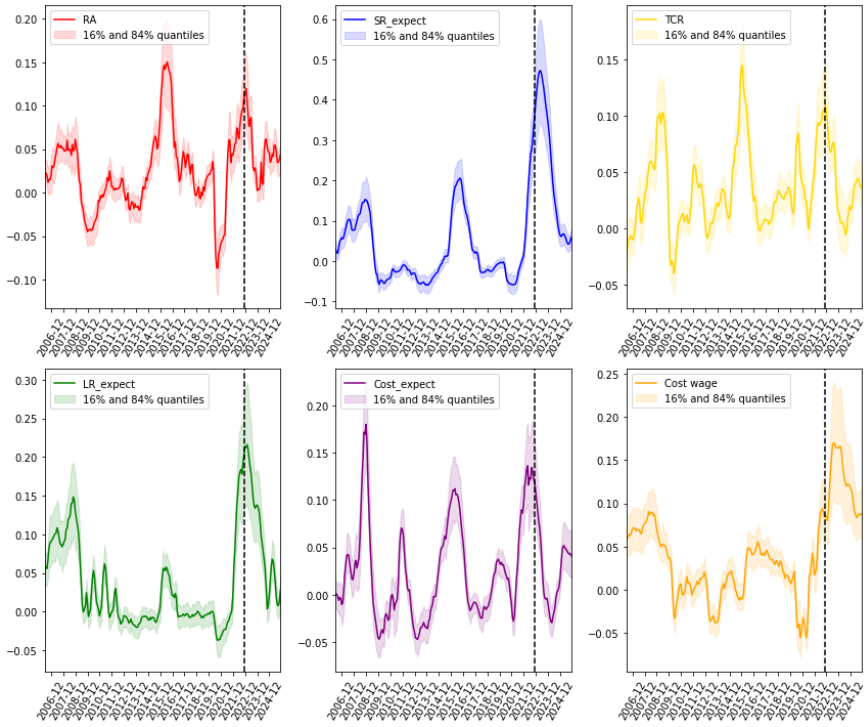

Figure C.2: Labor and Non-labor Costs extension: Latent factors plot

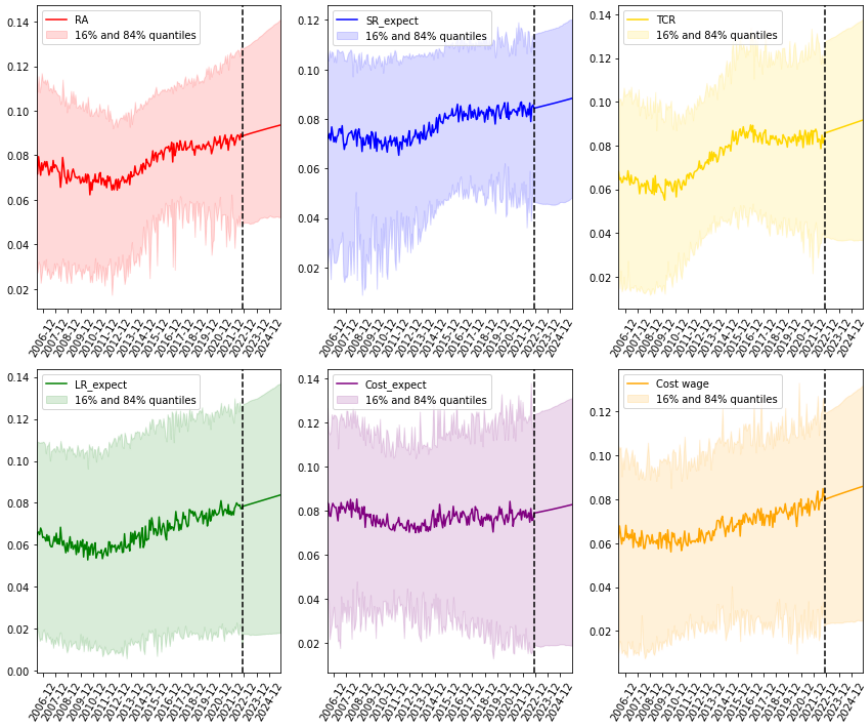

Figure C.3: Labor and Non-labor Costs extension: Coefficients plot

References

Abril-Salcedo, D. S., Melo-Velandia, L. F., and Parra-Amado, D. (2016). Impactos de los fenómenos climáticos sobre el precio de los alimentos en colombia. *Ensayos sobre política económica*, 34(80):146–158.

