Forecasting Brazilian GDP under Fiscal Foresight with a Noncausal Fiscal VAR

Luan Borelli
Brazilian School of Economics and Finance (FGV EPGE)

Christian Vonbun
Institute for Applied Economic Research (IPEA)

Abstract

Due to fiscal foresight, standard fiscal VAR models are inherently susceptible to issues of nonfundamentalness and noncausality, which can result in invalid estimates. While these problems have been extensively addressed in the fiscal literature, they have largely been overlooked in Brazilian fiscal VAR studies. To address this gap, we estimate a noncausal fiscal VAR model for Brazil—an alternative specification that may correct these issues—and use it to forecast Brazilian GDP. The results show that the noncausal VAR model outperforms the standard purely causal VAR in terms of forecasting performance, particularly when considering the typical Brazilian fiscal VAR dataset. This suggests that fiscal expectations may play a crucial role in shaping the dynamics of Brazilian GDP.

Keywords: VAR, fiscal foresight, nonfundamentalness, noncausality.

JEL Classification: E62, H30, C53.

1 Introduction

The seminal work by Sims (1980) sparked the widespread use of vector autoregressive (VAR) models in empirical macroeconomic studies, quickly becoming a cornerstone in the economics literature. It wasn't long before this class of models became a fundamental tool for a wide range of macroeconomic analyses, including fiscal ones. Fiscal VAR models have been widely adopted for forecasting and estimating fiscal multipliers. However, the empirical literature on these models has yet to reach a consensus that can guide policymakers, with significant disagreement remaining on issues such as the size of the multipliers and the effects of fiscal shocks on key macroeconomic aggregates.

One possible cause of this inability to reach a consensus may lie in the very nature of fiscal policy. Due to legislative and procedural delays, fiscal policies are typically implemented only after a significant lag following their public announcements, allowing agents' expectations to adjust immediately and giving rise to fiscal foresight. Unlike other types of macroeconomic policies, this gives fiscal policy the potential for noncausality, which, in turn, leads to the issue of nonfundamentalness in fiscal data.[1]

At a theoretical level, nonfundamentalness occurs when the MA representation of a VAR model is not invertible. This can arise, among other potential causes, when the information set available to agents in the real economy is broader than the information set available to the econometrician. Noncausality, on the other hand, occurs when the roots of the AR characteristic polynomial lie within the unit circle. As a result, to be recovered, the model's dynamics depend not only on past (lags) and present values of economic variables but also on their future values (leads).

Both phenomena are closely connected. An econometrician's inability to observe all the information available to agents is, in most cases, linked to the role of expectations. Economic agents can observe both economic fundamentals and their own expectations, making decisions based on this comprehensive information set. Econometricians, however, while having access to data records on economic fundamentals, rarely have access to reliable historical data on market expectations. Consequently, the information sets of econometricians and agents often diverge due to expectational components, leading to nonfundamentalness.[2] Since expectations can be interpreted as the process of anticipating the future to determine the present, the close connection between nonfundamentalness and noncausality becomes evident.

This connection is particularly relevant in the fiscal context, where nonfundamentalness is strongly associated with fiscal foresight, which is closely tied to expectations about future fiscal aggregates. According to Leeper et al. (2008), although macroeconomists acknowledge the possibility of fiscal foresight and theoretically develop its implications, empirical models are often not grounded in theory. This poses a significant challenge to

[1]Indeed, even Blanchard and Perotti (1999), in their seminal paper on fiscal multipliers, foreseen potential issues arising from the anticipation of fiscal policy through expectations: "Implicit in our approach is the assumption that the policy innovations we have estimated were indeed unanticipated by the private sector. While we share this assumption with the whole VAR literature, we recognize that it is particularly problematic here: most of the changes in tax and transfer programs are known at least a few quarters before they are implemented. We do not have a general solution to this problem." Blanchard and Perotti (1999), p. 18.

[2]Nonfundamentalness, however, can also arise from other causes. See Alessi et al. (2008) for a survey of these causes.

the econometric analysis of fiscal policy: the widespread occurrence of time series with a non-invertible moving average component.[3] In other words, fiscal foresight engenders nonfundamentalness.

The presence of such phenomena exposes traditional models considered by econometricians to potential biases of varying directions and magnitudes (Canova and Sahneh, 2018; Gourieroux and Jasiak, 2016; Leeper et al., 2013; Alessi et al., 2008; Fernández-Villaverde et al., 2007). Depending on the case, this can result in either an underestimation or overestimation of policy effects. The bias can be substantial enough to reverse the sign of impulse-response functions. Similarly, forecasts produced by econometricians using models affected by nonfundamentalness may be inconsistent and suboptimal. This problem spurred the development of testing procedures for both noncausality (Sahneh, 2015) and nonfundamentalness (Forni and Gambetti, 2014; Canova and Sahneh, 2018; Sahneh, 2015). Additionally, alternative model specifications have emerged in the literature to address these issues when they are identified. One approach to handling noncausality in data (and, consequently, nonfundamentalness) is the use of a noncausal VAR model.[4]

The noncausal VAR model extends the standard (purely causal) VAR specification by incorporating not only present and past values (lags) of economic variables but also their future values (leads), capturing expectational components. This framework was first introduced by Lanne and Saikkonen (2013), generalizing the literature on univariate noncausal autoregressive models to encompass general multivariate processes.[5] Building on this work, Nyberg and Saikkonen (2014) developed a forecasting procedure for noncausal VAR models, extending the univariate forecasting methodology of Lanne et al. (2012) to general multivariate settings. Further advancements include the work of Gourieroux and Jasiak (2016), who proposed a state-space representation for these models, and Lanne and Luoto (2016), who introduced a noncausal Bayesian VAR.

Despite the growth of this literature, Brazilian research on fiscal VARs still does not appear to show significant concern with the problem. Aside from the work of Vonbun and Lima (2020), who applied the tests by Forni and Gambetti (2014) and Canova and Sahneh (2018) to identify nonfundamentalness in typical Brazilian fiscal VAR datasets, no other studies seem to have effectively considered these phenomena. Although Vonbun and Lima (2020) diagnosed the presence of nonfundamentalness in Brazilian fiscal data, their attempt to address it by widening the information set—one of the potential methods for handling nonfundamentalness—was unsuccessful.[6]

[3]In their words: "Fiscal foresight poses a formidable challenge because, as Yang (2005) shows, it generates an equilibrium with a non-invertible VARMA representation. Non-invertibility, in turn, implies that the fundamental shocks to tax policy cannot be recovered from current and past observable data, a central assumption of conventional econometric methods." Leeper et al. (2008), page 2.

[4]Other approaches include, for example, augmenting the model with additional data, such as proxies for forward-looking variables. This strategy aims to expand the model's information set to make it informationally sufficient, thereby rendering the MA component invertible and resolving the issue. Identifying suitable proxies and validating their relevance, however, is far from straightforward. Vonbun and Lima (2020), for instance, attempted to address nonfundamentalness in Brazilian fiscal data using this method but were unsuccessful.

[5]For univariate noncausal models, see, e.g., Breid et al. (1991), Lanne and Saikkonen (2011), Lanne et al. (2012), Lanne et al. (2011), and Hecq et al. (2020).

[6]The objective of Vonbun and Lima (2020) was to test the representative fiscal VAR model from the Brazilian literature to assess the presence of nonfundamentalness, which, as previously discussed, is closely related to noncausality in the fiscal context. Their main analysis included variables such as GDP, public investment, public expenditure, and tax revenue. The authors also highlighted the

The objective of this paper is to properly address the issues of fiscal foresight, nonfundamentalness, and noncausality in Brazilian fiscal data by employing a noncausal fiscal VAR model using the fiscal dataset typically considered in the Brazilian fiscal VAR literature.

Although noncausal VAR models are better suited for addressing specification issues arising from noncausality, at the time of writing this paper, they unfortunately lacked a structural-form solution. Significant challenges remain in obtaining identification relations for their structural representations, which prevents the computation of impulse-response functions.[7] Nonetheless, the reduced-form estimates of noncausal VAR models can still be used for forecasting. Depending on their performance, these estimates may provide evidence of noncausality (and, by extension, nonfundamentalness) in fiscal data.

To evaluate the importance of considering noncausality and nonfundamentalness, we estimated a noncausal VAR model and used it to compute pseudo-out-of-sample forecasts for GDP. To achieve this, we submitted data to the model selection procedure proposed by Lanne and Saikkonen (2013) to determine the number of lags and leads that best characterize the dynamics of the variables we considered. A purely forward-looking noncausal VAR was selected. When estimated, the model displayed relevant statistical significance for all variables as determinants of fiscal dynamics, potentially evidencing the importance of considering nonfundamentalness and noncausality in fiscal data. We then compared the performance of a standard (purely causal) VAR model with that of the selected noncausal model. The forecasts obtained from the noncausal model consistently outperformed those of the standard purely causal model across almost all predictive accuracy metrics and forecast horizons considered. This superiority was notably pronounced for shorter horizons—especially for one-quarter ahead forecasts.

This paper is organized as follows. Section 2 motivates our discussion by formally introducing the concepts of fiscal foresight, nonfundamentalness, and noncausality, along with their interrelationships. Section 3 details the noncausal VAR model, discussing its selection, estimation, and forecasting procedures. Section 4 focuses on our empirical application, covering data description, model selection, and forecast results using our estimated model. Finally, Section 5 provides concluding remarks.

2 Fiscal Foresight, Nonfundamentalness and Noncausality

To contextualize the importance of noncausal models, this section formally introduces the concepts of fiscal foresight, nonfundamentalness, and noncausality, along with their interrelationship. The discussion presented here closely follows Nelimarkka (2019). The formal definition of the noncausal econometric model to be empirically estimated in this work is deferred to Section 3.

significance of public debt and the debt/GDP ratio as important variables. These variables were included experimentally to address nonfundamentalness (and, by extension, noncausality), but the attempt was unsuccessful. They also sought to incorporate market expectations using FOCUS data from the Central Bank of Brazil, though this effort did not yield positive results either.

[7] During the review process, we learned that Nelimarkka (2019) had developed a method for computing impulse-response functions using noncausal VAR models. However, due to time constraints, we were unable to implement this method. As a result, we defer the computation of impulse-response functions to future research, focusing this paper on forecasting.

