Testing for fundamental vector moving average representations

BIN CHEN
Department of Economics, University of Rochester

JINHO CHOI
AMRO and Bank of Korea

JUAN CARLOS ESCANCIANO
Department of Economics, Indiana University

We propose a test for invertibility or fundamentalness of structural vector autoregressive moving average models generated by non-Gaussian independent and identically distributed structural shocks. We prove that in these models and under some regularity conditions the Wold innovations are a martingale difference sequence (mds) if and only if the structural shocks are fundamental. This simple but powerful characterization suggests an empirical strategy to assess invertibility. We propose a test based on a generalized spectral density to check for the mds property of the Wold innovations. This approach does not require the specification and estimation of the economic agent's information flows or the identification and estimation of the structural parameters and the noninvertible roots. Moreover, the proposed test statistic uses all lags in the sample and it has a convenient asymptotic $N(0, 1)$ distribution under the null hypothesis of invertibility, and hence, it is straightforward to implement. In case of rejection, the test can be further used to check if a given set of additional variables provides sufficient informational content to restore invertibility. A Monte Carlo study is conducted to examine the finite-sample performance of our test. Finally, the proposed test is applied to two widely cited works on the effects of fiscal shocks by Blanchard and Perotti (2002) and Ramey (2011).

KEYWORDS. Fundamental representations, generalized spectrum, identification, invertible moving average.

JEL CLASSIFICATION. C5, C32, E62.

1. Introduction

A moving average (MA) process is invertible or fundamental when the underlying shocks driving the process can be recovered from linear combinations of present and past values of the observations.[1] The well known identification problems of Gaussian non-invertible MAs have prompted the econometrics and macroeconomics literatures to systematically impose invertibility or fundamentalness in general MA representations without reservation as the mean to identify economic shocks and their dynamic impact in structural vector autoregression (VAR) models. We show in this paper that this modeling strategy has no formal justification, as fundamental and nonfundamental MA representations generally lead to different joint distributions in the observable time series for *all* but for the Gaussian case. Moreover, we prove that within this setting Wold innovations contain all the relevant information to empirically discriminate between fundamental and nonfundamental representations. Our main result is that a vector MA (VMA) process with non-Gaussian independent and identically distributed (iid) structural shocks is invertible if and only if the associated Wold innovations are a martingale difference sequence (mds), provided some regularity conditions are satisfied. This characterization suggests a simple empirical strategy to test for fundamentalness by testing for the mds property of the Wold innovations. We propose a test based on a generalized spectral density that has good power properties in finite samples while being straightforward to implement, as it only requires as inputs the estimated Wold innovations.[2]

The main motivation for our work is the substantial theoretical and empirical evidence on economic models leading to nonfundamental representations. Hansen and Sargent (1980, 1991) discussed situations where nonfundamentalness arises in rational expectation models. In a comment to Blanchard and Quah (1989), Lippi and Reichlin (1993) provided a simple bivariate example where learning-by-doing dynamics in productivity yields noninvertible representations. In their response to the comment, Blanchard and Quah (1993) provided further examples of noninvertibility in the context of the permanent income Friedman–Muth model and cointegrated models. Lippi and Reichlin (1994) analyzed the problem more generally, and discussed further empirical examples. Empirical evidence of noninvertibility in univariate models can be found in Huang and Pawitan (2000) for U.S. unemployment and in Ramsey and Montenegro (1992) for prime rates and expenditures for new plants and equipment. In an important paper, Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007) provided a simple characterization of fundamentalness in a state-space framework. More recent evidence of nonfundamental representations can be found in models with heterogenous information, see, e.g., Rondina (2008), in models with technology shock anticipation (see, e.g., Blanchard, L'Huillier, and Lorenzoni (2013) and Forni, Gambetti, and Sala (2014)) or in models of fiscal foresight (see, e.g., Leeper (1989), Yang (2005), and

---

[1]In this paper we rule out unit roots in the MA components for the reasons described below, and hence we use the terms invertibility and fundamentalness interchangeably, as they are equivalent under our setting. Also, invertibility and fundamentalness are relative concepts (e.g., a process is invertible relative to certain shocks), but to simplify the exposition we simply say the process is invertible.

[2]A GAUSS code to implement our test is available from the authors upon request. Concrete practical recommendations to implement our test are given in Appendix A.3.

Leeper, Walker, and Yang (2013)). For a comprehensive survey of this literature and further empirical evidence, see Alessi, Barigozzi, and Capasso (2011).

Despite the compelling empirical and theoretical evidence pointing to nonfundamental representations in economic models, and the significant implications for empirical work and policy analysis, little is known about how to empirically assess the lack of invertibility. The only proposal that we are aware of is that of Forni and Gambetti (2014). These authors relate lack of invertibility with informational sufficiency, and propose a method to check for invertibility based on the Granger causality of certain estimated factors. Their method crucially relies on estimation of the economic agent's information flows through a set of factors. In contrast, our test does not require additional information not included in the set of variables under testing. The method proposed by Forni and Gambetti (2014) has no power when the structural shocks are nonfundamental with respect to the information set of the agents, as is the case in the recent "news noise" literature (see, e.g., Blanchard, L'Huillier, and Lorenzoni (2013)). In such a case, enlarging the information set cannot solve the fundamentalness problem, since it does not arise from a gap of information between the econometrician and the agents; see Forni, Gambetti, Lippi, and Sala (forthcoming). By contrast, the present method can in principle be used to detect this kind of nonfundamentalness. Nevertheless, the method of Forni and Gambetti (2014) and our method are complements rather than substitutes, as they are based on different assumptions. Forni and Gambetti's (2014) method uses information from outside the observable data, and thus can be applied to Gaussian non-iid shocks and allows for a unit root in the MA (which is ruled out in our setting). Our method relies on the non-Gaussianity but does not require the specification and estimation of the underlying informational structure, and it is based on simple primitives of the model, namely, the Wold innovations. These innovations can be easily estimated for invertible and noninvertible processes.

Our approach is simple because it follows the traditional modeling strategy of imposing invertibility. Hence, under the null of invertibility standard inference applies. Furthermore, we do not need to identify the structural model under invertibility. All we need in our method are consistent estimates of Wold innovations. Under the alternative hypothesis of lack of invertibility, we face the situation where the econometrician fits an invertible model to a noninvertible one, and we base our omnibus test on the non-mds property of the resulting Wold innovations. Considering the null of invertibility and using an omnibus approach (i.e., not accounting for the vector autoregressive moving average (VARMA) structure in the Wold innovations under the alternative) avoids dealing with the difficult problems of identification and estimation of noninvertible roots, for which solutions are not yet available. It is known that if the true model is noninvertible, imposing invertibility has the potential to mislead structural VAR-based inferences in several aspects. Econometricians may fail to correctly identify economic shocks with interpretations of "information carriers" or "news" observed by private agents. Moreover, the subsequent policy analysis is likely to be incorrect as the resulting impulse-response functions may become unreliable. For instance, Leeper, Walker, and Yang (2013) provide compelling empirical evidence on the misleading inferences from fitting invertible MA representations to noninvertible VARMA processes in the context of fiscal foresight.

The aforementioned literature has suggested two different empirical strategies to deal with lack of invertibility. First, estimate directly a fully specified structural model. Inferences within this strategy may be quite sensitive to the correct specification of the model (e.g., the dynamics of information flows; see, e.g., Schmitt-Grohé and Uribe (2008, 2012) for sensitivity to different specifications). A second and more popular strategy within the foresight literature consists in expanding the econometrician's information set, so as to restore invertibility; see, for example, Leeper, Walker, and Yang (2013) and Ramey (2011). Not only can our test be used to empirically identify the invertibility problem in the first place, but it can also be applied to check if a proposed solution solves it, that is, whether or not adding certain variables restores invertibility. We carry out our new testing procedure for two of the most widely cited empirical works on fiscal shocks: Blanchard and Perotti (2002; henceforth BP) and Ramey (2011). Our empirical results do not suggest evidence against invertibility in BP's application and little evidence in Ramey's (2011) application. The additional informational variables suggested in the literature increase the probability of passing the invertibility test, with the exception of the "defense news" variable in Ramey (2011), which leads to a decrease in the test's p-value of approximately 45% in Ramey's specification. Interestingly, in a more parsimonious bivariate specification of Ramey (2011), our test strongly rejects invertibility and supports adding the defense news variable to restore invertibility.

An important by-product of our analysis is the clarification of identification of nonfundamental representations. Partly explained by the negative result in the Gaussian case, it is generally believed that parameters and shocks in nonfundamental representations are not identified. However, Cheng (1992) has proved that noninvertible univariate autoregressive moving average (ARMA) models are identified when the innovations' distribution is non-Gaussian. See also Ramsey and Montenegro (1992), and, more recently, Lanne, Meitz, and Saikkonen (2013) and Gospodinov and Ng (2015) for further identification results in univariate MA models. Despite the important efforts made to obtain similar results in the multivariate case, this still remains an open unresolved problem; see Chan, Ho, and Tong (2006).[3] Our results can be seen as a first step toward solving the identification problem in structural noninvertible VARMA models, as we show that invertible and noninvertible representations are generally not observationally equivalent in the multivariate non-Gaussian case.[4]

Our empirical strategy for testing invertibility relies on a new characterization of noninvertibility in VARMA models with non-Gaussian iid structural shocks, extending previous results by Rosenblatt (2000, Section 5.4) to the multivariate case.[5] Proving that

---

[3]Recently, Gourieroux and Monfort (2014) have investigated identification in the multivariate case.

[4]Non-Gaussianity is well motivated in economics; see, for example, Geweke (1993, 1994). It holds generically. Many features observed in economic data, such as fat tails, make normality implausible. Other features such as asymmetries, threshold effects, precautionary behavior, time irreversibility, and other phenomena of interest in macroeconomics are difficult to reconcile with the assumption of normality. Among the recent studies, Cúrdia, Negro, and Greenwald (2014) show that Student's t distribution is strongly favored by the usual set of macro time series data over the 1964–2011 period using dynamic stochastic general equilibrium (DSGE) models. Below, we provide empirical evidence of non-Gaussianity in BP's and Ramey's (2011) applications.

[5]It may be possible to relax the assumption of identically distributed errors. After the first version of this paper was written, Sahneh (2015) provided an argument that is valid for the univariate case.

Wold innovations are not a mds is equivalent to proving that the conditional mean for noninvertible non-Gaussian VARMA processes is nonlinear. Hence, our results have also independent implications for prediction and impulse response functions interpretation. We then propose a test for the mds property of the Wold innovations building on the generalized spectrum approach of Hong (1999) and Hong and Lee (2005), which accounts for all lags in the sample and has a simple standard normal distribution under the null of invertibility.

The rest of the paper is organized as follows. Section 2 provides a formal statement of the testing problem—the characterization of invertibility—as well as a motivational example. Section 3 introduces formally the test statistic based on the generalized spectrum and Section 4 investigates its asymptotic properties. Section 5 examines the finite-sample performance of the test through some Monte Carlo simulation experiments in the context of the fiscal foresight model of Leeper, Walker, and Yang (2013). Section 6 reports the applications of our tests to the settings of BP and Ramey (2011). Section 7 concludes. The Appendix contains details on implementation and a new test for white noise, which is of independent interest. The Supplemental Appendix, available in a supplementary file on the journal website, http://qeconomics.org/supp/393/supplement.pdf, contains further Monte Carlo simulations and proofs of the main results.

