Output Gap Measurement after COVID for Colombia: Lessons from a Permanent-Transitory Approach *

Camilo Granados †
University of Texas at Dallas

Daniel Parra-Amado ‡
Banco de la República

July 8, 2025

Abstract

We estimate the output gap for the Colombian economy explicitly accounting for the COVID-19 period. Our estimates reveal a significant 20% decline in the output gap but with a faster recovery compared to previous crises. Our empirical strategy follows a two-stage Bayesian vector autoregressive (BSVAR) model where i) a scaling factor in the reduced form of VAR is used to model extreme data, such as those observed around the COVID-19 period, and ii) permanent and transitory shocks are structurally identified. As a result, we obtain that a single structural shock explains the potential GDP, while the remaining shocks within the model are transitory in nature and thus can be used to estimate the output gap. We elaborate on the relative strengths of our method for drawing policy lessons and show that the improved approximation accuracy of our method allows for inflation forecasting gains through the use of Phillips curves, as well as for rule-based policy diagnostics that align more closely with the observed behavior of the Central Bank.

JEL Codes: C11, C51, E3, E32, E37

Key words: Bayesian methods, business cycles, potential output, output gaps, structural estimation

*The authors thank Luca Benati for his guidance and feedback. This research was conducted through a coaching program under the Bilateral Assistance and Capacity Building for Central Banks (BCC), financed by SECO, and the Graduate Institute in Geneva. The authors also appreciate the comments of Nils Goernemann, Cédric Tille and Jean-Paul Renne, as well as those of the participants of the 2024 CEMLA XXIX Meeting of Central Bank Researchers' Network and the 9th International Conference on Applied Theory, Macro and Empirical Finance (AMEF 2025). The views expressed in this paper are solely those of the authors and do not necessarily reflect those of Banco de la República or its board of directors.

†School of Economic, Political and Policy Sciences, University of Texas at Dallas. Richardson, TX 75080, USA. Email: camilo.granados@utdallas.edu

‡Macroeconomic Modeling Department, Banco de la República, Colombia. Email: dparraam@banrep.gov.co

1 Introduction

The COVID-19 pandemic introduced an era of unprecedented economic uncertainty, exposing vulnerabilities in both global and national economies. Severe contractions in 2020, leading, for example, to a 7.0% GDP decline in Colombia, were followed by unexpectedly rapid rebounds in 2021 and 2022, mainly due to substantial vaccine rollouts and reopening efforts. These dramatic shifts in economic performance raised critical questions about the lasting effects on potential GDP and the output gap. Understanding these impacts is essential for policymakers seeking to assess whether the pandemic has permanently altered the productive capacity of economies or merely caused temporary distortions. This task is particularly daunting if we acknowledge that the measurement of the gap is subject to high uncertainty and controversy even during normal times, let alone in times of unprecedented downturns as the one we witnessed recently, an episode that, if something, only adds more noise to the task of disentangling the permanent and transitory forces driving this unobservable variable.

The traditional approach to gauging potential output has consisted of modelling it as driven by supply forces rather than demand ones, and therefore, rendering the latter forces as associated only with the transitory component of output that generates short-run output fluctuations and movements in the output gap (Blanchard and Quah, 1989; Barsky and Sims, 2011; Blinder and Rudd, 2013; Chen and Gornicka, 2020). This view has been revised in recent decades, where the role of demand - in explaining long-term production - has been vindicated due to the observed losses and prolonged crisis characterizing the Global Financial Crisis of 2008, where an anemic demand coupled with a troubled banking sector led to both a current and a future (expected) output deterioration that was substantial enough to shift down the path of potential growth (Fontanari, Palumbo, and Salvatori, 2020). This experience was enough to revive interest on the hysteresis hypothesis where different types of shocks —and not only supply ones— can have a permanent repercussion on output (Blanchard and Summers, 1986; Ball, 2009; Summers, 2015; Benati and Lubik, 2021).[1]

With this renewed view on the role of a variety of shocks in mind, we estimate the output gap for the Colombian economy by using a Bayesian Vector Auto-regressive model (BSVAR) with a relatively large variable set and adjust our estimates by the presence of extreme observations characterizing the COVID pandemic following Granados and Parra-Amado (2024). Our approach is structural, however, given the key limitations in SVAR frameworks, where the identification scheme tends to impose strong assumptions in order to isolate specific shocks such as supply, demand, or

[1] Fornaro and Wolf (2020) model a drop in the productivity growth rate as a consequence of pandemic shock under a new Keynesian model framework depicting a negative endogenous feedback between the current and the expected growth path due to a contraction in demand, which is itself prompted by anticipated future losses that cause a stagnation trap. Similarly, Guerrieri, Lorenzoni, Straub, and Werning (2022) show how a supply shock reducing potential GDP in one sector of the economy could diminish demand in other sectors, which in turn, can cause additional drops long-run production. In this case, the destruction of jobs and businesses exacerbates the initial shock and spreads the recession when the elasticity of substitution between sectors is relatively low, the intertemporal elasticity of substitution is relatively high, and markets are incomplete.

monetary policy shocks.[2] To circumvent this, and building on Granados and Parra-Amado (2024), we employ an agnostic identification as our empirical identification strategy along the lines of Uhlig (2003, 2004)[3]. At the same time, we adjust our estimation framework to allow for the extreme COVID observations as in Lenza and Primiceri (2022). With this strategy, we can handle the extreme volatility of the pandemic episode while still identifying the structural shocks that capture the bulk of the variability of the long-run Colombian GDP in the context of a flexible scheme that purposely avoids imposing strong restrictions (or shocks' labels) or other biases towards specific types of business cycle drivers.

As a first result, we find that a single shock is enough to describe the fluctuations in the Colombian business cycle, which is similar to other studies such as Angeletos et al. (2020) and Brignone and Mazzali (2022). We then build the output gap based on the accumulation of the dynamics implied by the remaining shocks that most closely resemble the cyclical component of output, which ultimately enables us to decompose, additively, the output dynamics into its potential and short-run components. From this exercise, we find that the COVID-19 pandemic caused a sharp decrease in the Colombian gap, reaching a low of -20% in the second quarter of 2020. However, this decline was temporary—unlike the persistent downturns seen in previous recessions— and then the gap bounced back quickly.

To elaborate on the properties of our estimations, we perform an evaluation and a series of policy-oriented exercises where we show that our approach yields more stable and reliable estimates of the potential output and output gap in the presence of abnormally large shocks given that, unlike standard alternatives, our specification prevents atypical events from unduly influencing the estimated long-run dynamics of output. Moreover, in forecasting exercises, the adjusted model outperforms the benchmark methods in predicting inflation at horizons of up to one year ahead, particularly in the turbulent post-COVID period.

These improvements also translate into more coherent policy signals when using the estimated output gap in a Taylor-rule framework. To see this, we explain how our adjusted approach better aligns the implied policy rates with the actual path taken by the central bank —that one can assume takes a rich information set and considerable expertise into account. This is the case both in "normal" times and during moments of substantial uncertainty. For example, unlike other methods, which might produce spurious policy fluctuations around large shocks, our approach delivers a steadier potential output measure and a more plausible inference of the economy's slack, thus offering policymakers a more reliable and economically meaningful guide for decision-making.

[2] Among the most common identification alternatives are contemporaneous effects as Choleski or short restrictions (Sims, 1980; Christiano et al., 1996, 1999), sign restrictions (Uhlig, 2005), long-run restrictions (Blanchard and Quah, 1989), through heteroskedasticity (Rigobon, 2003), instruments for shocks (Romer and Romer, 2004, 2010), and others (Beaudry and Portier, 2006; Gertler and Karadi, 2015; Lütkepohl and Netšunajev, 2017).

[3] This approach consists of finding the structural errors by maximizing the explained fraction of the long-horizon Forecast Error Variance (FEV) of a specific variable and horizon length chosen by the researcher, for example in our case, we consider the GDP and 25-years respectively.

Related literature As mentioned before, it is common to associate potential GDP with the trend component of observed GDP, which is then used to calculate the output gap defined as the deviation of real GDP from its potential level. From a policy perspective, many central banks use the output gap measure as a source or indicator of inflationary pressures, which could condition their monetary stance by assessing the response of observed GDP to shocks and how those fluctuations could lead to deviations from the optimal path of output. Likewise, the output gap indicates the economy's position along the business cycle, which enables us to evaluate how close or far the current fiscal deficit is from the one considered as neutral.

From an empirical perspective, potential GDP, as a latent variable from which the output gap is derived, is typically estimated using statistical and economic models. As noted by Kiley (2013), the underlying gap concept varies across estimation methods and can broadly be categorized into three types of approaches: statistical filters, the production function approaches, and microfounded models-based setups. From the latter, the New Keynesian framework is typically used. Similarly, Álvarez and Gómez-Loscos (2018) classify gap estimation methods based on factors such as complexity, decision variables, and modeling type (e.g., univariate or multivariate, for details, see Appendix A).