In order to formally illustrate the problem of nonfundamentalness, consider the equilibrium of a linearized macroeconomic model for k observable variables in y_t with a vector autoregressive moving average (VARMA) representation, as presented in Nelimarkka (2019):

$$A(L)y_t = B(L)u_t, \tag{1}$$

where $A(L) = I - A_1L - \cdots - A_pL^p$, $B(L) = B_0 + B_1L + \cdots + B_dL^d$, and u_t is a vector containing the k uncorrelated structural shocks driving the economy such that $\mathbb{E}_t[u_{t+j}] = 0$ when $j > 0$ and $\mathbb{E}_t[u_{t+j}] = u_{t+j}$ for $j \leq 0$. $\mathbb{E}[\cdot]$ denotes the expectation conditional on the information set of the agents. $A(L)$ and $B(L)$ are $k \times k$ matrix polynomials, with L being the usual lag operator, that determine the unique equilibrium of the model in terms of finite lags up to a truncation. $A(L)$ is assumed to be stable, implying an MA representation $y_t = A(L)^{-1}B(L)u_t$.

When the MA polynomial $B(L)$ in (1) is invertible in the past, i.e. $|B(z)|$ does not have roots within the unit circle, structural shocks and impulse response functions can be obtained from the reduced-form error term $\varepsilon_t = B_0u_t$ with a standard purely causal VAR(p) model $C(L)y_t = \varepsilon_t$, where $C(L) = I - C_1L - \cdots - C_pL^p$, after the imposition of identification restrictions in the matrix B_0.

However, as Nelimarkka (2019) points out, under nonfundamentalness the polynomial $B(L)$ is not invertible in the past, which implies that there is no VAR(∞) representation to recover shocks u_t only from the past history of y_t. In this case, fitting a standard VAR model to y_t produces a nonfundamental error term which is a linear combination of the past shocks (Lippi and Reichlin, 1994; Fernández-Villaverde et al., 2007), distorting conclusions drawn from the estimated impulse response functions.

The nonfundamentalness problem may arise for multiple reasons, but typically it is closely linked to the effects of expectations unobservable to the econometrician on the actions of agents, which in turn influence macroeconomic variables. This potential source of nonfundamentalness occurs when the econometrician's information set diverges from that of the economy's real agents due to insufficient information about future variables—namely, expectations. This divergence leads to (and defines) noncausality. Consequently, both phenomena—nonfundamentalness and noncausality—can be understood as informational issues that, in practice, often manifest as different forms of omitted variable bias.

To illustrate the relationship between fiscal foresight, nonfundamentalness, and noncausality, we follow the example provided by Nelimarkka (2019), which encompasses both fundamental and nonfundamental processes as particular cases. This framework assumes that the observable variables y_t include all state variables except for k uncorrelated exogenous variables in z_t.[8] From Sims (2002), when the MA component is noninvertible, y_t has a general forward-looking solution[9]

$$y_t = \Theta_1 y_{t-1} + \Theta_c + \Theta_0 z_t + \Theta_y \sum_{s=1}^{\infty} \Theta_f^{s-1} \Theta_z \mathbb{E}_t z_{t+s}. \tag{2}$$

[8]See Nelimarkka (2019), pages 104 and 107.

[9]As reported by Nelimarkka (2019), "matrices Θ_1, Θ_c, Θ_0, Θ_y, Θ_f, Θ_z are functions of parameters of the model of dimensions $k \times k$, $k \times 1$, $k \times k$, $k \times m$, $m \times m$ and $m \times k$, respectively. m is a dimension of the unstable block of the system, defined in Sims (2002)."

Exogenous variables are driven by unanticipated shocks when $z_t = u_t$. In this case, a standard purely causal VARMA representation is obtained as a consequence of the vanishing of the last term resulting directly in a VAR(1) representation. On the other hand, when agents have q-periods ahead foresight of exogenous variables, $z_t = u_{t-q}$ and the equilibrium is determined by

$$y_t = \Theta_1 y_{t-1} + \Theta_c + \Theta_0 u_{t-q} + \Theta_y \Theta_z u_{t-q+1} + \Theta_y \Theta_f \Theta_z u_{t-q+2} + \cdots + \Theta_y \Theta_f^{q-1} \Theta_z u_t. \quad (3)$$

This representation corresponds to the VARMA representation (1) of the model. As noted by Nelimarkka (2019), despite the fact that the more distant expected events of z_t obtain a weaker weight in the forward looking solution (2), the most recent innovation u_t informative about the future event z_{t+q} is discounted the heaviest by a factor of $\Theta_y \Theta_f^{q-1} \Theta_z$ in (3). This reverse discount easily causes the noninvertibility of the MA polynomial $B(L) = \Theta_0 L^q + \Theta_y \Theta_z L + \cdots + \Theta_y \Theta_f^{q-1} \Theta_z$, with the most recent shocks having the least influence on the overall dynamics of y_t.

The noninvertible solution prevents the recovery of the news shock contained in u_t based only on the present and past values of y_t. As demonstrated by Nelimarkka (2019), it is possible to rewrite the VARMA model (1) under noninvertibility as a noncausal autoregressive representation of y_t, representing structural shocks in terms of both past and future terms.[10] Let l roots of $B(z)$ lie within the unit circle. Then y_t can be written as the following noncausal model:

$$\bar{c}_l \beta(L)^{-1} B^{adj}(L) \alpha(L^{-1})^{-1} A(L) y_t = u_{t-l}, \quad (4)$$

where $\bar{c}_l$ is constant, $B^{adj}(L)$ is the adjoint matrix of $B(z)$, and $\alpha(z^{-1})^{-1}$ and $\beta(z)^{-1}$ are scalar convergent power series expansions in z^{-1} and z, respectively. Through the lead polynomial $\alpha(z^{-1})^{-1}$, the time-shifted structural shocks u_{t-l} are functions of the past, current and future terms of y_t.

While the history of y_t lacks enough information to capture the variation of u_t, movements of the lagged shocks are captured by a linear weighted sum of the past and future values of y_t. Consequently, both the lags and the leads of the observable variables are sufficient to recover the structural shocks that are now anticipated due to the time-shifting.

In this theoretical framework, agents' foresight capacity gives rise to nonfundamentalness, which, in this context, may manifest as noncausality. A model is considered noncausal when its dynamics depend not only on the present and past values (lags) of economic variables but also on future values (leads) that reflect expectational components. In the fiscal setting, nonfundamentalness is typically associated with fiscal foresight, which relates to expectations about the future of fiscal aggregates. As a result, the issue often arises from omitted variables tied to these expectations, highlighting a strong link between nonfundamentalness and noncausality. Diagnosing nonfundamentalness can thus provide a way to identify noncausality, while addressing noncausality offers a potential solution to nonfundamentalness.

[10] For further details, see Nelimarkka (2019), Appendix 4.A.

3 The Noncausal VAR model

This exposition closely follows Lanne and Saikkonen (2011, 2013).

3.1 Specification

As Nelimarkka (2019) pointed out, “direct inference on the noncausal representation (4) is infeasible.”[11] Lanne and Saikkonen (2013) solve this problem by considering an n-dimensional stochastic process y_t specified as

$$\Phi(L)\Psi(L^{-1})y_t = \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \tag{5}$$

where $\Phi(L) = I_n - \Phi_1 - \cdots - \Phi_r L^r$ and $\Psi(L^{-1}) = I_n - \Psi_1 L^{-1} - \cdots - \Psi_s L^{-s}$ are $n \times n$ matrix polynomials in the lag operator L and ϵ_t is a $n \times 1$ sequence of independent, identically distributed (continuous) random vectors with zero mean and finite positive definite covariance matrix. Moreover, to ensure stationarity and the existence of an MA representation, the matrix polynomials $\Phi(z)$ and $\Psi(z)$ are assumed to have their zeros outside the unit disc, so that

$$\det\Phi(z) \neq 0, |z| \leq 1 \quad \text{and} \quad \det\Psi(z) \neq 0, |z| \leq 1. \tag{6}$$

If $\Psi_j \neq 0$ for some $j \in \{1, \dots, s\}$, equation (5) defines a noncausal vector autoregression, referred to as purely noncausal when $\Phi_1 = \cdots = \Phi_r = 0$. The corresponding standard purely causal model is obtained when $\Psi_1 = \cdots = \Psi_s = 0$. In this case, the former condition in (6) guarantees the stationarity of the model. In the general setup of equation (5) the same is true for the process

$$u_t = \Psi(L^{-1})y_t. \tag{7}$$

Specifically, there exists a $\delta_1 > 0$ such that $\Phi(z)^{-1}$ has a well-defined power series representation $\Phi(z)^{-1} = \sum_{j=0}^{\infty} M_j z^j = M(z)$ for $|z| < 1 + \delta_1$. Consequently, the process u_t has the causal moving average representation

$$u_t = M(L)\epsilon_t = \sum_{j=0}^{\infty} M_j \epsilon_{t-j}. \tag{8}$$

Notice that $M_0 = I_n$ and that the coefficient matrices M_j decay to zero at a geometric rate as $j \to \infty$ (cf. Kohn, 1979, Lem. 3). When convenient, Lanne and Saikkonen (2013) assumes $M_j = 0$ for $j < 0$.

Writing $\Phi(z)^{-1} = (\det\Phi(z))^{-1}\Xi(z) = M(z)$, where $\Xi(z)$ is the adjoint polynomial matrix of $\Phi(z)$, the authors obtain the relation $\det\Phi(L)u_t = \Xi(L)\epsilon_t$ and, by the definition of u_t,

$$\Psi(L^{-1})w_t = \Xi(L)\epsilon_t, \tag{9}$$

[11]Nelimarkka (2019), p. 107.

where $w_t = \det(\Phi(L))y_t$. Notice that $\Xi(z)$ is a matrix polynomial with degree at most $(n-1)r$ and, since $\Phi(0) = I_n$, we have also $\Xi(0) = I_n$. By the latter condition in (6) one can find a $0 < \delta_2 < 1$ such that $\Psi(z^{-1})^{-1}\Xi(z)$ has a well-defined power series representation $\Psi(z^{-1})^{-1}\Xi(z) = \sum_{j=-(n-1)r}^{\infty} N_j z^{-j} = N(z^{-1})$ for $|z| > 1-\delta_2$. Thus, the process w_t has the representation

$$w_t = \sum_{j=-(n-1)r}^{\infty} N_j \epsilon_{t+j}, \tag{10}$$

where the coefficient matrices N_j decay to zero at a geometric rate as $j \to \infty$.