2. Characterization of fundamental representations

Let $\{\mathbf{x}_t\}_{t \in \mathbb{Z}}$ be a d-dimensional stationary solution of the causal VARMA model of order (p, q), satisfying the difference equations

$$\boldsymbol{\Phi}(L)\mathbf{x}_t = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\{\boldsymbol{\varepsilon}_t\}$ is a sequence of iid structural shocks with zero mean and identity variance–covariance matrix $\mathbf{I}_d$, and where

$$\boldsymbol{\Phi}(L) := \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_p L^p,$$

$$\boldsymbol{\Theta}(L) := \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 L + \cdots + \boldsymbol{\Theta}_q L^q$$

are the autoregressive and moving average polynomials, respectively. Henceforth, L is the lag operator (i.e., $L\mathbf{x}_t = \mathbf{x}_{t-1}$). We assume throughout that $\boldsymbol{\Phi}_p \neq \mathbf{0}$, $\boldsymbol{\Theta}_q \neq \mathbf{0}$, $\det \boldsymbol{\Phi}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, and that the equation $\det \boldsymbol{\Theta}(z) = 0$ has no roots on the unit circle of the complex plane (i.e., $\det \boldsymbol{\Theta}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$).[6] We assume that the VARMA representation is minimal. A sufficient condition for this is left coprimeness: if $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ have a left common factor $\mathbf{C}(L)$ such that $\boldsymbol{\Phi}(L) = \mathbf{C}(L)\boldsymbol{\Phi}^{(1)}(L)$ and $\boldsymbol{\Theta}(L) = \mathbf{C}(L)\boldsymbol{\Theta}^{(1)}(L)$, then $\det \mathbf{C}(L)$ is independent of L. Henceforth, we assume that the structural VARMA model (2.1) is correctly specified, and that all the variables involved have finite second moments.

---

[6]The case where $|z| = 1$ seems to be empirically less relevant in macroeconomics; see Watson (1986) for a discussion. Furthermore, there already exist methods for empirically detecting this case; see, for example, Tsay (1993).

The VARMA process $\{\mathbf{x}_t\}$ is said to be invertible if all the roots of the equation $\det \boldsymbol{\Theta}(z) = 0$ lie outside the unit circle in the complex plane, that is, $\det \boldsymbol{\Theta}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$ (see Brockwell and Davis (1991)). If the equation $\det \boldsymbol{\Theta}(z) = 0$ has a root inside the unit circle, we say the process $\{\mathbf{x}_t\}$ is noninvertible. We assume that $\{\mathbf{x}_t\}$ is defined on the probability space $(\Omega, \mathcal{F}, P)$ and let $L_2 := L_2(\Omega, \mathcal{F}, P)$ denote the Hilbert space of all real-valued measurable square-integrable functions on $(\Omega, \mathcal{F}, P)$. For a generic vector process $\{\mathbf{z}_t\}$, define $\mathcal{H}_t^{\mathbf{z}} = \overline{\text{span}}\{\mathbf{z}_s : s \leq t\}$ as the closed linear span of $\{\mathbf{z}_s : s \leq t\}$ in L_2. We say that $\{\boldsymbol{\varepsilon}_t\}$ is $\mathbf{x}_t$-fundamental if $\mathcal{H}_t^{\boldsymbol{\varepsilon}} = \mathcal{H}_t^{\mathbf{x}}$. Within our setting, $\{\mathbf{x}_t\}$ is invertible if and only if $\{\boldsymbol{\varepsilon}_t\}$ is $\mathbf{x}_t$-fundamental (see Rozanov (1967)). It is known that when $\{\boldsymbol{\varepsilon}_t\}$ is Gaussian, fundamental and nonfundamental representations are observationally equivalent and therefore cannot be discriminated based on data. When $\{\boldsymbol{\varepsilon}_t\}$ is non-Gaussian, fundamental and nonfundamental representations generally lead to different joint distributions of observables; see Breidt and Davis (1992) and Rosenblatt (2000). However, even in the univariate case, it is not known how to empirically discriminate between these two observationally different situations. This paper extends previous results in Rosenblatt (2000) to the multivariate case and proposes a simple test for empirically assessing this difference. The noninvertibility or nonfundamentalness problem often arises under circumstances when the econometricians' *linear* information set is smaller than the agents' information set, that is, $\mathcal{H}_t^x \subset \mathcal{H}_t^{\varepsilon}$.[7] The following example illustrates that the problem of lack of invertibility can be generic in settings of news or foresight.

EXAMPLE ((Fiscal Foresight)). Leeper, Walker, and Yang (2013) provide a simple growth model with two-quarter fiscal foresight, leading to the dynamics

$$k_t = \alpha k_{t-1} - \kappa \varepsilon_{t-1} - \kappa \theta \varepsilon_t, \tag{2.2}$$

where $\{k_t\}$ measures log deviations of capital stock from the steady state, $|\alpha| < 1$, $|\theta| < 1$, $\kappa \in \mathbb{R}$, and $\{\varepsilon_t\}$ is an iid sequence of tax news shocks, which agents will face after two periods. Considering the tax foresight, agents would respond to the news shocks that arrived at earlier periods much heavier than those that arrived today. Hence, the fact that more recent tax news is discounted heavier than older news makes the model in (2.2) noninvertible; see Leeper, Walker, and Yang (2013).

The Wold innovations in the model (2.2) are related to the structural shocks through the Blaschke filter (see Rozanov (1967))

$$u_t = \left(\frac{L + \theta}{1 + \theta L} \right) \varepsilon_t. \tag{2.3}$$

By definition, $u_t = k_t - L[k_t | \mathcal{H}_{t-1}^k]$, where, henceforth, $L[k_t | \mathcal{H}_{t-1}^k]$ denotes the optimal *linear* predictor of $\{k_t\}$ given its past. If $\{\varepsilon_t\}$ is Gaussian, then $\{u_t\}$, being uncorrelated and Gaussian, is an independent process. This paper builds on the observation that if $\{\varepsilon_t\}$ is non-Gaussian, then $\{u_t\}$ is not a mds. This follows from results by Rosenblatt

---

[7]Note that the econometrician's information set $\mathcal{F}_t^{\mathbf{x}} := \sigma(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots)$ is different from $\mathcal{H}_t^x$. To emphasize this crucial difference we refer to the latter as the *linear* information set.

(2000, Section 5.4), who showed that the optimal predictor $E[k_t|\mathcal{F}_{t-1}^k]$ is nonlinear when $\{\varepsilon_t\}$ is non-Gaussian. Nonlinearity means that $L[k_t|\mathcal{H}_{t-1}^k] \neq E[k_t|\mathcal{F}_{t-1}^k]$ or, equivalently, $E[u_t|\mathcal{F}_{t-1}^k] \neq 0$. Since $\mathcal{F}_{t-1}^u = \mathcal{F}_{t-1}^k$, we conclude that $\{u_t\}$ cannot be a mds.[8] Among other things, this implies that the standard errors for estimating the parameters in (2.2) based on the independence assumption of Wold innovations are in general invalid (see Francq, Roy, and Zakoïan (2005)), and, more importantly for policy analysis, the true impulse response function $E[k_{t+1}|\varepsilon_t = \delta, \mathcal{F}_t^k] - E[k_{t+1}|\varepsilon_t = 0, \mathcal{F}_t^k]$ will be different from $L[k_{t+1}|u_t = \delta, \mathcal{H}_t^k] - L[k_{t+1}|u_t = 0, \mathcal{H}_t^k]$ due to the nonlinearity.

Despite the evidence provided by Hansen and Sargent (1980, 1991), Lippi and Reichlin (1993), Leeper, Walker, and Yang (2013), and many others, the standard practice in empirical work is to rule out noninvertibility by restricting the parameter space to the invertible region. Imposing invertibility on a noninvertible model leads to misspecification and misleading inferences as illustrated in the previous example. We now extend the arguments of the previous example to VARMA models. This entails extending the results of Rosenblatt (2000) from the univariate to the multivariate case. Our arguments use properties of Blaschke matrices; see Rozanov (1967) and Lippi and Reichlin (1994). A Blaschke matrix is defined as a polynomial $\mathbf{A}(L) = \sum_{j=0}^{\infty} \mathbf{A}_j L^j$, where the $\mathbf{A}_j$'s are $(d \times d)$-dimensional matrices such that (i) $\mathbf{A}^*(L^{-1})\mathbf{A}(L) = \mathbf{I}_d$, where $\mathbf{A}^*$ is the conjugate transposed of $\mathbf{A}$, (ii) $\mathbf{A}(L)$ is one-sided in nonnegative powers of L, and (iii) $\boldsymbol{\Theta}(z)\mathbf{A}(z^{-1})$ has a power series expansion with square summable coefficients and $\det \boldsymbol{\Theta}(z)\mathbf{A}(z^{-1}) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. See Lippi and Reichlin (1994) for further details. Blaschke matrices have been used to characterize the mapping between the shocks $\{\varepsilon_t\}$ and the Wold innovations $\{\mathbf{u}_t\}$. These mappings correspond to the multivariate extensions of (2.3).

Suppose the true process is a noninvertible VARMA process (2.1), but we incorrectly fit an invertible VARMA as $\boldsymbol{\Phi}(L)\mathbf{x}_t = \widetilde{\boldsymbol{\Theta}}(L)\mathbf{u}_t$. Using Theorem 2 in Lippi and Reichlin (1994), one can show that Wold innovations are related to the original innovations $\{\boldsymbol{\varepsilon}_t\}$ through the equation

$$\mathbf{u}_t = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t, \tag{2.4}$$

which is an extension of (2.3). The matrix $\mathbf{A}(L) = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\boldsymbol{\Theta}(L)$ is a Blaschke matrix. Let r denote the number of noninvertible roots of $\boldsymbol{\Theta}(L)$, and assume these roots are simple. We will separately consider two cases: (i) $r = 1$ and (ii) $r > 1$.

For the case $r = 1$, Theorem 1 in Lippi and Reichlin (1994) shows that we can express

$$\mathbf{A}(L) = \boldsymbol{R}(b_1, L)\mathbf{K}_1,$$

where $\mathbf{K}_1$ is an orthogonal matrix, b_1 is the noninvertible root with $|b_1| < 1$, and

$$\boldsymbol{R}(\alpha, L) = \begin{pmatrix} \dfrac{L - \alpha}{1 - \alpha L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d-1} \end{pmatrix}.$$

---

[8]Henceforth, the concept of mds we use is with respect to its own history, that is, $E[u_t|\mathcal{F}_{t-1}^u] = 0$.

Define $\widetilde{\varepsilon}_t = \mathbf{K}_1 \boldsymbol{\varepsilon}_t$, and for a vector $\mathbf{a}_t$, let $a_{t,j}$ denote its jth component. From (2.4) and the discussion above,

$$u_{t,1} = \left(\frac{L - b_1}{1 - b_1 L} \right) \widetilde{\varepsilon}_{t,1},$$

which is similar to (2.3) and will allow us to apply results known from the univariate case.