A popular tool to measure the output gap is the production function approach, which estimates the output deviation from a level consistent with the current level of technology, and the available capital, and labor resources (CBO, 2001, 2014; Havik et al., 2014). From an empirical perspective, these measures are usually associated with the notion of a trend GDP where the gap thereby defined as cyclical output deviations from its trend value (e.g., filtering approaches such as Beveridge and Nelson (1981), Hodrick and Prescott (1997) (HP filter), Baxter and King (1999), Christiano and Fitzgerald (2003) (CF filter), among others). In line with this type of approach, and given its broad use by practitioners, the potential output is often conceived as a smooth trend (Basu and Fernald, 2009), which implies assuming that supply-side factors do not have substantial high-frequency fluctuations, rendering the output gap as driven mainly by demand shocks. The latter assumption can be considered too strong, in particular given the recent crises experiences (Cochrane, 1990).[4]

A similar view also took traction in studies based on multivariate models. A seminal example is in Blanchard and Quah (1989) who identify two structural shocks from an SVAR model with long-run restrictions. Among these, one shock is allowed to have permanent effects on output while the other only affects production transitorily. At the same time, the former shock is interpreted as supply-sided or supply-driven, and the latter as shaped by demand forces. Similarly, Cochrane (1994) explores the decomposition of permanent and transitory components in GNP by analyzing consumption behavior. This approach provides a useful measure of the trend in GNP, grounded in the permanent income hypothesis. According to the author, under the assumption of random walk dynamics in consumption, if consumption remains unchanged, consumers interpret any

[4] Cochrane (1990) compares univariate and multivariate filters and finds much larger cyclical components in the latter ones where including variable as consumption enable to assess whether shocks to GNP are persistent.

contemporaneous shock or fluctuation in GNP as transitory.

In a more closely related fashion to our approach, Uhlig (2003, 2004) identify the permanent and transitory shocks in US economy by maximizing the Forecast Error Variance (FEV) of real GNP over five-years horizon. As a result, it is possible to explain around 90% of the GNP variability by using the largest two shocks of the FEV. Although this method is entirely empirical and does not contemplate structural restrictions from economic theory in the identification process, the author suggests that the first shock resembles a productivity shock while the second could be attributed to inflationary forces. Following this identification scheme, Angeletos, Collard, and Dellas (2020) find that a single shock can be used to explain macroeconomic fluctuations, which they refer to as a Main Business Cycle shock (MBC). Brignone and Mazzali (2022) expand Angeletos et al. (2020) in a high dimensional framework by using dynamic factor model over a set of 136 variables. They found the the economy seems to be mainly driven in the long-run by just one supply-side permanent shock. In this aspect, our research follows this branch of the literature in which the identification mechanism is agnostic and data-driven and the business cycle is explained by a single permanent shock.

In recent years, empirical evidence suggests that the use of multivariate models produces more reasonable estimates of the output gap, since they include a broad set of relevant information to capture the reduced-form dynamics of macroeconomic variables and identify aggregate shocks (Jarociński and Lenza, 2018; Morley and Wong, 2020; Barigozzi and Luciani, 2021; Furlanetto et al., 2022, among others). In particular, Morley and Wong (2020) apply a Beveridge-Nelson type of decomposition based on a large BVAR to estimate the U.S. output gap by taking into account the FEV of the GDP one-step ahead. Similar to several of the aforementioned studies, and to this paper, the authors mitigate the possibility of over-fitting output growth by Bayesian shrinkage.

Our paper also relates to studies that are adjusted with the specific goal of accounting for the atypical COVID period data, on this front, Berger et al. (2023) build a nowcast model to US output gap by combining a Beveridge-Nelson decomposition and a mixed-frequency Bayesian VAR (Ghysels, 2016; Cimadomo et al., 2022). They estimate a steep decline of the US output gap in 2020Q2 between -10.1% and -8.3%, conditional on the set of information from April to June 2020. They show that monthly indicators such as consumer sentiment, credit risk spread, and the unemployment rate are useful for nowcasting the output gap in real time. Similar results for the German economy have been reported in Berger and Ochsner (2022). They find that the output gap dropped from 1.17 to -8.78% due to the COVID disruption, which, in contrast to the observed GDP, only marginally affected potential output, thereby generating a massive decline in the output gap. These articles take into account a large set of information to determine in real time the impacts of the COVID episode by using monthly variables, but do not explicitly model the COVID data—or atypical observations within the context of their frameworks. Conversely, Morley, Palenzuela, Sun, and Wong (2022) propose a two-step approach where the authors use a hybrid of Bayesian and maximum likelihood estimation (MLE) together with the adjustments developed by Lenza and

Primiceri (2022) to account for the COVID data, and secondly, they decompose the GDP growth following Beveridge and Nelson (1981) in order to fit the GDP trend-permanent component of the Eurozone. Our research is similar in that we use a two-stage method, but differs in that instead of using the above decomposition, we rely on Uhlig (2003, 2004) as our structural identification scheme.

This article is organized as follows. In the next section, we describe the methodology we follow as proposed in Granados and Parra-Amado (2024), which itself borrows from Lenza and Primiceri (2022) to adjust the model by the presence of the pandemic and achieves identification (of the permanent and transitory components of output) in the sense of Uhlig (2003, 2004). In Section 3, we introduce the data and our main estimates. In Section 4, we discuss a number of exercises for evaluating the model for both estimation and policy applications, and then we conclude.

2 Methodology

To calculate the output gap, we employ a permanent-transitory decomposition approach as outlined in Granados and Parra-Amado (2024), which is divided into two stages. First, we use a reduced-form Vector Autoregressive (VAR) model that includes a scale factor to account for COVID-19-induced volatility. As in Lenza and Primiceri (2022), the VAR setup includes a s_t term as follows:

$$Y_t = B_0 + B_1 Y_{t-1} + B_2 Y_{t-2} + \ldots + B_p Y_{t-p} + s_t u_t, \qquad u_t \sim N(0, \Sigma), \tag{1}$$

where s_t is set to one in the sample period before the COVID-19 shock (t^*), and subsequently, latent parameters ($\theta \equiv [\bar{s_0}, \bar{s_1}, \bar{s_2}, \rho]$) are estimated to capture the increased uncertainty during the COVID-19 period, which then diminishes as the economy recovers. As usual in this framework, we fit the scale factors for the first three quarters in the COVID period starting at the first quarter of 2020 ($t^* = 2020Q1$) and a decay parameter ρ for the next quarters. Thus, the unobserved parameters are $s_{t^*} = \bar{s_0}$, $s_{t^*+1} = \bar{s_1}$, $s_{t^*+2} = \bar{s_2}$ and $s_{t^*+j} = 1 + (\bar{s_2} - 1)\rho^{j-2}$ for $j \geq 3$.

Equation (1) can be estimated as in Giannone, Lenza, and Primiceri (2015) by assuming the prior distributions of the coefficients to be conjugate Normal-Inverse Wishart and by including the scale factors into the posterior hyperparameters. They are jointly estimated using Bayesian techniques by drawing those parameters in a Metropolis-Hastings procedure. The priors of β and Σ can be described as:

$$\Sigma \sim IW(\Psi, d),$$
$$\beta | \Sigma \sim N(b, \Sigma \otimes \Omega),$$

where $\beta \equiv vec([B_0, B_1, \ldots, B_p]')$ and $\gamma \equiv (\Psi, d, b \text{ and } \Omega)$ are the hyperparameter vectors. The prior for θ is defined analogously as a Normal distribution centered at 1.[5]

The posterior of θ is used to capture the dynamics of s_t, which is jointly evaluated with the posterior of γ as proposed by Lenza and Primiceri (2022):

$$p(\gamma, \theta|Y) \propto p(Y|\gamma, \theta) \cdot p(\gamma, \theta).$$

Second, we recast the model of equation (1) into an SVAR form by identifying the main shock explaining the Colombian business cycle in the long run, which is done, along the lines of Uhlig (2003, 2004), that is, by maximizing the explained fraction of the total FEV of the GDP at a long-run horizon (e.g. 15 or 25 years ahead). Recall that the structural errors (ε_t) are related to the reduced-form errors (u_t) in equation (1) through the impact matrix (A_0) that establishes $u_t = A_0\varepsilon_t$ and $\Sigma = A_0A_0'$. Uhlig (2003, 2004) use an alternative matrix $\ddot{A}_0$ which can be found by using an orthonormal matrix Q where $A_0 = \ddot{A}_0Q$ and $QQ' = I$ through maximization of the following expression:

$$q_1 = \text{argmax}\, q_1'Mq_1 \equiv q_1' \sum_{h=0}^{k} \ddot{A}_0{}' C_h'(e_je_j')C_h\ddot{A}_0q_1$$

$$\text{subject to} \qquad q_1'q_1 = 1,$$

where q_1 is a column of Q that explains the k-step-ahead forecast error of the j-th variable in Y_t (in our case, the log of GDP), whose variance is given by M. Simultaneously, as shown in Uhlig (2003), q_1 is the eigenvector associated with the largest eigenvalue of the matrix M. e_j is a selector vector with zeros everywhere and a 1 in the j-th position, and C_h is a component of the long-run impact matrix of the VAR associated to the horizon h.[6] The constraint guarantees that q_1 is a unit-length column vector that belongs to an orthonormal matrix.

Notably, the method recovers all eigenvalues and eigenvectors of M, which, given the decomposition method, are ordered from higher to lower fractions, explained by the FEV of the target variable. Thus, we can verify whether one or more shocks explain a larger component of the long-run FEV of the GDP. In other words, this approach identifies the shock that best explains the long-run component of the target variable. This is done in the following section.[7]

[5] This recognizes the non-stationary nature of our data. See Giannone et al. (2015) for details.

