

# BORRADORES DE ECONOMÍA



Measuring the Unmeasurable:  
Unraveling the complexities of  
real-time output gap estimation

By:  
Karen L. Pulido-Mahecha  
Sergio Restrepo- Ángel  
Franky Juliano Galeano-Ramírez

No. 1284  
2024



# Midiendo lo inmensurable: desentrañando las complejidades de la estimación de la brecha del producto en tiempo-real\*

Karen L. Pulido-Mahecha<sup>†§</sup>

Sergio Restrepo-Ángel<sup>‡</sup>

Franky Juliano Galeano-Ramírez<sup>†</sup>

Las opiniones contenidas en el presente documento son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva. Los errores y omisiones en este trabajo son responsabilidad de los autores.

## Resumen

Este artículo evalúa siete modelos de brecha de producto para estimaciones en tiempo-real, con base en tres criterios: estabilidad de las estimaciones ante nuevas observaciones, revisiones de datos y/o cambios metodológicos; precisión en el pronóstico de inflación y la respuesta del producto potencial ante choques económicos estructurales. Los resultados confirman que ningún modelo lidera en todos los criterios. Los VAR estructurales exhiben los mejores pronósticos de inflación, pero muestran una alta inestabilidad, mientras que los modelos semiestructurales producen respuestas de producto potencial teóricamente más consistentes. Para superar este trade-off, proponemos un nuevo enfoque de agrupación para construir conjuntos de modelos en función de su rendimiento en tiempo-real con el fin de obtener mejores estimaciones. Nuestros hallazgos resaltan el valor de este método para mejorar la medición de la brecha del producto en tiempo real e informar mejor las decisiones de política monetaria.

*Clasificación JEL:* E2, E3, E6

*Palabras clave:* brecha del producto; estimación en tiempo real; ciclos reales.

---

\*Agradecemos especialmente a Miguel Ángel Cely, Federico Meneses, Juan Felipe Carmona, Oscar Murillo y Alejandra Cotes por su excelente asistencia de investigación. También agradecemos los valiosos comentarios de Carlos Daniel Rojas, Hernando Vargas, Juan José Ospina y Carlos Huertas, así como al resto del equipo técnico del Banco de la República. Los autores son <sup>†</sup> Profesional especializado y <sup>‡</sup> Profesional líder del departamento de programación e inflación del Banco de la República. Cualquier error es responsabilidad exclusiva de los autores.

<sup>§</sup> Autor para correspondencia, correo electrónico: kpulidma@banrep.gov.co

# Measuring the Unmeasurable: Unraveling the complexities of real-time output gap estimation\*

Karen L. Pulido-Mahecha<sup>†§</sup>      Sergio Restrepo-Ángel<sup>‡</sup>  
Franky Juliano Galeano-Ramírez<sup>‡</sup>

The opinions, statements, findings and interpretations contained in this document are the sole responsibility of the authors and do not represent the position of Banco de la República or its Board of Directors. All errors and omissions in this work are our responsibility.

## Abstract

This paper evaluates seven output gap models for real-time estimates, based on three criteria: stability of estimations on new observations, data revisions and/or methodological changes; inflation forecasting accuracy; and potential output response to structural economic shocks. Results confirm no single model outperforms across all criteria. Structural VARs exhibit superior inflation forecasts but show high instability, while semi-structural models produce more theoretically consistent potential output responses. To overcome this trade-off, we propose a novel clustering approach to pool models based on their real-time performance, yielding improved estimates. Our findings highlight the value of this method for enhancing real-time output gap measurement and informing monetary policy decisions.

*JEL classification:* E2, E3, E6

*Keywords:* output gaps; real-time estimation; business cycles.

---

\*We especially thank Miguel Ángel Cely, Federico Meneses, Juan Felipe Carmona, Oscar Murillo and Alejandra Cotes for their excellent research assistance. We also appreciate the helpful comments of Carlos Daniel Rojas, Hernando Vargas, Juan José Ospina, Carlos Huertas, and Banco de la República's technical staff. The authors are <sup>†</sup> Specialized Analysts and <sup>‡</sup> Leading Analyst at the Inflation and Macroeconomic Programming Department of the Central Bank of Colombia. Any remaining errors are the exclusive responsibility of the authors.

<sup>§</sup> Corresponding author, e-mail: kpulidma@banrep.gov.co

# 1 Introduction

The output gap is a key input for monetary policy analysis by informing policymakers about the existence of inflationary pressures. Despite its importance, the output gap is often difficult to measure because potential output may have different interpretations and is unobservable. Therefore, its estimation has significant uncertainty that is more prominent when estimating it in real-time.

Real-time estimations have different challenges. First, although a wide range of techniques have been developed, there is no clearly superior methodology. Methods tend to produce heterogeneous results in both magnitude and direction, increasing the uncertainty in their results. Second, models estimate multiple parameters that can change over time because of structural changes or idiosyncratic shocks, leading to historical adjustments in the output gap and potential GDP figures. The uncertainty in real-time results also comes from the set of information used. Data revisions can affect model interpretation and, thus, their diagnosis of the economy's position in the business cycle. This is compounded by the lagged availability of the National Accounts information, which requires guiding models in the forecast horizon with GDP nowcasts that can be significantly distant from the observed.

The specialized literature has analyzed how informative and useful real-time output gap estimates are. Most studies have focused on their performance in forecasting inflation, sensitivity to data revisions, and the potential output response to economic shocks. The findings vary depending on the method used. Models that measure potential output by using a structural VAR have more accurate inflation forecasts, whereas those incorporating labor market figures provide more stable estimates. However, empirical evidence suggests an important sensitivity from real-time output gap estimates to data and model uncertainty, which

can mislead the diagnosis of the economic situation and, hence, monetary policy decisions.<sup>1</sup> Evidence also suggests that filtering-based models produce potential output estimates that respond to transitory shocks procyclically, whereas models such as structural VARs can perform better (Coibion et al. (2017); Chen and Gornicka (2020)).

This challenge is more complex for developing countries. These economies are subject to a wide range of both external and internal idiosyncratic shocks that are often of larger magnitudes and more frequent or persistent than those faced by developed economies. Further, revisions in the macroeconomic time-series are usually larger and, thus, can significantly distort the information that policymakers extract from real-time estimates of the output gap. However, only a few studies have evaluated its utility and stability in developing countries (Cusinato et al. (2013) for Brazil; Amador-Torres (2017) for Colombia; Figueroa et al. (2019) for Chile). The results suggest an important sensitivity from real-time estimates to the inclusion of the latest information and data revisions. For Colombia, most studies have focused on designing methodologies to measure the output gap and potential GDP but without evaluating their real-time performance and, hence, their implications for monetary policy decision-making.<sup>2</sup>

Notwithstanding the growing interest in analyzing real-time performance, there is no methodology that allows identifying the best estimates for monetary policy decision-making. We contribute to this research in two ways. First, we are the first to present an aggregate real-time evaluation of models for the criteria proposed in the literature: stability through time, core inflation forecast, and potential GDP response under economic shocks. Second,

---

<sup>1</sup>See Orphanides and van Norden (2002); Camba-Mendez and Rodriguez-Palenzuela (2003); Planas and Rossi (2004); Mitchel, 2007; Proietti et al. (2007); Garratt et al. (2008); Marcellino and Musso (2011); Fleischman and Roberts (2011); Clements and Galvão (2012); Edge and Rudd (2016); Kangur et al. (2019); Barbarino et al. (2020); Burlon and D’Imperio (2020)

<sup>2</sup>See Cobo (2005); Torres (2008); Sánchez-Beltrán and Melo-Velandia (2013); Amador-Torres (2017); González et al. (2013); Gonzalez et al. (2020); De Castro-Valderrama et al. (2021); Gómez-Pineda and Roa-Rozo (2023)

we propose a new method that provides a better way to nowcast the output gap and potential GDP by pooling models with cluster algorithms, according to their performance in the criteria mentioned. It also allows policymakers to identify the weaknesses of models to improve their performance. We believe that this approach could be useful for central banks to mitigate the misestimation of real-time results with an easy, generalized, and innovative implementation.

The results confirm the difficulty in estimating the output gap in real-time. We find that there is no method that is clearly superior to the others. Models with good accuracy in forecasting core inflation, within which SVAR is one of the most representative, are also characterized by major instability in their real-time estimates. In contrast, semi-structural models' estimation seems to have the potential GDP most consistent with economic theory since economic shocks with permanent effects on output significantly affect their levels. However, most of the potential figures evaluated do not respond adequately under a productivity shock. Finally, we found that the best real-time estimates for Colombia come from pooling models by assigning a major weight to a semi-structural model based on adaptive expectations and which includes labor market indicators. We consider that the new output gap and potential GDP figures can inform monetary policy decisions more meaningfully.

The remainder of this paper is organized as follows. Section 2 describes the real-time data and the output gap measures used. Section 3 presents the empirical strategy for the model evaluation and the main results for each criterion. In Section 4, we propose an original approach to pooling real-time estimates for short-run results. Finally, Section 5 concludes.

## 2 Empirical design

### 2.1 Models

We consider seven methodologies used by the technical staff of Colombia’s Central Bank for *short-term* estimates. The methods apply different estimation processes and capture diverse aspects of goods and labor markets. In this section, we briefly describe each methodology.<sup>3</sup>

The first methodology is the **univariate filter** proposed by Blagrove et al. (2015). It defines a stochastic process for annual real GDP that comprises three equations and is subject to three types of shocks: a growth rate shock to the trend, a level-shock to the trend and an output gap shock. This structure offers a major economic interpretation compared to other detrending methods and recognizes the existence of factors affecting the potential output (e.g., technological and productivity change, demographic trends, and the Covid-19 pandemic).

The second methodology is a growth accounting model that represents the productive capacity of the economy through a Cobb-Douglass **production function** with human capital (Rodriguez and Prieto, 1997; Rodríguez et al. (2004); Cobo (2005)).<sup>4</sup> The model expresses productive factors in terms of their effective levels: the observed labor force and all the capital stock available at a specific time adjusted by installed capacity utilization. The labor force dynamic incorporates the latest results of the 2018 Colombian National Population Census,<sup>5</sup> the new methodological framework of household surveys, and the effects of the Venezuelan migratory phenomenon on the Colombian population.<sup>6</sup> The same applies to po-

---

<sup>3</sup>A more detail description of the methodologies is in Appendix A.

<sup>4</sup>Following the methodology suggested by Inklaar and Timmer (2013), Cohen and Leker (2014).

<sup>5</sup>It incorporates the changes in the demographic evolution, territory organization and population distribution experienced by Colombia in the last decade.

<sup>6</sup>Since 2015, Colombia was the Latin American country that received the largest flow of migrants from Venezuela. According with information from household surveys, in 2019 close to 1.8 million Venezuelan

tential GDP, but the trend components of the series are used.

Following Chen and Gornicka (2020), we also include a **Blanchard-Quah structural VAR (SVAR)** for a small open economy. It generates a historical decomposition of real GDP through six shocks: supply, demand, monetary policy, exchange rates, and transitory and permanent global shocks. The potential GDP is the sum of those affecting long-run levels. This model imposes sign restrictions, according to Fry and Pagan (2011), and it uses Bayesian methods with Minnesota priors for its estimation.

The remaining four methodologies are semi-structural models, that express economic relationships through a system of simultaneous macroeconomic equations. The methods based their micro-foundation on New-Keynesian models and used Bayesian techniques for parameter estimation.

The first two are the **adaptive and rational expectation models**, proposed by González et al. (2013). These represent a small open economy without a labor market. The models differ in how agents form their expectations. Both models represent an IS curve, Phillips curve, policy rule, and uncovered interest rate parity equation. They also include definitions of the output gap, potential GDP, inflation expectations, equilibrium levels of the real domestic and foreign interest rates, and the real exchange rate.

The third semi-structural model represents a closed economy with a labor market, as proposed by Blagrove et al. (2015). It incorporates an equation relating the output to the unemployment rate (i.e., the **Okun's Law**). The model also includes a Phillips curve, permanent and transitory shocks in the goods and labor markets, and definitions of output and unemployment rate gaps.

---

migrants resided in Colombia. Empirical evidence suggests that this population shock could have short and long-term effects on the Colombian economy (see Tribín-Uribe et al. (2020).

Finally, we incorporate a version of the semi-structural model for monetary policy and macroeconomic forecasting proposed by Gonzalez et al. (2020).<sup>7</sup> The model called “4GM” is based on a New-Keynesian rational expectation framework for an oil-exporting small open economy. It consists of four equations: an IS curve, a set of four Phillips curves that characterize the inflation of the baskets of tradable, non-tradable, food and regulated goods, an uncovered interest rate parity condition, and a monetary policy rule.

Considering that models provide an approximation according to their structure and sources of information, we also evaluate the results derived from combining individual models. Alternative gap measures are constructed by pooling all combinations with simple and weighted averages (based on a sequence of weights that takes finite decimal values from up to one). In total, we evaluate 7,946 output gaps in Colombia. Table 1 summarizes the main characteristics.