For $r > 1$, we need conditions on the $d \times d$ matrices $\mathbf{H}_j$ with ikth element $h_{j,ik}$ given by

$$h_{j,ik} = E[\varepsilon_{t,j} \varepsilon_{t,i} \varepsilon_{t,k}].$$

ASSUMPTION A.1. (i) *Let $r = 1$ and $\{\widetilde{\varepsilon}_{t,1}\}$ be iid following a non-Gaussian distribution with $E|\widetilde{\varepsilon}_{t,1}|^\eta < \infty$, and with non-zero ηth cumulant, for some $\eta \geq 3$; (ii) Let $r > 1$ and the matrices $\mathbf{H}_j$ be finite and linearly independent for $j = 1, \ldots, d$. Furthermore, $\{\boldsymbol{\varepsilon}_t\}$ is iid and $\boldsymbol{\Theta}_q' \boldsymbol{\Theta}_0 \neq \mathbf{0}$.*

An assumption similar to Assumption A.1(i) is required in Rosenblatt (2000) for univariate noninvertible ARMA processes. This assumption allows for other components of $\widetilde{\boldsymbol{\varepsilon}}_t$ different from the first component to be Gaussian and/or serially dependent. More generally, the assumption of iid shocks cannot be dispensed with in our proofs, although Monte Carlo evidence below suggests that our results might be valid for mds structural shocks. For the case $r > 1$ we use Assumption A.1(ii). The linear independence of $\mathbf{H}_j$ follows, for example, if the components of $\boldsymbol{\varepsilon}_t$ are independent with nonzero third moments. This assumption allows us to extend the results of Rosenblatt (2000, Section 5.4) to the multivariate case. We could also extend our results for $r > 1$ to higher order moments in $\boldsymbol{\varepsilon}_t$ larger than three, as we do for $r = 1$, at the cost of a longer and more involved proof. Note that for the univariate case our assumptions boil down to those of Rosenblatt (2000, Corollary 5.4.3) (with the exception that Rosenblatt (2000) assumes $\boldsymbol{\Theta}_q \neq \mathbf{0}$ and $\boldsymbol{\Theta}_0 = 1$). The condition $\boldsymbol{\Theta}_q' \boldsymbol{\Theta}_0 \neq \mathbf{0}$ is often mild, particularly so if $\boldsymbol{\Theta}_0$ is nonsingular, as in this case it boils down to $\boldsymbol{\Theta}_q \neq \mathbf{0}$, which is needed for identification of q. Nevertheless, we only need this condition to hold for one set of observationally equivalent structural parameters. That is, if $\boldsymbol{\Theta}_q' \boldsymbol{\Theta}_0 = \mathbf{0}$ we can often multiply the whole VARMA process by a nonsingular matrix so that the resulting structural parameters satisfy the condition; see (5.4) and the discussion following that equation for an example.

The next result justifies our empirical strategy for testing invertibility.

THEOREM 1. *Let Assumption A.1(i) or A.1(ii) hold under noninvertibility. Then the causal non-Gaussian VARMA model (2.1) is noninvertible if and only if the Wold innovation process $\{\mathbf{u}_t\}$ is not a mds.*

3. GENERALIZED SPECTRUM-BASED TESTS

We aim to test the null hypothesis of invertibility,

$$\mathbb{H}_0 : \{\mathbf{x}_t\} \text{ is invertible in (2.1)}, \tag{3.1}$$

against the alternative given by the negation of $\mathbb{H}_0$, say $\mathbb{H}_A$. Based on Theorem 1, we propose to check for the condition

$$\{\mathbf{u}_t := \mathbf{u}_t(\boldsymbol{\theta}_0)\} \text{ is a mds for some } \boldsymbol{\theta}_0 \in \Omega \subset \mathbb{R}^s \tag{3.2}$$

versus the alternative that $\{\mathbf{u}_t(\boldsymbol{\theta}_0)\}$ is not a mds, where $\mathbf{u}_t(\boldsymbol{\theta}_0) = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\boldsymbol{\Phi}(L)\mathbf{x}_t \equiv \boldsymbol{\Pi}_{\boldsymbol{\theta}_0}(L)\mathbf{x}_t$ is the Wold innovation obtained from fitting an invertible VARMA representation (2.1) and $\boldsymbol{\theta}_0 = \text{vec}(\boldsymbol{\Phi}_0^{-1}\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_0^{-1}\boldsymbol{\Phi}_p, \boldsymbol{\Phi}_0^{-1}\widetilde{\boldsymbol{\Theta}}_0, \boldsymbol{\Phi}_0^{-1}\widetilde{\boldsymbol{\Theta}}_1, \ldots, \boldsymbol{\Phi}_0^{-1}\widetilde{\boldsymbol{\Theta}}_q)$ (vec denotes vectorization). We remark that we do not need to identify the structural parameters in $\boldsymbol{\Phi}$ or $\boldsymbol{\Theta}$. All we need are parametric consistent estimates of Wold innovations.

It is not straightforward to test the mds property of $\mathbf{u}_t : E[\mathbf{u}_t|\mathbf{u}_{t-1}, \mathbf{u}_{t-2}, \ldots] = 0$. First, the conditioning information set contains all past information and hence there is a "curse of dimensionality" problem associated with testing the mds property. Classical tests only check for a fixed number of lags, which may not be able to capture the dependence from all past history. Second, $\{\mathbf{u}_t\}$ may display serial dependence in higher order moments. The test should be robust to potential conditional heteroskedasticity and other time-varying higher order conditional moments. Third, given the unknown parameter $\boldsymbol{\theta}_0$, we need to construct a $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}$ for testing for the null (3.1), where T is the sample size. However, it is well established that, in general, estimation of unknown parameters gives rise to loss of the "nuisance parameter-free" property in the null limit distribution of statistical tests; see, for example, Durbin (1973). To overcome these problems while permitting for all lags in the sample, we consider a multivariate generalized spectral approach, which is an extension of the generalized spectrum method proposed by Hong (1999) and Hong and Lee (2005). Compared with other existing tests in the literature that check a growing number of lags as the sample size increases, such as Escanciano (2006), our test has the advantage of being asymptotically pivotal, with a standard normal limiting distribution and with the estimation uncertainty having no impact asymptotically.

Following Hong (1999) and Hong and Lee (2005), we define a generalized covariance function

$$\gamma_j(\mathbf{a}, \mathbf{b}) := \text{cov}[e^{i\mathbf{a}'\mathbf{u}_t}, e^{i\mathbf{b}'\mathbf{u}_{t-|j|}}] = \varphi_{|j|}(\mathbf{a}, \mathbf{b}) - \varphi(\mathbf{a})\varphi(\mathbf{b}),$$

where $\varphi_{|j|}(\mathbf{a}, \mathbf{b})$ is the joint characteristic function of $(\mathbf{u}_t, \mathbf{u}_{t-|j|})$, $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{2d}$, $\varphi(\mathbf{a})$ is the marginal characteristic function, and $i = \sqrt{-1}$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{i\mathbf{a}'\mathbf{u}_t(\boldsymbol{\theta})}\}$, where $\mathbf{a} \in \mathbb{R}^d$. The generalized spectral density is defined as the Fourier transform of $\gamma_j(\mathbf{a}, \mathbf{b})$,

$$f(\lambda, \mathbf{a}, \mathbf{b}) := \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j(\mathbf{a}, \mathbf{b})e^{-ij\lambda}, \tag{3.3}$$

where $\lambda \in [-\pi, \pi]$ is the frequency. Note that the function $f(\lambda, \mathbf{a}, \mathbf{b})$ can capture any type of pairwise serial dependence in $\{\mathbf{u}_t\}$, that is, dependence between $\mathbf{u}_t$ and $\mathbf{u}_{t-|j|}$ for any nonzero lag j. This is analogous to the higher order spectra (Brillinger and Rosenblatt (1967a, 1967b)) in the sense that $f(\lambda, \mathbf{a}, \mathbf{b})$ can capture the serial dependence in

higher order moments. An advantage of generalized spectral analysis is that it can capture cyclical patterns caused by both linear and nonlinear dependence (e.g., Hamilton and Lin (1996)).

The generalized covariance function $\gamma_j(\mathbf{a}, \mathbf{b})$ and the generalized spectrum $f(\lambda, \mathbf{a}, \mathbf{b})$ are not suitable for testing invertibility when $\{\mathbf{u}_t\}$ is a mds, because they can capture the serial dependence in mean and in higher order moments. However, just as the characteristic function can be differentiated to generate various moments, $\gamma_j(\mathbf{a}, \mathbf{b})$ and $f(\lambda, \mathbf{a}, \mathbf{b})$ can be differentiated to capture the serial dependence in various moments. To detect (and only detect) the serial dependence in conditional mean, we consider

$$f^1(\lambda, \mathbf{b}) := \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j^1(\mathbf{b}) e^{-ij\lambda}, \quad \text{where } \lambda \in [-\pi, \pi] \text{ and } \mathbf{b} \in \mathbb{R}^d,$$

(3.4)

$$\gamma_j^1(\mathbf{b}) := \partial \gamma_j(\mathbf{a}, \mathbf{b})/\partial \mathbf{a}|_{\mathbf{a}=0} = i \operatorname{cov}[\mathbf{u}_t, \exp(i\mathbf{b}'\mathbf{u}_{t-|j|})].$$

The function $\gamma_j^1(\mathbf{b})$ checks whether the autoregression function $E[\mathbf{u}_t|\mathbf{u}_{t-j}]$ is zero at lag j.

In the current context, $\{\mathbf{u}_t\}$ is unobservable and has to be estimated. Assume that we have T observations $\{\mathbf{x}_t\}_{t=1}^T$ of a process satisfying (2.1) and $\hat{\boldsymbol{\theta}}$ is such that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$, where $\boldsymbol{\theta}_0$ is the parameter that generates the Wold innovations $\mathbf{u}_t$ in (2.4). Given a $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, for example, a quasi-maximum likelihood estimator imposing the invertibility assumption (Boubacar Mainassara and Francq (2011)), we can compute residuals $\mathbf{u}_t(\hat{\boldsymbol{\theta}}) := \boldsymbol{\Pi}_{\hat{\boldsymbol{\theta}}}(L)\mathbf{x}_t$. We note that the lag polynomial $\boldsymbol{\Pi}_{\hat{\boldsymbol{\theta}}}(L)$ may involve an infinite number of lags and may not be feasible to compute. Thus, we may need to assume some initial values in computing $\mathbf{u}_t(\hat{\boldsymbol{\theta}})$ and we let $\hat{\mathbf{u}}_t := \hat{\mathbf{u}}_t(\hat{\boldsymbol{\theta}})$ denote the (approximated) residuals based on the observed information set $\{\bar{\mathbf{x}}_0, \mathbf{x}_1, \ldots, \mathbf{x}_T\}$, which contains some initial value $\bar{\mathbf{x}}_0 := (\mathbf{x}_0, \ldots, \mathbf{x}_{1-p}, \mathbf{u}_0, \ldots, \mathbf{u}_{1-q})$. We provide a condition (see Assumption A.3 in Section 4) to ensure that the use of initial values has no impact on the asymptotic distribution of the proposed test statistic. In applications, we recommend obtaining residuals $\mathbf{u}_t(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Pi}_{\hat{\boldsymbol{\theta}}}(L)\mathbf{x}_t$ from fitting a VAR model, as is commonly done, so $\hat{\boldsymbol{\theta}}$ is the vector of least squares estimates of a VAR model.

With the estimated residuals $\{\hat{\mathbf{u}}_t\}$, we can estimate $f^1(\lambda, \mathbf{b})$ by a smoothed kernel estimator

$$\hat{f}^1(\lambda, \mathbf{b}) := \frac{1}{2\pi} \sum_{j=1-T}^{T-1} \left(1 - |j|/T\right)^{1/2} k(j/h) \hat{\gamma}_j^1(\mathbf{b}) e^{-ij\lambda}, \quad \lambda \in [-\pi, \pi] \text{ and } \mathbf{b} \in \mathbb{R}^d,$$

where

$$\hat{\gamma}_j^1(\mathbf{b}) = \sum_{t=|j|+1}^{T} \frac{i}{T-|j|} \hat{\mathbf{u}}_t \widehat{\psi}_{t-|j|}(\mathbf{b}),$$

(3.5)

$\widehat{\psi}_t(\mathbf{b}) = \exp(i\mathbf{b}'\hat{\mathbf{u}}_t) - (T-|j|)^{-1} \sum_{t=|j|+1}^{T} \exp(i\mathbf{b}'\hat{\mathbf{u}}_t)$, $h = h(T)$ is a bandwidth, and $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel. Examples of $k(\cdot)$ include Bartlett, Daniell, Parzen, and quadratic spectral kernels (e.g., Priestley (1981, p. 442)). The factor $(1 - |j|/T)^{1/2}$ is a finite-sample correction and could be replaced by unity. Under certain conditions,

$\hat{f}^1(\lambda, \mathbf{b})$ is consistent for $f^1(\lambda, \mathbf{b})$. The lag order h is a smoothing parameter, and we will consider a data-driven choice of h and conduct sensitivity checks on the impact of the choice of h in our simulation and empirical study.