[6] Note that $C(L) = I + C_1L + C_2L^2 + C_3L^3 \cdots + C_hL^h + \ldots$ and the moving average representation of the model is given by $Y_t = B(L)^{-1}u_t = C(L)u_t$.

[7] It should be noted that our main concern is the clear identification of a structural set of exogenous shocks driving long-run fluctuations of output. Beyond this, we are not particularly focused on obtaining a set of exogenous and interpretable short-run shocks as opposed to other SVAR studies (e.g., Blanchard and Quah, 1989; Forbes et al., 2018, among others). It is due to these features that our identification can be labeled as agnostic.

3 Results

3.1 Data and empirical strategy

We set a nine-variable B-SVAR in levels for the period 2002Q3 to 2023Q4 using Colombian data.[8] The variables included are GDP, household consumption (CON), government consumption (GOV), investment (INV), consumer price index (CPI), exchange rate (ITCR), interbank interest rate (ITB), Brent oil price (OIL) and unemployment rate (UNR). The domestic account variables (first five in the VAR) and the unemployement rate were obtained from the Colombian National Statistics Department (DANE), the exchange rate and interest rate from the Central Bank of Colombia (Banco de la República), the oil price from Bloomberg.[9]

We choose two lags ($p = 2$) based on the Bayesian and Hannan–Quinn information criteria, and then estimate the VAR in levels via the hierarchical conjugate Normal–inverse–Wishart framework of Giannone, Lenza, and Primiceri (2015). In this setup we combine the three "classic" conjugate priors—(i) the Minnesota prior, (ii) the sum-of-coefficients prior, and (iii) the dummy-initial-observation prior—and treat the overall shrinkage coefficient λ as a model hyperparameter (i.e. it is estimated jointly with the VAR coefficients). As a result, the estimation automatically calibrates the degree of shrinkage—tighter when the number of parameters is large relative to the sample, looser when the data are more informative—without any ad hoc tuning.[10] We ran 20,000 draws and kept half for estimation after burn-in. In addition, we explicitly modelled the COVID-19 extreme observations, as in Lenza and Primiceri (2022). From this first stage, we obtain a reduced-form VAR that has already been adjusted by the scale factor st and incorporates the pandemic shock.

In the second stage, we identified the impact of the matrix of the SVAR by maximizing the explained share of the forecast variance error of the GDP for a 25-year horizon, as in Uhlig (2003, 2004). As part of the procedure, we restrict that the share of the FEV one step ahead of consumption explained by the first structural error, or (the majority) of the permanent component, is larger than that of the output, and for the latter to be larger than that of the investment. As explained by Cochrane (1994) and King, Plosser, Stock, and Watson (1987), this accounts for the fact that consumption is more closely aligned to the permanent component of GDP, while investment should reflect its most volatile and transitory components. After verifying these restrictions and keeping the draws that comply with them, we conducted PT decomposition and computed the permanent (and transitory) output component.[11]

[8] We assess the presence of a unit root in each of the series, as well as the presence of cointegration. It was found that the series are cointegrated, which justifies the choice to model the VAR in levels despite the individual unit root in each series. These tests can be provided upon request to the authors.

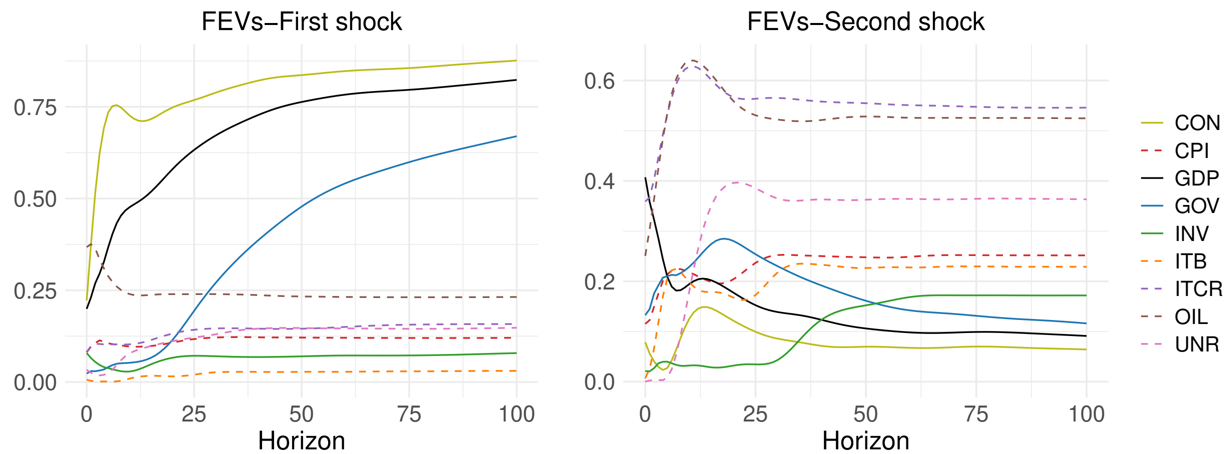
[9] We evaluated a range of alternative specifications—e.g. multiple commodity prices, a principal component of commodity indices, export and import price series, and terms-of-trade measures—but found no meaningful gains from increasing the variable set relative to our baseline.

[10] This parameter will indicate the relative importance of the prior and likelihood for the associated posterior and, therefore, should be on the lower side when the sample is deemed informative.

[11] As a check, we increased the number of draws to 100,000 and obtained similar results.

As aforementioned, the decomposition and resulting impact matrix already consider the ordering of structural shocks according to their share of the explained variance of the target variable. This can be verified in Figure 1, where we can see that only the first structural error is necessary to account for approximately 82% of the long-run (permanent) component of the GDP. Concurrently, the next most important shock explains the GDP's FEV in the short run, which is more resembling of the transitory output component. In light of these results, we compute the output gap based on the second to ninth structural shocks and use only the first one to recover the potential GDP.[12] On a related point, it should also be noted that the first structural error will explain the majority of the long-run FEV of the GDP (target variable), but not necessarily the largest share of the FEV for other variables.

Figure 1: Contribution of FEV explanation over each variable by the two shocks explaining the highest share of long-run Colombian GDP FEV



Note: The Figure shows the Forecast Error Variance (FEV) explained by the two structural shocks with the largest explained share for the long-term GDP. Given a single shock explains almost 82% of GDP for large horizons (left-panel), it is associated as the main driver of the permanent component of output. In contrast, the second highest (right panel) and remaining shocks are instead considered as driving the transitory component of output.

3.2 Baseline Results

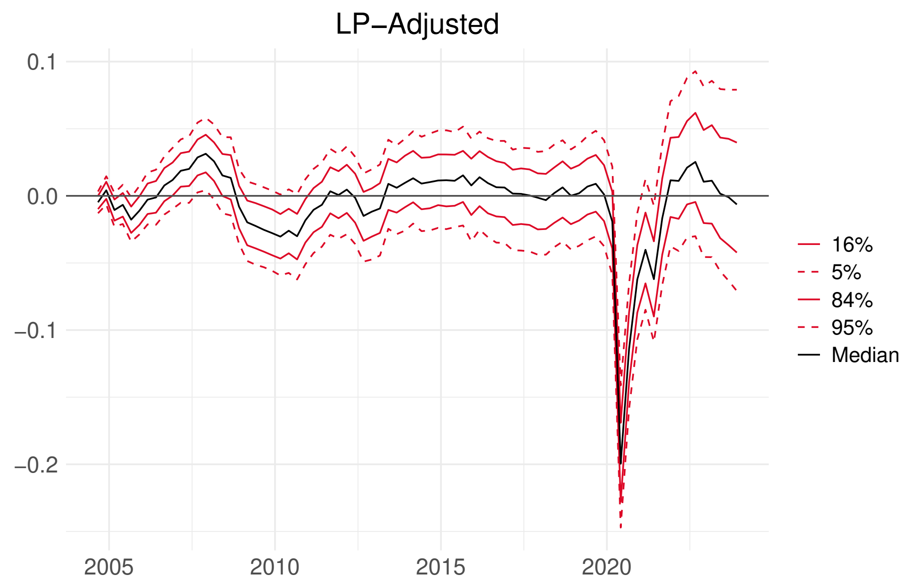
Figure 2 shows the output gap for the Colombian economy obtained from our proposed BSVAR, using a combined PT decomposition and a scale factor adjustment to include and adjust for the COVID-19 period observations (LP-Adjusted). Our estimate of the output gap begins the decade in negative territory following the severe mortgage crisis of 1999.[13] Other episodes of slight deterioration were observed in 2015-2016, driven by the drop in international oil prices and its

[12]Analogously, the potential GDP can be obtained as the original series minus the transitory component.

[13]For the Colombian case, our main downturns of reference are the 1999 and Global Financial Crisis of 2008 (GFC). The former is one of the worst recessions to date, while the latter is relatively mild compared with the dynamics of advanced economies.

impact on terms of trade.[14] In general, these dynamics are aligned with previous business cycle dating exercises carried out for Colombia (e.g., Alfonso et al., 2013), despite the high uncertainty one may expect to see around these estimates also reflected in the amplitude of the percentile intervals shown in the Figure 2.

Figure 2: Baseline results: Output gap for Colombia



Notes: The solid black line represents the median estimates. The solid and dotted red lines represent the percentiles of 5%, 95%, 16%, and 84%, respectively.