Table 1: Individual models for estimate output gap

Type	Model	Available	Frequency
Statistical filters	Univariate	1981Q1 – 2024Q1	Annual*
Growth accounting	Product function	1981Q1 – 2024Q1	Quarterly
Semi-structural models	Adaptative expectations	1994Q1 – 2024Q1	Quarterly
	Rational expectations	1994Q1 – 2024Q1	Quarterly
	4GM	2003Q1 – 2024Q1	Quarterly
	Okun’s Law	1984Q1 – 2024Q1	Annual*
Shock identification	SVAR	1998Q1 – 2024Q1	Quarterly
Combinations	Simple average	2003Q1 – 2024Q1	Quarterly
	Weighted average	2003Q1 – 2024Q1	Quarterly

\*For their quarterly values we apply an interpolation method that uses the simple average of the other individual models. Source: own elaboration.

<sup>7</sup>This version of 4GM model does not include any constraints and differs from the official version used by the technical staff of the Central Bank of Colombia.

## 2.2 Real-time data

We use real-time vintage data throughout this study. Following Orphanides and van Norden (2002), these estimates correspond to the output gap and potential GDP calculated using the set of all available information at period  $t$ . This relates to the first announcement of the output gap for a specific quarter  $t + 1$ , calculated using the observed data until quarter  $t$  and its short-run forecast, respectively.

To construct these vintage estimates from 2007Q1 until 2024Q1, we replicate the estimations of the Monetary Policy Report (MPR) led by the Central Bank of Colombia.<sup>8</sup> The sequence of real-time output gaps considers the information available in each quarter. This information includes both the GDP release from the National Accounts of the Colombian official statistical bureau (DANE) and the official GDP forecast produced by the Central Bank's technical staff.<sup>9</sup> We have replicated the MPR process since 2007Q1, maintaining the most recent structure and parameter calibration of the models.<sup>10</sup>

Figure 1 shows the sequence of real-time output gap estimates for individual models throughout the replicated vintages (69, since 2007Q1). The output gap results of the models are broadly procyclical and capture some important turning points such as the economic recession of the late 1990s, the economic boom between 2004 and 2007, the Global Financial Crisis of 2008, the sudden and sizeable GDP drop due to the Covid-19 pandemic and the subsequent recovery experienced in recent years. However, estimates at any time are heterogeneous in magnitude, which reinforces the remark that real-time output gap estimation widely differs across models.

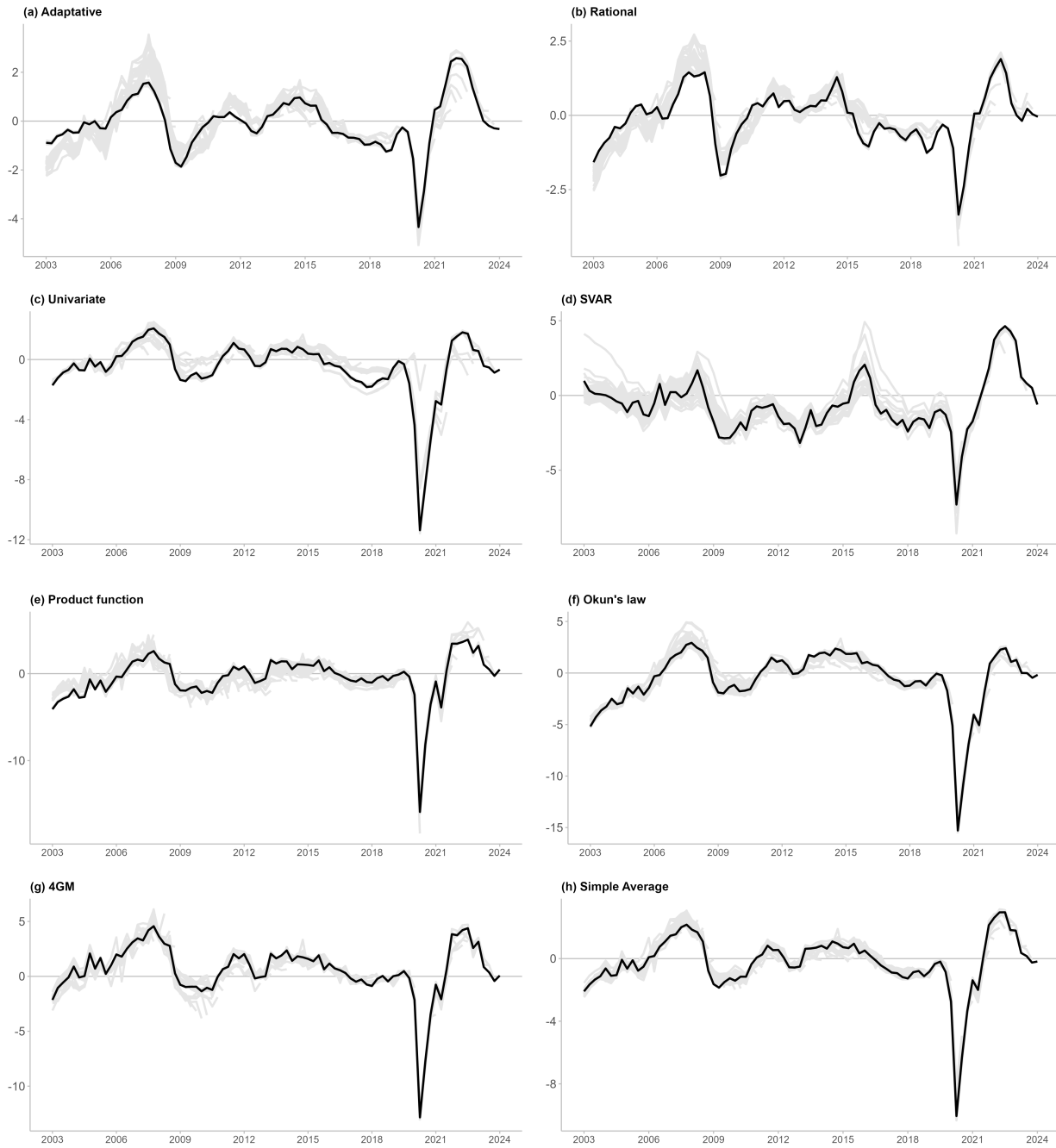
---

<sup>8</sup>MPR is updated every quarter and published in January, April, July, and October. For more information consult <https://www.banrep.gov.co/publicaciones-investigaciones/informe-politica-monetaria>

<sup>9</sup>We do not include forecasts for labor market indicators, internal and external inflation, policy interest rate, exchange rate and installed capacity since they are available at the time MPR is published.

<sup>10</sup>Not all methods have always been part of the technical staff estimations, some were included over time and other adopted new assumptions into its structure.

Figure 1: Real-time output gap estimates for Colombia  
Individual models, 2003Q1 - 2024Q1



Note: The black line corresponds to the latest estimation up to 2024Q1 and the grey ones are previous estimates. Each vintage includes data from 2003Q1. Source: own estimates.

### **3 Real-time evaluation**

To assess the reliability and usefulness of real-time estimates for policymaking, we evaluate the performance of the models with the properties suggested by Kangur et al. (2019).

Considering that real-time output gap estimates tend to change over time, with ex-post revisions that are at least as large as the estimates themselves, we expect to have models with a degree of stability in their results. The more stable the real-time estimates are, the smaller the monetary policy error associated with its misestimation. If this is satisfied, then future revisions should not be foreseeable because the models use all the information available at the time of the estimation. It is also suitable for real-time estimates to accurately signal inflationary pressures in the economy, which implies a high correlation with core inflation and other slack economic indicators. Another desirable property is that potential GDP levels respond only to idiosyncratic shocks with long-term output effects.

In this section, we define three criteria for quantifying and evaluating these properties using the 7,946 combinations described in Table 1. The first property is assessed by the historical stability of output gap estimates, while the latest are evaluated by the predictive power of output gap estimates to forecast core inflation and the response of potential GDP to a set of economic shocks previously identified. In the following sections, we present the main methodological aspects and results.

#### **3.1 Estimations stability**

As shown in Figure 1, real-time output gap estimates are often revised, highlighting the uncertainty surrounding the results. The specialized literature has documented three main sources of output gap revisions and potential GDP.

First, the inclusion of **new observations** occurs due to the passage of time. This can lead to a re-estimation of the parameters and shock variance, affecting the previous inference of the trend components. Second, methods are usually based on quarterly time series, which are subject to **data revisions** due to their provisional and preliminary nature for up to three years after their first publication. Finally, output gap estimates can vary due to **methodological changes**, including new assumptions incorporated into the model's structure or adjustments in the indicators used (e.g., changes in household or sector survey frameworks).

Considering that the output gap vintages were constructed with the current model structure and methodological survey frameworks, the changes illustrated in Figure 1 are mainly explained by the inclusion of new observations or data revisions.

Therefore, let be  $R_{m,q}^{RT}$  the output gap revisions of model  $m$  for at a specific quarter  $q$ . We define this as the difference between the real-time output gap estimates that result from using all available information at  $t + i$  and those estimated with the previous information set at  $t + i - 1$ . Because our database has a fixed rolling window structure, these revisions were divided into the  $N_q$  vintages obtained at each point in time.<sup>11</sup>

$$R_{m,q}^{RT} = \frac{\sum_{i=0}^I [y_{m,q|X_{t+i}} - y_{m,q|X_{t+i-1}}]}{N_q} \quad (1)$$

To understand the drivers of these revisions, we decompose them into the source of the instability mentioned above. The changes that turn out from the passage of time correspond to the revisions of the quasi-real time series ( $R_{m,q}^{QT}$ ), which are estimated according to the Equation (1) using the same replication process of the MPR but with the latest available

---

<sup>11</sup>By construction we have a matrix of revisions, which is different from other seminal papers that consider at least three or four benchmark revisions.

publication of the National Accounts.<sup>12</sup> Revisions associated with data corrections result from the difference between the real-time output gap revisions and quasi-real time revisions. Table 2 displays the sources of the output gap revisions according to equation’s (1) notation.

Table 2: Sources of revisions

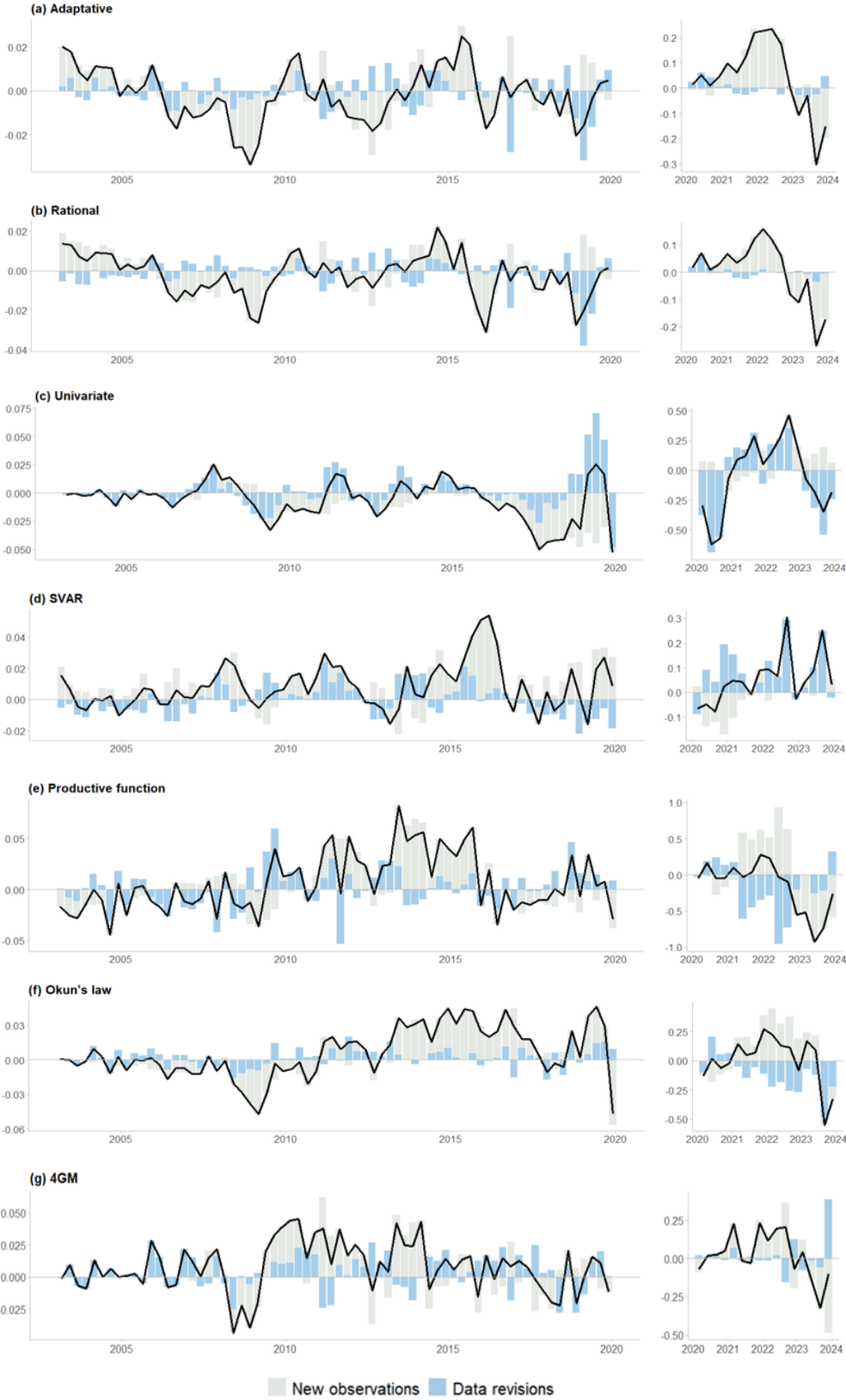
Source	State	Calculation
(i) Changes in trend (or potential output) because new observations	On	$R_{m,q}^{QT}$
(ii) Data revisions (National Accounts)	On	$R_{m,q}^{RT} - R_{m,q}^{QT}$
(iii) Methodological changes	Off	-
Total revisions = (i) + (ii) + (iii)		$R_{m,q} = (R_{m,q}^{RT} - R_{m,q}^{QT}) + R_{m,q}^{QT}$

Figure 2 shows the total average real-time output gap revisions ( $R_{m,q}$ ) for any quarter for the individual models and the decomposition between the sources of instability. Generally, the revisions’ size from 2020Q1 until 2023Q4 is greater than that during the pre-pandemic period. This characteristic of one side is due to the provisional and preliminary nature of the National Accounts, whose original and seasonally adjusted quarterly times series often show significant changes between vintages up to three years after their first publication.