From (6) it follows that the process y_t itself has the representation

$$y_t = \sum_{j=-\infty}^{\infty} \Pi_j \epsilon_{t-j}, \tag{11}$$

where Π_j is the $n \times n$ coefficient matrix of z^j in the Laurent series expansion of $\Pi(z) \equiv \Psi(z^{-1})^{-1}\Phi(z)^{-1}$ that exists for $1-\delta_2 < |z| < 1+\delta_1$ with Π_j decaying to zero at a geometric rate as $j \to \infty$.

It is easy to see that the representation (11) can be obtained by multiplying both sides of (10) by $\det(\Phi(L))^{-1}$ or by multiplying both sides of (8) by $\Psi(L^{-1})^{-1}$. Therefore, we arrive at the following equivalence of MA(∞) representations:

$$y_t = \det(\Phi(L))^{-1} \sum_{j=-(n-1)r}^{\infty} N_j \epsilon_{t+j} = \Psi(L^{-1})^{-1} \sum_{j=0}^{\infty} M_j \epsilon_{t-j} = \sum_{-\infty}^{+\infty} \Pi_j \epsilon_{t-j}. \tag{12}$$

This equivalence shows the possibility of separating forward looking and backward looking unilateral representations from bilateral representation (11).

Lanne and Saikkonen (2013) notice that the representation (11) implies that y_t is a stationary and ergodic process with finite second moments. We adopt the notation VAR(r,s) for the model defined by (5). In the causal case $s = 0$, the standard notation VAR(r) is also adopted.

In the noncausal case (i.e., $\Pi_j \neq 0$ for some $j < 0$) the connection between the noncausal VAR model and nonfundamentalness previously discussed is evidenced. In order to visualize greater implications of noncausality, consider the following illustration by Lanne and Saikkonen (2013). Denote by $\mathbb{E}_t[\cdot]$ the conditional expectation operator with respect to the information set $\{y_t, y_{t-1}, \dots\}$ and, from (5) and (11), write

$$y_t = \sum_{-\infty}^{s-1} \Pi_j \mathbb{E}_t[\epsilon_{t-j}] + \sum_{j=s}^{\infty} \Pi_j \epsilon_{t-j}. \tag{13}$$

In the standard causal case, $s = 0$ and $\mathbb{E}_t[\epsilon_{t-j}] = 0$ for $j \leq -1$, so that the right-hand side reduces to the moving average representation (8). However, in the noncausal case, this does not happen. In this case, $\Pi_j \neq 0$ for some $j < 0$, which in conjunction with the representation (11) shows that y_t and ϵ_{t-j} are correlated. Consequently, $\mathbb{E}[\epsilon_{t-j}] \neq 0$ for some $j < 0$, implying that future errors can be predicted by past values of the process y_t. This can be seen as an alternative characterization of nonfundamentalness.

In addition to the dependence on expected future errors, the process y_t can also be interpreted as being dependent on its expected future values. To visualize this, Lanne and Saikkonen (2013) focus, for simplicity, on the purely noncausal model, where $\Phi(L) = I_n$. In this case, the model (5) can be written as:

$$y_t = \Psi_1 y_{t+1} + \cdots + \Psi_s y_{t+s} + \epsilon_t, \tag{14}$$

and, taking conditional expectations with respect to the information set $\{y_t, y_{t-1}, \dots\}$, one obtains:

$$y_t = \Psi_1 \mathbb{E}_t[y_{t+1}] + \cdots + \Psi_s \mathbb{E}_t[y_{t+s}] + \mathbb{E}_t[\epsilon_t]. \tag{15}$$

This shows, according to Lanne and Saikkonen (2013), that the elements of the coefficient matrix Ψ_j give the effect of the expectation of y_{t+j} on y_t. In the general case ($\Psi(L) \neq I_n$), we obtain a similar expression for y_t with the exception that $\mathbb{E}_t[\epsilon_t]$ is replaced by $\mathbb{E}_t[u_t]$.

3.2 Gaussian Indistinguishability

A practical complication of noncausal models is that they cannot be distinguished from standard causal models through second-order properties of Gaussian likelihoods. This occurs because, assuming Gaussian errors, the spectral density of a VAR(r, s) model specified according to equation (5) is given by

$$(2\pi)^{-1}[\Psi(e^{i\omega})'\Phi(e^{-i\omega})'\mathbb{C}[\epsilon_t]^{-1}\Phi(e^{i\omega})\Psi(e^{-i\omega})]^{-1}, \tag{16}$$

where $\mathbb{C}[\cdot]$ denotes the covariance operator. In this expression, the matrix in the brackets is 2π times the spectral density matrix of a second order stationary process whose autocovariances are equal to zero at lags greater than $r+s$. As is well known, this process has an invertible MA representation of order $r+s$, implying that the same spectral density can be obtained from a purely causal VAR($r+s$) model specified as

$$\Phi(L)\Psi(L)y_t = \xi_t, \tag{17}$$

where ξ_t is a vector of sequences of stationary innovations, uncorrelated, but, in general, not independent, with zero mean and constant variance.

Specifically, by a slight modification of Theorem 10' of Hannan (1970), Lanne and Saikkonen (2013) get the unique representation

$$\Psi(e^{i\omega})'\Phi(e^{-i\omega})'\mathbb{C}[\epsilon_t]^{-1}\Phi(e^{i\omega})\Psi(e^{-i\omega}) = \left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{-i\omega}\right)' \left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{i\omega}\right), \tag{18}$$

where the $n \times n$ matrices $\mathcal{C}_0, \ldots, \mathcal{C}_{r+s}$ are real with $\mathcal{C}_0$ positive definite, and the zeros of $\det(\sum_{j=0}^{r+s} \mathcal{C}_j e^{i\omega})$ lie outside the unit disc. Thus, the spectral density matrix of y_t has the representation $(2\pi)^{-1}(\sum_{j=0}^{r+s} \mathcal{C}_j e^{ij\omega})^{-1}(\sum_{j=0}^{r+s} \mathcal{C}_j e^{ij\omega})'^{-1}$, which is the spectral density matrix of a causal VAR$(r+s)$ process.

The consequence of this equivalence of spectral densities is the impossibility of statistically distinguishing causal processes VAR$(r+s)$ from noncausal processes VAR(r,s) when y_t or, equivalently, the errors ϵ_t are Gaussian. In the Gaussian case, nothing is lost using a standard causal representation. However, if errors are non-Gaussian, using a causal representation for a truly noncausal process means using a VAR model whose errors can only be guaranteed as uncorrelated, but not as independent. In this case, future errors can be predicted by past values of the series considered and, as already discussed, this engendres the problem of nonfundamentalness, implying that the errors of the estimated causal VAR model do not represent fundamental economic shocks.

The primary conclusion of practical interest from the problem of Gaussian indistinguishability is that noncausal models must assume non-Gaussian errors to achieve identification.[12] A wide range of probability distributions satisfy this condition. Except for specific cases,[13] the most reasonable and commonly used alternative in the literature is the Student's t-distribution. Accordingly, this distribution will serve as the default for all exercises in the subsequent sections of this paper.

3.3 Model Selection and Estimation

The selection of orders r and s can be based on standard information criteria. The common approach was originally proposed by Breid et al. (1991), later reinforced by Lanne and Saikkonen (2011), and followed by subsequent literature. It consists in comparing all possible VAR(r,s) models that satisfy $r+s=p$ with $r, s \geq 0$, where p is the number of lags selected for a VAR$(r+s)$ model by standard information criteria, and then choosing the one with the largest likelihood value.[14]

Notice that by following this selection procedure, the possibility of selecting a causal model is not ruled out beforehand. A VAR model with a sufficient number of lags and leads covers both the possibility of causal and noncausal dynamics. Under fundamentalness, lead terms tend to become statistically insignificant and a standard purely causal model tends to be selected. Consequently, the selection of a model with $s > 0$ and statistically significant lead coefficients may suggest the inadequacy of the standard purely causal VAR model and its invertible MA representation to capture shocks underlying an economy.

In this regard, some authors argue that the estimation of coefficients associated with leads of variables can be viewed as a "test" for nonfundamentalness of data. However, this interpretation has faced some criticism in the literature. The arbitrary choice of a

[12]Several studies consider non-Gaussian distributions for macroeconomic data. See, e.g., Fagiolo et al. (2008), Cúrdia et al. (2014), Chib and Ramamurthy (2014), and Ascari et al. (2015).

[13]A relatively large literature on financial bubbles within the realm of noncausal models, for instance, employs the Cauchy distribution due to its effectiveness in simulating explosive time series behavior.

[14]Since the VAR(p) model is known to adequately capture the autocorrelations of the time series data and, due to the issue of indistinguishability, it is known that a noncausal VAR(r,s) model with $r+s=p$ will share the same autocorrelation function, then it is reasonable to assume that the noncausal model capturing the autocorrelations of the time series data is among those models where $r+s=p$.

non-Gaussian distribution, which is required for the statistical identification of the noncausal model, may undermine the diagnosis of noncausality for a given dataset, thereby invalidating the selection of a noncausal model. For this procedure to be a reasonable tool for diagnosing noncausality in a given dataset, it would be essential to ensure that the non-Gaussian distribution chosen is indeed appropriate.

For this reason, alongside the model selection procedure described above, rigorous statistical tests for noncausality or nonfundamentalness are desirable to provide a more robust foundation for selecting a noncausal model. In this regard, fundamentalness tests were proposed by Forni and Gambetti (2014) and by Canova and Sahneh (2018). Additionally, Sahneh (2015) proposed tests for noncausality. Fortunately, the dataset considered for the empirical applications of this paper has already been determined to be nonfundamental (Vonbun and Lima, 2020) using the tests proposed by Forni and Gambetti (2014) and Canova and Sahneh (2018).

Finally, regarding estimation, noncausal VAR models can be estimated by either maximizing an approximate log-likelihood function, as proposed by Lanne and Saikkonen (2013), or by employing Bayesian methods. In this paper, we adopt the approximate log-likelihood approach. Further technical details on deriving the approximate log-likelihood function and estimating the model can be found in Appendix A.

3.4 Forecasting

As previously discussed, the current state of the literature on noncausal autoregressive models does not yet allow for the development of structural analyses, making the computation of impulse-response functions unfeasible. Nevertheless, the estimates derived from these models still hold value. Thanks to the methodological advancements by Nyberg and Saikkonen (2014), these models can be used to produce forecasts.

