Graduate Institute of International and Development Studies
International Economics Department
Working Paper Series

Working Paper No. HEIDWP04-2023

Estimating the Output Gap After COVID: An Application to Colombia

Camilo Granados
University of Texas at Dallas
Daniel Parra-Amado
Banco de la República

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

Estimating the Output Gap After COVID: An Application to Colombia

Camilo Granados
University of Texas at Dallas
Daniel Parra-Amado
Banco de la República

Abstract

This study examines whether and how important it is to adjust output gap frameworks during the COVID-19 pandemic and similar unprecedentedly large-scale episodes. Our proposed modelling framework comprises a Bayesian Structural Vector Autoregresion with an identification setup based on a permanent-transitory decomposition that exploits the long-run relationship of consumption with output whose residuals are scaled up around the COVID-19 period. Our results indicate that (i) a single structural error is sufficient to explain the permanent component of the gross domestic product (GDP); (ii) the adjusted method allows for the incorporation of the COVID-19 period without assuming sudden changes in the modelling setup after the pandemic; and (iii) the proposed adjustment generates approximation improvements relative to standard filters or similar models with no adjustments or alternative ones, but where the specific rare observations are not known. Importantly, abstracting from any adjustment may lead to over- or underestimating the gap, too-quick gap recoveries after downturns, or too-large volatility around the median potential output estimations.

Keywords: *Bayesian methods, business cycles, potential output, output gaps, structural estimation.*

JEL: E2, E3, E32, E36, O41

The author(s) thank Luca Benati from the University of Bern for the academic supervision of this paper. 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 author(s) also appreciate the comments of Cédric Tille and Jean-Paul Renne. The views expressed in this paper are solely those of the author(s) and do not necessarily reflect those of Banco de la República or its board of directors.

1 Introduction

The COVID-19 pandemic unprecedentedly affected humanity, not only in terms of public health but also economically. Unlike other crises, economic deterioration has been globally synchronised. According to the International Monetary Fund, the world economy contracted by approximately 3.0% in 2020, with both developed and emerging economies falling by 4.8% and 1.9%, respectively. Colombia's economy was no exception to this pattern and reported the lowest growth rates historically during this episode (-7.0%). Ensuing recovery was relatively quick owing to the distribution of vaccines and the gradual opening during and after the lockdown. This allowed most economies to bounce back to positive growth rates in one to three quarters, which, given the magnitude of the initial downturn, led to the worldwide growth of 6.0% in 2021. In summary, we witnessed extreme macroeconomic data, which also led to high levels of uncertainty.

The fluctuations observed naturally raise questions about the macroeconomic effects of the COVID-19 shock, particularly on variables such as the potential output and gap. On simple inspection, it is difficult to label a downturn of that magnitude as trivial for long-run variables. However, the pace of recovery makes it challenging to deem the shock as highly influential.

With this in mind, we aim to answer whether estimations of the output gap should be adjusted to account for the COVID-19 shock. We intend to determine a way to reconcile the magnitude of the shock with its transitory nature when approximating the potential output.

Our approach must, therefore, cover two fronts: first, how to obtain a good econometric framework for estimating the output gap, and second, how this can be adjusted in a manner that allows incorporating the COVID-19 shock information but prevents it from influencing the model as if it was representative of the data-generating process. For the first point, we rely on a permanent-transitory (PT) decomposition framework to identify the fluctuations of a set of macroeconomic variables (where the output is included) in a Bayesian Structural Vector Autoregression (BSVAR) setting; this is done by following Uhlig (2003, 2004). Based on the resulting model, we recover a path for the potential output that covers the COVID-19 period. For the second point, we adjust the model estimation with a scale factor around the rare shock date along the lines of Lenza and Primiceri (2022)).

Primarily, we include an ample set of variables in our model that contrasts with the usual output gap estimation frameworks, such as univariate statistical filters or production function approaches.[1] This enables us to include additional sources of information in our setup and account for the permanent income hypothesis through the relationship between consumption and long-run output, which, as mentioned by Cochrane (1994), facilitates identifying the permanent component of the output.

Nonetheless, the identification task in the context of SVAR models can be challenging, as these

[1] See Álvarez and Gómez-Loscos (2018) for an overview of the gap estimation methods.

frameworks usually rely on imposing strong assumptions about the nature of shocks that can be too restrictive. For example, it is usual to impose that long-run output is driven only by supply shocks, while demand is only associated with transitory components (e.g., Barsky and Sims, 2011; Blinder and Rudd, 2013; Keating and Valcarcel, 2015; Chen and Gornicka, 2020). However, recent data has vindicated the potential long-term role of demand-driven phenomena; for example, in the Global Financial Crisis (GFC) and the protracted recovery that followed, a weak demand affected both the current output and its future expectations in such a persistent way that it shifted down the path of potential growth (Fontanari, Palumbo, and Salvatori (2020)).[2] The literature has followed suit and has recently pointed out that other shocks, such as demand (Furlanetto, Lepetit, Robstad, Rubio-Ramírez, and Ulvedal, 2021) and monetary (Jordà, Singh, and Taylor, 2020) shocks, can also have long-run effects.

The aforementioned consideration is even more valid in the context of the COVID-19 shock, which was considered a supply-driven shock, but eventually showed to involve demand-driven fluctuations.[3] We circumvent this issue of a separate identification of supply and demand shocks and their association with different terms by adopting an agnostic identification approach along the lines of Uhlig (2003, 2004), that is, based on the maximization of the explained fraction of long-horizon Forecast Error Variance (FEV) of the gross domestic product (GDP). Such an approach is particularly reasonable for gauging the potential output when we observe shocks such as the COVID-19 downturn that are perceived as a combination of supply- and demand-driven fluctuations (rather than either exclusively).

This identification scheme has been used by recent studies, such as Angeletos, Collard, and Dellas (2020) and Brignone and Mazzali (2022) and with the same objective of decomposing the permanent and transitory fluctuations of macroeconomic variables. We follow a similar approach while adjusting the econometric modelling along the lines of Lenza and Primiceri (2022), which allows us to incorporate the COVID-19 downturn in the sample but limits the impact of the rare event on the estimated parameters. The joint application of the identification setup and adjustment for high-magnitude shocks in the context of output gap estimations represents our contribution.

We apply our approach to the Colombian economy and find that a single structural shock is sufficient to characterise the long-run behaviour of GDP. By contrast, the remaining shocks tend to explain their transitory effects more significantly. This result aligns with the findings of Dieppe, Francis, and Kindberg-Hanlon (2021), Angeletos, Collard, and Dellas (2020) and Brignone and Mazzali (2022) for the US and European countries. Based on this result and the structural shocks, we approximate the GDP gap at each date as the weighted sum of the transitory shocks (and use the other shock to recover potential output).