Abril-Salcedo, D. S., Melo-Velandia, L. F., and Parra-Amado, D. (2020). Nonlinear relationship between the weather phenomenon el niño and colombian food prices. *Australian Journal of Agricultural and Resource Economics*, 64(4):1059–1086.

Atkeson, A., Ohanian, L. E., et al. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

Banco de la República (2006). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2007). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2008). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2014). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2015). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2016). Informe de política monetaria. Technical report, Banco de la República de Colombia. Diciembre.

Banco de la República (2020a). Informe de política monetaria. Technical report, Banco de la República de Colombia. Enero.

Banco de la República (2020b). Informe de política monetaria. Technical report, Banco de la República de Colombia. Octubre.

Banco de la República (2021a). Informe de política monetaria. Technical report, Banco de

la República de Colombia. Julio.

Banco de la República (2021b). Informe de política monetaria. Technical report, Banco de la República de Colombia. Octubre.

Banco de la República (2022). Informe de política monetaria. Technical report, Banco de la República de Colombia. Octubre.

Banco de la República (2023). Informe de política monetaria. Technical report, Banco de la República de Colombia. Octubre.

Banco de la República (2024). Informe de política monetaria. Technical report, Banco de la República de Colombia. Enero.

Bejarano-Salcedo, V., Caicedo-García, E., Lizarazo-Bonilla, N. F., Julio-Román, J. M., and Cárdenas-Cárdenas, J. A. (2020). Hechos estilizados de la relación entre el niño, la niña y la inflación en colombia. Technical Report 1105, Banco de la República de Colombia.

Boaretto, G. and Medeiros, M. C. (2023). Forecasting inflation using disaggregates and machine learning. arXiv:2308.11173.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Cárdenas-Cárdenas, J. A., Cristiano-Botía, D. J., and Martínez-Cortés, N. (2023). Colombian inflation forecast using long short-term memory approach. Borradores de Economía 1241, Banco de la República.

Cobo-Serna, A. L. (2005). Output gap in colombia: An eclectic approach. Borradores de Economía 327, Banco de la República.

Cote-Barón, J. P., Pulido-Mahecha, K. L., Rodríguez-Rodríguez, N. V., and Rojas-Martínez, C. D. (2023). Un indicador para monitorear la actividad económica colombiana en alta frecuencia. Borradores de economia.

Coulombe, P. G., Frenette, M., and Klieber, K. (2023). From reactive to proactive volatility modeling with hemisphere neural networks. arXiv:2311.16333. Revised 2024-04.

Coulombe, P. G., Leroux, M., Stevanovic, D., and Surprenant, S. (2021). Macroeconomic

data transformations matter. *International Journal of Forecasting*, 37(4):1338–1354.

Coulombe, P. G., Leroux, M., Stevanović, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*.

Eduardo, L.-R. C. and Gonzalo, O.-C. F. (2018). Nowcasting economic activity with electronic payments data: A predictive modeling approach. Borradores de economia.

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica*, 89(1):181–213.

Gáfaro-González, M. M., Cobo-Serna, A. L., Caicedo-García, E., and González-Ramírez, A. (2022). Recuadro 1: Inflación de alimentos: una comparación con otros países. *Informe de Política Monetaria*.

Galeano-Ramírez, F. J., Martínez-Cortés, N., and Rojas-Martínez, C. D. (2021). Nowcasting colombian economic activity: Dfm and factor-midas approaches. Borradores de economia.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy. *The Review of Economic Studies*, 72(3):707–734.