A key advantage of traditional (purely causal) autoregressive models in forecasting is that their problems are linear. This property allows for the derivation of optimal linear forecasts (in a mean-squared-error sense) using explicit, closed-form solutions. For noncausal models, in contrast, the problem is nonlinear—as highlighted in the literature by Rosenblatt (2000), and Lanne et al. (2012). This significantly complicates forecast construction, as computing them generally requires simulation-based methods. Nonetheless, Nyberg and Saikkonen (2014) developed a method to address this problem.

Nyberg and Saikkonen's (2014) approach begins with equation (10) and involves approximating the infinite sum it contains with a finite sum. Using equations (21) and (10), they propose approximating $\mathbb{E}_T[y_{T+h}]$ as

$$\mathbb{E}_T[y_{T+h}] \approx a_1 \mathbb{E}_T[y_{T+h-1}] + \cdots + a_{nr} \mathbb{E}_T[y_{T+h-nr}] + \mathbb{E}_T\left[\sum_{j=-(n-1)r}^{M-h} N_j \epsilon_{T+h+j}\right], \qquad (19)$$

where $M > 0$ is a sufficiently large number to make approximation errors negligible. Since $\mathbb{E}_T[y_{T+h-j}] = y_{T+h-j}$ for $j \geq h$, approximate predictions can be computed recursively starting from $h = 1$ if the conditional expectation of the last term on the right-hand side of (19) is computed for every $h \geq 1$. Therefore, the problem reduces to computing an approximation for this last term.

Nyberg and Saikkonen's (2014) idea is to approximate this term using simulations. A key difference between the multivariate and univariate cases is that, in the latter (as considered in Lanne, Luoto, and Saikkonen, 2012), this conditional expectation depends only on the errors $\epsilon_{T+1}, \ldots, \epsilon_{T+M}$. However, except for the purely noncausal case $r = 0$, this is not true for the multivariate case. In the multivariate case, the errors $\epsilon_{T+1-(n-1)r}, \ldots, \epsilon_T$ are also involved, and $\epsilon_{T-s+1}, \ldots, \epsilon_T$, for $s > 0$, cannot be expressed as functions of the observed data (Nyberg and Saikkonen, 2014). Fortunately, in the purely noncausal case—the relevant case for our empirical application—these errors disappear from the right-hand side of the equation (19), simplifying the situation and allowing the solution to be a direct extension of the method proposed by Lanne et al. (2012) for the univariate case. The general case, however, still requires a more nuanced approach. The details of these treatments are beyond the scope of this work but can be found in Nyberg and Saikkonen (2014).[15]

A final comment is in order. The construction of forecasts depends on specifying the truncation parameter M and the number of replications N for the Monte Carlo simulations used to approximate conditional expectations. Lanne et al. (2012) provide simulation evidence that even with relatively small values for the truncation parameter M and the number of replications N, the approximation in equation (19) remains accurate. Based on these simulations, they recommend $M = 50$ and $N = 10,000$ as reasonable values for these quantities. For the empirical applications in this paper, we adopt $M = 50$ as recommended by the existing literature; for robustness, however, we consider several alternative values for N.

4 Empirical Application

4.1 Data and Model Selection

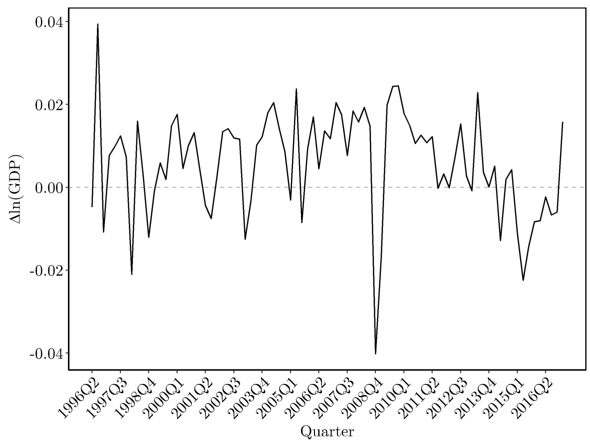
The dataset we consider is the same as that used by Vonbun and Lima (2020) to test for the occurrence of nonfundamentalness in Brazilian fiscal data. The time series are quarterly, covering the period from the first quarter of 1996 to the first quarter of 2017. This fiscal dataset was organized by Orair et al. (2016),[16] who consolidated the country's public finances and correctly classified them into government investment, I^G, government spending, G, and net tax burden, T. Additionally, we included real GDP at 1995 constant prices, Y, calculated by the Brazilian Institute of Geography and Statistics (IBGE) and obtained from its website.

All series are considered in real terms, seasonally adjusted using ARIMA-X13, and logarithmically transformed. The Augmented Dickey-Fuller (ADF) test was performed on all series, and the null hypothesis of a unit root could not be rejected for any of them. Given the substantial evidence that these variables may cointegrate, we conducted a

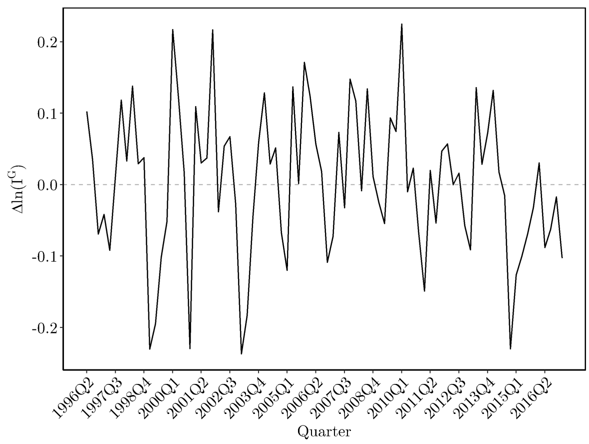
[15] Technical details on the methodological treatment of forecasting for the purely noncausal case—the relevant case for this work—can be found on page 8, Section 3.1 of Nyberg and Saikkonen (2014). Details on the general case can also be found on page 10, Section 3.2.

[16] Subsequently updated until 2017 by the public finance team of the Institute for Applied Economic Research's (IPEA) Directorate of Macroeconomic Studies and Policies (DIMAC).

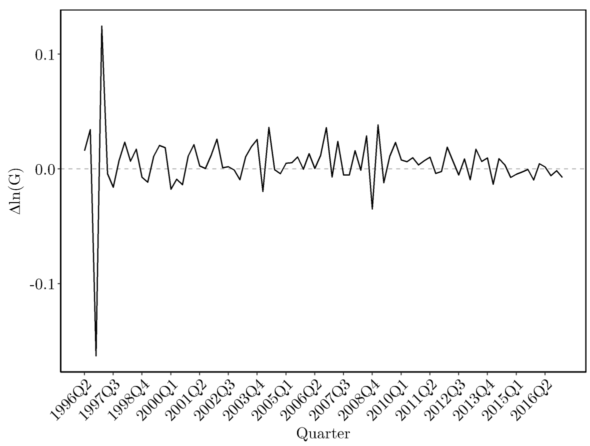
Figure 1: Time series used in modeling.



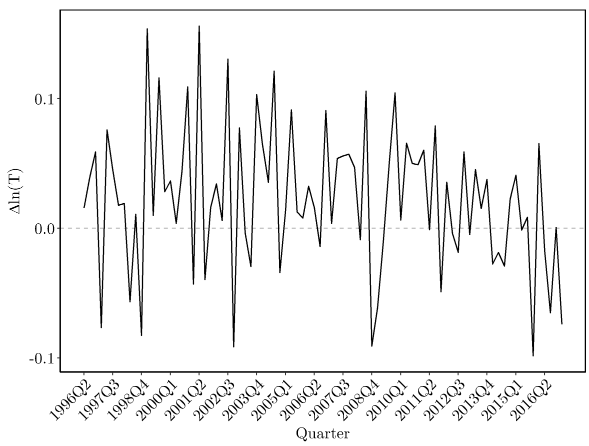
(a) Real GDP (at 1995 constant prices, log-differentiated), $\Delta\ln(Y)$.



(b) Real Public Investment (log-differentiated), $\Delta\ln(I^G)$.



(c) Real Government Spending (log-differentiated), $\Delta\ln(G)$.



(d) Real Net Tax Burden (log-differentiated), $\Delta\ln(T)$.

cointegration test. The results of the test are reported in Appendix B.1. For all relevant specifications, the test rejected cointegration. Therefore, all variables were differentiated to ensure stationarity. We also demeaned all variables.[17] Figure 1 displays the resulting time series used in our modeling.

Following the model selection procedure described in Section 3.3, a standard purely causal VAR model was estimated, and the optimal lag order $p = 1$ was chosen based on the Schwarz (SC) and Hannan-Quinn (HQ) information criteria. The results of the information criteria are reported in Appendix B.2. The normality of the residuals of the VAR(1) model was then tested using the Jarque-Bera test, which indicated signs of non-normality, justifying the use of non-Gaussian distributions for modeling. The results of the Jarque-Bera test are reported in Appendix B.4. Consistent with the literature, we chose the Student's t-distribution.

With the information criteria selecting $p = 1$, the only feasible noncausal specification respecting $r + s = 1$ was a VAR(0,1) with zero lags and one lead. The selection prob-

[17]The variables were prudently demeaned, as the implications of stochastic trends in noncausal models are still not fully understood.

lem was then reduced to choosing between a purely causal VAR(1,0) and a noncausal VAR(0,1). Since the dataset we considered was already previously verified as nonfundamental by Vonbun and Lima (2020), we selected the noncausal VAR(0,1) model, avoiding the need for log-likelihood comparisons. This selection approach relied on more rigorous statistical tests (Forni and Gambetti, 2014; Canova and Sahneh, 2018) compared to a simple log-likelihood comparison, which helps address the criticism discussed in Section 3.3. However, it is only feasible when there is only one possible noncausal model specification (i.e., when $p = 1$). For cases where $p > 1$, selecting the noncausal specification among multiple options would still require comparing the log-likelihoods of each model.

4.2 Estimation

The results of the estimated VAR(0,1) model are presented in Table 1. The parameter estimates of Ψ_1 do not appear to reject the presence of noncausality in Brazilian fiscal data. All four future variables showed a statistically significant coefficient in at least one of the system's reduced-form equations. Unfortunately, due to the inability to properly identify shocks, it is not feasible to isolate and quantify the effects of future variables on specific current variables. However, it is still possible to infer that the significance of these coefficients suggests the relevance of these future variables in the overall dynamics of the model, thus providing potential evidence for the occurrence of fiscal foresight, noncausality, and nonfundamentalness within the typical Brazilian fiscal VAR dataset.