During the COVID-19 pandemic, the gap underwent a steep decline (-20%) in the second quarter of 2020; however, unlike in the 1999 recession, the downturn was not persistent. Instead, it bounced back in the following quarters. As in most economies, the decrease is primarily explained by lockdown measures, while the gradual reopening of the economy elicited the subsequent recovery. In the late nineties, the potential GDP growth went negative, contrasting with the pandemic when it only decelerated (from 3.0% in 2019 to 2.3% in 2020). The recovery paths are also in contrast with the potential output trending upward and gap closing by 2022Q1. Subsequently, in 2023, a positive output gap was observed, which, along with the effects of global inflation, created a challenging scenario for the monetary stance.

3.2.1 Outlier observations around the COVID pandemic

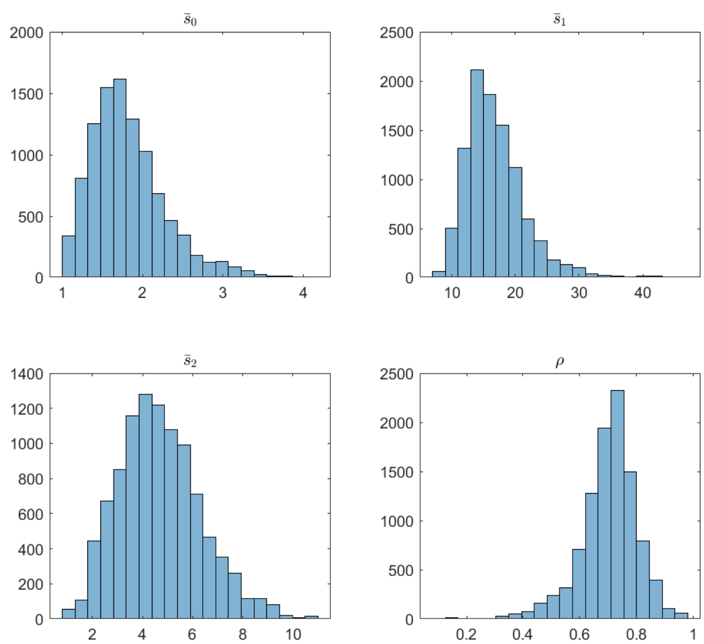
Given that our primary concern is to study the adjustment of potential output estimates to drastic magnitude shocks, such as those observed in the COVID-19 outbreak, verifying the estimates of the scale factors generated by our baseline estimates can be insightful. Principally, if scaling is

[14] Colombia's main export commodity is crude oil and related products.

irrelevant, the posterior estimates should suggest $\bar{s}_0 = \bar{s}_1 = \bar{s}_2 = 1$; otherwise, they should be sizeable. We estimate these parameters, as in Lenza and Primiceri (2022), and present our estimated scale factors around the onset of the pandemic in Figure 3.

The parameters posteriors are drawn based on a Metropolis-Hastings algorithm with a potentially Minnesota Prior as described before (and along the lines of Giannone et al., 2015). Thus, we estimated the scaling factors together with other hyperparameters in a hierarchical structure.[15] The resulting posteriors for $\bar{s}_0, \bar{s}_1, \bar{s}_2$ peak around 1.7, 15.7, and 4.5, respectively, indicating that, in effect, it is relevant for this sample to scale up the errors around the COVID-19 observations to account for the steep increase in volatility of that period, but that may not characterise its data-generating process, nor should it drastically influence the BVAR estimates. Nonetheless, the posterior of the decay coefficient (ρ) peaks around 0.73, which, together with $\bar{s}_2$, implies that the volatility scale factor falls by a third after 2020Q3 and then non-linearly towards one.[16]

Figure 3: Posterior distribution of the volatility scaling factors



At the same time, we can gauge the uncertainty around these scale factors and their dynamics by looking at them across a longer timeframe, as in Figure 4. As mentioned, the peak is located in the second quarter of 2020, the moment of greatest uncertainty during the shock. It is important

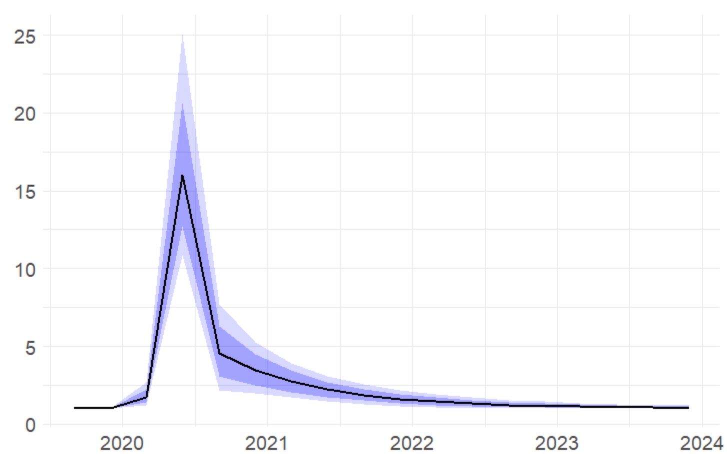
[15]We tested two alternatives, explicitly modeling both two and four quarters, and the results were similar and not statistically different with respect to these changes. In other words, we estimated one model using only $\bar{s}_0$, $\bar{s}_1$, ρ, and another using $\bar{s}_0$, $\bar{s}_1$, $\bar{s}_2$, $\bar{s}_3$, and ρ. Although the estimated median values change slightly, the dynamics of the scale factors are similar across both versions: the peak occurs in the second quarter of 2020 and then gradually fades, returning to levels considered normal by 2023. Regarding output gap estimates, these model versions do not show statistically significant differences.

[16]We also obtain the posterior for the shrinkage parameter of Minnesota prior (λ), depicting a mode around 0.18.

to note that the pandemic lockdowns began between February and March 2020, but the effects on production were more significant in the following months. On the other hand, the scale factor converges back to one by the fourth quarter of 2022. These results suggest that, as per our estimations, there was a substantial uncertainty taking place around the initial stages of the shutdown that persisted for about two years. Behind this, a plausible explanation is that production collapsed initially and for some months as the vaccine was under development. Afterwards, and once the vaccines became available (and a social immunity ensued), it was possible to resume production at near pre-pandemic levels. In terms of observed GDP, the low base effect prompted by these swings generated the temporary high output volatility hinted by the aforementioned figures (figures 3 and 4).

In support of this story, we can also compare the scale factor with other relevant pandemic-related indicators. For this, we report a figure (Figure 9 in Appendix B) with the scale factor and two COVID indexes for Colombia, namely the estimated number of deaths, and a contingency and health index. In the plot we can see that, even if the scale factor did not fully capture the effects of additional waves of infections observed in 2021 and 2022, it still comoved with the increases in the indexes (particularly the contingency one) around the early stages of the lockdown.

Figure 4: Scale factors over time

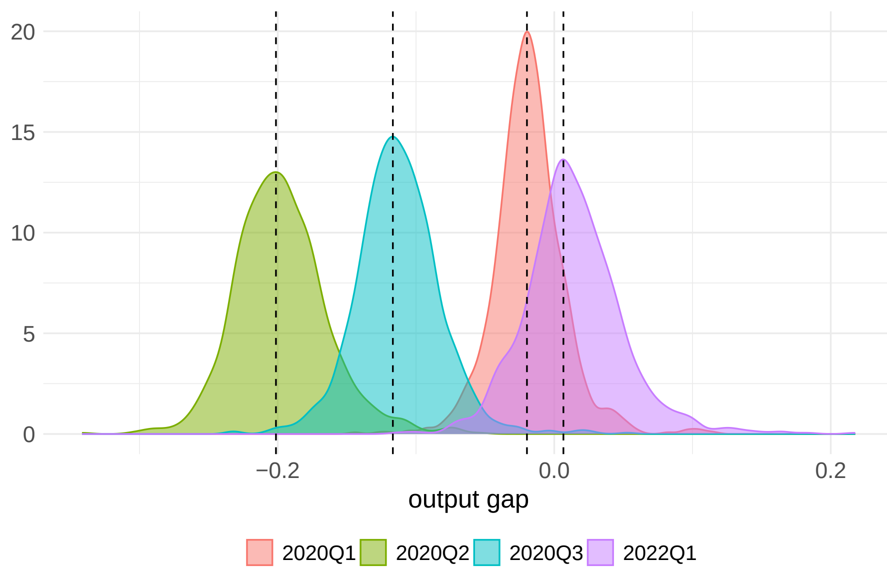


Notes: The solid black line represents the median estimates. The shade areas represent the percentiles of 5%, 95%, 16%, and 84%, respectively.

To further illustrate the impact of the COVID-19 shock on the output gap, we can depict the distributions of the draw estimates for dates around the episode, as shown in Figure 5. We reveal the quarter of the shock (2020Q1), the subsequent two quarters, and the first quarter of 2022 as a reference for a date when the potential output dynamics are, in principle, back to normal (here implicitly recognize the transitory nature of the pandemic shock).

As we can see in the figure, the distribution of the gap has a large shift to the left, implying that the potential GDP was not largely affected by the downturn (and instead, the gap lowered in line with the observed GDP). In addition, the distribution spread increased, reflecting an increase in uncertainty around the estimate during the pandemic. Afterwards, we observe the distribution shifts back to pre-COVID-19 levels, although it still reflects increased volatility. In summary, we can see that the impact on the mean gap was transitory, although a somewhat larger uncertainty remains. Nonetheless, the larger uncertainty is approximately one percentage point higher than before, rather than orders of magnitude larger, as may be induced by a model without a scale factor adjustment for the COVID-19 downturn.