[2] This experience even led to revisiting the literature on hysteresis, such as Cerra, Fatás, and Saxena (forthcoming), Benati and Lubik (2021) and Aikman et al. (2022).

[3] See Guerrieri, Lorenzoni, Straub, and Werning (2022) and Fornaro and Wolf (2020) for further discussion on demand to output spillovers and stagnation traps.

Our estimates suggest that the potential output in Colombia was only marginally affected by the COVID-19 episode, which generated a large decrease in the output gap (-18.9% in 2020Q2) and the volatility of the estimates increased starting that quarter and relative to the pre-COVID-19 period. In addition, we find that the output gap decrease around this downturn is not persistent, which contrasts with previous recessions and instead bounces back rapidly in the following quarters. Finally, we obtained a reduction of 1.4% in the potential output during 2020 owing to the lockdown.

We address our main research question and compare the potential output (and gap) estimates with standard gap estimation methods and the BSVAR counterpart with no COVID-19 adjustment. We find that our proposal (PT identification with COVID-19 adjustment) prevents the potential output from falling too rapidly at the onset of the shock and does not induce a fast recovery in subsequent periods, a known drawback of the usual univariate filtering techniques.

Then, we compare our method with an alternative BSVAR with the same identification scheme but a stochastic volatility setup. This alternative is, in principle, also adjusting for the effect of the COVID-19 episode on the model. However, in contrast to our scaling method around the shock date, the adjustment is entirely endogenous because the variance is time-varying. This model leads to a stronger decrease in the potential output, but even by 2022 shows no sign of recovery, suggesting that the large magnitude of the shock could persistently affect the estimates. In light of this, our model represents a more appropriate alternative for a shock of large magnitude but small persistence that is less representative of the data-generating process of the sample.

Finally, we evaluate our method in a simulation setting to compare it with alternatives in a more general light, not only with the specific COVID-19 episode and economy of reference in mind. The results reveal that the proposed method can increase the cross-correlation between the target (economic model simulated potential output) and the estimates from 0.5 to 0.72 at its peak, compared with standard filtering techniques.

In summary, these results and exercises consistently indicate that the model's performance is not only associated with the structural identification of the BSVAR but also with the adjustment of the model to include the large shock. The benefits of adjusting the gap estimates in the presence of shocks of unprecedented magnitude are non-trivial. The performance gains terms are present even in models, successful at approximating the potential output. In addition, our setup prevents a strong decrease in the potential output after the outlying downturn and a quick recovery once its transitory nature is made evident; that is, it improves on the drawbacks of complex counterparts and standard filters.

Related literature Our paper is related to various strands of the literature. At large, this paper belongs to the large literature on the estimation of the output gap, and to a greater extent to those based on multivariate approaches.[4] More specifically, our paper is related to studies using PT

[4]For an overview of this literature see Álvarez and Gómez-Loscos (2018), Guisinger et al. (2018); and for a discussion about multivariate approaches see Cochrane (1990)

decomposition-type of methods for estimating the output gap; among these papers, Angeletos, Collard, and Dellas (2020), Brignone and Mazzali (2022) and Dieppe, Francis, and Kindberg-Hanlon (2021) use the same approach of this paper, that is, based on explaining the highest possible share of the FEV of the output in the long-run, while studies such as Morley, Palenzuela, Sun, and Wong (2022), Berger, Morley, and Wong (2023), and Berger and Ochsner (2022) use a Beveridge-Nelson (BN) type of decomposition based on the optimal forecast at long horizons. Our contribution relative to the first group of these studies is that we adjust our baseline model along the lines of Lenza and Primiceri (2022) to include the COVID-19 period in the sample. Simultaneously, relative to the second group, rather than calling the optimal long-run forecast (obtained via BN decomposition) the potential GDP, we structurally identify a model where the share of the long-run variance is explained by a limited number of shocks.

This study also relates to the literature on the adjustment of econometric models to include COVID-19 periods. In particular, it closely follows the work of Lenza and Primiceri (2022) by scaling the model information around a researcher-specified date but allowing the scale factor parameters to be obtained in a Bayesian setting. Other studies proposing alternative adjustments in this direction are Hartwig (2022), Carriero, Clark, Marcellino, and Mertens (2022), and Ng (2021).

Finally, in an even more closely related study, Morley, Palenzuela, Sun, and Wong (2022) adjusted the model to include COVID-19 information and used a PT type of decomposition to estimate the output gap for the Eurozone using a VAR-X setup. Our study uses a different identification setup in a Bayesian-SVAR setting (based on Uhlig (2003, 2004)) rather than BN decomposition. In that sense, this study contributes to the literature by showing the effects of adjusting potential output models by the COVID-19 shock in the context of structural models.

The remainder of this paper is organised as follows. We explain the methodology in Section 2. Section 3 describes our data and main results, including a comparison of the proposed estimates with those yielded by other methods. In Section 4, we evaluate the performance of the proposed method in a simulation exercise and conclude the paper.

2 Methodology

Our empirical strategy was divided into two stages. First, we fit a reduced-form Vector Autoregressive (VAR) model with a scale factor adjustment around the COVID-19 crisis as in Lenza and Primiceri (2022). This allows us to account for the increased variance in the macroeconomic variables around the shock date. Second, we recast our model 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[5] at a long-run horizon (e.g. 15 or 25 years ahead).

[5]It is also possible on other variables, such as household consumption.

In the first stage, following Lenza and Primiceri (2022), a scale factor s_t is added to the VAR model's reduced-form residuals to capture the increased uncertainty during the COVID-19 crisis. s_t is set to one in the sample period before the COVID-19 shock (t^*), $s_{t^*} = \bar{s_0}$, $s_{t^*+1} = \bar{s_1}$, $s_{t^*+2} = \bar{s_2}$ and $s_{t^*+j} = 1 + (\bar{s_2} - 1)\rho^{j-2}$ for $j \geq 3$.[6] The scaled (reduced-form) VAR model is given by:

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

The COVID-19 outbreak dates back to the first quarter of 2020 ($t^* = 2020Q1$); therefore, $\bar{s}_0$ is estimated for that date, and $\bar{s}_1$, $\bar{s}_2$ for the next two quarters. Then, the scale factor decays at a rate ρ for all future periods.[7] Thus, $\theta \equiv [\bar{s_0}, \bar{s_1}, \bar{s_2}, \rho]$ is the vector of additional parameters to be estimated together with those of the VAR $(B_0, B_1, \ldots, B_p, \Sigma)$. 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-Hasting 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 $\gamma \equiv (\Psi, d, b$ and $\Omega)$ are the hyperparameter vectors. 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):

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

When (1) is estimated, we proceed with the second state, consisting of identifying structural shocks (ε_t) linked to the reduced-form errors by an impact matrix A_0such that $u_t = A_0\varepsilon_t$ and $\Sigma = A_0 A_0'$.. It should be noted that there is not a unique A_0 that satisfies these relationships. For any candidate matrix A_0 an alternative matrix $\ddot{A}_0$ exists that can be derived using an orthonormal matrix Q where $A_0 = \ddot{A}_0 Q$ and $QQ' = I$; in that sense, our approach also falls within the "set-identification" category.