Table 1: Estimation results for the fiscal VAR(0,1): GDP, Y, Government Investment, I^G, Government Spending, G, and Net Tax Burden, T.

	0.2264	0.0123	-0.1745*	0.083*		0.9285*	2.7604*	0.1399	1.7486*
	(0.1225)	(0.0124)	(0.0463)	(0.023)		(0.1759)	(1.1066)	(0.1495)	(0.5933)
	-0.1276	0.2802*	-0.1575	0.141		2.7604*	80.1325*	0.7769	-1.2667
Ψ_1	(1.0953)	(0.112)	(0.4097)	(0.2024)	Σ	(1.1066)	(15.0369)	(1.367)	(4.8483)
	0.2217	0.0118	-0.2378*	0.0195		0.1399	0.7769	1.6033*	0.6746
	(0.1521)	(0.0157)	(0.0624)	(0.0292)		(0.1495)	(1.367)	(0.3415)	(0.6882)
	1.409*	-0.0028	-0.354	-0.2703		1.7486*	-1.2667	0.6746	20.1141*
	(0.5635)	(0.0589)	(0.2051)	(0.1059)		(0.5933)	(4.8483)	(0.6882)	(3.7312)
λ	5.5482*				log$L$	-836.084			
	(1.5373)								

Note: The values in parentheses are the standard errors based on the Hessian of the log-likelihood function. Asterisks indicate statistical significance at 5%. Ψ_1 is the coefficient matrix of the variables one period ahead. Σ is the variance-covariance matrix. λ is the parameter of degrees of freedom for the multivariate t distribution. log$L$ is the value of the log-likelihood function.

The following reduced-form coefficients were statistically significant: government expenditure one quarter ahead on current GDP ($\Psi_{1,13}$) and current government expenditure ($\Psi_{1,33}$); net tax burden one quarter ahead on current GDP ($\Psi_{1,14}$) and current net tax burden ($\Psi_{1,44}$); government investment one quarter ahead on current government investment ($\Psi_{1,22}$); and GDP one quarter ahead on current net tax burden ($\Psi_{1,41}$).[18]

[18]It is important to emphasize that these relationships do not indicate inter-causal economic effects between the variables. To make interpretations of this nature, identification and structural analysis

4.3 Forecasts

In this section, we evaluate whether a correctly specified noncausal model can provide predictive gains over a standard misspecified purely causal model by computing forecasts and comparing the predictive performance of the selected noncausal VAR(0,1) model with that of the standard purely causal VAR(1) model. To this end, we compute pseudo out-of-sample forecasts for four quarters ahead, covering the period from the second quarter of 2017 to the first quarter of 2018.[19]

The predictive accuracy metrics used in the comparison are the Mean Squared Forecast Error (MSFE), the Root Mean Squared Forecast Error (RMSFE), the Mean Absolute Forecast Error (MAFE), and the Mean Average Percentage Error (MAPE). For robustness, we construct forecasts using varying numbers of replications, N, enabling an analysis of how predictive accuracy evolves as both the forecast horizons extend and the number of replications increases. Across all metrics, the forecasting errors remained stable as the number of replications increased from 10,000 to 500,000.

The relative predictive accuracy metrics (calculated as the ratio of the predictive accuracy metrics for the noncausal and causal VAR models) are presented in Table 2. Values below unity indicate the predictive superiority of the noncausal model over its causal counterpart. Generally, the VAR(0,1) model exhibited consistently smaller errors than the VAR(1) model across the various setups considered, including differing numbers of replications and forecast horizons.

Exceptions were observed in five cases, specifically with a low number of replications and a long forecast horizon (10,000 or 100,000 replications and a four-quarter horizon). However, increasing the number of replications to 200,000 or more led to better forecasts for the VAR(0,1) model across all horizons and cases considered. These results underscore the predictive superiority of the noncausal VAR(0,1) model over the standard purely causal VAR(1) model in nearly all scenarios. Of the 64 relative metrics calculated, 59 favored the noncausal VAR model. Strikingly, in no instance did simulations with 200,000 or more replications produce worse predictions than those of the VAR(1).

The predictive superiority of the noncausal VAR is illustrated in Figures 2-5. These figures present bar charts showing the predictive accuracy metrics for each model and forecast horizon. Each figure corresponds to a specific forecast horizon and predictive accuracy metric, displaying the metrics for the noncausal VAR across different numbers of replications, N. The final bar in each chart, positioned on the far right, represents the metric for the standard purely causal VAR(1). Bars shorter than this final bar indicate the superiority of the noncausal VAR(0,1) for the corresponding horizon and metric. To aid comparisons, dashed lines are included at the level of the VAR(1) metric.

In general, noncausal VAR forecasts are markedly superior for short horizons, especially

methods would be necessary, which are not yet readily available in the noncausal VAR literature. The only interpretation that can be drawn from these coefficients is a purely statistical one: they represent the effects that each variable will have on the *predictions* made by the model for the other variables (see equation 15). This interpretation is adopted, for instance, in Lanne and Saikkonen (2013). However, from a causality perspective, the magnitudes of these coefficients are meaningless.

[19] This forecast window was selected to exclude data from the second quarter of 2018, which was affected by the Truck Drivers' Strike. Forecasts generated for intervals including this period exhibited significant distortions due to entirely exogenous factors. Consequently, we chose to discard those forecasts and focus on periods of greater stability to ensure more reliable statistical comparisons.

Table 2: Relative predictive accuracy metrics.

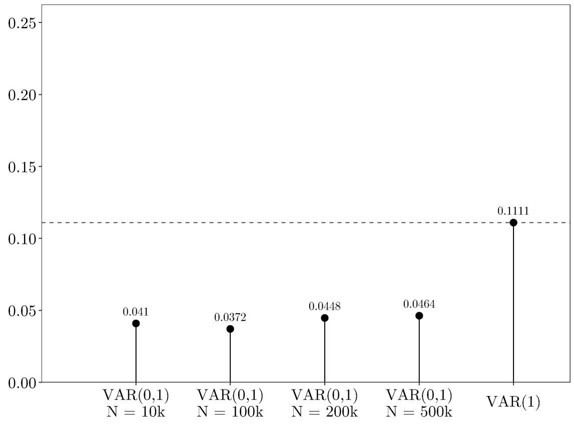
Relative MSFE				
Horizon	**VAR(0,1)/VAR(1)**			
	Number of replications			
	10000	**100000**	**200000**	**500000**
1	0.3695	0.3349	0.4030	0.4177
2	0.8040	0.8371	0.9271	0.9461
3	0.7691	0.9761	0.9135	0.9291
4	1.0184	1.0223	0.9353	0.9205

Relative RMSFE				
Horizon	**VAR(0,1)/VAR(1)**			
	Number of replications			
	10000	**100000**	**200000**	**500000**
1	0.6079	0.5787	0.6349	0.6463
2	0.8967	0.9149	0.9629	0.9727
3	0.8770	0.9880	0.9558	0.9639
4	1.0092	1.0111	0.9671	0.9594

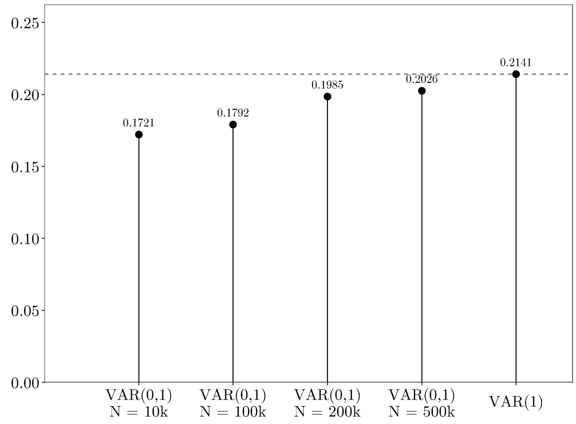
Relative MAFE				
Horizon	**VAR(0,1)/VAR(1)**			
	Number of replications			
	10000	**100000**	**200000**	**500000**
1	0.6079	0.5787	0.6349	0.6463
2	0.8452	0.8596	0.8945	0.9056
3	0.8413	0.9452	0.8999	0.9118
4	0.9831	0.9874	0.9301	0.9055

Relative MAPE				
Horizon	**VAR(0,1)/VAR(1)**			
	Number of replications			
	10000	**100000**	**200000**	**500000**
1	0.6079	0.5787	0.6349	0.6463
2	0.8921	0.9095	0.9541	0.9645
3	0.8614	0.9823	0.9434	0.9558
4	0.9936	1.0069	0.9637	0.9515

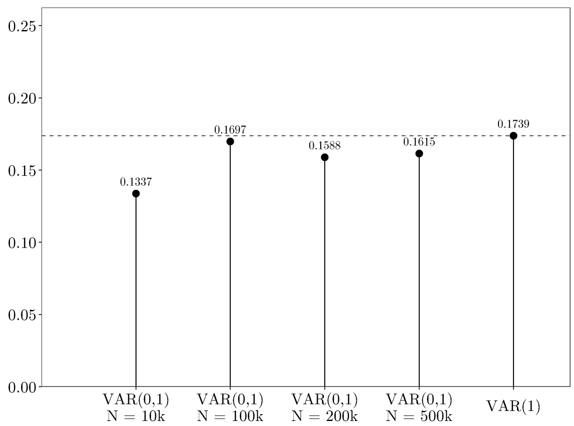
Figure 2: VAR(0,1) and VAR(1) — Mean Squared Forecast Errors (MSFE).