Figure 5: Distribution of the output gap estimation during COVID-19 shock and 2022Q1.



4 Evaluation of the framework and policy implications

Evaluating the performance of output gap estimation methods poses significant challenges, primarily because the potential GDP, our target variable, remains unobserved in real-world data. This lack of a definitive benchmark complicates the direct comparison of various estimation techniques. Due to this, studies usually follow either of two approaches: first, they leverage simulations (usually based on economic models) to generate a target to estimate, or secondly, they assess the usefulness of competing output gap estimations as an input for policy.

In this section, we carry out both types of exercises. First, we conduct a simulation-based evaluation framework as in Granados and Parra-Amado (2024) but in a single-country context. Then, we compare the relative performance of our method both as an input for predicting inflation using a Phillips Curve, and as an input for generating policy guidelines based on a Taylor-rule setup.

4.1 Evaluation based on model simulations

To evaluate the method we verify the co-movement of estimations of a competing set of methods —including our baseline— and an output gap target generated within the context of an general equilibrium model. This implies conducting a Monte Carlo simulation, where within each epoch, a set of economic variables and potential output is simulated based on the economic model, the competing methods are estimated, and the fitted gaps are compared to the target.

Analogously to Granados and Parra-Amado (2024), we use a standard three-equation New Keynesian DSGE model along the lines of Benati (2008), but where the (potential) output is assumed to have a unit root component that behaves as a random walk with a drift:

$$y_t^P = \delta + y_{t-1}^P + v_t, \qquad v_t \sim WN(0, \sigma_v^2) \tag{2}$$

The log-linearized version of the model consists of the following equations:

$$\pi_t = \frac{\beta}{1+\alpha\beta}\pi_{t+1|t} + \frac{\alpha}{1+\alpha\beta}\pi_{t-1} - \kappa\hat{y}_t + u_t, \qquad u_t \sim WN(0, \sigma_u^2) \tag{3}$$

$$\hat{y}_t = \gamma\hat{y}_{t+1|t} + (1-\gamma)\hat{y}_{t-1} - \sigma^{-1}(R_t - \pi_{t+1|t}) - (1-\gamma)\Delta y_t^P \tag{4}$$

$$R_t = \rho R_{t-1} + (1-\rho)\left[\phi_\pi \pi_t + \phi_y \hat{y}_t\right] + \epsilon_{R,t}, \qquad \epsilon_{R,t} \sim WN(0, \sigma_R^2) \tag{5}$$

Here, π_t represents inflation, R_t the nominal interest rate, and $\hat{y}_t = \ln(Y_t/Y_t^P)$ denotes the output gap, which is the deviation of actual output from potential output. The model parameters, $\Theta = \{\sigma_R^2, \sigma_u^2, \sigma_v^2, \kappa, \sigma, \alpha, \gamma, \rho, \phi_\pi, \phi_y\}$, are estimated using Bayesian methods tailored to the specific characteristics of the single-country dataset. The posterior distributions of these parameters are obtained via a Random-Walk Metropolis-Hastings algorithm as in An and Schorfheide (2007), ensuring robust parameter estimation aligned with the model's likelihood. The result of the estimation for the Colombian case is shown in Table 1.

Table 1: Prior and Posterior modes and standard deviations for the parameters

		Prior		Posterior	
Parameter	Prior Density	Mode	Standard Deviation	Mode	68% coverage percentiles
σ_R^2	Inverse Gamma	0.01	0.01	0.004	[0.0013, 0.0017]
σ_u^2	Inverse Gamma	0.01	0.01	0.004	[0.0013, 0.0018]
σ_v^2	Inverse Gamma	0.01	0.01	0.004	[0.0030, 0.0044]
κ	Gamma	0.10	0.10	0.058	[0.0355, 0.0836]
σ	Gamma	1	2	24.611	[16.9698, 24.9687]
α	Beta	0.90	0.05	0.906	[0.8266, 0.9301]
γ	Beta	0.50	0.25	0.732	[0.5239, 0.5480]
ρ	Beta	0.7500	0.10	0.742	[0.6260, 0.7297]
ϕ_π	Gamma	1.50	0.25	1.751	[1.7818, 2.2218]
ϕ_y	Gamma	0.50	0.15	0.466	[0.3700, 0.6076]

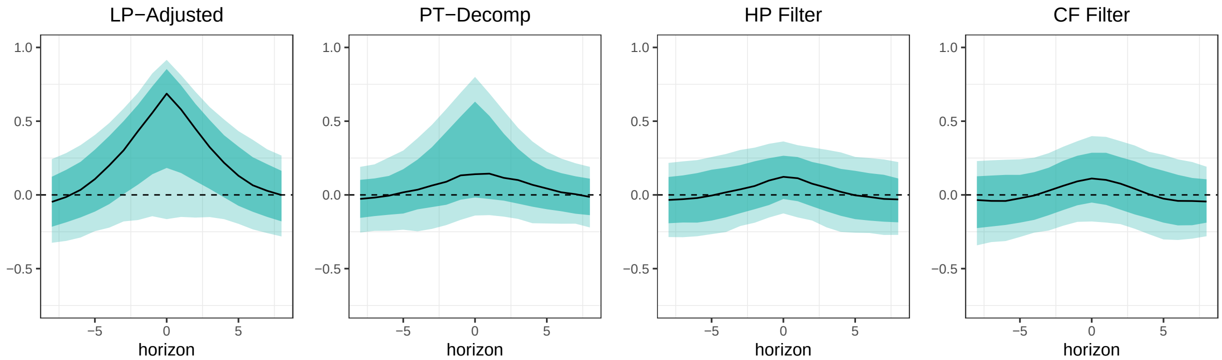
Note: The acceptance ratio of the Metropolis algorithm is 0.219.

Evaluation method Based on the estimated New Keynesian model, a Monte Carlo simulation is carried out, where, in each iteration, a sample (33 years long) of the model variables is simulated, and a corresponding output gap is obtained. The simulated economic variables are used as inputs for a set of competing econometric methods that estimate the output gap of the simulated model. At the last step of each iteration, the cross-correlation between each econometric estimate of the output gap and the simulated output gap (target) of the model is calculated and recorded.

The output gap frameworks compared are: (i) our proposal, a Permanent-transitory decomposition with a Primiceri-Lenza type adjustment for large shocks episodes with known date (LP-Adjusted), (ii) a Permanent-Transitory decomposition via a BSVAR (PT), (iii) a Christiano and Fitzgerald (2003) Band Pass type of filter (CF), and (iv) a Hodrick and Prescott (1997) filter (HP). The latter two filters are more frequent and widely available methods of estimation of the potential output, while the Permanent-Transitory decomposition is relatively more complex as it aims to achieve a structural identification for an SVAR based on the long-run forecasts of the output. Finally, our proposed method combines the structural long-run forecast identification method with an adjustment of the estimation to account for the presence of very large shocks whose date is known.

The results are reported in Figure 6. We see that the HP and CF filters have a similarly low performance, with the former barely outperforming the latter for low lag orders of correlations. This lack of performance is not surprising given these filters have end-of-sample biases, and thus allow the simulated shocks (put at the end of the sample to mimic the COVID episode) to disproportionately affect the associated output gap estimations for the whole sample. Conversely, the structural method with an adjustment —at a known date— generates substantially better estimations of the output gap, yielding larger cross-correlations. This result aligns with the findings reported in Granados and Parra-Amado (2024) for the G7 countries.

Figure 6: Cross-correlation between the output gap estimates and their simulated target



Note: median (black), 68% coverage (darker range) and 90% coverage percentiles of the cross-correlations between each output gap estimate of each method and the simulated output gap of the economic model.

4.2 Evaluation as an input for policy exercises

We can alternatively acknowledge that we do not know our estimation target, but still would like to evaluate how the competing estimates perform as inputs for other relevant measures used in guiding policy. For this, we will (1) assess the usefulness of our method for predicting inflation based on a Phillips curve, and (2) compare the policy guidelines the methods convey when used in a Taylor rule.

Predictive performance We also compare our estimations with those generated by usual filtering techniques, namely HP and CF filters, as well as with an estimation computed using a production function approach (PF).[17] These traditional techniques are univariate, unlike our model which includes a relatively large set of variables that enables us to recognize the hypothesis of permanent income and the key relationship between consumption and GDP growth long-run path in order to find a permanent-transitory decomposition (Cochrane, 1994). Also, we take into account variables like exchange rate and oil prices for the fact the Colombian economy is a small open economy dependent on oil as its main export product.