[6]This setup allows the scale factor to take three different values in the first three periods after the outbreak and then decay at a rate ρ in subsequent periods. This assumption seems in line with empirical evidence for the year after the onset of the pandemic.

[7]As mentioned, alternative adjustments to COVID-19 data for VAR models have emerged in the literature in both frequentist and Bayesian frameworks, several of which are based on the inclusion of additional pandemic-related variables as controls (dummies or indicators). See Ng (2021), Carriero, Clark, Marcellino, and Mertens (2022), and Hartwig (2022).

In this context (which is common to most BSVAR identification setups) we apply our specific identification strategy; that is, the maximum fraction of the long-horizon FEV, along the lines of Uhlig (2003, 2004). This method seeks a target q_1 that satisfies:

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

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

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.[8] 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.[9] This is done in the following section.

3 Results

3.1 Data and empirical strategy

We set an eight-variable B-SVAR in levels for the period 1995Q2 to 2022Q1 using Colombian data.[10] The variables included are GDP, household consumption (CON), government consumption (GOV), investment (INV), inflation (CPI), real exchange rate (RER), interbank interest rate (ITB), and Brent oil price.

The domestic account variables (first five in the VAR) 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), and the oil price from Bloomberg.

We select a lag length of two (p = 2) following the Bayesian and Hannan-Quinn Information

[8] Note that $C(L) = I + C_1 L + C_2 L^2 + C_3 L^3 \cdots + C_h L^h + \dots$ and the moving average representation of the model is given by $Y_t = B(L)^{-1} u_t = C(L) u_t$.

[9] It is not necessarily capable of replicating its entire FEV, but it explains a large proportion of it. In other words, even more shocks could be used to increase the percentage of explained FEV if a second or third shock is also found to explain the permanent component of the target.

[10] 10We report unit root and cointegration tests that are consistent with our model choice in Appendix A

criteria, and estimate the VAR in levels using a hierarchical modelling approach that allows us to make inferences about the informativeness of the prior distribution of the BSVAR, as proposed by Giannone, Lenza, and Primiceri (2015) which automatically determines a suitable measure of the shrinkage by considering a combination of conjugate priors such as a Minnesota prior and tighter priors when the model includes many coefficients relative to the number of observations. 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 years 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]

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 90% of the long-run (permanent) component of the GDP. Concurrently, the next most important shock explain the GDP's FEV in the short run which is more resembling the transitory output component.

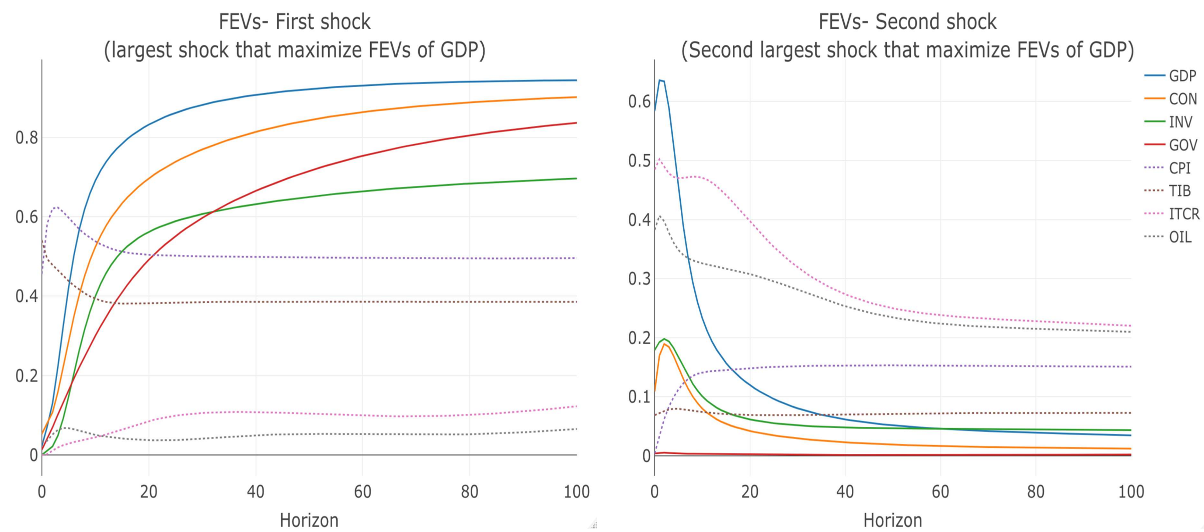
In light of these results, we compute the output gap based on the second to eighth structural shocks and use only the first one to recover the potential GDP.[12] On a related point, it should also be noted that the first structural error will explain the majority of the long-run FEV of the GDP (target variable), but not necessarily the largest share of the FEVfor other variables. The relative importance of the shocks to the other variables can be seen in the FEV decomposition per variable, as shown in Figure 9 in the Appendix C.1.[13]

[11]As a check, we increased the number of draws to 100.000 and obtained similar results.

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

[13]We leave additional results that are related to other variables and shocks for the appendix, as we are only concerned with approximating the target variable, but also because the trade-off of this method is that you compromise a structural interpretation of the shocks as separable types of drivers (e.g. monetary, financial, global, local, supply, or demand, among others), as by construction, the method only gauges the overall importance of shocks at different horizons.

Figure 1: Contribution of FEV explanation over each variable (first two largest shocks that maximize FEV of GDP)



(a) First shock (permanent) **(b)** Second shock (transitory)

Note: The left panel shows the contribution to the FEV of each variable by the structural error identified as the one with the highest percentage of explanation for the FEV of GDP. The right panel shows the second-largest shock that explains the GDP's FEV.