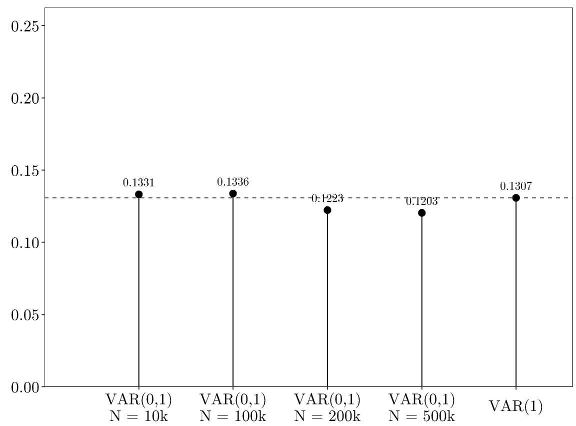
(a) $h = 1$



(b) $h = 2$



(c) $h = 3$

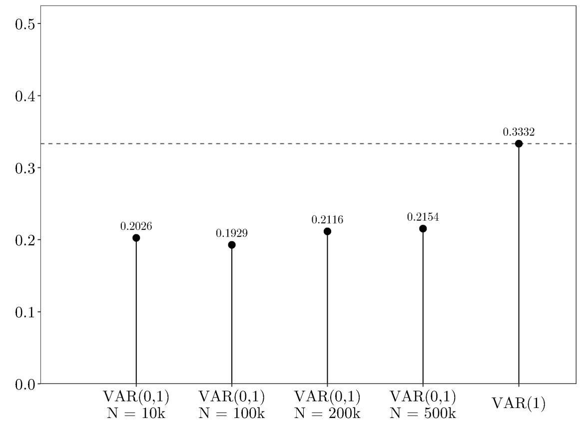


(d) $h = 4$

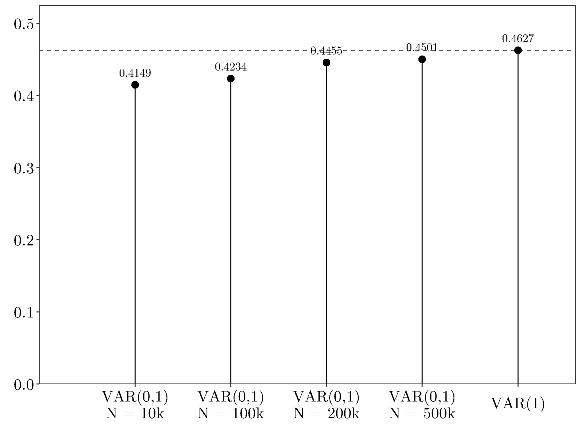
for $h = 1$ (one quarter ahead). For longer horizons ($h = 2$ or more), however, the predictive accuracy of the causal and noncausal models seems to converge. For horizons of two, three, and four quarters ahead, the Diebold-Mariano test shows no statistical evidence that the predictions of the causal and noncausal models differ. The results of the Diebold-Mariano test are detailed in Appendix B.3. Unfortunately, for the horizon that appeared particularly promising based on graphical inspection—one quarter ahead ($h = 1$)—the test cannot be conducted. Thus, for this horizon, we must rely on the qualitative graphical assessment previously described.

In summary, the overall evidence indicates the predictive superiority of noncausal VAR models over standard purely causal VAR models—particularly for shorter horizons—when examining the dataset and the analyzed period. This empirical evidence highlights, once again, the importance of addressing the issue of nonfundamentalness (and noncausality) in fiscal VAR models.

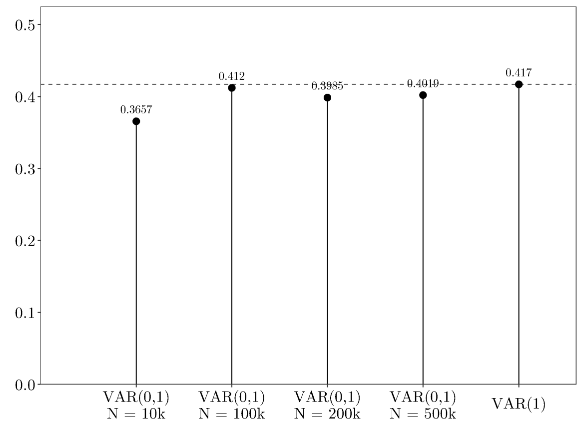
Figure 3: VAR(0,1) and VAR(1) Root — Mean Squared Forecast Errors (RMSFE).



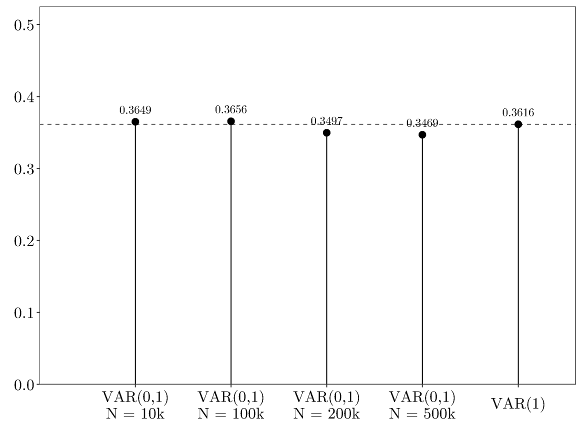
(a) $h = 1$



(b) $h = 2$

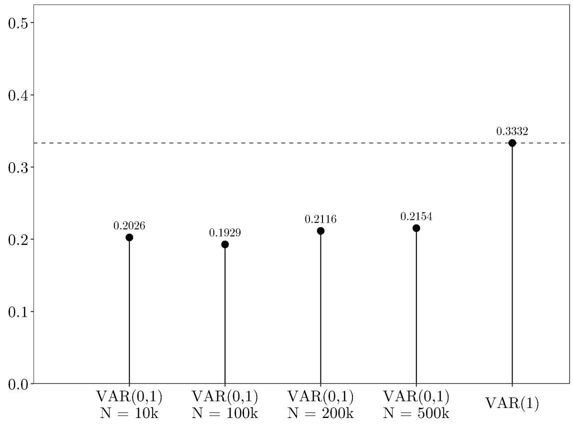


(c) $h = 3$

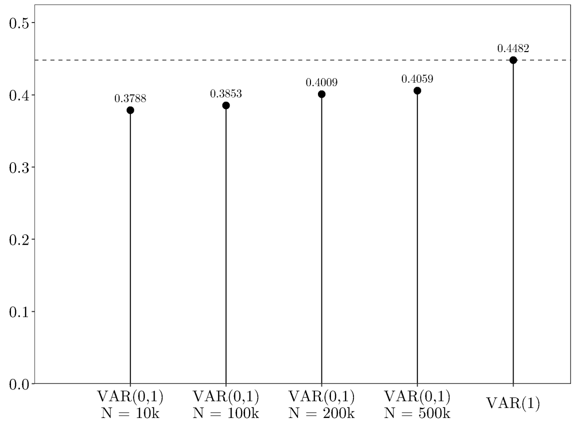


(d) $h = 4$

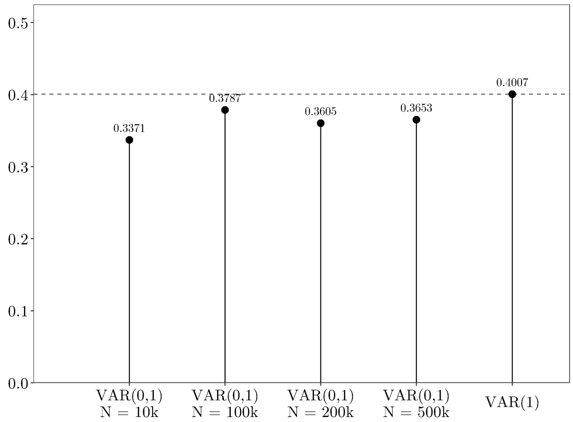
Figure 4: VAR(0,1) and VAR(1) — Mean Absolute Forecast Errors (MAFE).



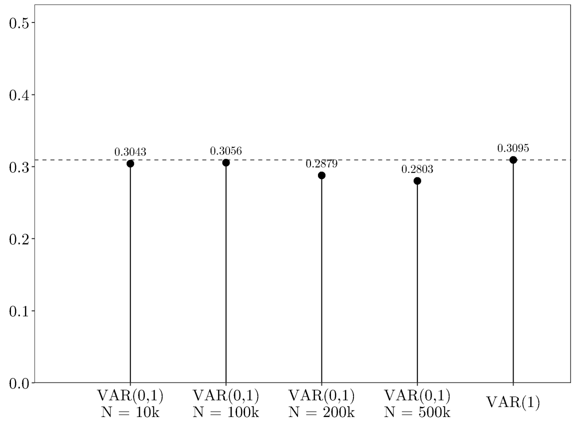
(a) $h = 1$



(b) $h = 2$

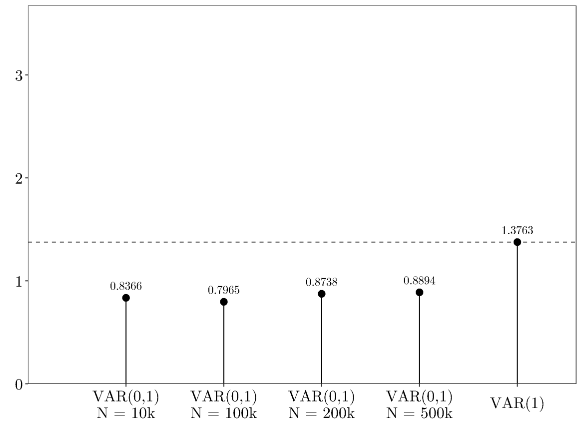


(c) $h = 3$

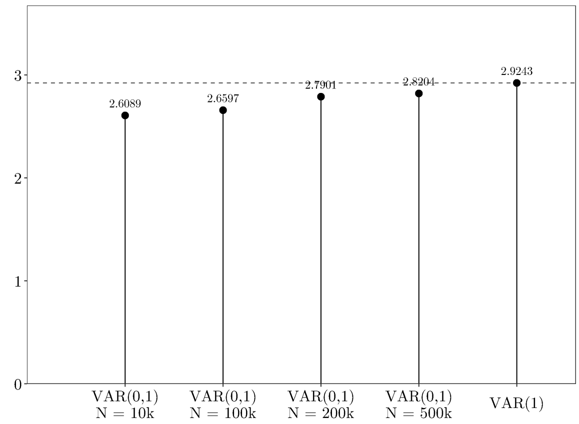


(d) $h = 4$

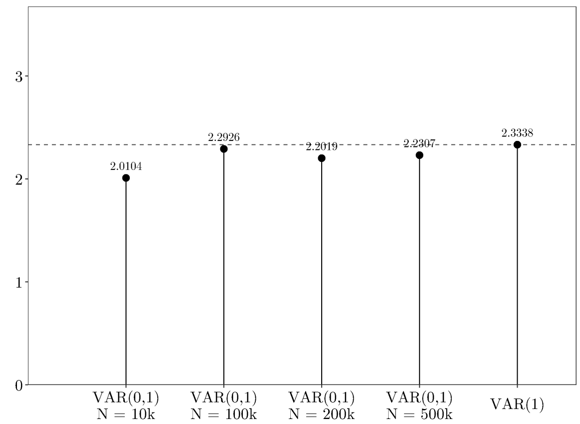
Figure 5: VAR(0,1) and VAR(1) — Mean Absolute Percentage Errors (MAPE).