Consequently, noise is introduced into the real-time output gap estimates at the end of the sample. In addition, the fixed rolling window structure of our database, which expands one observation at a time, has a limitation for those points to compare few estimates of available vintages  $N_q$  (i.e., the short run nowcast of the output gap for 2024Q1 had been estimated only one in our sample), which can lead to a greater variance in the revisions.

<sup>12</sup>At the time of publication of this document, the latest National Accounts publication used was as of February 2024.

Figure 2: Average decomposition of real-time output gap estimates for individual models.



Source: own estimates.

In terms of the source of revisions, we found that semi-structural models, such as Adaptive and Rational ones, are the most sensitive to incorporating new observations. This is even higher during a crisis and post-crisis periods, like 2001-05, 2008 and 2022. One explanation is that the models re-estimate the shock variance associated with the trend components according to the set of information available. When the economy suffers a disturbance that has never been observed (e.g., the Covid-19 pandemic) models will update those variances and, therefore, their previous diagnosis of the output gap and potential GDP. This behavior is not desirable when we use real-time estimates because the effects of the disturbance at one specific point should not affect the estimates of one decade behind.

In contrast, the sensitivity of univariate and SVAR estimates is significantly influenced by data revisions. However, the Production function and Okun's law models exhibit greater stability due to their incorporation of labor market indicators from household surveys, which remain consistent across publications. On the other hand, the output gap revisions in the 4GM are affected by both data revisions and the inclusion of new observations over recent years.

We focus on analyzing the statistical properties of the output gap revisions following Aruoba (2008). First, note that throughout this document, we have assumed that real-time output gap estimates use all available information at  $t+i$ , which in turn implies that revisions result from the arrival of further details in our data that could not have been observed previously. If this assumption is correct, revisions should have a mean of zero, because the first output gap estimation of model  $m$  for quarter  $q$  is an unbiased estimator of the final value.

Second, we want models with the lowest possible revisions volatility. To control for the size of the revisions, we evaluate whether the ratio of the standard deviation of the total revisions ( $R_{m,q}$ ) to the standard deviation of the output gap estimates tends to zero. This

measure is known as *noise-to-signal ratio* (NSR). As we use more benchmark revisions than in other studies, we focus on the median standard deviation for both variables.

An additional property is the *sign agreement*. It is common for models to change their diagnosis of the economic stance in the business cycle. This means that sometimes with all the sets of information  $X_{t+i-1}$ , one model might suggest a negative value for a specific quarter  $q$ , but when using the set  $X_{t+i}$ , it changes its estimation to a positive one and vice versa. The results from these models should be undertaken with great care because they can provide a misleading diagnosis and, most importantly, monetary policy mistakes. We construct the “*sign agrees*” measure by assigning a value of one or minus one to positive or negative output gaps,<sup>13</sup> respectively, for each quarter  $q$  through the  $N_q$  available vintages. For each quarter we sum all the ones and minus ones. The indicator consists of the average of those absolute values for the observed  $q$  quarters. It can take values from 0 to 100, where 100 indicates that model  $m$  always diagnoses the same sign for a quarter  $q$  through the  $N_q$  vintages.

Table 3 presents the results for the individual models, their simple average, and some combinations. The mean of the output gap revisions is significantly different from zero for 12,3% of the models evaluated, with SVAR and production function models being the most representative. In general, revisions are usually negative, suggesting that real-time estimations often overestimate the output gap. In contrast, the 4GM and some of its combinations result in an output gap that, on average, are underestimated.

The column labeled ‘NSR’ represents the *noise-to-signal ratio*. Among the individual models, SVAR and production function exhibit the lowest performance, with a ratio exceeding one. In contrast, the semi-structural models yield values close to zero, indicating greater real-time stability. The weighted average of the Rational, Univariate, and 4GM models yields the low-

---

<sup>13</sup>To avoid not significant changes, we omit those values that are between -0.2 and 0.2.

est NSR value. Only 17,5% of all evaluated combinations registered an NSR higher than one, with the weighted average between SVAR and Okun's law having the highest value.

Table 3: Summary statistics of output gap revisions

Model	Mean	Revision (std dev)	Ouput gap (std dev)	NSR	Sign agree
Adaptative	0.002	0.004	0.968	0.391	97.3
Rational	-0.005	0.004	0.948	0.394	93.7
Univariate	-0.004	0.003	0.987	0.325	93.0
SVAR	0.023*	0.017	1.061	1.588	90.7
Product function	-0.030*	0.020	1.342	1.458	88.9
Okun's Law	0.000	0.013	1.986	0.631	96.2
4GM	0.010	0.007	1.630	0.460	91.7
Simple average	-0.001	0.007	1.136	0.628	96.8

(\*) Null average hypothesis equal to zero is rejected at a level of significance of 5%.

	Model	NSR	Sing agree
Maximun	Weighted combination between Adaptative and Okun's Law models		99.6
	Weighted combination between the SVAR and Okun's law	1.789	
Minimun	Weighted combination between the Rational, Univariate and 4GM	0.258	
	Weighted combination between the SVAR and 4GM		84.2

Source: own estimates.

In evaluating the criteria for sign agreement, the models exhibit less heterogeneity. Generally, real-time estimates consistently maintain a sign diagnosis for over 80% of the available

$N_q$  estimations. The evidence suggests that combined models outperform individual ones, with the weighted average of Adaptive and Okun’s Law models proving to be the most stable. Conversely, the weighted average of the SVAR and 4GM models yields the poorest performance. Contrary to prevailing literature on developing countries (Chen and Gornicka (2020)), our analysis does not support the notion that all models incorporating labor market indicators inherently offer greater stability. In fact, the product function model underperforms when compared to semi-structured methods.

### 3.2 Inflation forecasting

For our second test, we use a Phillips Curve model to evaluate the accuracy of the output gap in forecasting core inflation one step ahead. Following a specification commonly proposed in the literature,<sup>14</sup> for each output gap we consider:

$$\pi_t = \alpha + \sum_{i=1}^4 \lambda_i \pi_{t-i} + \beta y_{m,t} + \gamma z_t + \varepsilon_t \quad (2)$$

Where,  $\pi_t$  is the quarterly annualized core inflation rate measured by CPI excluding food and regulated items<sup>15</sup> minus the inflation target (3.0% for Colombia);  $y_t$  is the real-time output gap estimate from model  $m$  with information available at period  $t + i$  and  $z_t$  corresponds to the real exchange rate gap.<sup>16</sup> We also add a constant term  $\alpha$  to control for any remaining mean inflation effects.

We estimate equation (2) using ordinary least squares (OLS) with the restriction:  $\sum_{i=1}^4 \lambda_i = 1$ . We are interested in the out-of-sample properties through the one-step ahead root-mean-square error (RMSE). For this, an iterative expanding window procedure is applied for the

---

<sup>14</sup>See Marcellino and Musso (2011); Figueroa et al. (2019) and Barbarino et al. (2020).

<sup>15</sup>We focus on core inflation because output gap estimates should account for price dynamics related with demand pressures, and not for those related with weather shocks affecting food prices or regulated prices ruled by laws and decrees.

<sup>16</sup>Estimates for Colombia corresponds to Salazar-Díaz et al. (2023).

period 2007Q1 to 2024Q1, using the set of real-time output gap estimates for  $N_q$  vintages.

A benchmark model that only includes the lagged values<sup>17</sup> of the core inflation is also considered to compare the RMSE with the model of Equation (2). Thus, we can evaluate whether the output gap estimates bring value added to forecast core inflation:

$$\pi_t = \alpha + \sum_{i=1}^4 \lambda_i \pi_{t-i} + \varepsilon_t \quad (3)$$

Table 4 presents the one-step ahead RMSE for individual models alongside their simple average. Additionally, it presents an improvement measure, defined as the ratio between the RMSE of the Phillips Curve model and that of the benchmark model. We segment the analysis into three distinct periods commencing from 2007Q1: the full sample period (up to 2024Q1), the pre-Covid-19 pandemic period (up to 2019Q4), and the post-pandemic period (from 2020Q1 to 2024Q1). For robustness, the years 2020 and 2021 are included in the post-pandemic category due to the limited observations in the latter period.

The one-step ahead RMSE across all samples does not indicate significant methodological disparities, with the Univariate, Production function, Okun’s Law, and 4GM models registering the highest errors among individual models. For the pre-pandemic era, the metrics were comparable across models. On average, forecasts from individual models were positioned 1.2 percentage points above or below the observed quarterly annualized core inflation rate, with no notable improvements over the benchmark specification.

In the post-pandemic timeframe, RMSE values surged markedly. These years were characterized by various unique shocks impacting the primary inflation drivers: from the repercussions of the Covid-19 pandemic, which led to the implementation of price relief measures

---

<sup>17</sup>Results do not vary when aggregating other control variables.

in Colombia,<sup>18</sup> to exchange rate fluctuations, global supply chain bottlenecks, increased shipping costs, and rising oil and fertilizer prices. Additionally, social unrest and road blockades in Colombia during 2021 further compounded these effects.

Table 4: Forecast performance one-step ahead for individual models.

<b>Model</b>	<b>RMSE</b>			<b>Improvement*</b>	
	All sample	Pre Covid-19	Post Covid-19	Pre Covid-19	Post Covid-19
Adaptative	1.562	1.201	2.342	0.934	0.962
Rational	1.541	1.217	2.261	0.946	0.929
Univariate	1.652	1.259	2.496	0.979	1.026
SVAR	1.599	1.294	2.293	1.006	0.942
Product function	1.620	1.272	2.389	0.990	0.982
Okun's Law	1.621	1.222	2.469	0.950	1.015
4GM	1.606	1.257	2.374	0.977	0.976
Simple average	1.598	1.236	2.386	0.962	0.981

(\*) It is calculated as the ratio between the RMSE of the CP model and that of a benchmark model

Source: own estimates.

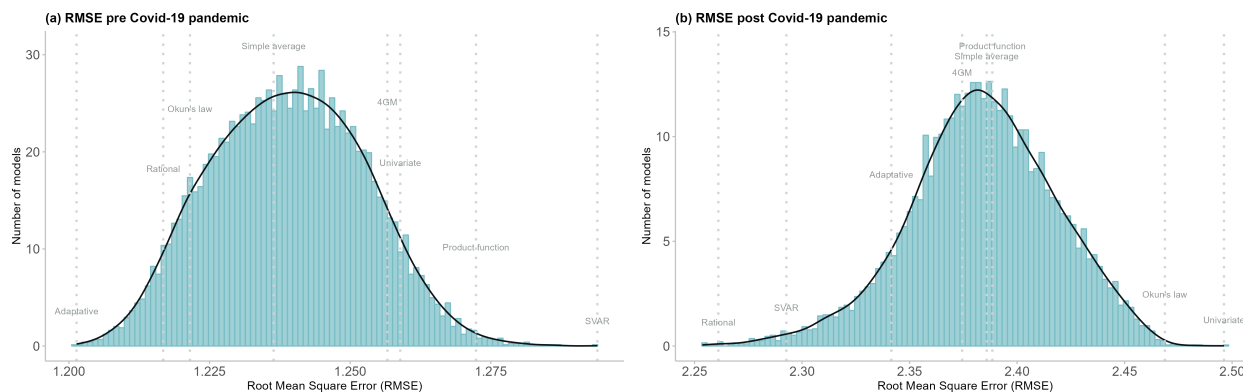
Most individual models saw a doubling of their RMSE figures, with the Univariate and Okun's Law models being notably impacted. Although the SVAR model also deteriorated as a predictor, it recorded the smallest increase in RMSE and maintained lower error figures. It also exhibited one of the smallest improvement measures, with a lower RMSE for the post-pandemic period compared to our benchmark approximation. However, caution is

<sup>18</sup>Caicedo-García et al. (2021) describes some of the main relief measures adopted in Colombia during the pandemic: (i) freeze the rental payments, (ii) increase the utilities subsidies for lower socioeconomic levels, (iii) gasoline price reduction, (iv) transitory VAT elimination for mobile phone and hostelry and tourism services, and for hygiene, health care and cleaning goods, (v) transitory VAT reduction for airfares and (vi) tax cut for food out home. Also, the government established some days in 2020, 2021 and 2022 without the VAT for some goods of the next categories: apparel, footwear, appliances, computers and communications equipment, sports elements, toys and games, school supplies, agricultural inputs, among others.

advised when interpreting these results for the years following 2020, as the small number of observations included may not guarantee a robust fit from Equations (2) and (3). Despite the duplication of RMSE results for the post-pandemic period, they are confined to a narrow range between 2.2pp and 2.5pp.