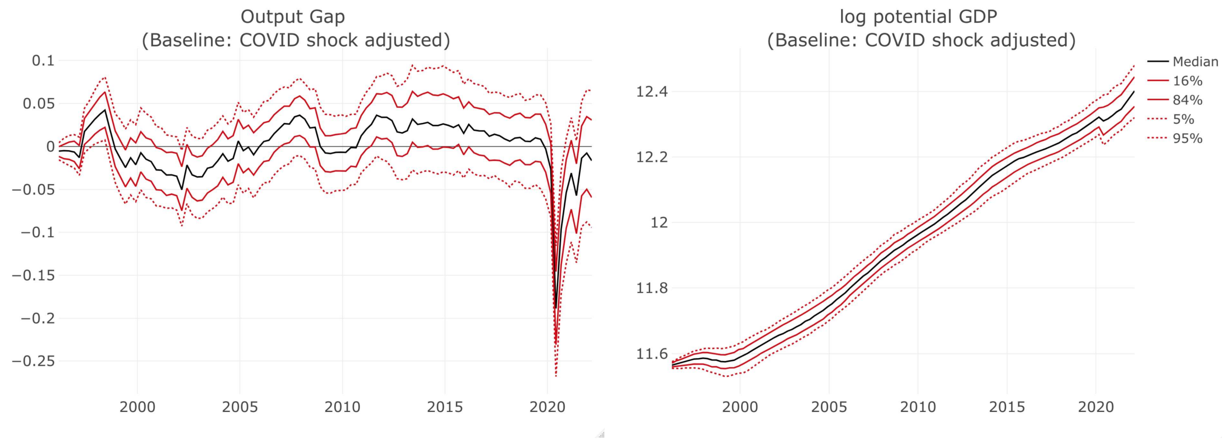
3.2 Baseline Results

Figure 2 shows the output gap and potential GDP 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. Before COVID-19, our estimated gap and potential output reflected the recession of the late 1990s, a slight deterioration during the GFC.[14] In both cases, we can see decreases in the gap dynamics and dips in the potential output. An additional gap decrease occurred during 2016, reflective of a reduction in terms of trade due to exogenous shocks in the international price of oil.[15] 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.

[14]For the Colombian case, our main downturns of reference are the 1999 and GFC crises. The former is one of the worst recessions to date, while the latter is relatively mild compared with the dynamics of advanced economies.

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

Figure 2: Baseline results: Output gap and potential GDP for Colombia



(a) Output gap

(b) Potential GDP

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 (−18.9%) 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 largely explained by lockdown measures, while the recovery is induced by the gradual reopening of the economy. The potential output also displays different dynamics than the 90s recession as it dips down, but more mildly and less persistently during the lockdown quarters. In the late nineties, the potential GDP growth went negative, contrasting with the pandemic when it only decelerated (from 3.5% in 2019 to 2.2% in 2020). The recovery paths are also in contrast with the potential output trending upward and gap closing by 2022Q1.

3.2.1 Comparison with alternative estimation methods

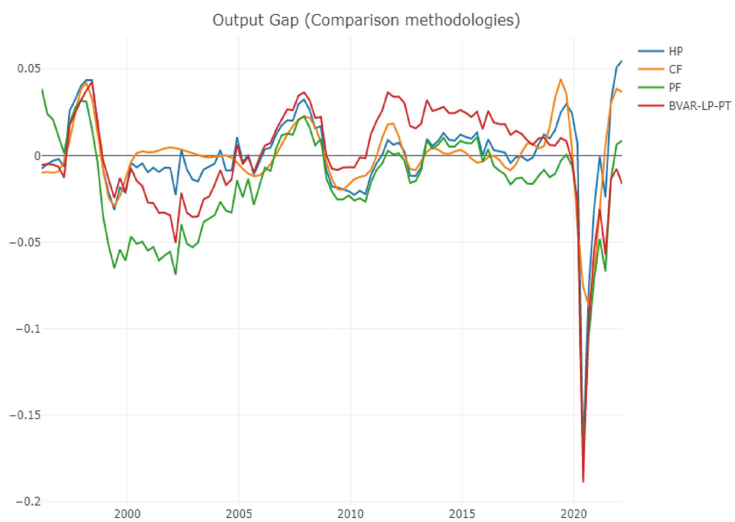
We also compare our estimations with those generated by usual filtering techniques, namely the Hodrick-Prescott (HP) and Christiano-Fitzgerald (CF) filters, as well as to an estimation computed using a production function approach (PF).[16] The output gap estimates for the compared methods and our proposal are shown in Figure Figure 3. We can see that 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 onto 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

[16] 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

the period, which is related to adjusting the 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 behaviour 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 labour inputs to the short-run behaviour of the GDP, which is counterfactual to recent studies on hysteresis and the scarring effects of protracted recessions (e.g., Cerra, Fatás, and Saxena, forthcoming; Aikman, Drehmann, Juselius, and Xing, 2022).

Figure 3: Comparison methodologies for output gap estimation



Finally, we compared the proposed BSVAR model with two models of the same type. That is models with the same identification setup (PT decomposition as in Uhlig, 2003). We consider an alternative BSVAR without a scaling adjustment for the COVID-19 episode and another where, instead of using a scale factor, the model is allowed to feature stochastic volatility (BSVAR-SV).

The associated output gaps and potential outputs of the two B-SVAR alternatives are shown in Figure 4 and 5. A large contrast between the baseline and the alternatives emerges at first sight: both the BSVAR (unadjusted) and the BSVAR-SV generate a less negative gap during the COVID-19 outbreak, which implies that the potential output is affected more drastically relative to our baseline model. In that sense, as with some of the simpler filters, the alternatives tend to overestimate the impact of the shock on the long-run output.

Regarding the volatility around the estimates, the BSVAR (unadjusted) displays the largest uncertainty, as reflected by percentile ranges that are twice as large as in the baseline. Nevertheless, the BSVAR-SV successfully mitigates volatility (at a similar range span as the baseline); however, it is the method where the potential output is affected the most during the downturn.