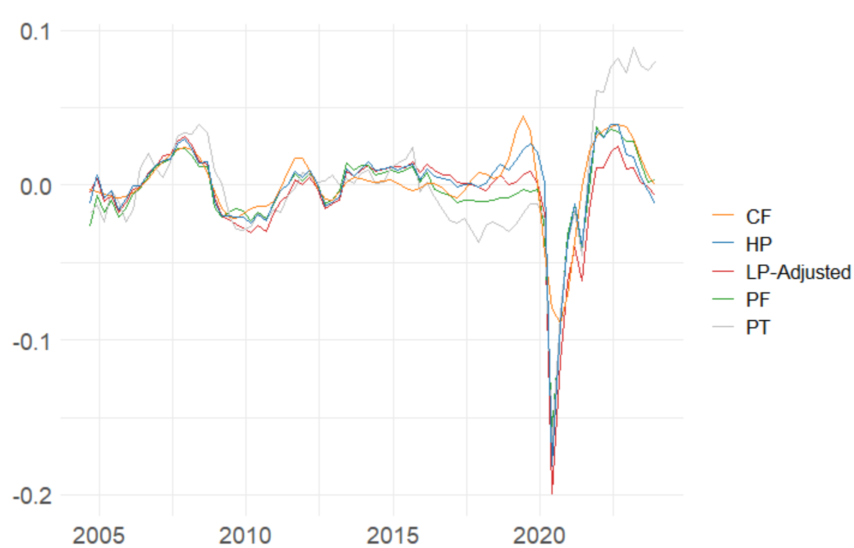
The output gap estimates for the compared methods and our proposal are shown in Figure 7. In this case, the univariate filters (HP, CF) tend to deliver a large gap right before COVID-19 and rapid and sizeable subsequent recovery, which sends that gap into positive territory (and at or beyond 5%) in a few quarters. These features may indicate an overestimation of the gap, specifically when we see that the other estimates, including our proposal, do not display such behaviour, and instead suggest a dynamic yet more moderate recovery. Notably, when tying these results to the associated potential output dynamics, these results indicate that our proposal does not lower the potential output significantly during the period, which is related to adjusting the

[17]The PF approach reconstructs the potential output from the individual inputs of GDP, aside from the total productivity, in the context of a Cobb-Douglas technology setup

model to incorporate COVID-19 observations in the estimation sample without assuming drastic changes in its data-generating process.

By contrast, the PF function seems to draw the gap in the opposite direction and could indicate its underestimation. First, it is below all competing methods throughout the sample, but primarily, it lowers the gap too steeply during every downturn (1999, 2008, 2016, COVID-19). These patterns also contrast with our proposal; thus, we see our method as a middle point. In particular, concerning the PF method, our proposal has the advantage of including more information in the model and pinpointing the long-run behavior of the GDP through its link to consumption. While the PF, conversely, can be too quick to associate the bulk, if not all, of the fluctuations in capital and labor inputs to the short-run behavior of the GDP, which is counterfactual to recent studies on hysteresis and the scarring effects of protracted recessions (e.g., Cerra, Fatás, and Saxena, 2023; Aikman, Drehmann, Juselius, and Xing, 2022).

Figure 7: Comparison methodologies for output gap estimation



In terms of forecasting performance, two pseudo-out-of-sample forecasting exercises were conducted for the periods from the first quarter of 2016 to the last quarter of 2019, and from the first quarter of 2020 to the fourth quarter of 2023. To achieve this, we used a traditional Phillips curve, $\pi_t = \phi_0 + \sum_{i=1}^{p} \phi_i \pi_{t-i} + \gamma \tilde{y}_t + \varepsilon_t$, incorporating different measures of the output gap ($\tilde{y}_t$) as inputs and evaluating their predictive capacity across different forecasting horizons (ranging from one quarter to one year ahead) for the quarterly seasonally adjusted core inflation (π_t).[18] As expected, forecasting errors increased significantly during the period of high uncertainty following COVID-19, which was marked by elevated inflation. Overall, the proposed model (LP-Adjusted) shows improved forecasting performance for horizons of $h = 1$ to $h = 4$ compared to the three

[18]Similar results are obtained with headline inflation, although the forecasting errors are significantly larger and more volatile.

other techniques commonly used to estimate the trend and cyclical components of GDP and the BSVAR with no COVID-type adjustment (Table 2).

Table 2: Root Mean Squared Error (RMSE) for quarterly core inflation (π_t) and different forecast horizons (one to four quarters ahead).

Sample: 2020Q1 - 2023Q4				
Output gap measure	$h = 1$	$h = 2$	$h = 3$	$h = 4$
BSVAR (LP-Adjusted)	1.018716	1.048559	1.073412	1.11187
PT	1.073134	1.103567	1.098208	1.139404
HP filter	1.121717	1.153062	1.113327	1.15189
CF filter	1.048939	1.079876	1.080275	1.121033
Production Function (PF)	1.166364	1.198608	1.163344	1.203933
Sample: 2016Q1 - 2019Q4				
Output gap measure	$h = 1$	$h = 2$	$h = 3$	$h = 4$
BSVAR (LP-Adjusted)	0.5844528	0.6002063	0.6215973	0.6469785
PT	0.6241584	0.6413413	0.6632999	0.68825
HP filter	0.6208985	0.6385685	0.6617766	0.6887832
CF filter	0.6170853	0.635219	0.6586359	0.685402
Production Function (PF)	0.5956912	0.6120511	0.6340404	0.6599295

Note: This table includes the RMSE for a set of competing models, including our benchmark, BSVAR (LP-Adjusted), a BSVAR with no COVID-type adjustment (PT), and two univariate filters (HP and CF).

When comparing the forecasting performance using the Diebold and Mariano test (Table 3), we can see that in the pre-COVID sample, or in "normal times", the forecasting gains are not significantly greater than those of the other methods. However, when comparing the 2020-2023 period, the LP-Adjusted outperforms all competing methods, namely, the HP and CF at all horizons and significance levels, and the PF for most horizons —although in a weaker sense, or at the 10% level. Therefore, these results indicate that accounting for the increased model uncertainty due to the COVID-19 episode is paramount. Finally, a comparison of the BSVAR with and without the COVID adjustment shows performance gains for forecast horizons up to two steps ahead; beyond that point, the differences become negligible.[19]

[19]For understanding the predictive performance difference across these periods, it is also relevant to note that any good prediction relies on a good model and good inputs (to be used by the model). Here, we are only inquiring about how beneficial it is to account for an output gap robust to COVID disruptions. However, from the tables 2 and 3, we can also see how all models' performance worsens during and after the onset of the lockdown. This issue is likely related to disruptions in the relationships posited by a Phillips curve in its usual format (as the one we consider). Changes in the shape of the Phillips curve during recent years have actually been suggested in the literature (see for example McLeay and Tenreyro, 2020; Hazell et al., 2022; Benigno and Eggertsson, 2023, among others). Despite this, however, the results also suggest that providing an output gap robust to these disruptions mitigates these predictive issues to a certain extent.

Table 3: Diebold and Mariano test for forecast accuracy
Comparison between model 1 = LP-Adjusted and Model 2 = (PT, HP, CF, PF)

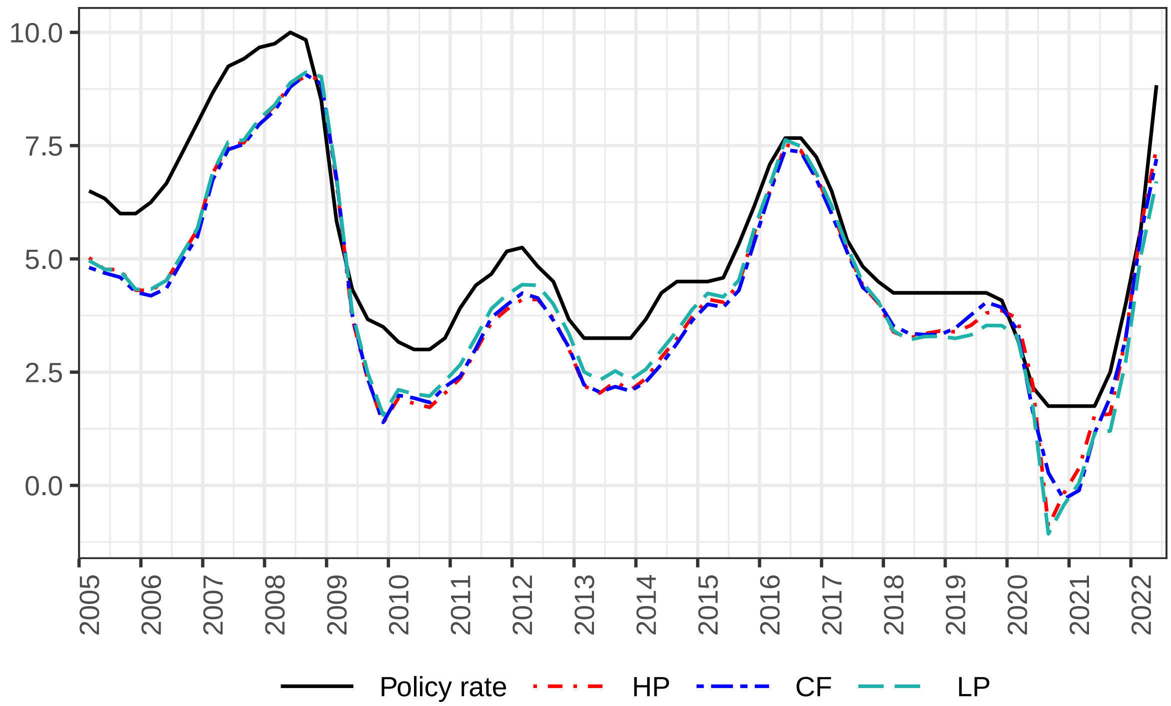
Sample	h	Ha: two sided*				Ha: greater**				Ha: less***			
		PT	HP	CF	PF	PT	HP	CF	PF	PT	HP	CF	PF
2020Q1-2023Q4	1	0.0100	0.0049	0.0037	0.1463	0.995	0.9975	0.9982	0.9268	0.0050	0.0025	0.0018	0.0732
	2	0.0973	0.0003	0.0039	0.0711	0.9513	0.9999	0.9980	0.9644	0.0486	0.0001	0.0020	0.0356
	3	0.1626	0.0032	0.0059	0.1370	0.9187	0.9984	0.9970	0.9315	0.0813	0.0016	0.0030	0.0685
	4	0.2313	0.0115	0.0066	0.2508	0.8843	0.9942	0.9967	0.8746	0.1157	0.0058	0.0033	0.1254
2016Q1-2019Q4	1	0.0008	0.0268	0.6836	0.0005	0.0004	0.0134	0.6582	0.9998	0.9996	0.9866	0.3418	0.0002
	2	0.0059	0.1675	0.7725	0.0040	0.0029	0.0838	0.6138	0.9980	0.997	0.9162	0.3862	0.0020
	3	0.0138	0.2832	0.8062	0.0188	0.0069	0.1416	0.5969	0.9906	0.9931	0.8584	0.4031	0.0094
	4	0.0062	0.3623	0.8125	0.0431	0.0031	0.1811	0.5937	0.9784	0.9969	0.8189	0.4063	0.0216