Figure 3 illustrates the RMSE distribution for all models and combinations as detailed in Table 1. A significant shift in the RMSE distribution is observed, transitioning from a leptokurtic shape to a concentration around the 50th percentile. As previously mentioned, the SVAR model improved its core inflation forecasting performance, becoming one of the models with the lowest RMSE out of 7,946 measures during the post-pandemic period. In contrast, the Univariate model is situated at the far end of the distribution. However, the most salient finding is that structural models consistently registered the lowest RMSE in both pre and post-pandemic periods. This underscores the importance of employing methods with a higher degree of inference for short-term output gap estimates.

Figure 3: Root Mean Square Error distribution by period.



Note: We displayed RMSE distributions by period. Dashed lines highlight the position of individual models and simple averages into the distribution. Source: own elaboration.

### 3.3 Potential output performance.

It is also desirable to have potential GDP estimates consistent with the main drivers of long-term fluctuations. We focus on identifying which models bring about measures that

respond significantly to economic shocks with permanent effects over output, rather than those with transitory effects. We utilize quarterly time series data of the various types of shocks as identified by other researchers. Table 5 delineates these shocks and outlines the classification criteria employed in our assessment.

Transitory perturbations include monetary policy and fiscal spending shocks. The first is identified through the semi-structural model of an oil-exporting small open economy proposed by Gonzalez et al. (2020). The latter is derived from the discrepancy between the Central National Government’s projected spending and the actual expenditure reported by official statistics at year-end, in line with the methodologies of Restrepo-Ángel et al. (2022).

Table 5: Economic shocks previously identified

<b>Economic shock</b>	<b>Type</b>	<b>Available</b>	<b>Source</b>
Monetary policy	Temporal	2003Q1 - 2024Q1	Gonzalez et al. (2020)
Oil supply	Permanent	1975Q2 - 2023Q4	Baumeister and Hamilton (2019)
Fiscal policy (spending)	Temporal	1999Q1 - 2023Q4	Restrepo-Ángel et al. (2022)
Productivity (global)	Permanent	1947Q2 - 2024Q1	Fernald (2012)
Productivity (internal)	Permanent	1951Q4 - 2023Q4	Own elaboration

For permanent shocks, we use the oil supply shock estimated by Baumeister and Hamilton (2019), which results from a BVAR for international oil prices. For global productivity perturbances, we used the shock estimated by Fernald (2012) as the annual growth in the Total Factor Productivity (TFP) for the United States business sector. A domestic productivity shock is also considered, applying Fernald’s approach to Colombia’s TFP growth, as per the annual estimates by DANE and those published by the Conference Board.<sup>19</sup>

<sup>19</sup>The measurement of TFP follows the proposed by the OECD and Hofman et al. (2017). While the Conference Board used a production function which considers the average hours worked.

Following Coibion et al. (2017), the response of potential GDP can be defined as:

$$\Delta \ln \hat{Y}_{m,t} = \alpha + \sum_{h=0}^H \phi_h \varepsilon_{t-h} + u_t \quad (4)$$

where  $t$  indexes time (quarters);  $\Delta \ln \hat{Y}_{m,t}$  is the quarterly annualized growth of potential output estimated by the model  $m$  at the last available vintage  $N_q$ ; <sup>20</sup>  $\varepsilon_t$  is a previously identified shock; and  $u_t$  is the residual. The local projection described by equation (4) uses up to eight lags of shocks and considers the Newey–West standard errors.<sup>21</sup> Impulse responses (IRFs) come directly from  $\phi_h$  and to recover the effect over the potential GDP level, we use the cumulative response up to a given horizon.<sup>22</sup>

Figures 4 to 6 illustrate the impulse-response of potential output estimates derived from individual models and the aggregated simple average in response to permanent economic disturbances: oil supply (Figure 4), U.S. productivity (Figure 5), and domestic technological shocks (Figure 6). The supplementary IRFs are detailed in Appendix B.

Upon examining the oil supply shocks, it is observed that an unanticipated increase in oil prices exerts a negative and statistically significant impact on the potential output across all individual models. The extent of this impact varies among models, with a range between -4.0 and -1.0 percentage points, and this effect persists beyond two years following the initial shock. These findings align with the anticipated negative and significant response, as a sudden hike in oil prices can escalate production costs in industries dependent on oil (Balke et al. (1999)), curtail investment, diminish labor productivity, and consequently, reduce the potential output (Le Roux et al. (2022)).

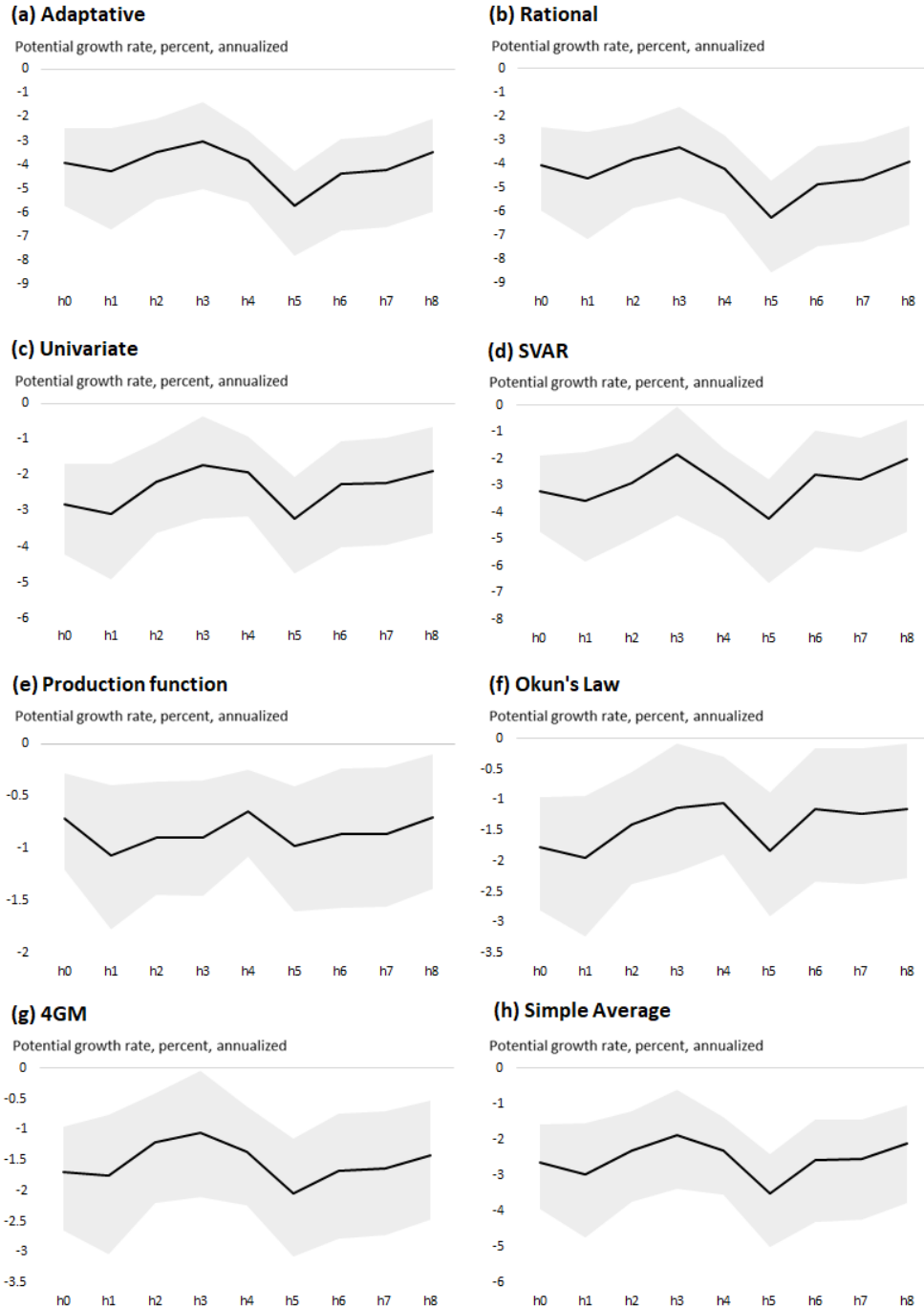
---

<sup>20</sup>Here we are not strictly using the concept of real-time evaluation applied in the above criteria.

<sup>21</sup>The residual structure is assumed to be heteroskedastic and autocorrelated up to some lag.

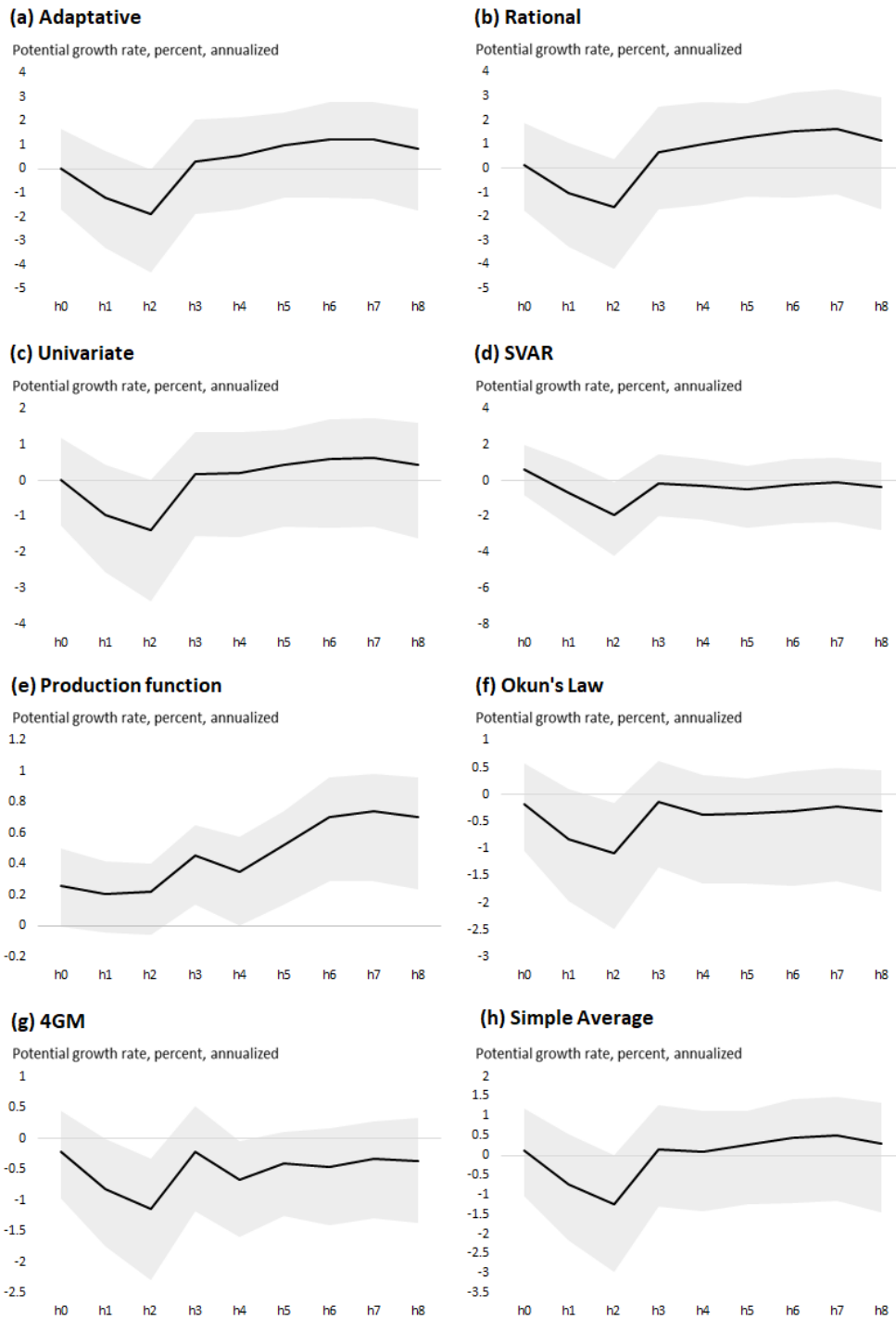
<sup>22</sup>For example,  $\phi_0$  for current period,  $\phi_1 + \phi_2$  for one period after a shock, and so on.