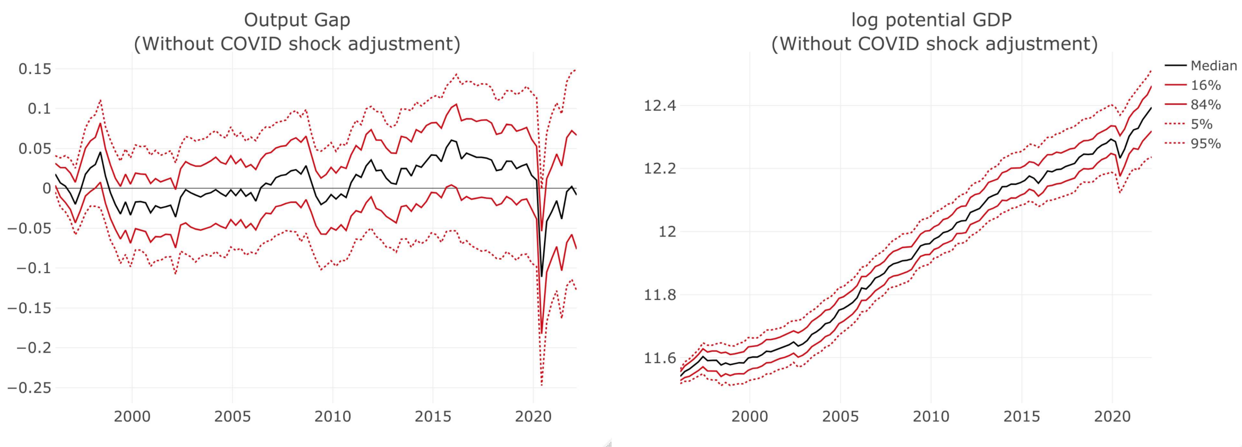


Figure 4: Output gap and Potential GDP- BSVAR without scale factor

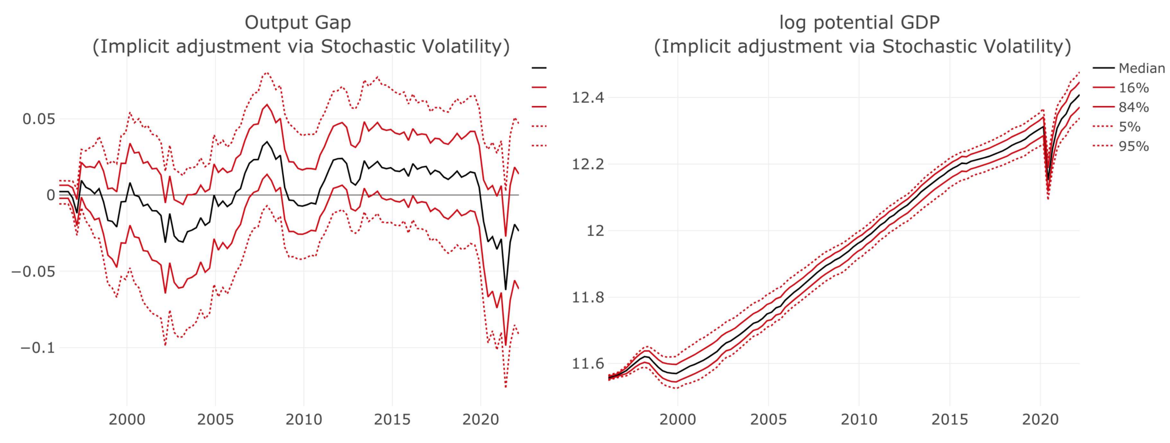


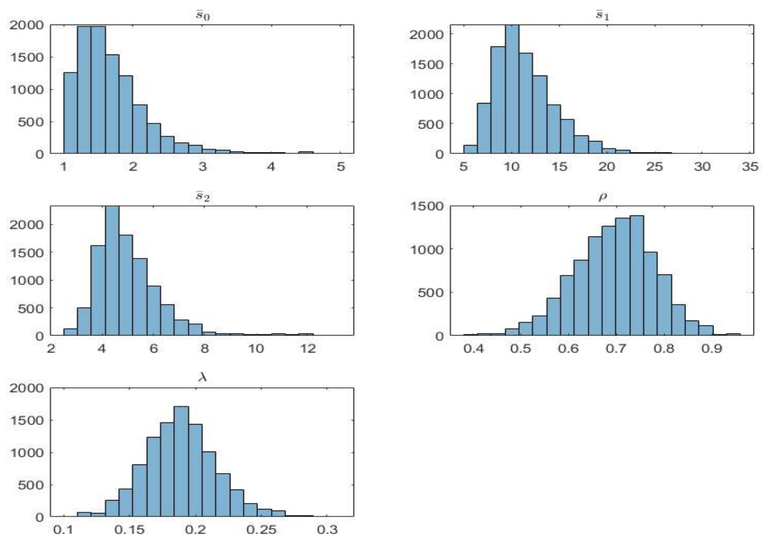
Figure 5: Output gap and Potential GDP-Stochastic volatility BSVAR without scale factor

3.2.2 Outlier observations around the COVID pandemic

Given that our main 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 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 estimate of scale factors (and shrinkage) in Figure 6.

The parameters posteriors are drawn based on a Metropolis Hastings algorithm with a Minnesota Prior. Thus, we estimated the scaling factors together with other hyperparameters in a hierarchical structure. The resulting posteriors for $\bar{s}_0, \bar{s}_1, \bar{s}_2$ peak around 1.5, 10, 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 0.75, which, together with $\bar{s}_2$, implies that the volatility scale factor falls by a third after 2020Q3 and then non-linearly towards one.[17]

Figure 6: Posterior distribution of the overall standard deviation of Minnesota prior and volatility scaling factors



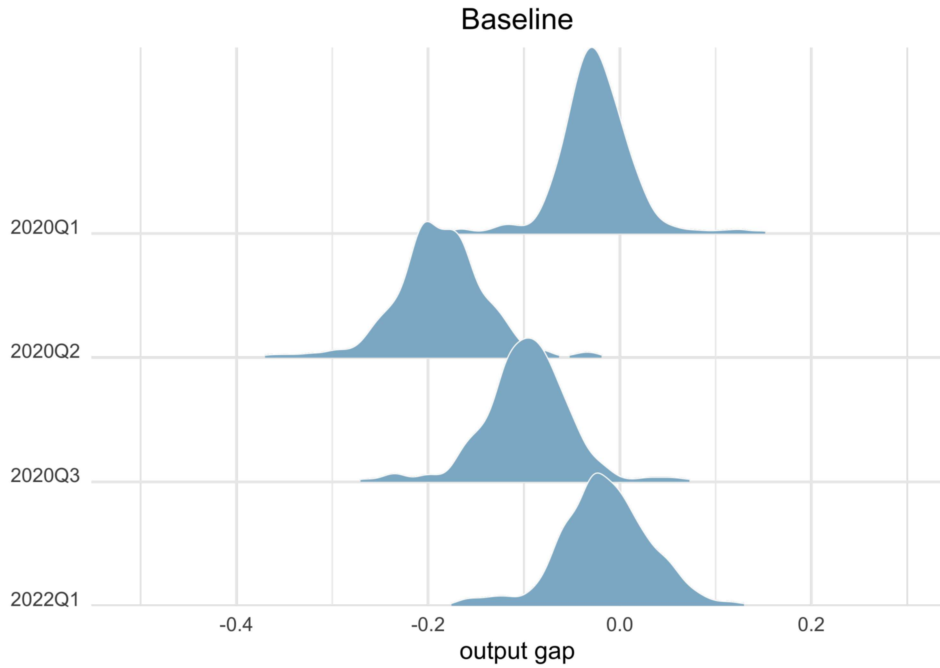
To further illustrate the impact of the COVID-19 shock on the output gap, we can depict the distributions of the draw estimates for dates around the episode, as shown in Figure 7. We reveal the quarter of the shock (2020Q1), the subsequent two quarters, and the first quarter of 2022 as a reference for a date when the potential output dynamics are, in principle, back to normal (here implicitly recognise the transitory nature of the pandemic shock).[18]

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.