Under $\mathbb{H}_0$, we have

$$\gamma_j^1(\mathbf{b}) = 0 \quad \text{a.s. } \forall j \geq 1 \text{ and } \mathbf{b} \in \mathbb{R}^d. \tag{3.6}$$

Consequently, the generalized spectral derivative $f^1(\lambda, \mathbf{b})$ becomes flat as a function of λ,

$$f^1(\lambda, \mathbf{b}) = f_0^1(\lambda, \mathbf{b}) := \frac{1}{2\pi} \gamma_0^1(\mathbf{b}), \quad \lambda \in [-\pi, \pi] \text{ and } \mathbf{b} \in \mathbb{R}^d, \tag{3.7}$$

which can be consistently estimated by

$$\hat{f}_0^1(\lambda, \mathbf{b}) := \frac{1}{2\pi} \hat{\gamma}_0^1(\mathbf{b}), \quad \lambda \in [-\pi, \pi] \text{ and } \mathbf{b} \in \mathbb{R}^d. \tag{3.8}$$

The estimators $\hat{f}^1(\lambda, \mathbf{b})$ and $\hat{f}_0^1(\lambda, \mathbf{b})$ converge to the same limit under $\mathbb{H}_0$ and generally converge to different limits under $\mathbb{H}_A$. Thus, any significant divergence between them amounts to evidence of the violation of the mds property and, hence, of the invertibility of the process. We can measure the distance between $\hat{f}^1(\lambda, \mathbf{b})$ and $\hat{f}_0^1(\lambda, \mathbf{b})$ by the quadratic form

$$\begin{aligned}
\widehat{L} &:= \frac{\pi T}{2} \int \int_{-\pi}^{\pi} \left\| \hat{f}^1(\lambda, \mathbf{b}) - \hat{f}_0^1(\lambda, \mathbf{b}) \right\|^2 d\lambda \, dW(\mathbf{b}) \\
&= \sum_{j=1}^{T-1} k^2(j/h)(T-j) \int \left\| \hat{\gamma}_j^1(\mathbf{b}) \right\|^2 dW(\mathbf{b}),
\end{aligned} \tag{3.9}$$

where $\| \cdot \|$ denotes the Euclidean norm, the second equality follows by Parseval's identity, and $W(\mathbf{b}) = \prod_{c=1}^d W_0(b_c)$ with $W_0 : \mathbb{R} \to \mathbb{R}^+$ a nondecreasing weighting function that weighs sets symmetric about the origin equally. Examples of $W_0(\cdot)$ include the cumulative distribution function (CDF) of any symmetric probability distribution, either discrete or continuous.

The proposed test statistic for the invertibility hypothesis is an appropriately standardized version of $\hat{L}$,

$$\hat{Q}^1(h) = \left[\widehat{L} - \hat{C}^1(h)\right]/\sqrt{\hat{D}^1(h)}, \tag{3.10}$$

where

$$\hat{C}^1(h) = \sum_{j=1}^{T-1} k^2(j/h)(T-j)^{-1} \sum_{t=j+1}^{T} \|\hat{\mathbf{u}}_t\|^2 \left[\int |\hat{\psi}_{t-j}(\mathbf{b})|^2 dW(\mathbf{b})\right],$$

$$\hat{D}^1(h) = 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/h)k^2(l/h)$$

$$
\times \sum_{m=1}^{d} \sum_{q=1}^{d} \int \int \left| \frac{1}{T - \max(j, l)} \right.
$$

$$
\left. \times \sum_{t=\max(j,l)+1}^{T} \hat{u}_{t,m} \hat{u}_{t,q} \widehat{\psi}_{t-j}(\mathbf{b}_1) \widehat{\psi}^*_{t-l}(\mathbf{b}_2) \right|^2 dW(\mathbf{b}_1)\, dW(\mathbf{b}_2).
$$

Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. The factors $\hat{C}^1(h)$ and $\hat{D}^1(h)$ are approximately the mean and the variance of the quadratic form $\hat{L}$. Note that $\hat{Q}^1(h)$ involves d-dimensional numerical integrations, which can be computationally cumbersome when d is large. In practice, we recommend using a d-dimensional Gaussian CDF as W, since for this choice there is a closed-form expression for the test statistic, which is given in Appendix A.1.

4. Asymptotic properties

To derive the null asymptotic distribution of the test $\hat{Q}^1(h)$, we impose the following regularity conditions.

ASSUMPTION A.2. *The estimator $\hat{\boldsymbol{\theta}}$ is such that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$, where $\boldsymbol{\theta}_0$ is the parameter that generates the Wold innovations $\{\mathbf{u}_t\}$ in (2.4). Furthermore, $\boldsymbol{\theta}_0$ is in the interior of a compact parameter space Θ.*

ASSUMPTION A.3. *Let $\bar{\mathbf{x}}_0 := (\mathbf{x}_0, \ldots, \mathbf{x}_{1-p}, \mathbf{u}_0, \ldots, \mathbf{u}_{1-q})$ be some assumed initial values. Then $E\|\bar{\mathbf{x}}_0\|^2 < \infty$.*

ASSUMPTION A.4. (i) *Let $k : \mathbb{R} \to [-1, 1]$ be a symmetric function that is continuous at zero and all points in $\mathbb{R}$ except for a finite number of points;* (ii) $k(0) = 1$; (iii) $\int_0^\infty k^2(z)\, dz < \infty$; (iv) $k(z) \le c|z|^{-b}$ *for some $b > \frac{1}{2}$ as $|z| \to \infty$.*

ASSUMPTION A.5. *Let $W : \mathbb{R}^d \to \mathbb{R}^+$ be a nondecreasing weighting function that weighs sets symmetric about the origin equally, with $\int \|\mathbf{u}\|^4\, dW(\mathbf{u}) < \infty$.*

ASSUMPTION A.6. (i) *We have $E\|\boldsymbol{\varepsilon}_t\|^4 < \infty$;* (ii) *Let $\{\mathbf{u}_t\}$ be a strictly stationary α-mixing process with mixing coefficients satisfying $\sum_{j=0}^{\infty} \alpha(j)^{\frac{\nu}{2+\nu}} < \infty$ and $E\|\mathbf{u}_t\|^{4+2\nu} < \infty$ for some $\nu > 0$.*

In Assumption A.2, we permit but do not require $\hat{\boldsymbol{\theta}}$ to be a quasi-maximum likelihood or a least squares estimator. Any $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}$ suffices. Boubacar Mainassara and Francq (2011) provided primitive conditions for this assumption to hold. Assumption A.3 is a startup value condition. It ensures that the impact of initial values assumed in the observed information set is asymptotically negligible. Assumption A.4 covers most commonly used kernels. For kernels with bounded support, such as the Bartlett and Parzen kernels, we have $b = \infty$. For kernels with unbounded support, b is some finite positive real number. For example, we have $b = 2$ for the quadratic spectral

kernel. Assumption A.5 imposes mild conditions on the weighting function $W(\cdot)$. Any CDF with finite fourth moments satisfies Assumption A.5. Assumption A.6(ii), which is only required for Theorem 3, restricts the degree of temporal dependence in $\{\mathbf{u}_t\}$ under the alternative. This assumption is satisfied under mild conditions on the distribution of $\boldsymbol{\varepsilon}_t$, as shown in Theorem 3.1 of Pham and Tran (1985). A similar assumption was imposed in Francq and Zakoïan (1998), Boubacar Mainassara and Francq (2011), and Chen and Hong (2011).

THEOREM 2. *Suppose Assumptions* A.2–A.6(i) *hold and* $h = cT^\eta$ *for* $0 < \eta < 1$, *where* $0 < c < \infty$. *Then as* $T \to \infty$, $\hat{Q}^1(h) \xrightarrow{d} N(0, 1)$ *under* $\mathbb{H}_0$ *where* $\{\mathbf{u}_t\}$ *is a mds.*

An important feature of $\hat{Q}^1(h)$ is that the use of the estimated errors $\{\hat{\boldsymbol{u}}_t\}$ in place of the unobservable $\{\mathbf{u}_t\}$ has no impact on the limiting distribution of $\hat{Q}^1(h)$. One can proceed as if the true parameter value $\boldsymbol{\theta}_0$ were known and equal to $\hat{\boldsymbol{\theta}}$. The reason is that the convergence rate of the estimator $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ is faster than that of the nonparametric kernel estimator $\hat{f}^1(\lambda, \mathbf{b})$ to $f^1(\lambda, \mathbf{b})$. Consequently, the limiting distribution of $\hat{Q}^1(h)$ is solely determined by $f^1(\lambda, \mathbf{b})$, and replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$ has no impact asymptotically. This delivers a convenient procedure, because any $\sqrt{T}$-consistent estimator can be used. We remark that no distributional assumptions are needed in Theorem 2, that is, the theorem is also valid for Gaussian innovations.

Next, we establish the consistency of $\hat{Q}^1(h)$ for a large class of alternatives (i.e., non-invertible processes) under a weak dependence condition imposed by Assumption A.6.

THEOREM 3. *Suppose Assumptions* A.1–A.6 *hold, and* $h = cT^\eta$ *for* $0 < \eta < 1$, *where* $0 < c < \infty$. *Then as* $T \to \infty$,

$$\frac{h^{\frac{1}{2}}}{T}\hat{Q}^1(h) \xrightarrow{p} \frac{1}{\sqrt{D^1}} \sum_{j=1}^{\infty} \int \left\| \gamma_j^1(\mathbf{b}) \right\|^2 dW(\mathbf{b}), \tag{4.1}$$

where

$$D^1 = 2 \int_0^\infty k^4(z)\, dz \sum_{m=1}^{d} \sum_{q=1}^{d} \left[E(u_{t,m} u_{t,q}) \right]^2 \sum_{j=-\infty}^{\infty} \int \int \left| \gamma_j(\mathbf{a}, \mathbf{b}) \right|^2 dW(\mathbf{a})\, dW(\mathbf{b}).$$

Following a reasoning analogous to Bierens (1982) and Stinchcombe and White (1998), we have that for $j > 0$, $\gamma_j^1(\mathbf{b}) = 0$ for all $\mathbf{b} \in \mathbb{R}^d$ if and only if $E(\mathbf{u}_t | \mathbf{u}_{t-j}) = 0$. Thus, the generalized covariance derivative $\gamma_j^1(\mathbf{b})$ can capture various departures from the invertibility. Suppose $E(\mathbf{u}_t | \mathbf{u}_{t-j}) \neq 0$ at some lag $j > 0$. Then we have $\int \| \gamma_j^1(\mathbf{b}) \|^2 dW(\mathbf{b}) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing, and continuous, with unbounded support on $\mathbb{R}^d$. Consequently, $P[\hat{Q}^1(h) > C(T)] \to 1$ for any sequence of constants $\{C(T) = o(T/h^{1/2})\}$. Thus $\hat{Q}^1(h)$ has asymptotic unit power at any given significance level $\alpha \in (0, 1)$, whenever $E[\mathbf{u}_t | \mathbf{u}_{t-j}]$ is nonzero at some lag $j > 0$ under $\mathbb{H}_A$. However, notice that the hypothesis in (3.2) that the Wold innovation process is a mds is not the same as the hypothesis that $E(\mathbf{u}_t | \mathbf{u}_{t-j}) = 0$ for all $j > 0$. The latter

implies the former but not vice versa. Hence, our test will not be consistent against all alternatives. This is the price we need to pay to deal with the difficulty of the so-called curse of dimensionality problem. Nevertheless, the examples where $E(\mathbf{u}_t|\mathbf{u}_{t-j}) = 0$ for all $j > 0$ but $\{\mathbf{u}_t\}$ is not a mds may be rare in practice.