Figure 4: Responses of real-time potential output estimates to an oil supply shock.



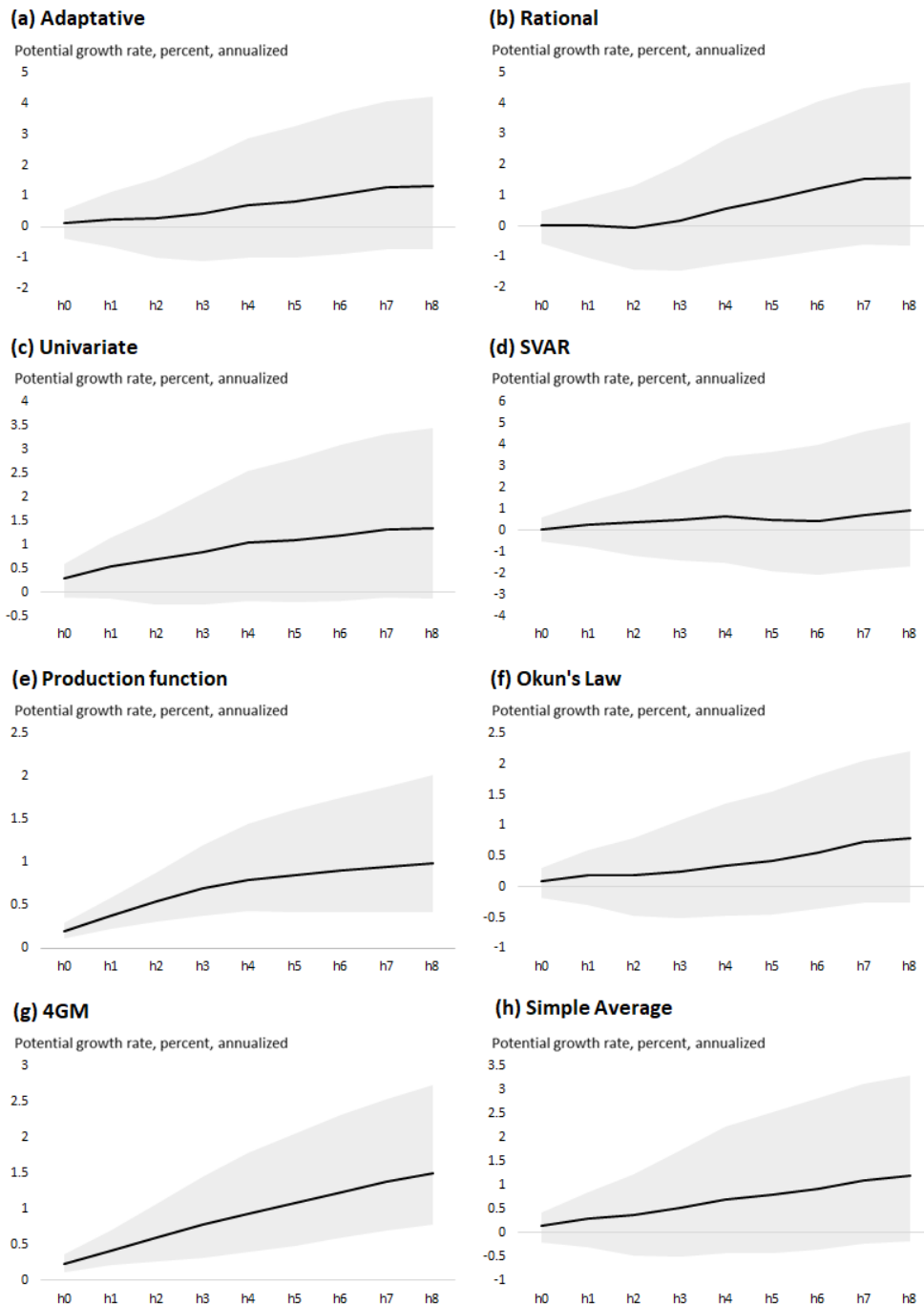
Note: This figure reports cumulative impulse response functions (IRFs) estimated using equation (4). The estimation sample covers the longest possible period with no missing observations for shocks and potential output. This is why we use estimates associated with the  $N_q$  available vintage. The dashed lines show 66 percent confidence intervals calculated using bootstrapping over the cumulated response. Source: own elaboration.

Figure 5: Response of real-time potential output estimates to an U.S. productivity shock.



Note: This figure reports cumulative impulse response functions (IRFs) estimated using equation (4). The estimation sample covers the longest possible period with no missing observations for shocks and potential output. This is why we use estimates associated with the  $N_q$  available vintage. The dashed lines show 66 percent confidence intervals calculated using bootstrapping over the cumulated response. Source: own elaboration.

Figure 6: Response of real-time potential output estimates a domestic productivity shock.



Note: This figure reports cumulative impulse response functions (IRFs) estimated using equation (4). The estimation sample covers the longest possible period with no missing observations for shocks and potential output. This is why we use estimates associated with the  $N_q$  available vintage. The dashed lines show 66 percent confidence intervals calculated using bootstrapping over the cumulated response. Source: own elaboration.

In contrast, the response to a U.S. productivity shock is found to be non-significant in six models, as depicted in Figure 5. Only the production function exhibited a significant and positive response. One possible explanation is that increases in global productivity affect the Colombian economy indirectly and with some degree of lag through international trade and technology spillovers (Coe and Helpman (1995)). Some authors also argue that a U.S productivity shock can lead to higher financial market volatility and exchange rate movements that affect the initial effects (Rodrik (2014)). Factors that can affect the model's interpretation and thus, in a non-expected response of potential GDP figures. However, the results can also be interpreted as a weakness of models identifying permanent shocks, and thus, their real-time diagnosis ability.

The impact of a domestic productivity shock on Colombia's output gap is substantial and positive, as evidenced by two models: the Production function and the 4GM, with both showing a similar magnitude of approximately 1.0 percentage points that persists for two years following the shock (Figure 6). Contrary to most of the literature which suggests that positive domestic productivity shocks—like technological improvements or enhanced labor efficiency—tend to permanently elevate potential output by increasing the economy's efficiency in producing goods and services (Rezai et al. (2008)), our model responses diverge. This discrepancy may stem from the potential noise introduced by revisions in the annual National Accounts, which serve as an input for the KLEIMS methodology,<sup>23</sup> thereby affecting the measurement of domestic productivity disturbances.

Under a monetary policy shock, the response of most models, with the exception of the production function and 4GM, is not statistically significant, aligning with the established economic principle that productive factors, rather than central bank actions, primarily influ-

---

<sup>23</sup>The PTF is measured by the National Administrative Department of Statistics (DANE) based on the OECD and LAKLEMS methodologies. The KLEMS methodology uses a framework of growth accounting with sectorial information of production, capital, employment, intermediate consumption (energy, raw materials, and services), among others.

ence potential GDP dynamics (Lucas 1996; Cerra et al. (2023)) (Figure B.1). The production function and 4GM models, however, do show a response, which Jorda et al. (2020) suggest may be due to higher interest rates causing firms to reduce investment in areas such as R&D, thereby affecting potential GDP levels. Additionally, our analysis indicates that a positive fiscal spending shock of one standard deviation has non-significant effects in five individual models (Figure B.2). The significant effects observed in the rational and production function models could be attributed to the composition of spending (infrastructure versus consumption) and the economic state impacting potential GDP levels.

To quantitatively assess the properties of the IRFs, we create an indicator according to the cumulative response two years after the initial perturbation ( $\phi^k = \phi_1 + \phi_2 + \dots + \phi_8$ ). For each shock ( $k$ ) we create a dummy variable that takes the value of one when  $\phi^k$  of model  $m$  yields the expected significance and zero otherwise. The indicator used to evaluate model performance corresponds to the average of these dummy variables. If a model registers a value of one in the aggregated indicator, its potential GDP figures responds significantly to all permanent shocks and not to transitory shocks.

$$\psi_m = \frac{\sum_{k=0}^K \chi_m^k}{K} \quad \text{with} \quad \chi_m^k = \begin{cases} 1 & \text{if } \phi_m^k \text{ is an expeted response} \\ 0 & \text{if } \phi_m^k \text{ is a contrary response} \end{cases} \quad (5)$$

The analysis derived from Equation (5) reveals a diversity of responses across the models. Notably, none of the models or their combinations consistently react to the five economic shocks under consideration. Supply oil shocks are observed to have a significant impact on potential GDP, in contrast to transitory shocks, which leave 98.2% of the models unaffected. This outcome deviates from existing literature that typically does not associate filtering-based models with procyclical reactions to transitory shocks. A notable limitation of these models is their inability to accurately discern the effects of productivity shocks; in fact, only

approximately 13.0% of the 7,946 models evaluated can capture their long-term influence on output.

## 4 How to combine real-time results?

Until now, we have evaluated models using different criteria with widely heterogeneous results, without a unique model or combination that offers the best performance. To identify those satisfying the desirable properties for real-time estimates, we use cluster analysis, which is an unsupervised machine learning technique that groups data points based on their similarity. Specifically, we apply partitioning around the medoids method (PAM) proposed by Kaufman and Rousseeuw (1990).

Utilizing the set of models  $M$  outlined in Table 1, the Partitioning Around Medoids (PAM) algorithm initiates by randomly selecting a subset  $S$  to serve as medoids. The remaining models ( $R = M - S$ ) are then categorized based on the dissimilarity between each model  $p \in R$  and the nearest medoid in  $S$ . This dissimilarity is quantified using the Manhattan distance<sup>24</sup> over the scaled indicators established previously: RMSE for all samples, noise-to-signal ratio, sign agreement, and aggregate shock response. To mitigate the effects of the initial random selection and enhance clustering quality, the algorithm iteratively swaps medoids by evaluating all pairs  $(i, h) \in S \times M$  until it achieves the lowest average dissimilarity within each group. PAM is favored over other clustering methods due to its superior performance in maintaining stability of assignments,<sup>25</sup> which measures the consistency of partitions and clusters despite perturbations to the original dataset (Liu et al. (2022)).

Upon grouping our 7,946 models and combinations into ten clusters, Table 6 presents the total count of models assigned to each cluster along with their median performance based

---

<sup>24</sup>Corresponds to the sum of absolute distances between objects.

<sup>25</sup>These measures include the average distance (AD), the average distance between means (ADM), and the figure of merit (FOM).

on the established criteria. An initial observation reveals that the utility of output gaps for forecasting core inflation remains inconclusive. However, when examining other properties, certain clusters emerge with models that excel in potential GDP criteria and exhibit stable real-time estimates, particularly evident in the noise-to-signal ratio outcomes observed in the 5th and 10th clusters.

Table 6: Performance by criterion and cluster (median)

Cluster	Number of models	RMSE (all sample)	NSR	Sign agree	Shock's indicator
9	1092	1.60	0.51	95.09	0.80
8	732	1.61	0.84	94.80	0.80
5	777	1.59	0.51	97.23	0.60
4	1139	1.61	0.67	96.86	0.60
6	838	1.60	0.67	96.01	0.60
1	590	1.58	0.45	95.48	0.60
2	743	1.60	0.46	95.15	0.60
10	723	1.61	0.92	94.24	0.60
7	587	1.59	1.03	92.30	0.60
3	725	1.61	1.27	91.67	0.60

Note: Clusters are organized according with their performance in the shock's indicator, following by the sign agree, the noise-to-signal ratio and the RMSE. Source: own elaboration.

In organizing the clusters, we give precedence to the potential criteria, as monetary policy decisions should be informed by potential output figures that are devoid of cyclical variations in output and capable of identifying long-term changes. Secondly, it is advantageous to have output gap estimates that are less affected by periodic revisions to the National Accounts data. Lastly, the models' performance in forecasting core inflation is considered less critical, as there is no significant disparity in results across the models.

In our analysis of individual models, we observed that the SVAR and production function models were categorized into the 3rd cluster, which ranks as the lowest-performing group. The Okun’s Law and the simple average of the seven methods were allocated to the 4th and 6th clusters, respectively, which comprise models with stable real-time estimates but a lower shock indicator. The Adaptive and Rational models were placed in the 1st cluster, while the Univariate and 4GM models were in the 2nd cluster, both demonstrating higher results in stability and RMSE criteria, yet with lower shock indicator values. The most proficient clusters were the 9th and 8th, encompassing 1,824 potential combinations.