[17] We also obtain the posterior for the shrinkage parameter of Minnesota prior (λ), depicting a mode around 0.19.

[18] As an additional exercise, we present a counterfactual exercise in Appendix D.1, where we discuss the gap and potential output that would have been observed in the absence of the pandemic shock.

Figure 7: Distribution of the output gap estimation during COVID-19 shock and 2022Q1.



4 Evaluation of the method

Evaluating the relative performance of our estimates is a hard task, given that our target, the potential GDP, is an unobserved variable. There is no well-defined target against which to perform a "horse race" using a set of competing methods.

However, an assessment of these methods is in order, and alternative evaluation methods can be proposed. These usually imply assuming knowledge of relevant features of the actual potential output that can be tested. One route taken by the literature (e.g., Chen and Gornicka (2020)) comprises setting up a Phillips Curve with the output gap on the right-hand side of the equation; subsequently, an estimation method of the output gap is assessed according to its capacity to forecast inflation in the context of the Phillips Curve. Here, we assume that the output gap is a relevant variable for determining inflation and that the relationship captured in the Phillips curve is stable over time; that is, the curve setup is an appropriate device for testing the relationship between the output gap and inflation.

Although that is a feasible venue, it also opens discussions about how stable the Phillips Curve in each country is considered, whether such a relationship exists (e.g., McLeay and Tenreyro (2020)) or if its slope has flattened over time (Hazell, Herreño, Nakamura, and Steinsson (2022)). These discussions are relevant more in recent times when the trade-off between output stabilisation and inflation is strongly felt worldwide. However, such debates are beyond the scope of our study and may divert attention from what we aim for in this study, approximating the potential output.

Alternatively, we take a more direct approach and assume to count with a real measure of the potential output and then approximate it with a set of methods whose estimates are assessed based on their co-movement with the actual output gap. We do this in the context of a Monte Carlo simulation, where the set of economic variables and potential output is simulated based on an economic model taken as given.

4.1 The model used to simulate the output gap

We consider a standard three-equation New Keynesian DSGE model along the lines of Benati (2008), but where the 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 associated log-linearized model is given by:

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

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

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

The first two equations, the hybrid Phillips curve, and dynamic IS feature both backward- and forward-looking components, whereas the monetary policy rule, is given by a Taylor rule with smoothing. π_t is inflation, R_t is the nominal rate, and the real GDP is Y_t which in the model is rescaled by its unit root component (Y^P) as $\hat{y}_t = \ln\left(Y_t/Y_t^P\right)$ to achieve stationarity. The latter implies that $\hat{y}_t$ is the output gap or the output as a deviation of the potential GDP given by its stochastic trend. The other variables were set as log deviations of their non-stochastic steady-state values.

The parameters of the model, $\Theta = \{\sigma_R^2, \sigma_u^2, \sigma_v^2, \kappa, \sigma, \alpha, \gamma, \rho, \phi_\pi, \phi_y\}$, were estimated using Bayesian methods. The posterior mode is found via simulated annealing, as in Benati (2008), and the posterior distribution of Θ is characterized by implementing a Random-Walk Metropolis-Hastings algorithm, as in An and Schorfheide (2007). Both simulated annealing and Metropolis simulations require the evaluation of the likelihood (and posterior) of the model based on its Sims canonical form and associated state-space representation.

Table 1 shows the parameters' priors, posterior modes, and percentiles obtained in our estimations. An additional step in the simulations is the scale factor adjustment of the variances, which is revised every 10% of the iterations and adjusted depending on the fraction of accepted draws in the subset draws. With that, the acceptance ratio of the simulation is 0.219.

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.

4.2 Evaluation method of the output gap estimations

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. For each iteration, the cross-correlation between each econometric estimate of the output gap and the simulated output gap of the model is calculated and recorded.

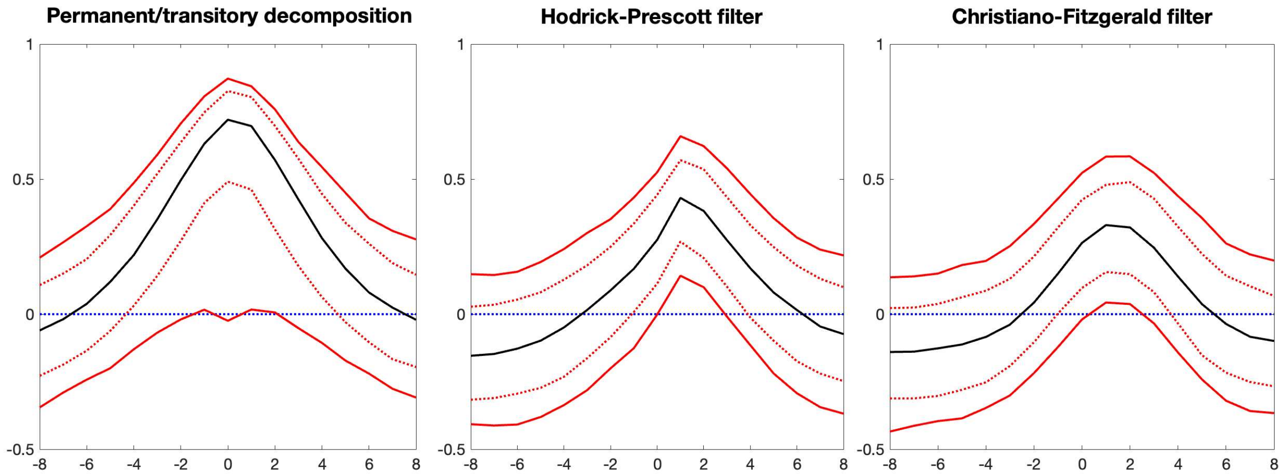
In other words, in the last simulation, we construct an output gap and feed the econometric methods with a set of other economic variables that are consistent with this gap. Then, we assess the methods using the co-movement between the estimated gap and the actual gap (simulated).

The methods compared are (i) the PT decomposition, (ii) a CF Band Pass type of filter, and (iii) an HP filter. The two filters are frequently used and widely available methods for estimating the potential output, whereas the PT decomposition method represents a relatively more complex alternative set to achieve an SVAR model structural identification as a function of its long-run forecasting performance.