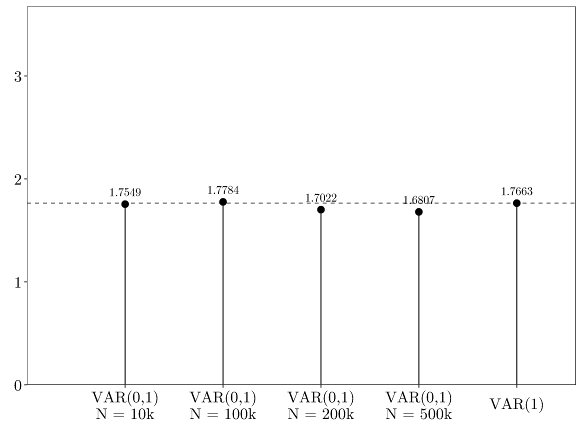
(a) $h = 1$



(b) $h = 2$



(c) $h = 3$



(d) $h = 4$

5 Conclusion

In this paper, we presented an application of Lanne and Saikkonen's (2013) noncausal VAR model to the "typical" dataset used in the Brazilian fiscal VAR literature, which Vonbun and Lima (2020) previously found to be nonfundamental. Nonfundamentalness—often arising in fiscal data due to fiscal foresight—leads to standard fiscal VAR models being misspecified, resulting in biased and inconsistent forecasts. Noncausal VAR models address this issue by allowing the model dynamics to depend not only on past and current values of the variables but also on their future values, thereby potentially correcting for misspecification.

We implemented a model selection procedure similar to that proposed by Lanne and Saikkonen (2013). The selected model was a purely forward-looking noncausal model with one lead—meaning it depended only on the current values of the variables and their expectations one quarter ahead. This model was estimated, and all future variables were found to be statistically significant in determining the dynamics of at least one of the system's equations. We then compared the predictive performance of this noncausal fiscal VAR to that of its purely causal counterpart.

The results were favorable to the predictive ability of the noncausal VAR model, particularly for larger numbers of replications and shorter horizons. Only in five instances—with a low number of replications and a long forecast horizon—did the noncausal model perform worse than the standard purely causal. As the number of replications increased, the noncausal specification consistently outperformed the causal specification. This superior performance was especially pronounced for the shortest forecast horizon (one quarter ahead, $h = 1$). For horizons of two quarters or more ($h \geq 2$), however, the predictive performance of the two models appeared to converge, with the Diebold-Mariano test indicating no statistically significant differences in their predictive accuracies.

The statistical significance of the estimated coefficients and the predictive superiority of the forecasts provided additional evidence of fiscal foresight and noncausality as sources of nonfundamentalness in the Brazilian fiscal dataset analyzed, partially corroborating the results of Vonbun and Lima (2020). This evidence underscores the importance of accounting for these phenomena in Brazilian fiscal data, which are particularly prone to such issues. Although these phenomena have been widely recognized as challenges in the fiscal literature, they have so far been overlooked in the Brazilian VAR fiscal literature.

While nonfundamentalness and noncausality can be addressed through methods other than employing a noncausal VAR, the approach presented in this paper appears to be the first to achieve some degree of success. A natural alternative is to enhance the informational sufficiency of purely causal models by incorporating proxies for forward-looking variables. This approach was previously attempted by Vonbun and Lima (2020), but unfortunately, it proved unsuccessful. Consequently, the noncausal VAR model emerges as a promising alternative for addressing these issues in Brazilian fiscal studies.

Suggestions for future research include applying noncausal models to different economic contexts where nonfundamentalness is present. Additionally, significant improvements in empirical studies can be achieved with advances in techniques for identifying noncausal VAR models. More ambitious projects could involve developing and applying techniques for identification and structural analysis to these models, enabling the construction of impulse response functions and the analysis of foresight-robust fiscal multipliers.

References

L. Alessi, M. Barigozzi, and M. Capasso. A review of nonfundamentalness and identification in structural var models. 2008.

G. Ascari, G. Fagiolo, and A. Roventini. Fat-tail distributions and business-cycle models. *Macroeconomic Dynamics*, 19(2):465–476, 2015.

O. Blanchard and R. Perotti. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. 1999.

F. J. Breid, R. A. Davis, K.-S. Lh, and M. Rosenblatt. Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198, 1991.

F. Canova and M. Sahneh. Are small-scale svars useful for business cycle analysis? revisiting nonfundamentalness. *Journal of the European Economic Association*, 16(4): 1069–1093, 2018.

S. Chib and S. Ramamurthy. Dsge models with student-t errors. *Econometric Reviews*, 33(1-4):152–171, 2014.

V. Cúrdia, M. Del Negro, and D. L. Greenwald. Rare shocks, great recessions. *Journal of Applied Econometrics*, 29(7):1031–1052, 2014.

G. Fagiolo, M. Napoletano, and A. Roventini. Are output growth-rate distributions fat-tailed? some evidence from oecd countries. *Journal of Applied Econometrics*, 23(5): 639–669, 2008.

J. Fernández-Villaverde, J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson. Abcs (and ds) of understanding vars. *American economic review*, 97(3):1021–1026, 2007.

M. Forni and L. Gambetti. Sufficient information in structural vars. *Journal of Monetary Economics*, 66:124–136, 2014.

C. Gourieroux and J. Jasiak. Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3):405–430, 2016.

E. Hannan. Multiple time series john wiley and sons inc new york usa. 1970.

A. Hecq, J. V. Issler, and S. Telg. Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics*, 35(3):328–343, 2020.

M. Lanne and J. Luoto. Noncausal bayesian vector autoregression. *Journal of Applied Econometrics*, 31(7):1392–1406, 2016.

M. Lanne and P. Saikkonen. Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3), 2011.

M. Lanne and P. Saikkonen. Noncausal vector autoregression. *Econometric Theory*, 29 (3):447–481, 2013.

M. Lanne, H. Nyberg, and E. Saarinen. Forecasting us macroeconomic and financial time series with noncausal and causal ar models: a comparison. 2011.

M. Lanne, J. Luoto, and P. Saikkonen. Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3):623–631, 2012.

E. M. Leeper, T. B. Walker, and S.-C. S. Yang. Fiscal foresight: Analytical issues. In *Society for Economic Dynamics 2008 Meeting Papers*, volume 786, 2008.

E. M. Leeper, T. B. Walker, and S.-C. S. Yang. Fiscal foresight and information flows. *Econometrica*, 81(3):1115–1145, 2013.

M. Lippi and L. Reichlin. Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, 63(1):307–325, 1994.

J. Nelimarkka. Causal effects in macroeconomics through higher moments. 2019.

H. Nyberg and P. Saikkonen. Forecasting with a noncausal var model. *Computational statistics & data analysis*, 76:536–555, 2014.

R. Orair, F. Siqueira, and S. GOBETTI. Política fiscal e ciclo econômico: uma análise baseada em multiplicadores do gasto público. *XXI Prêmio do Tesouro Nacional*, 41, 2016.

M. Rosenblatt. *Gaussian and non-Gaussian linear time series and random fields*. Springer Science & Business Media, 2000.

M. H. Sahneh. Testing for noncausal vector autoregressive representation. In *Proceedings of Economics and Finance Conferences*, volume 2204921. International Institute of Social and Economic Sciences, 2015.

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

C. A. Sims. Solving linear rational expectations models. *Computational economics*, 20 (1-2):1, 2002.

C. Vonbun and E. C. R. Lima. Há fundamentalidade nos modelos de var fiscal típicos para o brasil? Technical report, Texto para Discussão, 2020.

S.-C. S. Yang. Quantifying tax effects under policy foresight. *Journal of Monetary Economics*, 52(8):1557–1568, 2005.

Appendix

A The Log-likelihood Function and Estimation

To carry out the estimation, the first step is to derive the transformation representative of the equivalence relationship between the process' information sets. For this, Lanne and Saikkonen (2013) use the definition $w_t = \det(\Phi(L))y_t$ from (9). In this definition, $\det\Phi(L)$ is a polynomial. Denoting this polynomial as

$$a(z) \equiv \det(\Phi(z)) = 1 - a_1 z - \cdots - a_{nr} z^{nr}, \tag{20}$$

it is possible to write

$$w_t = a(L)y_t. \tag{21}$$

Using this process together with the equation $u_t = \Psi(L^{-1})y_t$ from (7), Lanne and Saikkonen (2013) obtain:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \begin{bmatrix} y_1 - \Psi_1 y_2 - \cdots - \Psi_s y_{s+1} \\ \vdots \\ y_{T-s} - \Psi_1 y_{T-s+1} - \cdots - \Psi_s y_T \\ nT - s + 1 - a_1 y_{T-s} - \cdots - a_{nr} y_{T-s-nr+1} \\ \vdots \\ y_T - a_1 y_{T-1} - \cdots - a_{nr} y_{T-nr} \end{bmatrix} = \mathbf{H}_1 \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}, \tag{22}$$

or simply $\mathbf{x}_1 = \mathbf{H}_1\mathbf{y}$.

From the definition of u_t and (5) it follows that $\Phi(L)u_t = \epsilon_t$ so that, from the above equality, the authors obtain:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} - \Phi_1 u_r - \cdots - \Phi_r u_1 \\ \vdots \\ u_{T-s} - \Phi_1 u_{T-s-1} - \cdots - \Phi_r u_{T-s-r} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \mathbf{H}_2 \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} \\ \vdots \\ u_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix}, \tag{23}$$

or $\mathbf{x}_2 = \mathbf{H}_2\mathbf{x}_1$.

Lanne and Saikkonen (2013) also perform a third transformation that transforms the variables $w_{T-s+1}, \ldots, w_T$ into x_2. For this purpose, they define:

$$v_{k,T-s+k} = w_{T-s+k} - \sum_{j=-(n-1)r}^{-k} N_j \epsilon_{T-s+k+j}, \quad k = 1, \ldots, s, \tag{24}$$

where the sum is interpreted as zero when $k > (n-1)r$, that is, when the lower bound exceeds the upper bound. Note also that, from (5) and (10), $v_{k,T-s+k}$ can be expressed as a function of the observed data $y_1, \ldots, y_T$ and that the representation $v_{k,T-s+k} = \sum_{j=-k+1}^{\infty} N_j \epsilon_{T-s+k+j}$ holds, which shows that $v_{k,T-s+k}$ $(k = 1, \ldots, s)$ are independent of ϵ_t, $t \leq T-s$. It is now possible to introduce the following transformation:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ v_{1,T-s+1} \\ \vdots \\ v_{s,T} \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ w_{T-s+1} - N_{-(n-1)r}\epsilon_{T-s+1-(n-1)r} - \cdots - N_{-1}\epsilon_{T-s} \\ \vdots \\ w_T - N_{-(n-1)r}\epsilon_{T-(n-1)r} - \cdots - N_{-s}\epsilon_{T-s} \end{bmatrix} = \mathbf{H}_3 \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_t \end{bmatrix}, \tag{25}$$

or $\mathbf{z} = \mathbf{H}_3\mathbf{x}_2$.