Figure 7: Colombian output gap figures.  
 Quarterly, 2003Q1 - 2024Q1



Note: black line corresponds to the median of 1,869 models and gray area shows the 5th and 95th quantiles from the total distribution of 7,946 models. For estimation, we used the results from the 5th and 10th cluster. Source: own elaboration.

For nowcast estimation, we recommend utilizing the median values from the combinations within the top two clusters (9th and 8th) as a measure of the output gap. It is noteworthy that many combinations forming these groups include the Adaptive, Univariate, and Okun’s Law models, which appear to yield more stable and theoretically consistent results for real-time estimation. Acknowledging that the outcomes from this methodology may vary if a model enhances its performance based on the third criterion, we propose an interval that encapsulates the uncertainty at each point in time. This interval is delineated by the 5th and

95th percentiles from the total distribution of the 7,946 models assessed. Figure 7 illustrates these findings. The suggested quarterly output gap effectively captures the pivotal turning points of the Colombian economy, albeit with heightened uncertainty in estimates during periods marked by distinct idiosyncratic shocks that influence economic growth.

## 5 Conclusions

The estimation of the output gap and potential output in real-time poses significant challenges, particularly for developing economies that are subject to frequent and persistent idiosyncratic shocks. Despite the importance of these measures for monetary policy analysis, there is no clear superior methodology that consistently outperforms others across various evaluation criteria. This study contributes to the existing literature by providing a comprehensive real-time evaluation of different models, considering their stability, inflation forecasting accuracy, and potential output response to economic shocks.

The results confirm the challenge of real-time output gap estimation. Models that perform well in forecasting core inflation during post pandemic period, such as structural vector autoregressive (SVAR) models, tend to exhibit greater instability in their real-time estimates. Conversely, semi-structural models provide more stable estimates but may be less accurate in forecasting inflation. We also found that most of the evaluated potential output figures do not respond adequately to productivity shocks, which is inconsistent with economic theory. However, pooling models seem to capture the potential output response to economic shocks more accurately.

We contribute to the literature by proposing an approach that uses cluster algorithms, based on their performance across various criteria, to provide a better way to nowcast the output gap and potential output. We believe that this approach offers a valuable tool for central

banks to mitigate real-time estimation errors. It is easy to implement and allows central banks to incorporate more meaningful real-time estimates into their monetary policy decisions.

Future research could explore the application of this pooling methodology to other developing economies, as well as investigate the potential benefits of incorporating additional evaluation criteria or alternative model specifications. Furthermore, as new data becomes available and economic conditions evolve, it will be crucial to regularly re-evaluate and refine the real-time estimation techniques to ensure their continued relevance and accuracy for monetary policy decision-making.

## References

- Amador-Torres, J. S. (2017). Finance-neutral potential output: An evaluation in an emerging market monetary policy context. *Economic Systems*, 41(3):389–407.
- Aruoba, S. B. (2008). Data revisions are not well behaved. *Journal of Money, Credit and Banking*, 40(2-3):319–340.
- Balke, N. S., Brown, S. P. A., and Yücel, M. K. (1999). Oil price shocks and the U.S. economy: where does the asymmetry originate? Technical report.
- Barbarino, A., Berge, T. J., Chen, H., and Stella, A. (2020). Which Output Gap Estimates Are Stable in Real Time and Why? Finance and Economics Discussion Series 2020-102, Board of Governors of the Federal Reserve System (U.S.).
- Baumeister, C. and Hamilton, J. D. (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review*, 109(5):1873–1910.
- Blagrave, P., Garcia-Saltos, R., Laxton, D., and Zhang, F. (2015). A simple multivariate filter for estimating potential output. IMF Working Papers 2015/079, International Monetary Fund.
- Bonilla, L., A., F. L., Hermida, D., Lasso, F., and Morales, L. F. (2004). Dinámica reciente del empleo y metodologías para el cálculo de la tasa de desempleo no inflacionaria. Reporte de mercado laboral, Banco de la Republica de Colombia.
- Burlon, L. and D’Imperio, P. (2020). Reliable real-time estimates of the euro-area output gap. *Journal of Macroeconomics*, 64:103191.
- Caicedo-García, E., Hernández-Ortega, R., and Martínez-Cortés, N. (2021). Alivios de pre-

- cios en la pandemia del Covid-19: ejercicio sobre el impacto en la inflación de Colombia. Borradores de Economía 1172, Banco de la Republica de Colombia.
- Camba-Mendez, G. and Rodriguez-Palenzuela, D. (2003). Assessment criteria for output gap estimates. *Economic Modelling*, 20(3):529–562.
- Cerra, V., Fatás, A., and Saxena, S. C. (2023). Hysteresis and Business Cycles. *Journal of Economic Literature*, 61(1):181–225.
- Chen, M. J. and Gornicka, L. (2020). Measuring Output Gap: Is It Worth Your Time? IMF Working Papers 2020/024, International Monetary Fund.
- Clements, M. and Galvão, A. (2012). Improving real-time estimates of output and inflation gaps with multiple-vintage models. *Journal of Business Economic Statistics*, 30(4):554–562.
- Cobo, A. (2005). Output Gap In Colombia: An Eclectic Approach. Borradores de Economía 3310, Banco de la Republica.
- Coe, D. T. and Helpman, E. (1995). International rd spillovers. *European Economic Review*, 39(5):859–887.
- Cohen, D. and Leker, L. (2014). Health and Education: Another Look with the Proper Data. *CEPR Discussion Paper*, 33(9940).
- Coibion, O., Gorodnichenko, Y., and Ulate, M. (2017). The Cyclical Sensitivity in Estimates of Potential Output. NBER Working Papers 23580, National Bureau of Economic Research, Inc.
- Cusinato, R., Minella, A., and Júnior, S. S. P. (2013). Output gap in Brazil: a real-time data analysis. *Empirical Economics*, 44(3):1113–1127.

- De Castro-Valderrama, M., Forero-Alvarado, S., Moreno-Arias, N., and Naranjo-Saldarriaga, S. (2021). Unraveling the exogenous forces behind analysts' macroeconomic forecasts. Technical report.
- Edge, R. M. and Rudd, J. B. (2016). Real-time properties of the federal reserve's output gap. *The Review of Economics and Statistics*, 98(4):785–791.
- Fernald, J. G. (2012). A quarterly, utilization-adjusted series on total factor productivity. Technical report.
- Figuroa, C., Fornero, J., and García, P. (2019). Hindsight vs. Real time measurement of the output gap: Implications for the Phillips curve in the Chilean Case. Working Papers Central Bank of Chile 854, Central Bank of Chile.
- Fleischman, C. A. and Roberts, J. M. (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. Technical report.
- Garratt, A., Lee, K., Mise, E., and Shields, K. (2008). Real-time representations of the output gap. *The Review of Economics and Statistics*, 90(4):792–804.
- Gonzalez, 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 de Colombia.
- González, A., Ocampo, S., Pérez, J., and Rodríguez, D. (2013). Output Gap and Neutral Interest Measures of Colombia. *Monetaria*, 0(2):231–286.
- Gómez-Pineda, J. and Roa-Rozo, J. (2023). A trend-cycle decomposition with hysteresis. Borradores de economía, Banco de la República de Colombia.
- Hofman, A., Mas, M., Aravena, C., and de Guevara, J. F. (2017). Crecimiento económico y

- productividad en latinoamérica. el proyecto la-klems / economic growth and productivity in latin america: La-klems. *El Trimestre Económico*, 84(334 (2)):259–306.
- Jorda, O., Singh, S., and Taylor, A. (2020). The long-run effects of monetary policy. NBER Working Papers 26666, National Bureau of Economic Research, Inc.
- Kangur, M. A., Kirabaeva, K., Natal, J.-M., and Voigts, S. (2019). How Informative Are Real Time Output Gap Estimates in Europe? IMF Working Papers 2019/200, International Monetary Fund.
- Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction To Cluster Analysis*.
- Le Roux, J., Szörfi, B., and Weißler, M. (2022). How higher oil prices could affect euro area potential output. *Economic Bulletin Boxes*, 5.
- Liu, T., Yu, H., and Blair, R. H. (2022). Stability estimation for unsupervised clustering: A review. *WIREs Computational Statistics*, 14(6):e1575.
- Marcellino, M. and Musso, A. (2011). The reliability of real-time estimates of the euro area output gap. *Economic Modelling*, 28(4):1842–1856.
- Nigrinis-Ospina, M. (2004). ¿es lineal la curva de phillips en colombia? Borradores de economía, Banco de la Republica de Colombia.
- Orphanides, A. and van Norden, S. (2002). The Unreliability of Output-Gap Estimates in Real Time. *The Review of Economics and Statistics*, 84(4):569–583.
- Planas, C. and Rossi, A. (2004). Can inflation data improve the real-time reliability of output gap estimates? *Journal of Applied Econometrics*, 19(1):121–133.
- Proietti, T., Musso, A., and Westermann, T. (2007). Estimating potential output and the output gap for the euro area: a model-based production function approach. *Empirical Economics*, 33(1):85–113.

- Restrepo-Ángel, S., Rincón-Castro, H., and Ospina-Tejeiro, J. J. (2022). Multipliers of taxes and public spending in Colombia: SVAR and local projections approaches. *Latin American Journal of Central Banking (previously Monetaria)*, 3(3).
- Rezai, A., Gang, G., Chen, P., and Sr, W. (2008). Productivity shocks in the short and long-run: An intertemporal model and estimation. *Investigación Económica*, 67.
- Rodrik, D. (2014). The past, present, and future of economic growth. *Challenge*, 57(3):5–39.
- Rodríguez, J. I., Perilla, J. R., and Reyes, J. D. (2004). Cálculo del PIB Potencial en Colombia.:1970-2003. Archivos de Economía 2078, Departamento Nacional de Planeación.
- 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-Beltrán, P. M. and Melo-Velandia, L. F. (2013). Combinación de brechas del producto colombiano. Technical report.
- Torres, J. (2008). La estimación de la brecha del producto en colombia. In *Estimación y Uso de Variables no Observables en la Región*, volume 1, chapter 4, pages 103–132. Centro de Estudios Monetarios Latinoamericanos, CEMLA, 1 edition.
- Tribín-Uribe, A. M., Adhvaryu, A., Anzola-Bravo, C., Ávila Montealegre, O., Bonilla-Mejía, L., Castro-Fernández, J. C., Flórez, L. A., Grajales-Olarte, , and Guarín, A. (2020). Migración desde Venezuela en Colombia: caracterización del fenómeno y análisis de los efectos macroeconómicos. *Revista ESPE - Ensayos Sobre Política Económica*, (97):1–74.

# Appendix

## A. Models

### a) Univariate filter

In this model, proposed by Blagrove et al. (2015), the stochastic process for output (annual real GDP) is comprised of three equations, and subject to three types of shocks:

$$\bar{Y}_t = \bar{Y}_{t-1} + g_t + \varepsilon_t^{\bar{Y}} \quad (\text{A.a.1})$$

$$g_t = \theta g^{ss} + (1 - \theta)g_{t-1} + \varepsilon_t^g \quad (\text{A.a.2})$$

$$y_t = \phi_1 y_{t-1} + \phi_2 \varepsilon_t^g - \phi_3 \varepsilon_t^{\bar{Y}} + \varepsilon_t^y \quad (\text{A.a.3})$$

Where,  $y_t$  is the output gap, defined as the difference between the logarithms of GDP ( $Y_t$ ) and potential GDP ( $\bar{Y}_t$ );  $g_t$  is the potential annual growth rate,  $g^{ss}$  is the potential annual growth rate of the steady state;  $\varepsilon_t^{\bar{Y}}$  is a level-shock to potential output,  $\varepsilon_t^g$  is a growth rate shock to potential output, and  $\varepsilon_t^y$  is the output gap shock. First, the level of potential output ( $\bar{Y}_t$ ) develops according to the potential growth ( $g_t$ ) and level-shock term ( $\varepsilon_t^{\bar{Y}}$ ). Second, potential growth is also subject to shocks ( $\varepsilon_t^g$ ), and the adjustment to steady-state growth rate following a shock depends on  $\theta$  parameter. Finally, the output-gap is also subject to shocks ( $\varepsilon_t^y$ ), which effectively cause demand shocks.