The table shows p-values for Diebold and Mariano Test.
The null hypothesis is that the two methods have the same forecast accuracy.
*Ha: two sided is that method 1 and method 2 have different levels of accuracy.
**Ha: greater is that method 2 is more accurate than method 1.
***Ha: less is that method 2 is less accurate than method 1.

Use in policy rules Although one of the main features of our baseline is that it prevents unprecedented events from affecting the gauge of uncertainty around the output gap estimates, it also generates more stable point estimates of potential output —due to both the identification and adjustment— as a temporary large shock is less likely to affect the long-run component of GDP that the method targets. Other methods, on the contrary, can be heavily influenced by any shocks, and in some cases by fluctuations at the end of the sample.

Importantly, different point estimations can lead to distinct policy lessons and to a varied usefulness of the gap for providing information to the Central bank. To see this, we compute a Taylor rule-based interest rate using (the deterministic component) of Equation (5) and using the posterior parameters in Table 1. The results are displayed in Figure 8. In each case, the inflation gap (actual inflation minus target) is the same, and the only difference will be the provided output gap.

Figure 8: Taylor rules based policy rates from competing output gap methods



Note: This figure plots the predicted policy rates based on a Taylor rule as in Equation 5 and different output gap methods. The methods are: HP: Hodrick-Prescott filter, CF: Christiano-Fitzgerald filter, LP: BSVAR with COVID-19 adjustment as in Lenza and Primiceri (2022). The plot also shows the actual policy rate in Colombia (quarterly average).

We can see that our baseline aligns more closely with the actual policy rate during normal times. However, there are even more notable differences among the output methods around the COVID-19 episode period. First, prior to the shock, the non-adjusted methods indicate a sudden increase in the rate that does not align with either inflationary pressures or the stance of the bank, then, during the COVID period, the HP filter gap incorrectly attributes the shock to a potential output decrease (despite being short-lived), the CF filter gap does not falter in that case, however, both non-adjusted methods imply a steep jump of the rate right at the start of the recovery period —an effect elicited by the issue of an almost instantaneous and too strong rebound of the gap onto positive territory that adversely affects the filters.

At the same time, we can compare quantitatively the similarity of the implied Taylor-rule rates from each method, for which we report measures of fit in Table 7 the Appendix D, we can see that for six out of seven measures our baseline model (LP-Adjusted) aligns more closely to the observed rate.

Taking stock, the stability of the potential output that characterizes our baseline method is translated into a policy rate rule guideline that prevents swings in the interest rate prior to and after large shocks, and it also has the advantage of being more aligned with actual policy implementations in normal times.

5 Concluding remarks

We study the dynamics of the output gap in Colombia during and after the COVID recession and the relevance of accounting explicitly for the presence of unprecedented economic variations relative to more traditional and simplified alternatives. Special attention is paid to the implications of informing policy exercises with COVID-adjusted estimates. In line with other studies (Granados and Parra-Amado, 2024; Morley et al., 2022) we find that it is important to adjust the model so that it does not consider the atypical episode as part of the data-generating process of the model and that, ideally, the scale of adjustment should be informed by the data itself.

To conclude this, we consider a baseline model incorporating ample information sources into a structural framework that allows for the application of an identification strategy that exploits the relationship between consumption and output to recover the permanent and transitory components of GDP, as in Uhlig (2003, 2004). Based on this setup, we adjusted the model with a scaling factor of the residuals around the COVID-19 pandemic outbreak along the lines of Lenza and Primiceri (2022). Our results indicate that only one structural error is enough to account for most of the long-run behavior of GDP (and potential output) and that the remaining shocks mainly explain transitory fluctuations (i.e., the gap). Our setup prevents quick output gap reversals after downturns or drastic changes in the potential output after high-magnitude transitory observations. In particular, during COVID times, our findings indicate an 18.8% decline in the output gap during the pandemic. Furthermore, our statistical analysis reveals a 1.4 percentage point reduction in potential GDP due to the COVID-19 crisis.

Evaluations of the model for policy applications indicate improvements from considering adjusted output gap estimates as input for policy exercises. Although more complex, these methods provide estimated gaps with higher predictive performance, and that can yield rule-based policy rates that better represent the state of the economy and thus align more closely with the actual policy rates implemented by central banks, where gap estimates are included as one of several inputs for their decisions.

While our identification strategy has its limitations, particularly in terms of decomposing output dynamics in economic structural drivers (e.g., monetary, financial, global, supply, and demand), it provides a simple way to find an agnostic structural identification that explains the FEV of the GDP in the long run. Although we can estimate the output gap with reasonable accuracy, future research can explore these drivers in more detail, building on our findings to refine the approximation and minimize the drawbacks of arbitrary shock horizon adjustments common in other approaches.

References

Aikman, D., M. Drehmann, M. Juselius, and X. Xing (2022, October). The Scarring Effects of Deep Contractions. BIS Working Papers 1043, Bank for International Settlements.

Alfonso, V., L. E. Arango, F. Arias, G. Cangrejo, and J. A. D. Pulido (2013, 06). Ciclos de negocios en Colombia: 1975-2011. *Lecturas de EconomÃa*, 115 – 149.

Álvarez, L. J. and A. Gómez-Loscos (2018). A menu on output gap estimation methods. *Journal of Policy Modeling 40*(4), 827–850.

An, S. and F. Schorfheide (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews 26*(2-4), 113–172.

Angeletos, G.-M., F. Collard, and H. Dellas (2020). Business-cycle anatomy. *American Economic Review 110*(10), 3030–70.

Ball, L. M. (2009). Hysteresis in unemployment: old and new evidence. Technical report, National Bureau of Economic Research.

Barigozzi, M. and M. Luciani (2021). Measuring the output gap using large datasets. *The Review of Economics and Statistics*, 1–45.

Barsky, R. B. and E. R. Sims (2011). News shocks and business cycles. *Journal of monetary Economics 58*(3), 273–289.

Basu, S. and J. G. Fernald (2009). What do we know (and not know) about potential output? *Federal Reserve Bank of St. Louis Review 91*(July/August 2009), 187–213.

Baxter, M. and R. G. King (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics 81*(4), 575–593.

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

Benati, L. (2008). Investigating Inflation Persistence Across Monetary Regimes. *The Quarterly Journal of Economics 123*(3), 1005–1060.

Benati, L. and T. A. Lubik (2021). Searching for hysteresis. Technical report, Federal Reserve Bank of Richmond.

Benigno, P. and G. B. Eggertsson (2023, April). It's Baaack: The Surge in Inflation in the 2020s and the Return of the Non-Linear Phillips Curve. NBER Working Papers 31197, National Bureau of Economic Research, Inc.

Berger, T., J. Morley, and B. Wong (2023). Nowcasting the output gap. *Journal of Econometrics 232*(1), 18–34.

Berger, T. and C. Ochsner (2022). Robust real-time estimates of the german output gap based on a multivariate trend-cycle decomposition. Technical report, Deutsche Bundesbank Discussion Paper No. 35/2022, available at SSRN working paper series.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics 7*(2), 151–174.

Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review 79*(4), 655–673.

Blanchard, O. J. and L. H. Summers (1986). Hysteresis and the european unemployment problem. *NBER macroeconomics annual 1*, 15–78.

Blinder, A. S. and J. B. Rudd (2013, June). *The Supply-Shock Explanation of the Great Stagflation Revisited*, pp. 119–175. University of Chicago Press.

Brignone, D. and M. Mazzali (2022). Evidence on the confounding nature of the main business cycle driver. Technical report, Available at SSRN working paper series.

CBO (2001). CBO's method for estimating potential output: An update. Technical report, Congressional Budget Office.

CBO (2014). Revisions to CBO's projection of potential output since 2007. Technical report, Congressional Budget Office.

Cerra, V., A. Fatás, and S. C. Saxena (2023, March). Hysteresis and business cycles. *Journal of Economic Literature 61*(1), 181–225.

Chen, M. J. and L. Gornicka (2020, February). Measuring Output Gap: Is It Worth Your Time? IMF Working Papers 2020/024, International Monetary Fund.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1996). Identification and the effects of monetary policy shocks. In *Financial Factors in Economic Stabilization and Growth*, pp. 36–74. Cambridge University Press.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics 1*, 65–148.