5. Monte Carlo evidence

5.1 *Simulation design: Bivariate fiscal foresight model*

This section presents our simulation exercises using a bivariate economic model of fiscal foresight studied in Leeper, Walker, and Yang (2013). Based on this model, we generate a battery of invertible and noninvertible MA processes, followed by the conventional VAR estimation procedure. We apply our proposed test to the VAR residuals, reporting the size and power of our test based on $\hat{Q}^1(h)$. Last, we conduct a variety of sensitivity checks.

To illustrate the effects of fiscal foresight, Leeper, Walker, and Yang (2013) provide a simple standard growth model with a representative household where log utility and complete depreciation of capital are assumed. The maximization of the agent's expected log utility leads to the equilibrium condition for capital stock in log-linearized form,

$$k_t = \alpha k_{t-1} + \varepsilon_{A,t} - (1-\theta)\left(\frac{\tau}{1-\tau}\right)\sum_{i=0}^{\infty}\theta^i E_t\widehat{\tau}_{t+i+1}, \tag{5.1}$$

where k_t, $\widehat{\tau}_t$, and $\varepsilon_{A,t}$ denote capital, the income tax rate, and the exogenous iid technology shock, respectively. The parameter α measures capital stock's persistence, with $0 < \alpha < 1$, β is the discount factor, with $0 < \beta < 1$, and τ is the steady state tax rate.[9] The parameter $\theta := \alpha\beta(1-\tau)$ governs the noninvertibility of the equilibrium MA representation in the presence of foresight. Leeper, Walker, and Yang (2013, p. 1122) show formally that "as θ^2 approaches unity (zero), the difference between the agent's and econometrician's information sets gets smaller (larger)." This implies that the problem of noninvertibility becomes increasingly more serious as the value of θ becomes smaller. Our simulation results support this claim.

In addition, to model the q-period foresight, a simple tax policy rule is specified as

$$\widehat{\tau}_t = \varepsilon_{\tau,t-q}, \tag{5.2}$$

which implies that agents are assumed to receive the tax news q periods before the tax shock $\varepsilon_{\tau,t}$ realizes.

Combining the equilibrium condition (5.1) with the tax rule (5.2) yields a set of equations to be used for our data simulation. Following Leeper, Walker, and Yang (2013), we examine the case of no foresight ($q = 0$) and the case of two-quarter foresight ($q = 2$). For

---

[9]Notice that the variables in the equilibrium condition (5.1) are expressed in terms of percentage deviations from steady state values. that is, $k_t := \log(K_t) - \log(K)$ and $\widehat{\tau}_t := \log(\tau_t) - \log(\tau)$, where K_t and τ_t denote capital and income tax rate at time t, respectively.

the former case, the equilibrium conditions (5.1) and (5.2) can be represented by a VAR process

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 - \alpha L \end{bmatrix} \begin{bmatrix} \widehat{\tau}_t \\ k_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{A,t} \end{bmatrix}.
\tag{5.3}
$$

In contrast, under the two-period foresight, the equilibrium dynamics can be characterized by a noninvertible VARMA process

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 - \alpha L \end{bmatrix} \begin{bmatrix} \widehat{\tau}_t \\ k_t \end{bmatrix} = \begin{bmatrix} L^2 & 0 \\ -\kappa(L + \theta) & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{A,t} \end{bmatrix},
\tag{5.4}
$$

where $\kappa := (1 - \theta)(\tau/(1 - \tau))$.[10] The joint law of motion of taxes and capital in (5.4) illustrates that the presence of fiscal foresight creates the seemingly perverse case where recent tax news receives heavier discounts compared to older news in determining capital accumulation. This is because already arrived (in period $t - 2$) tax news affects the contemporaneous tax rate (in period t), whereas the contemporaneous tax news would adjust the future tax rate (in period $t + 2$). For more details, see Leeper, Walker, and Yang (2013).

For our simulation experiments, we focus on non-Gaussian, iid structural shocks.[11] Given the foresight models, first, under the null hypothesis, we generate the bivariate, invertible MA representation with no foresight (5.3) using the following data generating processes (DGPs).

DGP1. No foresight model with iid shocks $\{\varepsilon_{\tau,t}, \varepsilon_{A,t}\}$, mutually independent, and distributed as a Student's t variable with 3 degrees of freedom; in short, $\boldsymbol{\varepsilon}_t \sim$ iid $t(3)$.

DGP2. No foresight model with iid shocks $\{\varepsilon_{\tau,t}, \varepsilon_{A,t}\}$, mutually independent, and distributed as a standardized chi-squared variable with 3 degrees of freedom; in short, $\boldsymbol{\varepsilon}_t \sim$ iid $\chi^2(3)$.

Next, under the alternative, we generate the noninvertible MA representation with the tax foresight (5.4) in combination with the following DGPs.

DGP3. Two-period foresight model with $\boldsymbol{\varepsilon}_t \sim$ iid $t(3)$.

DGP4. Two-period foresight model with $\boldsymbol{\varepsilon}_t \sim$ iid $\chi^2(3)$.

For the baseline calibration we use $\alpha = 0.4$, $\beta = 0.99$, and $\tau = 0.25$ (i.e., $\theta = 0.297$) to simulate the bivariate processes (5.3) and (5.4) under DGP1–DPG4, which is consistent with Leeper, Walker, and Yang (2013) and Forni and Gambetti (2014).

---

[10]Here the condition $\boldsymbol{\Theta}'_q \boldsymbol{\Theta}_0 \neq \mathbf{0}$ does not hold in (5.4). But we can multiply both sides of this equation by a nonsingular matrix with rows (a, b) and (c, d), respectively, such that $ab + cd \neq 0$, so that the resulting parameters satisfy $\boldsymbol{\Theta}'_q \boldsymbol{\Theta}_0 \neq \mathbf{0}$.

[11]Additional simulation results using non-Gaussian errors, conditional heteroskedastic errors as well as Gaussian errors are reported in the Supplemental Appendix.

5.2 *Simulation results*

Given the simulated data, we need to implement the conventional VAR estimation procedure to obtain the Wold residuals. This entails choosing a lag length p of the VAR(p) model. In our Monte Carlo experiments, we explore two alternatives. First, we consider using a fixed value for p through simulations, which is chosen following Kilian's (2001) method based on the finite-sample distribution of the lag order estimates for each lag order selection criteria: the Schwarz information criterion (SIC) and the Akaike information criterion (AIC), all of which lend strong support to the first-order VAR (i.e., $p = 1$).[12] Second, we consider choosing p by SIC, $\widehat{p}_{SIC}$ say, and by AIC, $\widehat{p}_{AIC}$ say, for each simulated sample. For the sake of space, we only report results for $\widehat{p}_{SIC}$ in the main text. The results for $\widehat{p}_{AIC}$ are slightly inferior to those for $\widehat{p}_{SIC}$, and they are reported in the Supplemental Appendix for completeness. Thus, in our simulations Wold innovations are computed from least squares estimators of the VAR(1) and VAR($\widehat{p}_{SIC}$) specifications.

For the choice of kernel $k(\cdot)$, we compare the performance of Bartlett kernel $k_B(z) := (1 - |z|)\mathbf{1}(|z| \leq 1)$ and Parzen kernel $k_P(z) := (1 - 6z^2 + 6|z|^3)\mathbf{1}(|z| \leq 1/2) + 2(1 - |z|^3)\mathbf{1}(1/2 \leq z \leq 1)$. For the weighting function $W(\cdot)$, we employ the standard normal distribution function $N(0, 1)$ (the closed-form solution for the test statistic is given in the Appendix). To select a proper bandwidth h for computing $\hat{Q}^1(h)$, we employ a data-driven lag order $\hat{h}$. Following Hong and Lee (2005), we use the plug-in bandwidth, which involves the choice of a preliminary bandwidth $\overline{h}$. To investigate the sensitivity of the choice of preliminary bandwidth $\overline{h}$ on the size and power of $\hat{Q}^1(\hat{h})$, we consider a wide range of the bandwidth $\overline{h} \in \{10, 11, \ldots, 40\}$. We implement our simulations with 1000 Monte Carlo iterations for each of sample sizes: $T = 100, 250,$ and 500.

Table 1 reports the empirical rejection probabilities of $\hat{Q}^1(h)$ under invertible (no foresight) models under DGP1 and DGP2 at the 10%, 5%, and 1% levels. As our simulation results are robust to the kernel choice, we report only the results using Bartlett kernel to save space. Under DGP1 and DPG2, $\hat{Q}^1(h)$ shows an excellent empirical size performance for both implementations, with fixed and data-driven choices of p. The size is sensitive to the bandwidth $\overline{h}$ for small sample sizes ($T = 100$), with smaller values of $\overline{h}$ leading to more accurate size results. However, the empirical size becomes stable as a function of the bandwidth for moderate and large values of T (i.e., $T = 250$ and 500).

Table 2 reports the empirical power of our proposed test against noninvertible (two-period foresight) models under DGP3 and DPG4 at the 10%, 5%, and 1% levels. For the noninvertible processes DGP3 and DPG4, our test has nontrivial power, particularly for errors with asymmetric distributions such as the χ^2. The power increases with the sample size, as expected. It also decreases with the initial bandwidth $\overline{h}$, which suggests that for both size accuracy and high power a small value of $\overline{h}$ such as $\overline{h} = 10$ is preferred. The additional simulation results using conditional heteroskedastic errors in the Supplemental Appendix suggest that our theory is also valid with conditional heteroskedastic structural errors. The empirical sizes are accurate for typical sample sizes used in macroeconomics (e.g., $T = 250$), although we observed some size distortions for the $t(3)$

---

[12]The simulation results for lag order distribution are reported in the Supplemental Appendix.