The median cross-correlations are depicted in Figure 8. The results suggest that the HP filter is slightly better than the CF filter for low orders of correlation (up to two lags); however, the two methods perform similarly. Nonetheless, there is a noticeable improvement when using the PT decomposition. For example, the best of the other two methods' median is not even within the one standard confidence interval of the PT decomposition for lags of order 5 and up to leads of order one; moreover, the 68% confidence intervals between the best alternative (HP) and the PT decomposition do not overlap for the contemporaneous correlation and two first-order lags (0, -1, -2 in the horizontal axis of the plot).

Finally, at their peaks, the highest cross-correlations between the HP and CF filters do not exceed 0.5, whereas the PT at its peak features a correlation of 0.72%.

Figure 8: Cross-correlation between the output gap estimates and their simulated target



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

Given these results, we observe a substantial improvement in the output gap estimates when performing a PT decomposition to identify structural shocks in the BSVAR.

5 Concluding remarks

This study examined whether potential output models should be adjusted to account for rare, large-magnitude shocks, such as those experienced during the COVID-19 lockdown in 2020. It aimed to include a complete set of observations in the model while preventing observations of unprecedented magnitudes (that do not resemble the sample data-generating process) from affecting the quality of the resulting econometric modelling framework.

To address this question, we considered 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 behaviour of GDP (and potential output) and that the remaining shocks majorly explain transitory fluctuations (i.e. the gap). However, simulation exercises show that the adjusted model outperforms both simple filtering alternatives and similarly complex models that abstract from adjusting the large shock periods or that do so in alternative setups that do not explicitly account for outlying observations at the specific dates of the high-magnitude shocks (e.g. models with stochastic volatility). Concurrently, our setup prevents quick output gap reversals after downturns or drastic changes in the potential output after high-magnitude transitory observations. In that sense, while our setup aligns with the findings of recent studies on the scarring effects of economic downturns (e.g., Cerra, Fatás, and Saxena (forthcoming) Aikman, Drehmann, Juselius, and Xing, 2022), it still prevents the unprecedented-magnitude observations from affecting the resulting model substantially.

It is relevant to mention that we can make a good approximation of the potential GDP (and gap) by trading off the possibility of disentangling output dynamics into separate drivers (e.g. monetary, financial, global, supply, and demand). Not being able to carry out such a type of exercise is the cost of accessing our identification strategy, which is strictly concerned with an endogenous determination of the horizon profile of the structural shocks. In that spirit, a separation of the output dynamics into fundamental drivers is left for future research, where we can draw lessons from the results of this study that allow mitigating the approximation costs of *ad-hoc* changes in the term horizon of the shocks, common to other identification setups.

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.

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

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.

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.

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.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022, 06). Addressing COVID-19 Outliers in BVARs with Stochastic Volatility. *The Review of Economics and Statistics 105*, 1–38.

Cerra, V., A. Fatás, and S. Saxena (forthcoming). Hysteresis and Business Cycles. *Journal of Economic Literature*.

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

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.

Dieppe, A., N. Francis, and G. Kindberg-Hanlon (2021). The identification of dominant macroeconomic drivers: coping with confounding shocks. Technical report, ECB Working Paper No. 2534.

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.

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

Furlanetto, F., A. Lepetit, Robstad, J. F. Rubio-Ramírez, and P.-b. Ulvedal (2021, September). Estimating Hysteresis Effects. CEPR Discussion Papers 16558, C.E.P.R. Discussion Papers.

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

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.

Guisinger, A. Y., M. T. Owyang, and H. Shell (2018). Comparing Measures of Potential Output. *Review 100*(4), 297–316.

Hartwig, B. (2022). Bayesian vars and prior calibration in times of covid-19. *Studies in Nonlinear Dynamics Econometrics*.

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.

Jordà, O., S. R. Singh, and A. M. Taylor (2020, January). The Long-Run Effects of Monetary Policy. NBER Working Papers 26666, National Bureau of Economic Research, Inc.

Keating, J. W. and V. J. Valcarcel (2015). The time-varying effects of permanent and transitory shocks to real output. *Macroeconomic Dynamics 19*(3), 477–507.

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.

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).

Ng, S. (2021). Modeling macroeconomic variations after covid-19. Technical report, National Bureau of Economic Research.

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.

A Additional descriptive data

Table 2: Unit root test

Variable-Test	ADF		PP		KPSS		ERS	
	Level	First Diff	Level	First Diff	Level	First Diff	Level	First Diff
GDP	0.9609	0.0000	0.9684	0.0000	1.16079	0.101328	198.0309	0.505899
CON	0.9962	0.0000	0.9984	0.0000	1.152888	0.313152	158.1545	0.478252
INV	0.8669	0.0000	0.842	0.0000	0.971256	0.145345	26.53758	0.495762
GOV	0.8951	0.0000	0.6729	0.0000	1.188665	0.136647	715.5667	0.679222
CPI	0.2782	0.0006	0.4451	0.0000	0.826933	0.075889	45.25559	0.030919
TIB	0.0948	0.0000	0.0664	0.0000	0.77068	0.322704	9.37788	0.866045
ITCR	0.3569	0.0000	0.2941	0.0000	0.723053	0.222996	75.03317	0.92045
OIL	0.1299	0.0000	0.2149	0.0000	0.246352	0.069921	1.960358	0.556965

Note: * For ADF and PP the data in table corresponds to p-values, and the test statistic are reported for KPSS (1%: 0.739, 5%:0.463, 10%: 0.347) and ERS (1%: 1.9472, 5%:3.1142, 10%: 4.1812)
Source: Authors' calculations.

Table 3: Cointegration test

	GDP CON INV				GDP CON INV GOV CPI TIB ITCR OIL			
	Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	p-value*	Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	p-value*
	None	0.191276	30.49843	0.0415				
	At most 1	0.081933	8.844121	0.38	None *	0.556582	260.3158	0
Unrestricted Cointegration Rank Test (Trace)	At most 2	0.001222	0.124712	0.724	At most 1 *	0.455536	177.3651	0
					At most 2 *	0.342743	115.3538	0.0012
					At most 3 *	0.238327	72.54651	0.0298
					At most 4	**0.199753**	**44.7783**	**0.0946**
					At most 5	0.131228	22.0492	0.2958
					At most 6	0.066799	7.700448	0.498
					At most 7	0.00634	0.648709	0.4206
	Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	p-value*	Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	p-value*
	None	0.191276	21.6543	0.0422	None *	0.556582	82.95069	0
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)	**At most 1**	**0.081933**	**8.719409**	**0.3103**	At most 1 *	0.455536	62.01128	0.0005
	At most 2	0.001222	0.124712	0.724	At most 2 *	0.342743	42.80733	0.024
					At most 3	**0.238327**	**27.7682**	**0.2244**
					At most 4	0.199753	22.72911	0.1853
					At most 5	0.131228	14.34875	0.3371
					At most 6	0.066799	7.051739	0.483
					At most 7	0.00634	0.648709	0.4206
*MacKinnon-Haug-Michelis (1999) p-values								

Source: Authors' calculations.