Christiano, L. J. and T. J. Fitzgerald (2003). The band pass filter. *international economic review 44*(2), 435–465.

Cimadomo, J., D. Giannone, M. Lenza, F. Monti, and A. Sokol (2022). Nowcasting with large bayesian vector autoregressions. *Journal of Econometrics 231*(2), 500–519.

Cochrane, J. H. (1990). Univariate vs. multivariate forecasts of gnp growth and stock returns: Evidence and implications for the persistence of shocks, detrending methods.

Cochrane, J. H. (1994). Permanent and transitory components of gnp and stock prices. *The Quarterly Journal of Economics 109*(1), 241–265.

Fontanari, C., A. Palumbo, and C. Salvatori (2020). Potential output in theory and practice: a revision and update of okun's original method. *Structural Change and Economic Dynamics 54*, 247–266.

Forbes, K., I. Hjortsoe, and T. Nenova (2018). The shocks matter: Improving our estimates of exchange rate pass-through. *Journal of International Economics 114*(C), 255–275.

Fornaro, L. and M. Wolf (2020). Covid-19 coronavirus and macroeconomic policy. Technical report, CEPR Discussion Paper No. DP14529.

Furlanetto, F., K. Hagelund, F. Hansen, and Ø. Robstad (2022). Norges bank output gap estimates: Forecasting properties, reliability, cyclical sensitivity and hysteresis. *Oxford Bulletin of Economics and Statistics*.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294–314.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Granados, C. and D. Parra-Amado (2024). Estimating the output gap after covid: How to address unprecedented macroeconomic variations. *Economic Modelling 135*, 106711.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning (2022). Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? *American Economic Review 112*(5), 1437–74.

Havik, K., K. Mc Morrow, F. Orlandi, C. Planas, R. Raciborski, W. Röger, A. Rossi, A. Thum-Thysen, V. Vandermeulen, et al. (2014). The production function methodology for calculating potential growth rates & output gaps. Technical report, Directorate General Economic and Financial Affairs (DG ECFIN).

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022, 02). The Slope of the Phillips Curve: Evidence from U.S. States*. *The Quarterly Journal of Economics 137*(3), 1299–1344.

Hodrick, R. J. and E. C. Prescott (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking 29*(1), 1–16.

Jarociński, M. and M. Lenza (2018). An inflation-predicting measure of the output gap in the euro area. *Journal of Money, Credit and Banking 50*(6), 1189–1224.

Kiley, M. T. (2013). Output gaps. *Journal of Macroeconomics 37*, 1–18.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1987). Stochastic trends and economic fluctuations.

Lenza, M. and G. E. Primiceri (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics 37*(4), 688–699.

Lütkepohl, H. and A. Netšunajev (2017). Structural vector autoregressions with smooth transition in variances. *Journal of Economic Dynamics and Control 84*, 43–57.

McLeay, M. and S. Tenreyro (2020). Optimal Inflation and the Identification of the Phillips Curve. *NBER Macroeconomics Annual 34*(1), 199–255.

Morley, J., D. R. Palenzuela, Y. Sun, and B. Wong (2022, Aug). Estimating the euro area output gap using multivariate information and addressing the covid-19 pandemic. Working Paper 2716, European Central Bank (ECB).

Morley, J. and B. Wong (2020). Estimating and accounting for the output gap with large bayesian vector autoregressions. *Journal of Applied Econometrics 35*(1), 1–18.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review 94*(4), 1055–1084.

Romer, C. D. and D. H. Romer (2010). The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review 100*(3), 763–801.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society 48*(1), 1–48.

Summers, L. H. (2015). Demand side secular stagnation. *American economic review 105*(5), 60–65.

Uhlig, H. (2003). What moves real gnp? Unpublished manuscript.

Uhlig, H. (2004). What moves gnp? In *Econometric Society 2004 North American Winter Meetings*, Number 636 in 1. Econometric Society.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics 52*(2), 381–419.

A Survey: methods

Table 4: Univariate estimation methods

	Model based	Decision variables	Complexity	Need or advisability of using forecats
Hodrick & Prescott	No	Smoothness parameter	Low	Yes
Baxter & King	No	Pass band Filter length	Low	Yes
Butterworth filtering	No	Pass band Filter length	High	Yes
Wavelet-based methods	No	Wavelet basis	High	Yes
Linear detrending	Yes	None	Low	No
Beveridge & Nelson	Yes	ARIMA model	High	Yes
Structural time series	Yes	STS model	High	No
Hamilton	Yes	Regime switching model	High	No
Kim & Nelson	Yes	Regime switching model	High	No

Source: Álvarez and Gómez-Loscos (2018).

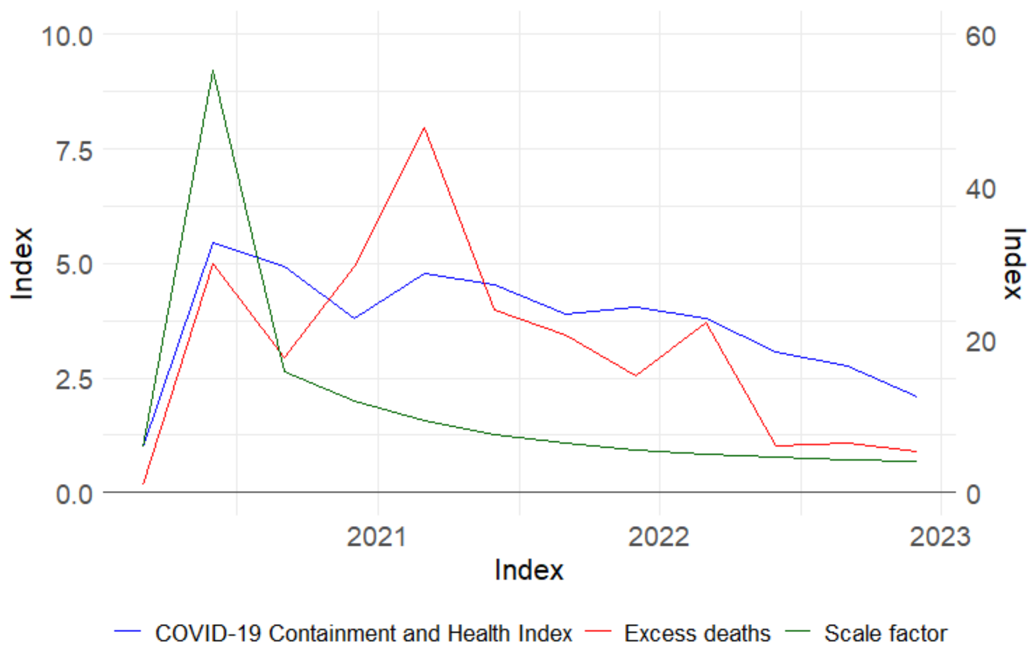
Table 5: Multivariate estimation methods

	Underlying economic theory	Decision variables	Complexity
Okun's Law	Okun's Law	VAR model	Medium
Production function	Production function	Production function Cyclically adjusted inputs	High
Blanchard & Quah	Supply and demand shocks	SVAR model	High
Phillips curve	Phillips curve	Output gap time series process	High
Natural rate of interest	Natural rate of interest	Lags in the Phillips curve, Output gap time series process	High
RBC model	General equilibrium	VECM model	High
DSGE model	General equilibrium	Model specification	High

Source: Álvarez and Gómez-Loscos (2018).

B Comparison Between Estimated Scale Factors and COVID-19-Related Indicators

Figure 9: Comparison of Scale Factor Estimates and COVID-19-Related Measures



The blue line is an index for COVID-19 containment and health, which is a composite measure based on thirteen policy response indicators, including school closures, workplace closures, travel bans, testing policy, contact tracing, face coverings, and vaccine policy. The red line is an index of estimated daily excess deaths per 100000 people during COVID-19 in Colombia (right axis). The green line is our estimate for the scale factors after the pandemic period. Source: Our World in Data, author's calculations.

C VAR Lag Order Selection Criteria

VAR Lag Order Selection Criteria			
Lag	**AIC**	**SC**	**HQ**
0	110.4471	110.6706	110.5377
1	89.41091	92.77367	90.31716
2	88.52696	**91.64601***	**90.24884***
3	**88.46951***	94.72781	91.00702
4	88.78933	97.05923	92.14247
5	89.17346	99.45497	93.34224
6	88.60526	100.8984	93.58966

Table 6: VAR Lag Order Selection Criteria

D Additional measures of fit of implies Taylor Rules

Table 7: Performance measures for rule-based approximations

measure	HP	CF	LP-Adjusted
RMSE	1.230	1.348	1.202
MAE	1.062	1.146	1.024
MAPE	25.212	24.784	24.847
THEIL	0.118	0.131	0.115
R^2	0.896	0.830	0.901
CORR	0.947	0.911	0.949
SSE	105.848	127.160	101.055

Notes: This table shows measures of fit between the observed policy rate and Taylor-rule rates implies by the output gap approximated by the Hodrick-Prescott filter (HP), the Christiano-Fitzgerald filter (CF), the BSVAR with COVID scaling adjustment (LP-Adjusted) depicted in Figure 8. The measures reported are the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Mean Absoltute Percentage Error (MAPE), the Theil's U inequality coefficient (bounded between 0 and 1, with 0 representing an perfect fit, the R^2 from a simple regression, the correlation, and the sum of squares.