TABLE 1. Empirical size of the $\hat{Q}^1$ test.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
		DGP1: Invertible with non-Gaussian errors $\varepsilon_t \sim$ iid Student's $t(3)$								
VAR(1)	10	7.9	4.7	1.7	9.9	6.7	2.8	9.7	6.3	3.4
	20	6.2	2.8	0.7	9.1	5.4	1.3	9.8	6.5	2.4
	30	3.7	1.9	0.3	8.0	3.9	0.7	8.4	5.5	2.3
	40	2.8	1.0	0.2	5.8	3.1	0.6	7.9	4.9	1.7
VAR($\widehat{p}_{\text{SIC}}$)	10	7.4	4.7	1.5	9.6	7.1	3.4	10.5	7.2	3.4
	20	5.7	2.5	0.3	8.3	4.9	2.0	10.8	6.9	3.0
	30	3.6	1.4	0.0	6.5	4.0	1.1	10.0	5.9	2.2
	40	2.3	1.0	0.0	5.3	3.0	0.7	7.8	4.9	2.0
		DGP2: Invertible with non-Gaussian errors $\varepsilon_t \sim$ iid standardized $\chi^2(3)$								
VAR(1)	10	5.8	3.7	1.4	7.6	4.8	2.4	6.9	5.4	2.6
	20	6.3	3.3	1.2	8.2	5.3	2.1	9.2	6.4	2.9
	30	5.3	2.8	0.7	8.4	3.8	1.8	9.3	6.6	2.4
	40	3.9	1.9	0.4	7.7	3.9	1.2	9.2	6.2	2.3
VAR($\widehat{p}_{\text{SIC}}$)	10	5.1	3.4	0.9	7.7	5.1	2.1	6.4	4.6	2.3
	20	5.0	2.3	0.4	9.2	5.2	2.1	8.4	5.9	2.3
	30	4.1	1.8	0.2	7.7	4.9	1.9	9.5	6.4	2.4
	40	3.1	1.2	0.1	7.0	4.1	1.1	9.8	5.6	2.7

TABLE 2. Empirical power of the $\hat{Q}^1$ test.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
		DGP3: Noninvertible with non-Gaussian errors $\varepsilon_t \sim$ iid Student's $t(3)$								
VAR(1)	10	14.0	8.8	4.7	31.7	24.4	14.1	58.0	51.8	40.4
	20	9.6	5.5	1.7	22.7	15.4	8.2	48.4	40.5	26.0
	30	6.9	3.0	0.6	16.2	10.1	4.4	40.2	29.5	17.3
	40	4.3	1.7	0.2	13.3	7.7	2.5	31.8	21.9	11.2
VAR($\widehat{p}_{\text{SIC}}$)	10	13.2	9.6	4.3	32.2	26.9	15.7	57.6	51.1	39.9
	20	8.4	5.1	1.4	25.7	17.3	7.7	47.2	38.8	26.7
	30	5.7	2.8	0.5	19.8	12.2	4.5	39.0	29.5	17.9
	40	4.2	1.2	0.1	15.3	8.7	2.6	32.4	23.0	11.6
		DGP4: Noninvertible with non-Gaussian errors $\varepsilon_t \sim$ iid standardized $\chi^2(3)$								
VAR(1)	10	33.9	25.9	15.1	79.6	75.9	66.0	98.7	98.4	96.8
	20	23.4	14.4	7.0	70.1	62.6	46.5	96.2	95.5	91.6
	30	17.1	10.3	3.6	59.7	49.7	32.8	93.3	90.5	83.8
	40	13.2	6.7	1.4	51.1	39.2	22.8	89.6	85.2	75.0
VAR($\widehat{p}_{\text{SIC}}$)	10	33.2	26.3	16.3	78.9	74.7	65.9	99.0	98.7	97.0
	20	23.5	16.3	6.6	69.5	63.0	48.5	96.2	95.0	91.4
	30	16.9	9.8	3.1	61.1	51.5	35.0	93.3	90.6	84.0
	40	11.9	6.2	1.4	53.5	42.7	24.4	89.2	84.7	76.5

distribution. The fact that for heteroskedastic errors with a $\chi^2(3)$ distribution the empirical size is quite accurate suggests that this distortion may be due to the fat tails of the process rather than due to the presence of conditional heteroskedasticity.

In sum, these Monte Carlo simulations show a satisfactory finite-sample performance for our test under the null and under the alternative. Nevertheless, to gain further insights into the finite-sample properties of our test, we carried out a detailed sensitivity analysis.

5.3 *Sensitivity analysis*

We implement sensitivity checks with the degree of noninvertibility, the choice of VAR lag orders, the degree of persistence in the process, and the degree of non-Gaussianity using the parametrized model of Leeper, Walker, and Yang (2013). To quantify the degree of invertibility we follow Leeper, Walker, and Yang (2013, p. 1122) and use θ as a metric, $0 < \theta < 1$, with larger values of θ corresponding to cases closer to invertibility.

We examine the effects of the severity in the noninvertibility problem on the finite-sample performance of our test by considering values of θ different from the baseline value of 0.297 used in Tables 1 and 2. First we do this in conjunction with the sensitivity to the persistence in the process by varying the persistence parameter $\alpha \in [0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99]$, setting the discount factor and the steady state tax rate to $\beta = 0.99$ and $\tau = 0.25$, respectively, and computing the corresponding θ with $\theta = \alpha\beta(1 - \tau)$, leading to the values in $\theta \in [0.007, 0.074, 0.149, 0.223, 0.297, 0.371, 0.446, 0.520, 0.594, 0.668, 0.735]$. We then simulate DGP4 in (5.4) for each (α, θ, κ), where recall $\kappa = (1 - \theta)(\tau/(1 - \tau))$. To isolate the effect on the test of the change in the degree of invertibility from the change of persistence, we also run another set of experiments where the persistence parameter is fixed at $\alpha = 0.4$, β and τ are set at the same values, $\beta = 0.99$ and $\tau = 0.25$, and θ varies independently of other parameters with the same values as above (i.e., we do not use the relation $\theta = \alpha\beta(1 - \tau)$).

First, we evaluate the sensitivity to the degree of noninvertibility for a fixed degree of persistence ($\alpha = 0.4$). Figure 1 depicts the empirical rejection probabilities of $\hat{Q}^1(h)$ under the noninvertible foresight model with standardized chi-squared errors (DGP4) at the 5% level for $T = 100$, 250, and 500. The displayed results are obtained with Bartlett kernels and preliminary bandwidth $\overline{h} = 10$. Our test becomes increasingly powerful against DGP4 as the problem of noninvertibility becomes more serious, that is, the value of θ becomes smaller, as is highlighted with the case of $T = 500$. This empirical evidence supports the formal results given in Leeper et al. (2013, p. 1122). It also supports Sims's (2012) argument that the problem of nonfundamentalness is not a *binary*—either/or—proposition. The results varying the degree of persistence are reported in the Supplemental Appendix, and suggest that the power results of Figure 1 are not sensitive to the persistence parameter. We also simulated under the null DGP2, varying the degree of persistence. Large values of persistence such as $\alpha = 0.99$ do seem to have an effect on the empirical size of the test, which suggests a different asymptotic theory for nonstationary processes. Establishing this theory is beyond the scope of this study. Hence, practitioners are highly recommended to transform the data to induce stationarity by
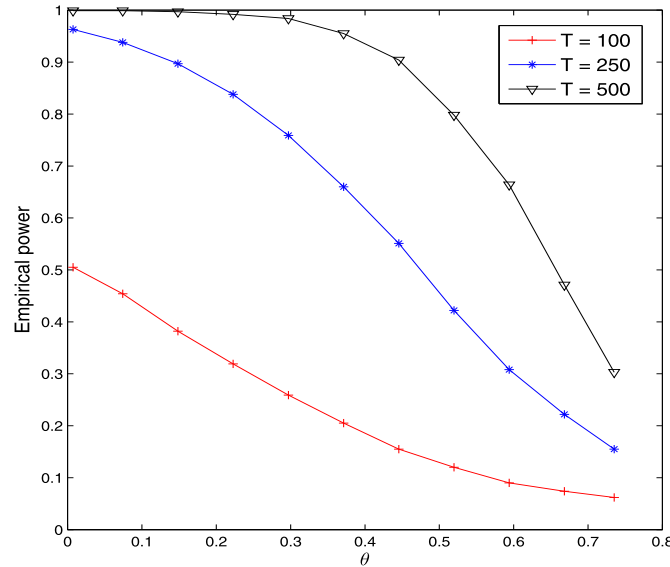
FIGURE 1. Sensitivity of the $\hat{Q}^1$ test power performance to the noninvertibility governing parameter (θ) under the DGP4 with $\alpha = 0.4$, $\beta = 0.99$, and $\tau = 0.25$. As the value of θ becomes smaller, the problem of noninvertibility becomes increasingly serious.

the conventional methods, for example, differencing, allowing for trends, or dividing by another variables, before applying our test.

Next we investigate the sensitivity to the lag order choice of VAR. Table 3 reports the finite-sample performance of $\hat{Q}^1(h)$ for different VAR lag lengths under DGP1 and DPG2, and DGP3 and DPG4. The size and power of $\hat{Q}(h)$ decrease with the lag lengths selected. For the most likely choice selected by the SIC and AIC criteria, which was $p = 1$, the empirical size and power results are best. This suggests that using selection criteria for the choice of VAR lag order is an important part of our procedure. Overfitting the model may lead to a significant reduction in power for our test. This also may explain why AIC performs slightly worse than SIC in our simulations, as it tends to choose larger values of p. Table 4 summarizes the sensitivity of the empirical size and power of $\hat{Q}^1(h)$ at the 5% level to the degree of non-Gaussianity using the Student's t and the standardized χ^2 distributions. Under the Student's t distribution with degrees of freedom (dof) $dof \geq 3$, the empirical level and power of $\hat{Q}^1(h)$ are most satisfactory when the degree of freedom is 3. Moreover, under the standardized χ^2 distribution, the empirical power of $\hat{Q}^1(h)$ increases as the degree of freedom declines, that is, the degree of non-Gaussianity increases. For the size, $\hat{Q}^1(h)$ exhibits satisfactory empirical levels uniformly in the values of dof. To sum up, our sensitivity analysis suggests that the finite-sample power of $\hat{Q}^1(h)$ improves as the noninvertibility problem becomes more severe and the innovation distribution is further away from the Gaussian. Moreover, a selection criteria for choosing the VAR lag order such as SIC should be used, as overfitting may lead to a decrease in power for our test.

TABLE 3. Lag sensitivity of the $\hat{Q}^1$ test performance.

| | $\varepsilon_t \sim$ iid Student's $t(3)$ | | | | | |
| | Size (DGP1) | | | Power (DGP3) | | |
Lag	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
1	4.7	6.7	6.3	8.8	24.4	51.8
2	3.0	3.4	4.1	5.2	18.0	40.7
3	2.0	2.9	2.5	4.2	15.7	34.3
4	1.1	2.6	2.7	2.5	13.9	35.5
	$\varepsilon_t \sim$ iid standardized $\chi^2(3)$					
	Size (DGP2)			Power (DGP4)		
1	3.7	4.8	5.4	25.9	75.9	98.4
2	0.8	1.1	1.8	15.4	68.0	96.2
3	0.1	0.8	1.1	10.6	58.2	95.7
4	0.3	0.7	0.6	9.3	54.3	95.3

Note: The significance level is at 5%.

TABLE 4. Sensitivity of the $\hat{Q}^1$ test performance to the degree of non-Gaussianity.

| | Size | | | Power | | |
dof	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
	$\varepsilon_t \sim$ Student's t distributed with dof degrees of freedom					
3	4.7	6.7	6.3	8.8	24.4	51.8
4	4.9	6.4	6.5	7.5	17.4	35.1
5	4.9	5.8	6.5	6.5	13.4	24.0
6	5.0	5.5	6.4	5.7	11.0	19.3
9	4.8	4.6	5.7	5.2	9.0	12.6
12	4.7	4.6	5.6	4.5	8.4	10.5
	$\varepsilon_t \sim$ Standardized χ^2 distributed with dof degrees of freedom					
3	3.7	4.8	5.4	25.9	75.9	98.4
4	3.9	4.7	5.1	19.8	62.9	94.2
5	3.5	5.1	5.1	15.9	52.1	88.0
6	3.6	4.7	5.1	14.2	44.7	81.6
9	3.7	4.8	5.0	10.3	31.9	63.6
12	3.9	4.9	5.2	8.4	25.2	51.1

Note: The significance level is at the 5%.