Combining the three transformations, the authors obtain the equation:

$$\mathbf{z} = \mathbf{H}_3\mathbf{H}_2\mathbf{H}_1\mathbf{y}, \tag{26}$$

where the nonstochastic matrices $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$ are nonsingular. The nonsingularity of $\mathbf{H}_2$ and $\mathbf{H}_3$ comes from the fact that $\det(\mathbf{H}_2) = \det(\mathbf{H}_3) = 1$, which can be easily verified. The demonstration of the nonsingularity of $\mathbf{H}_1$ is a more complicated task and can be found in Lanne and Saikkonen (2013).

Here we will observe once more the usefulness of the equivalent MA(∞) representations, exposed in the equation (12). From the equations (8) and (10), it can be seen that the components of $\mathbf{z}$ given by $\mathbf{z}_1 = (u_1, \ldots, u_r)$, $\mathbf{z}_2 = (\epsilon_{r+1}, \ldots, \epsilon_{T-s})$ and $\mathbf{z}_3 = (v_{1,T-s+1}, \ldots, v_{s,T})$ are independent. Therefore, under true parameter values, the joint probability density function of $\mathbf{z}$ can be expressed as:

$$h_{\mathbf{z}_1}(\mathbf{z}_1) \left(\prod_{t=r+1}^{T-s} f_\Sigma(\epsilon_t; \lambda) \right) h_{\mathbf{z}_3}(\mathbf{z}_3), \tag{27}$$

where $h_{\mathbf{z}_1}(\cdot)$ and $h_{\mathbf{z}_3}(\cdot)$ stand for the probability density functions of $\mathbf{z}_1$ and $\mathbf{z}_3$, respectively. Using the general specification (5) of the VAR(r, s) models and the fact that the determinants of $\mathbf{H}_2$ and $\mathbf{H}_3$ are equal to one, we can write the joint probability density function of the data vector $\mathbf{y}$ as:

$$h_{\mathbf{z}_1}(\mathbf{z}_1(\vartheta))\left(\prod_{t=r+1}^{T-s} f_{\Sigma}(\Phi(L)\Psi(L^{-1})y_t;\lambda)\right)h_{\mathbf{z}_3}(\mathbf{z}_3(\vartheta))|\det(\mathbf{H}_1)|, \tag{28}$$

where the argument $\mathbf{z}_1(\vartheta)$ is defined by replacing u_t in the definition of z_1 by $\Psi(L^{-1})y_t$ $(t = 1,\ldots,r)$ and $\mathbf{z}_3(\vartheta)$ is defined similarly by replacing $v_{k,T-s+k}$ in the definition of $\mathbf{z}_3$ by an analogue with $a(L)y_{T-s+k}$ and $\Phi(L)\Psi(L^{-1})y_{T-s+k+j}$ used in place of w_{T-s+k} and $\epsilon_{T-s+k+j}$ $(j = -(n-1)r,\ldots,-k,\ k = 1,\ldots,s)$, respectively.

Lanne and Saikkonen (2013) note that the determinant of the block $(T-s)n \times (T-s)n$ in the upper left corner of $\mathbf{H}_1$ is equal to one and, therefore, using the well-known formula for the determinant of partitioned matrices, it is possible to see that the determinant of $\mathbf{H}_1$ is independent of the size T of the sample. As a consequence of this result, the approximation of the joint density of $\mathbf{y}$ reduces to the problem of approximating only the second term of the previous expression. The approximate log-likelihood function can then be described as:

$$l_T = \sum_{t=r+1}^{T-s} g_t(\theta) = \sum_{t=r+1}^{T-s} \log f(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda) - \frac{1}{2}\log\,\det(\Sigma) \tag{29}$$

where

$$\epsilon_t(\vartheta) = u_t(\vartheta_2) - \sum_{j=1}^{r} \Phi_j(\vartheta_1)u_{t-j}(\vartheta_2) \tag{30}$$

and

$$u_t(\vartheta_2) = y_t - \Psi_1(\vartheta_2)y_{t+1} - \cdots - \Psi_s(\vartheta_2)y_{t+s}. \tag{31}$$

Maximizing $l_T(\theta)$ for the permissible values of θ yields an approximate maximum likelihood estimator for θ.

B Statistical Tests

B.1 Cointegration Test

Table 3: Johansen Cointegration Test for Y, I^G, G and T.

Selected (0.05 level*) Number of Cointegrating Relations by Model					
Data trend:	**None**	**None**	**Linear**	**Linear**	**Quadratic**
Test Type	**No Intercept, No Trend**	**Intercept, No Trend**	**Intercept, No Trend**	**Intercept, Trend**	**Intercept, Trend**
Trace	2	3	2	2	4
Max-Eig	1	1	1	1	1
*Critical values based on MacKinnon-Haug-Michelis (1999).					
Information Criteria by Rank and Model					
Data Trend:	**None**	**None**	**Linear**	**Linear**	**Quadratic**
Rank or No. of CEs	**No Intercept No Trend**	**Intercept, No Trend**	**Intercept, No Trend**	**Intercept, Trend**	**Intercept, Trend**
Log-Likelihood by Rank (rows) and Model (columns)					
0	-2999.429	-2999.429	-2992.832	-2992.832	-2992.620
1	-2986.109	-2981.939	-2978.072	-2975.271	-2975.060
2	-2977.552	-2972.886	-2969.093	-2963.908	-2963.713
3	-2972.169	-2965.827	-2962.853	-2955.224	-2955.142
4	-2972.004	-2962.691	-2962.691	-2951.685	-2951.685
Akaike Information Criteria by Rank (rows) and Model (columns)					
0	71.32774	71.32774	71.26663	71.26663	71.35576
1	71.20257	71.12798	71.10758	71.06519	71.13083
2	71.18945	71.12673	71.08454	71.00961*	71.05206
3	71.25104	71.17239	71.12594	71.01704	71.03863
4	71.43538	71.31038	71.31038	71.14553	71.14553
Schwarz Criteria by Rank (rows) and Model (columns)					
0	72.24732*	72.24732*	72.30116	72.30116	72.50524
1	72.35205	72.30620	72.37201	72.35836	72.51021
2	72.56883	72.56359	72.57887	72.56141	72.66134
3	72.86031	72.86788	72.85017	72.82748	72.87780
4	73.27455	73.26450	73.26450	73.21460	73.21460

Note: Lags interval of 1 to 2.

B.2 Information Criteria

Table 4: Information Criteria Results.

Lags	**LR**	**FPE**	**AIC**	**SC**	**HQ**
0	-	$1.76 \cdot 10^{29}$	78.69	78.81	78.74
1	651.26	$4.44 \cdot 10^{25}$	70.40	71.00*	70.65*
2	36.17	$4.00 \cdot 10^{25}$	70.30	71.37	70.73
3	29.69	$3.86 \cdot 10^{25}$	70.25	71.80	70.87
4	20.93	$4.19 \cdot 10^{25}$	70.32	72.35	71.14
5	34.43*	$3.57 \cdot 10^{25}$	70.14	72.64	71.14

Note: Asterisks (*) indicate the lag order selected by the criterion. LR: Sequential Modified LR Test Statistic (each test at the 5% level). FPE: Final Prediction Error. AIC: Akaike Information Criterion. SC: Schwarz Information Criterion. HQ: Hannan-Quinn Information Criterion.

B.3 Diebold-Mariano Test

Table 5: D-M test p-values.

Horizon	**Test Specification**	**Number of replications (N)**			
		10k	**100k**	**200k**	**500k**
2	**Less**	0.7887	0.7562	0.6711	0.6557
	Greater	0.2113	0.2438	0.3289	0.3443
	Two Sided	0.4226	0.4875	0.6579	0.6886
3	**Less**	0.8237	0.663	0.7293	0.7276
	Greater	0.1763	0.337	0.2707	0.2724
	Two Sided	0.3527	0.674	0.5414	0.5448
4	**Less**	0.5356	0.5381	0.7273	0.7799
	Greater	0.4644	0.4619	0.2727	0.2201
	Two Sided	0.9289	0.9238	0.5454	0.4402

Note: The null hypothesis H_0 is that the two methods have the same forecast accuracy. For the "*Less*" test specification, the alternative hypothesis H_1 is that the noncausal VAR(0,1) is less accurate than the standard purely causal VAR(1,0). For the "*Greater*" test specification, in turn, the alternative hypothesis H_1 is that the noncausal VAR(0,1) is more accurate than the standard purely causal VAR(1,0). Finally, for the "*Two Sided*" test specification, the alternative hypothesis H_1 is that the noncausal VAR(0,1) and the standard purely causal VAR(1,0) have different levels of accuracy. All tests consider a 5% statistical significance level.

B.4 VAR Residual Normality Test

Table 6: VAR Residual Normality Test.

Component	**Skewness**	**Chi-squared**	**df**	**Prob.**
1	-2.316679	77.82155	1	0.0000
2	-0.138122	0.276628	1	0.5989
3	0.080499	0.093962	1	0.7592
4	-0.986608	14.11425	1	0.0002
Joint		92.30639	4	0.0000
Component	**Kurtosis**	**Chi-squared**	**df**	**Prob.**
1	16.08080	620.2643	1	0.0000
2	3.834746	2.525903	1	0.1120
3	4.255636	5.715255	1	0.0168
4	4.798800	11.72934	1	0.0006
Joint		640.2348	4	0.0000

Component	**Jarque-Bera**	**df**	**Prob.**
1	698.0859	2	0.0000
2	2.802530	2	0.2463
3	5.809218	2	0.0548
4	25.84359	2	0.0000
Joint	732.5412	8	0.0000

Note: Orthogonalization: Cholesky (Lütkepohl). H_0: Residuals are multivariate normal.