### b) Production function

This model assumes that output can be represented by a constant return-to-scale Cobb-Douglas production function, augmented by human capital and where each factor is adjusted

by "the effective levels", i.e., considering the spare capacity. Thus, the output is as follows:

$$Y_t = A_t \times (ICU_t \times K_t)^\alpha \times (L_t \times H_t)^{1-\alpha} \quad (\text{A.b.1})$$

$$L_t = WAP_t \times PR_t \times (1 - U_t) \quad (\text{A.b.2})$$

Where,  $Y_t$  is the observed real GDP,  $A_t$  is the Total Factor Productivity (TFP),  $ICU_t$  is the percentage of installed capacity utilization,  $WAP_t$  is the ratio of working age population to total population,  $PR_t$  is the ratio of labor force and  $PET$ , and  $U_t$  is the urban unemployment rate. The labor factor includes human capital ( $H_t$ ), where the accumulation of  $H_t$  follows (see Inklaar and Timmer, 2013):

$$H = e^{\varphi(S)} \quad (\text{A.b.3})$$

with,

$$\varphi(S) = \begin{cases} 0.134s & \text{if } s < 4 \\ 0.134(4) + 0.101(s - 4) & \text{if } 4 < s < 8 \\ 0.134(4) + 0.101(4) + 0.068(s - 8) & \text{if } s > 8 \end{cases} \quad (\text{A.b.4})$$

The average years of schooling ( $s$ ) are constructed based on the Integrated Household Survey (GEIH by its acronym in Spanish), following Cohen and Leker (2014). The potential output was obtained as follows:

$$Y_t^* = A_t^* \times (NAICU_t \times K_t)^\alpha \times (L_t^* \times H_t)^{1-\alpha} \quad (\text{A.b.5})$$

$$L_t^* = WAP_t \times PR_t^* \times (1 - NAIRU_t) \quad (\text{A.b.6})$$

The long-run trends of total factor productivity ( $A_t^*$ ) and rate of participation ( $PR_t^*$ ) are obtained by smoothing the series using the Hodrick-Prescott filter with priors (HPP).  $NAICU_t$  is the non-accelerating inflation capacity utilization, which is estimated using a Kalman Filter and considers the relationship between the ICU and inflation (see Nigrinis-Ospina (2004)).

The  $NAIRU_t$  is the non-accelerating inflation rate of unemployment estimated by the Banco de la República (see Bonilla et al. (2004)). Finally, we used the series of working-age population and Human Capital used previously.

### c) Blanchard-Quah structural VAR (SVAR)

The Blanchard-Quah structural VAR (SVAR) for a small open economy aims to estimate supply shocks using identification restrictions (short-run, long-run, and sign) and to reconstruct potential output based on these shocks (see Chen and Gornicka (2020)). The straightforward SVAR specification consists of GDP growth ( $\Delta y_t$ ) and the unemployment rate ( $u_t$ ):

$$\begin{bmatrix} 1 & B_{0,12} \\ B_{0,21} & 1 \end{bmatrix} \begin{bmatrix} \Delta y_t \\ u_t \end{bmatrix} = \begin{bmatrix} A_{0,1} \\ A_{0,2} \end{bmatrix} + \begin{bmatrix} B_{1,11} & B_{1,12} \\ B_{1,21} & B_{1,22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} B_{2,11} & B_{2,12} \\ B_{2,21} & B_{2,22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-2} \\ u_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^s \\ \varepsilon_t^d \end{bmatrix} \quad (\text{A.c.1})$$

Where  $A_0$  is a vector of intercepts, and  $B_j$  is a matrix of coefficients for  $j$  lags ( $j = 0, 1, 2$ ). The potential output is obtained from the historical decomposition of shocks, adding supply shocks over time, following the MA representation of SVAR:

$$\bar{y}_t = y_0 + \sum_{i=1}^t \varepsilon_{t-i}^s \quad (\text{A.c.2})$$

where  $y_0$  is real GDP in the initial period. This model is extended to other shocks: monetary policy ( $MP$ ), exchange rate ( $ER$ ), transitory ( $GT$ ) and permanent global ( $GP$ ) shocks. Therefore, historical decomposition is used to construct the dynamic of the observed real GDP as follows:

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_{t-i}^s + \sum_{i=1}^t \varepsilon_{t-i}^d + \sum_{i=1}^t \varepsilon_{t-i}^{MP} + \sum_{i=1}^t \varepsilon_{t-i}^{ER} + \sum_{i=1}^t \varepsilon_{t-i}^{GT} + \sum_{i=1}^t \varepsilon_{t-i}^{GP} \quad (\text{A.c.3})$$

The dynamic of the potential output is obtained from a subset of shocks that affect long-term GDP:

$$\bar{y}_t = y_0 + \sum_{i=1}^t \varepsilon_{t-i}^s + \sum_{i=1}^t \varepsilon_{t-i}^{GP} \quad (\text{A.c.4})$$

## d) Adaptive and Rational Expectation Models

In this section we describe the main equations for the two semi-structural models proposed by González et al. (2013). These models represent a small open economy and differ in how agents form their inflation and real exchange rate expectations, like the dynamics of nominal domestic interest rates.

### Adaptive Expectations Model

The definition of the output gap and potential output are given by the following equations:

$$\tilde{y}_t = y_t - \bar{y}_t \quad (\text{A.d.1})$$

$$\bar{y}_t = \bar{y}_{t-1} + g_t + \varepsilon_t^{\bar{y}} \quad (\text{A.d.2})$$

$$g_t = (1 - \tau)\bar{g}_{ss} + \tau g_{t-1} + \varepsilon_t^g \quad (\text{A.d.3})$$

Furthermore, according to the IS curve, the output gap  $\tilde{y}_t$  is determined by its past values, the real interest rate gap (with  $\bar{r}_{t-1}$  being the neutral interest rate), the exchange rate gap ( $\bar{q}_{t-1}$ ) and demand shocks ( $z_t^y$ ).

$$\tilde{y}_t = \beta_1 \tilde{y}_{t-1} - \beta_3 (r_{t-1} - \bar{r}_{t-1}) + \beta_4 \bar{q}_{t-1} + z_t^y \quad (\text{A.d.4})$$

The Phillips curve for the quarterly annualized inflation is given by:

$$\pi_t = \pi_{t+1|t}^e + \lambda_2 \tilde{y}_{t-1} + \lambda_3 (q_t - q_{t-1}) + z_t^\pi \quad (\text{A.d.5})$$

where  $\pi_{t+1|t}^e$  denotes inflation expectations in  $t$  about  $t+1$ ,  $q_t$  is the level of the real exchange rate and  $z_t^\pi$  represents supply shocks. The dynamics of demand shocks and supply shocks are given by the AR (1) processes:

$$z_t^y = \rho_y z_{t-1}^y + \varepsilon_t^y \quad (\text{A.d.6})$$

$$z_t^\pi = \rho_\pi z_{t-1}^\pi + \varepsilon_t^\pi \quad (\text{A.d.7})$$

The inflation expectations are defined as the average between the target of inflation ( $\bar{\pi}_t$ ) and one lag of annual inflation ( $\pi_{4,t-1}$ ). In addition, in a monetary policy framework with changes in the inflation target, it is determined by the inflation rate in the steady state and by its own lag.

$$\pi_{t+1|t}^e = \lambda_1 \bar{\pi}_t + (1 - \lambda_1) \pi_{4,t-1} + \varepsilon_t^{E\pi} \quad (\text{A.d.8})$$

$$\bar{\pi}_t = (1 - \omega) \bar{\pi}_{ss} + \omega \bar{\pi}_{t-1} + \varepsilon_t^T \quad (\text{A.d.9})$$

$$\pi_{4,t} = \frac{1}{4} (\pi_t + \pi_{t-1} + \pi_{t-2} + \pi_{t-3}) \quad (\text{A.d.10})$$

The dynamics of the domestic real interest rate, foreign real interest rate and real exchange rate are also described. Fisher equation and Uncovered Interest Parity (UIP) condition:

$$r_t = i_t - \pi_{t+1|t}^e \quad (\text{A.d.11})$$

$$r_t - r_t^* = (\bar{r}_t - \bar{r}_t^*) + 4(q_{t+1|t}^e - q_t) + \varepsilon_t^r \quad (\text{A.d.12})$$

The neutral interest rate follows an AR (1) process. In equilibrium, the neutral interest rate  $\bar{r}_t$  depends implicitly on the foreign interest rate, the real exchange rate, and the general state of the country's economy.

$$\bar{r}_t = \rho_r \bar{r}_{t-1} + (1 - \rho_r) \bar{r}_{ss} + \varepsilon_t^{\bar{r}} \quad (\text{A.d.13})$$

The real exchange rate gap ( $\tilde{q}_t$ ) is defined by the difference in its trend ( $\bar{q}_t$ ), which follows a random walk process. It is assumed that the real exchange rate expectation is an average between its trend and its lag.

$$\tilde{q}_t = q_t - \bar{q}_t \quad (\text{A.d.14})$$

$$\bar{q}_t = \bar{q}_{t-1} + \varepsilon_t^q \quad (\text{A.d.15})$$

$$q_{t+1|t}^e = \varphi \bar{q}_t + (1 - \varphi)q_{t-1} \quad (\text{A.d.16})$$

Finally, the nominal interest rate replies to a contemporaneous Taylor rule, which intercept the neutral interest rate ( $\bar{r}_t$ ) plus the inflation target. Furthermore, it is supposed that foreign neutral interest rates and foreign interest rate gaps evolve exogenously following an AR (1) process.

$$i_t = \gamma_1 i_{t-1} + (1 - \gamma_1)(\bar{r}_t + \bar{\pi}_t + \gamma_2(\pi_{4,t} - \bar{\pi}_t) + \gamma_3 \tilde{y}_t) + \varepsilon_t^i \quad (\text{A.d.17})$$

$$\bar{r}_t^* = \rho_{r^*} \bar{r}_{t-1}^* + (1 - \rho_{r^*}) \bar{r}_{ss}^* + \varepsilon_t^{\bar{r}^*} \quad (\text{A.d.18})$$

$$r_t^* - \bar{r}_t^* = k(r_{t-1}^* - \bar{r}_{t-1}^*) + \varepsilon_t^{r^*} \quad (\text{A.d.19})$$

## Rational Expectations Model

The second model introduces changes that modify the IS curve, inflation, and exchange rate expectations:

$$\tilde{y}_t = \beta_1 \tilde{y}_{t-1} + \beta_2 \mathbb{E}_t \{ \tilde{y}_{t+1} \} - \beta_3 (r_{t-1} - \tilde{r}_{t-1}) + \beta_4 \tilde{q}_{t-1} + z_t^y \quad (\text{A.d.20})$$

$$\pi_{t+1|t}^e = \lambda_1 \mathbb{E}_t \{ \pi_{4,t+4} \} + (1 - \lambda_1) \pi_{4,t-1} + \varepsilon_t^{E\pi} \quad (\text{A.d.21})$$

$$q_{t+1|t}^e = \varphi \mathbb{E}_t \{ q_{t+1} \} + (1 - \varphi) q_{t-1} \quad (\text{A.d.22})$$

In addition, the Fisher equation is defined with rational inflation expectations and the Taylor rule is modified to include the value of inflation expectation for the fourth periods, considering

the lagged impact of monetary policy.

$$r_t = i_t - \mathbb{E}_t\{\pi_{t+1}\} \quad (\text{A.d.23})$$

$$i_t = \gamma_1 i_{t-1} + (1 - \gamma_1)(\bar{r}_t + \bar{\pi}_t + \gamma_2 \mathbb{E}_t(\pi_{4,t+4} - \bar{\pi}_{t+4}) + \gamma_3 \tilde{y}_t) + \varepsilon_t^i \quad (\text{A.d.24})$$

All  $\varepsilon^j$  variables, with  $j \in \{y, \pi, r, q, \bar{r}, \bar{r}^*, r^*\}$ , are assumed to be i.i.d. gaussian disturbances with zero mean and constant variance.

### e) Okun's Law

The semi-structural model proposed by Blaggrave et al. (2015) represents a closed economy with a labor market. It draws from the definition of the output gap, growth, and level of potential output. The model also includes a Phillips curve and a set of equations that describe unemployment dynamics:

$$\pi_t = \lambda \pi_{t+1}^e + (1 - \lambda) \pi_{t+1} + \beta y_t + \varepsilon_t^\pi \quad (\text{A.e.1})$$

$$\bar{U}_t = (\tau_4 \bar{U}_{ss} + (1 - \tau_4) \bar{U}_{t-1}) + g_t^{\bar{U}} + \varepsilon_t^{\bar{U}} \quad (\text{A.e.2})$$

$$g_t^{\bar{U}} = (1 - \tau_3) g_{t-1}^{\bar{U}} + \varepsilon_t^{g^{\bar{U}}} \quad (\text{A.e.3})$$

$$u_t = \tau_2 u_{t-1} + \tau_1 y_t + \varepsilon_t^u \quad (\text{A.e.4})$$

$$u_t = \bar{U}_t - U_t \quad (\text{A.e.5})$$

Where  $\bar{U}_t$  is the NAIRU and is subject to shocks ( $\varepsilon_t^{\bar{U}}$ ); the trend of  $\bar{U}_t$  ( $g_t^{\bar{U}}$ ) is given by its lag and a shock ( $\varepsilon_t^{g^{\bar{U}}}$ ); which specifies Okun's law relationship between the unemployment gap ( $u_t$ ) and output gap. Some inflation and GDP growth expectations are added to improve the accuracy of estimates and help to identify shocks.

## f) 4GM Model

The 4GM model proposed by Gonzalez et al. (2020) is a semi-structural framework developed for analyzing monetary policy and macroeconomic trends. It consists of four interrelated blocks. The initial block focuses on analyzing the IS Curve and examining potential output growth. The second block displays Phillips curves for each CPI basket, detailing their respective prices and performing CPI aggregation. The third block discusses the impact of various factors on domestic demand. Lastly, the fourth block elucidates the methodology used to determine the foreign exchange rate.