6. Empirical application

This section applies our invertibility test to two widely cited studies on the effects of fiscal policy: BP and Ramey (2011). To this end, we begin with replicating the baseline analysis in these studies without modifications.[13] Then we apply our test procedure to

---

[13] The replication data and programs are accessible in a supplementary file on the journal website, http://qeconomics.org/supp/393/code_and_data.zip.

the resulting residuals, also examining how test results change as we vary the lag length of the VAR. Last, we assess whether existing studies that incorporate informational variables into VARs are more likely to pass our test than studies excluding such variables.

In the macro fiscal literature, the dynamic effects of government taxes and/or spending on key macro variables such as consumption and real wages have long been of interests to academic researchers and policymakers because, in particular, the success of fiscal stimulus packages to boost the economy may be highly dependent on whether government spending boosts private consumption, investment, and so forth.

Among prevailing VAR methods in the literature to address this issue, a so-called *narrative* approach developed by Ramey and Shapiro (1998) exploits exogenous "war dates" to identify government spending shocks, concluding that consumption and real wages fall after a positive spending shock hits the economy, which supports a neoclassical view of the government spending effect; see, for example, Edelberg, Eichenbaum, and Fisher (1999), Burnside, Eichenbaum, and Fisher (2004), and Ramey (2011).

In contrast, a statistical innovation-based approach proposed by BP has been more standard in the related literature, yielding a conclusion to the opposite of the narrative approach that consumption and real wages tend to rise after the shock, which is consistent with new Keynesians' economic models; see, for example, Perotti (2005, 2008), Galí, López-Salido, and Vallés (2007), and Mountford and Uhlig (2009).

To reconcile these different identification strategies, Ramey (2011) argues that the presence of fiscal foresight gives rise to a mistiming of the news in the statistical innovation-based VAR approach, suggesting the use of a defense news variable as a proxy for the expected discounted value of government spending changes so as to improve the Ramey–Shapiro approach.[14] In a similar vein, Leeper, Walker, and Yang (2013) proposed adding asset prices to VAR models with foresight, seeking to fill the information gap between the econometrician and the agent. As an illustration, they augment BP's VARs with the implicit tax rates, capturing information flows on pending tax changes based on the spreads between the U.S. municipal and treasury bonds.

6.1 *Blanchard and Perotti (2002)*

BP constructed trivariate VAR models including quarterly taxes, spending, and gross domestic product (GDP) in log real per capita terms to gauge the dynamic effects of discretionary fiscal policy shocks on the economy. BP took quarterly data from 1947:1 to 1997:4 allowing for alternative model specifications with deterministic or stochastic trends. To apply our test, we compute BP's least squares estimates for the trivariate VAR under the specification with deterministic (quadratic) trends and dummies, which is termed BP's baseline model. We assess which lag length would be preferable based on information-based criteria as implemented in the simulation study. According to the information criteria for BP's specification, SIC supports the lag choice $p = 1$. Notice that in the subsequent analysis we also construct two information variable augmented BP models inspired by Leeper, Walker, and Yang (2013). More specifically, we add the implicit tax rates

---

[14]For recent empirical studies that consider the "anticipated" nature of government spending shocks, see, for example, Mertens and Ravn (2010), Fisher and Peters (2010), and Leeper, Walker, and Yang (2013).

Table 5. The p-values of residual tests for Blanchard and Perotti (2002).

| | | | Information Variable Augmented | |
Test	Lag	Baseline	1-yr Spread	5-yr Spread
Normality (λ_{sk})	1	0.00	0.00	0.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
No heteroskedasticity (LMF)	1	0.67	0.92	0.93
	2	0.91	0.74	0.94
	3	0.85	0.70	0.93
	4	0.69	0.79	0.74
No serial correlation ($\hat{Q}^{1,1}$)	1	0.00	0.01	0.00
	2	0.24	0.78	0.63
	3	0.78	0.82	0.82
	4	0.75	0.77	0.78

using the 1-year or the 5-year municipal bond spreads to the baseline BP models for our comparisons.[15]

Table 5 reports some residual tests for the baseline and augmented BP models, allowing for lag orders up to $p = 4$ to diagnose the presence of non-Gaussianity, heteroskedasticity, and autocorrelation in VAR residuals, which would be helpful in interpreting the subsequent invertibility test results. The multivariate normality test (λ_{sk}) proposed by Lütkepohl (1991, pp. 155–158), which compares the third and fourth moments of the residuals to those from the Gaussian distribution, strongly rejects the null of normality in the residuals for all the model specifications considered, which is a prerequisite for applying our tests. In contrast, the multivariate extension of White's (1980) heteroskedasticity test, which is an F-approximation to the likelihood ratio form of the Lagrange multiplier test (LMF) developed by Kelejian (1982) and Doornik (1996), fails to reject the null of no heteroskedasticity across all the lags considered.[16] Furthermore, our correlation test ($\hat{Q}^{1,1}$), which is proposed in Section A.2, with the Bartlett kernel and preliminary bandwidth $\bar{h} = 10$ shows that we reject the null of no correlation in Wold residuals with one lag for all the models considered, but the test does not reject for lag orders larger than or equal to 2. Thus, to avoid dynamic misspecification in Wold innovations we need to fit VAR models with at least two lags. BP's specification used $p = 4$. Our results suggest that a more parsimonious model with $p = 2$ seems to capture well the linear dynamics of the process. Table 6 presents our invertibility test results for the BP's VAR models allowing for lags $p = 2$ through 4. Guided by Leeper, Walker, and Yang (2013), we add the implicit tax rates using the 1-year or the 5-year municipal bond spreads to the baseline BP models to assess whether incorporating such information variables would

---

[15]The implicit tax rate data used in Leeper, Walker, and Yang (2013) were kindly provided by Todd B. Walker.

[16]The aforementioned VAR residual tests for non-Gaussianity and heteroskedasticity can be readily computed using widely used econometrics software such as EViews or R.

TABLE 6. Tests of invertibility for Blanchard–Perotti (2002) specifications.

| | | Baseline | | Information Variable Augmented | | | |
| | | | | 1-yr Spread | | 5-yr Spread | |
p-Values	Lag	Bartlett	Parzen	Bartlett	Parzen	Bartlett	Parzen
$\hat{Q}^1$	2	0.488	0.463	0.760	0.757	0.531	0.526
	3	0.702	0.695	0.957	0.949	0.887	0.888
	4	**0.783**	**0.781**	**0.973**	**0.969**	**0.964**	**0.965**

Note: The boldface type indicates BP's original specification.

increase the possibilities to pass our tests. The reported results are based on the preliminary bandwidth $\bar{h} = 10$ for the Bartlett and the Parzen kernels. Under BP's original specification with $p = 4$, our test fails to reject the null of invertibility. We also find little empirical evidence against invertibility in more parsimonious specifications with $p = 2$ and $p = 3$, suggesting that overparametrization is not the cause of the lack of rejection. Likewise, concerns with the low power of our test when the distribution of shocks is close to Gaussian are not empirically supported by the results of the Gaussianity test in Table 5. Adding the informational variables significantly increases the already high probability of passing the null of invertibility, particularly with the 1-year municipal bond spread.

6.2 *Ramey (2011)*

Next, we apply our test procedures to Ramey's (2011) seven-variable VAR models including government spending, GDP, total hours worked, nondurable plus service consumption, private fixed investment, tax rates, and real wages. For the baseline model, Ramey took quarterly data from 1947:1 to 2008:4, allowing for a quadratic time trend. Again we obtain the Ramey's original least squares estimates with four lags allowed. The SIC information-based criteria favor a lag order of $p = 1$. Notice that as with the BP application, we also construct two information variable augmented models—the one with the war dates variable used in Ramey and Shapiro (1998) and the other one with the defense news variable constructed by Ramey (2011).[17] As our normality tests (upper panel) in Table 7 strongly reject the null of Gaussianity for all the specifications, we proceed to apply our test to the estimated residuals from Ramey's (2011) specification. Furthermore, the White tests (middle panel) provide substantial evidence of heteroskedasticity in the residuals. The correlation test $\hat{Q}^{1,1}$ proposed in Section A.2, with the Bartlett kernels and preliminary bandwidth $\overline{h} = 10$ shows that consistent estimation of Wold innovations requires at least two lags. Ramey's (2011) specification used $p = 4$. Our empirical results suggest that $p = 3$ provides a similar fit while being more parsimonious.

---

[17]For our replication purposes, we exactly followed Ramey's different data spans for each of the specifications considered: 1947:1–2008:4 for the baseline and the war dates augmented VAR models; 1939:1–2008:4 for the defense news augmented VAR model, which consists of the defense news, government spending, per capita GDP, three-month T-bill rate, the income tax rate, and real wages.

TABLE 7. The p-values of residual tests for Ramey (2011).

| | | | Information Variable Augmented | |
Test	Lag	Baseline	War dates	Defense news
Normality (λ_{sk})	1	0.00	0.00	0.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
No heteroskedasticity (LMF)	1	0.00	0.00	0.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
No serial correlation ($\hat{Q}^{1,1}$)	1	0.00	0.00	0.00
	2	0.37	0.43	0.25
	3	0.74	0.77	0.77
	4	0.70	0.74	0.78

TABLE 8. Tests of invertibility for Ramey (2011) specifications.

| | | | | Information Variable Augmented | | | |
| | | Baseline | | War Dates | | Defense News | |
p-Values	Lag	Bartlett	Parzen	Bartlett	Parzen	Bartlett	Parzen
$\hat{Q}^1$	2	0.106	0.142	0.344	0.369	0.351	0.397
	3	0.273	0.296	0.365	0.369	0.278	0.355
	4	**0.369**	**0.420**	**0.469**	**0.481**	**0.212**	**0.274**

Note: The boldface type indicates Ramey's original specification.

Table 8 shows the test results of the Ramey's (2011) VAR models with lags $p = 2$–4. Ramey's (2011) baseline model is compared to two extended versions augmented with information variables on anticipated government spending: The war dates variable used in Ramey and Shapiro (1998) and the defense news variable constructed by Ramey (2011). Our results for the baseline model show that under Ramey's (2011) original specification with $p = 4$, the test fails to reject the null of invertibility, although the evidence is not as pronounced as with BP's application. Surprisingly, adding the defense news variable does not increase the probability of passing the invertibility test, whereas the war dates have more informational content according to our test. In more parsimonious specifications than those considered in Ramey (2011) there is more evidence against invertibility, and incorporating anticipated spending news variables helps significantly to increase the probability of passing our invertibility test. To illustrate these points, we apply our testing procedure to a bivariate VAR specification consisting of government spending and GDP. Table 9 summarizes the test results of no serial correlation $\hat{Q}^{1,1}$ and of invertibility $\hat{Q}^1$ for the parsimonious version of Ramey's (2011) specification, compared with those for the information-augmented models. For all of the three specifica-

TABLE 9. Tests of invertibility for bivariate specifications of Ramey (2011).

| | | Simple (Bivariate VAR) | | Information Variable Augmented | | | |
| | | | | War Dates | | Defense News | |
p-Values	Lag	Bartlett	Parzen	Bartlett	Parzen	Bartlett	Parzen
$\hat{Q}^{1,1}$	2	0.352	0.332	0.215	0.220	0.785	0.785
	3	0.862	0.871	0.832	0.833	0.799	0.800
	4	0.862	0.869	0.879	0.872	0.805	0.806
$\hat{Q}^1$	2	0.000	0.001	0.021	0.028	0.473	0.535
	3	0.153	0.170	0.234	0.240	0.324	0.391
	4	0.231	0.247	0.367	0.374	0.061	0.102

Note: The bivariate model (simple) includes government spending and GDP.

tions considered, the correlation test $\hat{Q}^{1,1}$ suggests at least two lags for consistent estimation of the Wold innovations. In addition, our test $\hat{Q}^1$ strongly rejects the null of invertibility for the bivariate model with lag $p = 2$. More interestingly, $\hat{Q}^1$ also rejects the null for the war dates augmented model at the 5% level, whereas failing to reject for the defense news augmented model.