B 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).

C Baseline model: Results

C.1 Forecast Error Variance decomposition (FEVs)

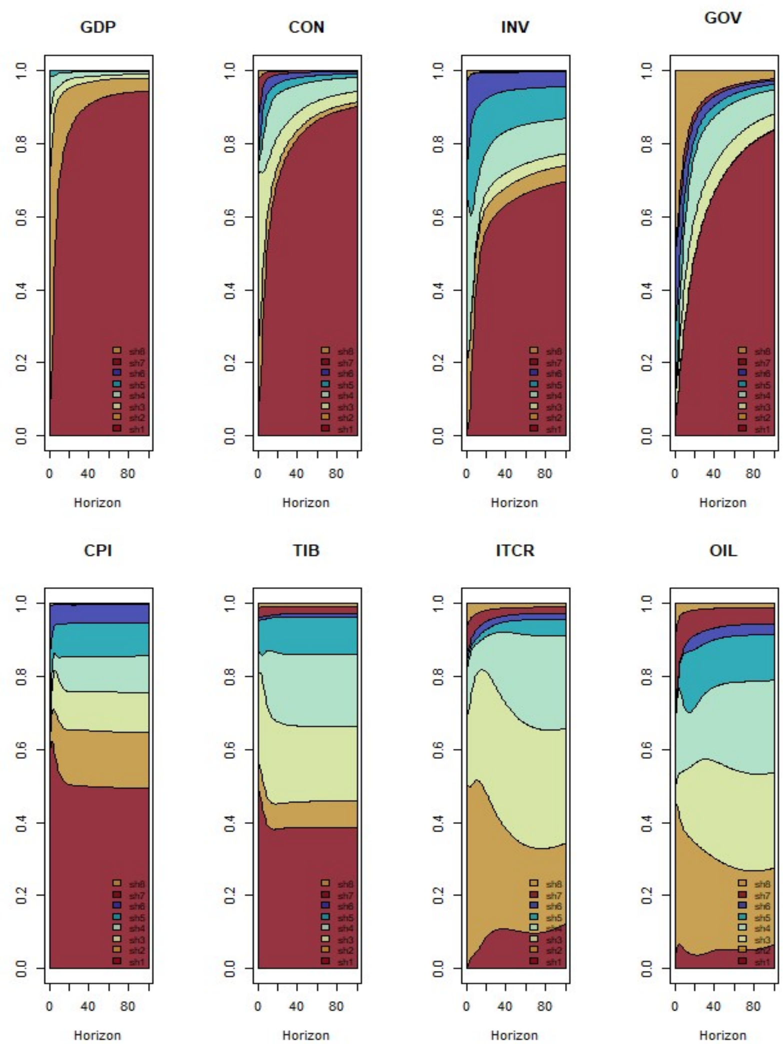


Figure 9: Forecast Error Variance decomposition (FEVs) by variable

D Other models: Results

D.1 Counterfactual exercise (no covid) vs baseline model

In this appendix we compute a counterfactual output gap where the COVID downturn is abstracted from. We run the B-SVAR model (with the proposed identification setup) until December 2019 and forecast nine quarters ahead. Then we obtain the gap as the transitory component of the GDP.

In a world without COVID we obtain the potential GDP would have grown 3.6% in 2020, i.e., 1.4% higher than in our baseline model with COVID (which grows by 2.2%). In terms of the output gap, the counterfactual shows a median around zero which contrasts with the estimates of the COVID period that show a median of -18.9.

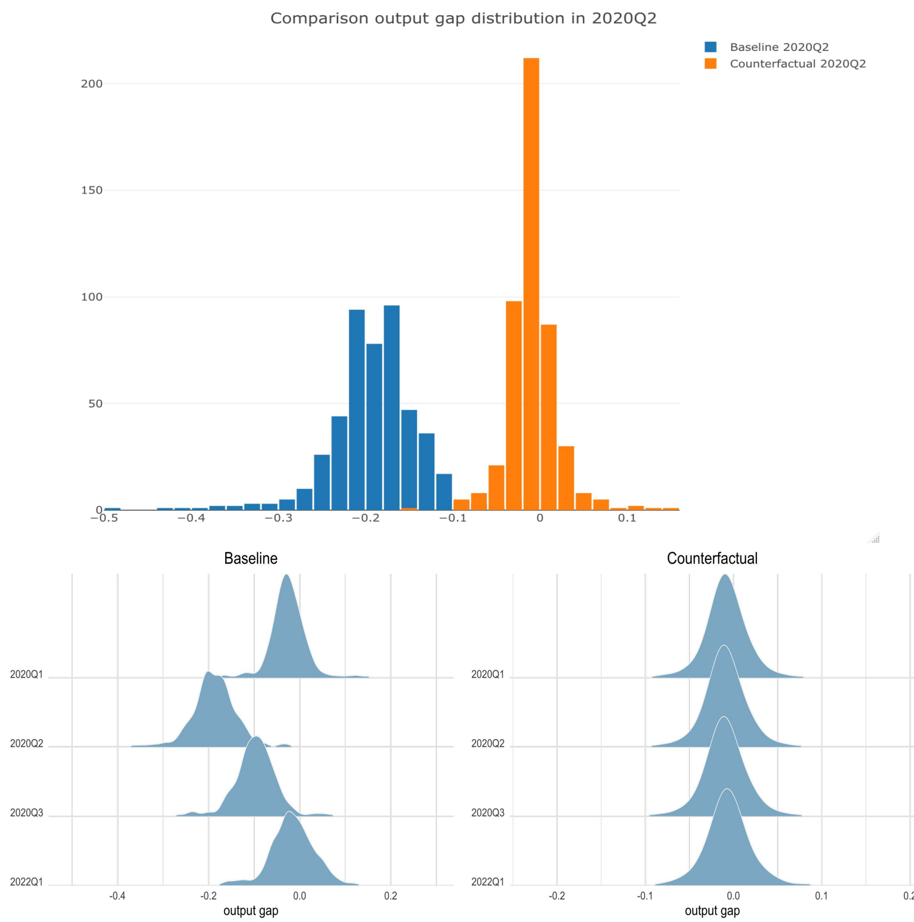


Figure 10: Counterfactual exercise vs baseline model