The IS curve and potential GDP growth in the 4GM model are represented by the following equation:

$$\hat{y}_t = \beta_1 \hat{y}_{t-1} + \beta_2 \mathbb{E}_t \{ \hat{y}_{t+1} \} - \beta_\phi \Phi_t + \beta_{\hat{y}} \hat{y}_t^* + \beta_{\hat{r}p_t^{oil}} \hat{r}p_t^{oil} + n_t^{\hat{y}} \quad (\text{A.f.1})$$

Where  $\hat{y}_t$  denotes the output gap, reflecting the difference between actual ( $y_t$ ) and potential output ( $y_t^*$ ). The equation incorporates the lagged output gap ( $\hat{y}_{t-1}$ ) to capture economic persistence, a forward-looking expectations component ( $\mathbb{E}_t \hat{y}_{t+1}$ ) to account for future output influences, and a term for monetary and fiscal factors that negatively impact the output gap ( $\Phi_t$ ). Additionally, it includes potential output to indicate the sustainable level of output without inflationary pressures, the relative oil price change ( $\hat{r}p_t^{oil}$ ) to highlight the significance of oil revenues in driving economic activity, and an error term to account for unexpected shocks.

In the 4GM model, the Phillips curve is used to analyze inflation dynamics by linking the inflation rates of different Consumer Price Index (CPI) baskets (tradables, non-tradables, food, and regulated goods) to economic variables such as the output gap and the real exchange

rate gap. The Phillips curve for each basket is expressed as:

$$\pi_t^j = \alpha_{\pi^j} \pi_{t-1}^j + (1 - \alpha_{\pi^j}) \mathbb{E}_t \{ \pi_{t+1}^j \} + \alpha_{rmc^j}^{rmc^j} rmc_t^j + \varepsilon_j^{\pi^j}; j = T, NT, F, R \quad (\text{A.f.2})$$

Where  $\pi_t^j$  is the annualized quarterly inflation,  $\pi_{t-1}^j$  reflects inflationary inertia,  $E_t \pi_{t+1}^j$  represents the forward-looking component, and  $\varepsilon_t^{\pi^j}$  is a supply shock. The real marginal cost is given by:

$$rmc_t^j = \begin{cases} \alpha_{\hat{y}}^{rmc^j} \hat{y}_t + (1 - \alpha_{\hat{y}}^{rmc^j}) (\hat{z}_t - \hat{r}p_t^j) & \text{for } j = T, NT \\ \alpha_{\hat{y}}^{rmc^j} \hat{y}_t + (1 - \alpha_{\hat{y}}^{rmc^j}) (\hat{r}p_t^{F*} + \hat{z}_t - \hat{r}p_t^j) & \text{for } j = F \\ \alpha_{\hat{y}}^{rmc^j} \hat{y}_t + (1 - \alpha_{\hat{y}}^{rmc^j}) (\hat{r}p_t^{oil} + \hat{z}_t - \hat{r}p_t^j) & \text{for } j = R \end{cases} \quad (\text{A.f.3})$$

The real marginal cost depends positively on the output gap  $\hat{y}_t$  and the Real Effective Exchange gap  $\hat{z}_t$ , and negatively on the relative price gap specific to each sector  $\hat{r}p_t^j$  also considers the relative price gaps of world food and world oil, which reflect changes in imported food prices or global oil prices. These adjustments in prices ensure that deviations in the relative price gap, whether positive or negative, exert pressure on real marginal costs, inflation, and prices within each sector until equilibrium is restored.

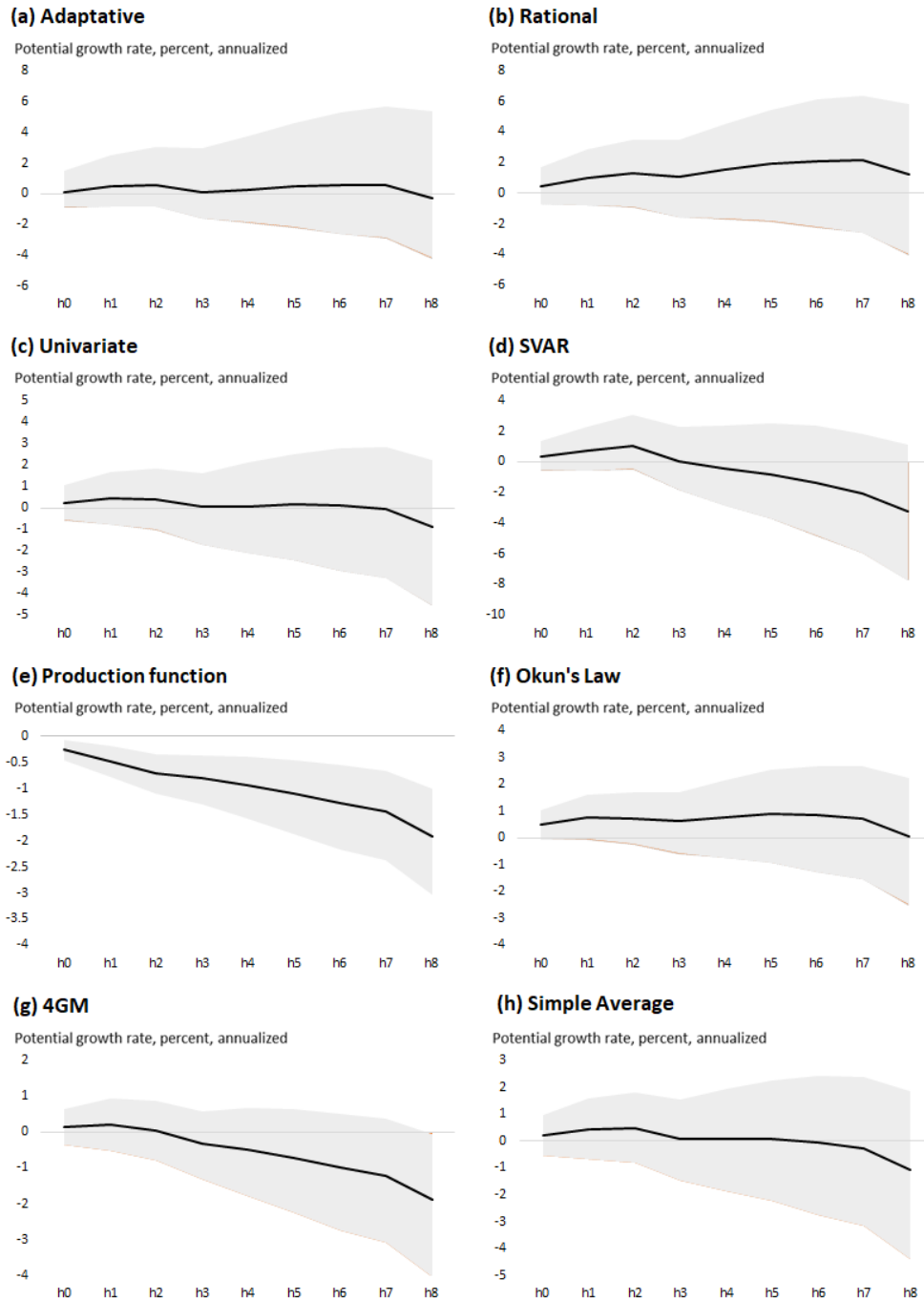
The monetary policy rate is determined based on a reaction function structured in the following form:

$$i_t = \rho_i i_{t-1} + (1 - \rho_i) (\bar{i}_t + \varphi_{\pi} (\mathbb{E}_t \pi_{t+3}^A - \mathbb{E}_t \bar{\pi}_{t+3}^A) + \varphi_{\hat{y}} \hat{y}_t) + \varepsilon_t^i \quad (\text{A.f.4})$$

This rule incorporates the lagged interest rate  $i_{t-1}$  to reflect persistence in monetary policy, the neutral nominal interest rate  $\bar{i}_t$  as a benchmark, and a response to inflation deviations through the term  $\varphi_{\pi} (\mathbb{E}_t \pi_{t+3}^A - \mathbb{E}_t \bar{\pi}_{t+3}^A)$ , which captures the central bank's commitment to price stability. Additionally, it includes a response to the output gap with the coefficient  $\varphi_{\hat{y}}$ , indicating how much the interest rate adjusts based on economic activity.

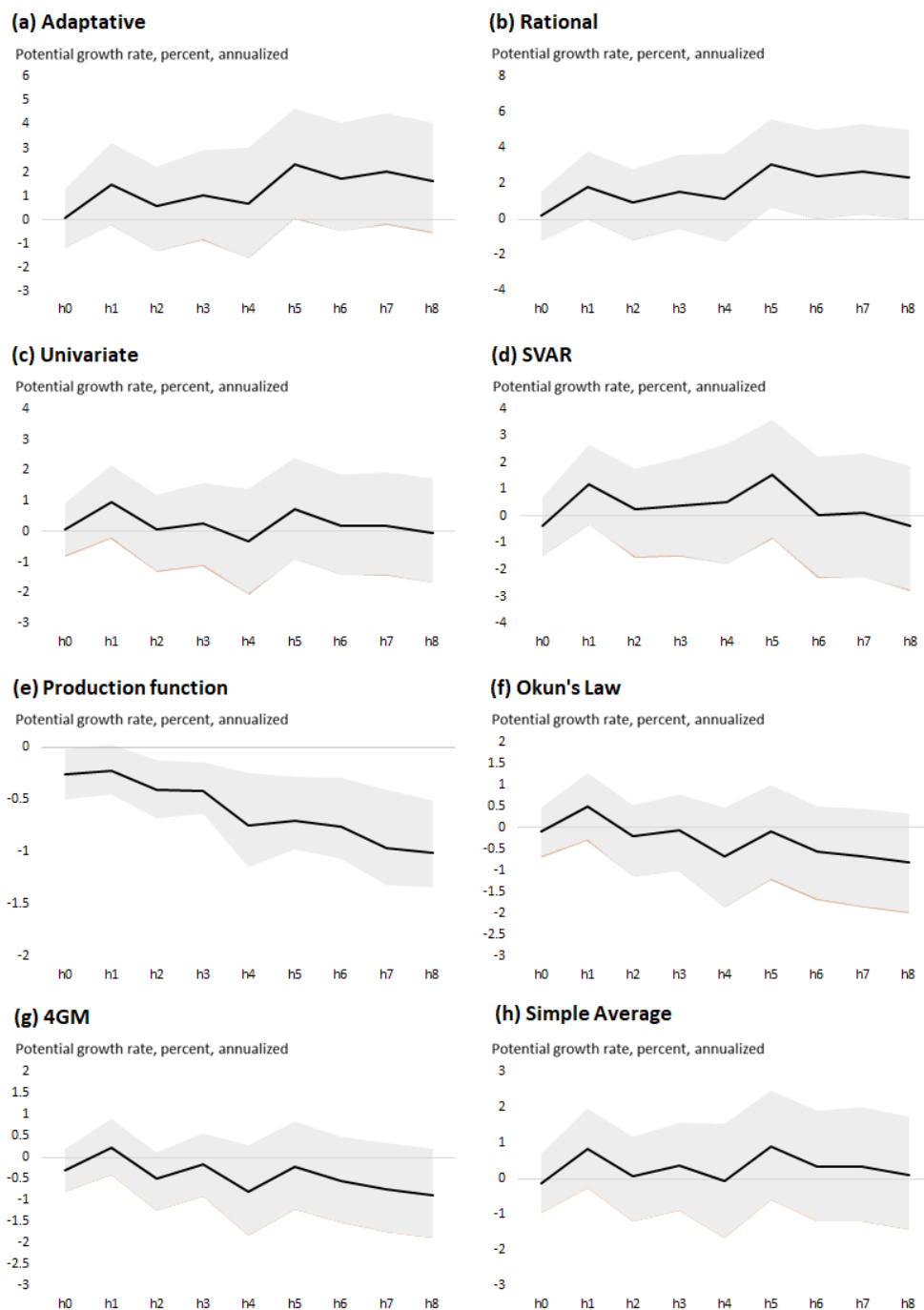
## B. Impulse-response for transitory shocks

Figure B.1: Response of real-time potential output estimates a monetary policy shock.



Note: This figure reports cumulative impulse response functions (IRFs) estimated using equation (4). The estimation sample covers the longest possible period with no missing observations for shocks and potential output. This is why we use estimates associated with the  $N_q$  available vintage. The dashed lines show 66 percent confidence intervals calculated using bootstrapping over the cumulated response. Source: own elaboration.

Figure B.2: Response of real-time potential output estimates a spending fiscal shock.



Note: This figure reports cumulative impulse response functions (IRFs) estimated using equation (4). The estimation sample covers the longest possible period with no missing observations for shocks and potential output. This is why we use estimates associated with the  $N_q$  available vintage. The dashed lines show 66 percent confidence intervals calculated using bootstrapping over the cumulated response. Source: own elaboration.