Gómez, J. and Julio, J. M. (2000). An estimation of the nonlinear phillips curve in colombia. Borradores de Economía 160, Banco de la República.

González, A. et al. (2013). Output gap and neutral interest measures for colombia. *Monetaria (CEMLA)*.

González, A., Guarín, A., Rodríguez-Guzmán, D. A., and Vargas-Herrera, H. (2020). 4gm: A new model for the monetary policy analysis in colombia. Borradores de Economía 1106, Banco de la República.

Gonzalez, A., Guarin, A., Rodriguez-Guzman, D. A., and Vargas-Herrera, H. (2020). 4gm: A new model for the monetary policy analysis in colombia. Borradores de economia.

González, A., Mahadeva, L., Prada, J. D., and Rodríguez, D. (2011). Policy analysis tool applied to colombian needs: Patacon model description. Borradores de economia.

Goulet Coulombe, P. (2025). A neural phillips curve and a deep output gap. *Journal of*

Business & Economic Statistics, 43(3):669–683.

Hastie, T. J. (2017). Generalized additive models. *Statistical models in S*, pages 249–307.

Lanau, S., Robles, A., and Toscani, F. (2018). Explaining inflation in colombia: A disaggregated phillips curve approach. IMF Working Paper WP/18/106, International Monetary Fund.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Muñoz-Martínez, J. A. (2026). A level-dependence approach for assessing de-anchoring of inflation expectations: Evidence from colombia. *Latin American Journal of Central Banking*, page 100199.

Nigrinis-Ospina, M. (2004). ¿es lineal la curva de phillips en colombia? Borradores de Economía 282, Banco de la República.

Pulido-Mahecha, K. L., Ángel, S. R., and Galeano-Ramírez, F. J. (2024). Measuring the unmeasurable: Unraveling the complexities of real-time output gap estimation. Borradores de economia.

Ramos-Veloza, M. A., Naranjo-Saldarriaga, S., and Pulido, J. (2026). Wage dynamics, unemployment, and inflation: Enhancing 4gm semi-structural model. Borradores de economia.

Roldán-Ferrín, F. and Parra-Polanía, J. A. (2025). Enhancing inflation nowcasting with online search data: a random forest application for colombia. Borradores de Economía 1318, Banco de la República.

Romero, J. V. and Naranjo-Saldarriaga, S. (2024). Weather shocks and inflation expectations in semi-structural models. *Latin American Journal of Central Banking*, 5(2):100112.

Salazar-Díaz, A., Garavito-Acosta, A. L., Restrepo-Ángel, S., and Arcila-Agudelo, L. V. (2023). Real equilibrium exchange rate in colombia: Thousands of vec models approach. *Lecturas de Economía*, (99):33–78.

Sánchez-Jabba, A. and Villabón-Hinestroza, E. (2024). The measure matters: Differences in the passthrough of inflation expectations in colombia. BIS Working Papers 1205, Bank

for International Settlements.

Sánchez-Jabba, A., Villabón-Hinestroza, E., and Romero-Torres, B. (2023). Inflation expectations measurement and its effect on inflation dynamics in colombia. Borradores de Economía 1257, Banco de la República.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Stock, J. H. and Watson, M. W. (2008). Phillips curve inflation forecasts.

Tschuchnig, M., Tschuchnig, P., Ferner, C., and Gadermayr, M. (2023). Inflation forecasting with attention based transformer neural networks. arXiv:2303.15364.

Vargas, H. (2016). Inflation expectations and a model-based core inflation measure in colombia. *BIS Paper*, (89i).

Vargas-Herrera, H., González-Gómez, A., González-Molano, E. R., Romero-Chamorro, J. V., Rojas, L. E., et al. (2009). Assessing inflationary pressures in colombia. *Borradores de Economía; No. 558*.

Zhao, Q. and Hastie, T. (2021). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281.