7. Conclusions

This paper provides a simple empirical tool for evaluation of the conventional and key invertibility or fundamentalness assumption in macroeconomic models. We convert the invertibility testing problem into one of testing for the mds property of the Wold innovations. To test this property we employ a nonparametric smoothing method based on a multivariate extension of Hong's (1999) generalized spectral density. Our proposed test has a convenient asymptotic $N(0, 1)$ distribution under invertibility and the estimation uncertainty has no impact on the limiting distribution. Our Monte Carlo study reports satisfactory finite-sample performance of our proposed test. The applications to two widely cited studies on the effects of fiscal shocks illustrate the use of the new test, in combination with other diagnostic tests of Gaussianity and correct dynamic specification of the Wold innovations (i.e., testing for white noise). The proposed test of white noise appears to be new in the literature and is of independent interest (see Section A.2).

Existing recommendations in the presence of non-fundamentalness in the data include fitting a full DSGE model (Hansen and Sargent (1980) and Fernández-Villaverde et al. (2007)), using a large dimensional dynamic factor model (Forni, Giannone, Lippi, and Reichlin (2009)), and searching for informational variables to restore invertibility (see, e.g., Leeper, Walker, and Yang (2013)). Another strategy would be to identify and estimate the noninvertible model. This would allow for appropriate estimation of the impulse response functions and policy analysis. The present paper can be considered as a first attempt to solve the identification problem, but further research on identification and estimation is guaranteed.

<center>Appendix</center>

A.1 Closed-form expression for the test statistic

When $W(\cdot)$ is chosen as the d-dimensional Gaussian CDF, we can obtain a closed-form expression for the test statistic which is given by

$$
\hat{Q}_n^1(h) = \Bigg\{ \sum_{j=1}^{T-1} k^2(j/h) \bigg[T_j^{-1} \sum_{t_1,t_2=j+1}^{T} \hat{\mathbf{u}}_{t_1}' \hat{\mathbf{u}}_{t_2} \bigg(M_{t_1,t_2,j} - 2T_j^{-1} \sum_{t_3=j+1}^{T} M_{t_1,t_3,j}
$$
$$
+ T_j^{-2} \sum_{t_3,t_4=j+1}^{T} M_{t_3,t_4,j} \bigg) \bigg] - \hat{C}_n^1(h) \Bigg\} \Big/ \sqrt{\hat{D}_n^1(h)},
$$

where

$$
\hat{C}_n^1(h) = \sum_{j=1}^{T-1} k^2(j/h) \bigg[T_j^{-1} \sum_{t_1=j+1}^{T} \|\hat{\mathbf{u}}_{t_1}\|^2 \bigg(1 - 2T_j^{-1} \sum_{t_2=j+1}^{T} M_{t_1,t_2,j}
$$
$$
+ T_j^{-2} \sum_{t_2,t_3=j+1}^{T} M_{t_2,t_3,j} \bigg) \bigg],
$$

$$
\hat{D}_n^1(h) = 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/h) k^2(l/h)
$$
$$
\times \sum_{m=1}^{d} \sum_{q=1}^{d} \sum_{t_1,t_2=\max(j,l)+1}^{T} \bigg[T_{jl}^{-1} \hat{u}_{t_1,m} \hat{u}_{t_1,q} \hat{u}_{t_2,m} \hat{u}_{t_2,q} \bigg(M_{t_1,t_2,j} M_{t_1,t_2,l}
$$
$$
- 4T_{jl}^{-1} \sum_{t_3=\max(j,l)+1}^{T} M_{t_1,t_2,j} M_{t_1,t_3,l} + 2T_{jl}^{-2} \sum_{t_3,t_4=\max(j,l)+1}^{T} M_{t_1,t_3,j} M_{t_1,t_4,l}
$$
$$
+ 2T_{jl}^{-2} \sum_{t_3,t_4=\max(j,l)+1}^{T} M_{t_1,t_2,j} M_{t_3,t_4,l} + 2T_{jl}^{-2} \sum_{t_3,t_4=\max(j,l)+1}^{T} M_{t_1,t_3,j} M_{t_2,t_4,l}
$$
$$
- 4T_{jl}^{-3} \sum_{t_3,t_4,t_5=\max(j,l)+1}^{T} M_{t_1,t_3,j} M_{t_4,t_5,l}
$$
$$
+ T_{jl}^{-4} \sum_{t_3,t_4,t_5,t_6=\max(j,l)+1}^{T} M_{t_3,t_4,j} M_{t_5,t_6,l} \bigg) \bigg],
$$

$M_{t_1,t_2,j} = \exp(-0.5\|\hat{\mathbf{u}}_{t_1-j} - \hat{\mathbf{u}}_{t_2-j}\|^2)$, $T_j = T - j$ and $T_{jl} = T - \max(j,l)$.

A.2 The test statistic for serial correlation

The Wold innovations are uncorrelated by definition. In applications, however, Wold innovations are estimated, and to avoid dynamic misspecification, it is important to

check whether estimated Wold innovations are consistently estimating a white noise process. To this end, we propose a new correlation test statistic $\hat{Q}^{1,1}(h)$, which accounts for the estimation uncertainty of model parameters: it is an extension of the classical Box–Pierce (1970) test to multivariate VARMA models; see Escanciano, Lobato, and Zhu (2013) and references therein for the related literature. Like $\hat{Q}^1$, and in contrast to classical portmanteau tests, the correlation test statistic $\hat{Q}^{1,1}(h)$ accounts for an increasing number of lags and is asymptotically pivotal. The test statistic for serial correlation is given by

$$\hat{Q}^{1,1}(h) = \left[\sum_{j=1}^{T-1} k^2(j/h)(T-j)\hat{R}_j^2 - \hat{C}^{1,1}(h)\right]\bigg/\sqrt{\hat{D}^{1,1}(h)},$$

where $\hat{R}_j = (T-j)^{-1}\sum_{t=j+1}^{T-1}(\hat{\mathbf{u}}_t - \bar{\mathbf{u}})'(\hat{\mathbf{u}}_{t-j} - \bar{\mathbf{u}})$,

$$\hat{C}^{1,1}(h) = \sum_{j=1}^{T-1} k^2(j/h)(T-j)^{-1}\sum_{m=1}^{d}\sum_{t=j+1}^{T} \hat{u}_{t,m}^2\hat{u}_{t-j,m}^2,$$

$$\hat{D}^{1,1}(h) = 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/h)k^2(l/h)$$

$$\times \sum_{m=1}^{d}\sum_{q=1}^{d}\left[\frac{1}{T-\max(j,l)}\sum_{t=\max(j,l)+1}^{T}\hat{u}_{t,m}\hat{u}_{t,q}\hat{u}_{t-j,m}\hat{u}_{t-l,q}\right]^2,$$

and $\bar{\mathbf{u}} = T^{-1}\sum_{t=1}^{T}\hat{\mathbf{u}}_t$. We use this test in the empirical application to check for potential dynamic (linear) misspecification of the VAR fits.

The asymptotic properties of $\hat{Q}^{1,1}(h)$ are established below.

THEOREM A.1. *Suppose Assumptions* A.2–A.6(i) *hold and* $h = cT^\eta$ *for* $0 < \eta < 1$, *where* $0 < c < \infty$. *Then as* $T \to \infty$, $\hat{Q}^{1,1}(h) \xrightarrow{d} N(0,1)$ *if* $\{\mathbf{u}_t\}$ *is a mds.*

THEOREM A.2. *Suppose Assumptions* A.1–A.6 *hold and* $h = cT^\eta$ *for* $0 < \eta < 1$, *where* $0 < c < \infty$. *Then as* $T \to \infty$,

$$\frac{h^{\frac{1}{2}}}{T}\hat{Q}^{1,1}(h) \xrightarrow{p} \frac{1}{\sqrt{D^{1,1}}}\sum_{j=1}^{\infty}\left[E(\mathbf{u}_t'\mathbf{u}_{t-j})\right]^2,$$

where

$$D^{1,1} = 2\int_0^\infty k^4(z)\,dz\sum_{m=1}^{d}\sum_{q=1}^{d}\left[E(u_{t,m}u_{t,q})\right]^2\sum_{j=-\infty}^{\infty}\left[E(u_{t,m}u_{t-j,q})\right]^2.$$

A.3 *Practical implementation of our test*

This section provides practical recommendations for the implementation of our test based on the Monte Carlo simulations and the empirical applications. These are the recommended steps.[18]

Step 1. Use the Schwarz information criterion (SIC) to select the lag order p of a VAR(p) fit, $\widehat{p}_{SIC}$ say.

Step 2. Compute Wold residuals by least squares in the VAR($\widehat{p}_{SIC}$).

Step 3. Check for dynamic linear misspecification of Wold residuals with the white noise test of Section A.2. If rejected, change p accordingly and repeat Step 2.

Step 4. Check for Gaussianity with the normality test proposed by Lütkepohl (1991, pp. 155–158), and check for heteroskedasticity using the multivariate extension of White's (1980) test developed by Kelejian (1982) and Doornik (1996).

Step 5. Compute our test using the closed-form solution in Section A.1 with a data-driven plug-in-bandwidth, as in Hong and Lee (2005), and with a preliminary bandwidth $\overline{h} = 10$.

References

Alessi, L., M. Barigozzi, and M. Capasso (2011), "Non-fundamentalness in structural econometric models: A review." *International Statistical Review*, 79, 16–47. [151]

Bierens, H. J. (1982), "Consistent model specification tests." *Journal of Econometrics*, 20, 105–134. [161]

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013), "News, noise, and fluctuations: An empirical exploration." *American Economic Review*, 103, 3045–3070. [150, 151]

Blanchard, O. J. and R. Perotti (2002), "An empirical characterization of the dynamic effects of changes in government spending and taxes on output." *The Quarterly Journal of Economics*, 117, 1329–1368. [149, 152, 170, 171]

Blanchard, O. J. and D. Quah (1989), "The dynamic effects of aggregate demand and supply disturbances." *American Economic Review*, 79, 655–673. [150]

Blanchard, O. J. and D. Quah (1993), "The dynamic effects of aggregate demand and supply disturbances: Reply." *American Economic Review*, 83, 653–658. [150]

Boubacar Mainassara, Y. and C. Francq (2011), "Estimating structural VARMA models with uncorrelated but non-independent error terms." *Journal of Multivariate Analysis*, 102, 496–505. [158, 160, 161]

Box, G. E. P. and D. A. Pierce (1970), "Distribution of residual autocorrelations in autoregressive integrated moving average time series models." *Journal of the American Statistical Association*, 65, 1509–1526. [175]

---

[18]A GAUSS code to implement our test is available from the authors upon request.

Breidt, F. J. and R. A. Davis (1992), "Time-reversibility, identifiability and independence of innovations for stationary time series." *Journal of Time Series Analysis*, 13, 377–390. [154]

Brillinger, D. R. and M. Rosenblatt (1967a), "Asymptotic theory of estimates of k-th order spectra." In *Spectral Analysis of Time Series* (B. Harris, ed.). Wiley, New York. [157]

Brillinger, D. R. and M. Rosenblatt (1967b), "Computation and interpretation of the k-th order spectra." In *Spectral Analysis of Time Series* (B. Harris, ed.). Wiley, New York. [157]

Brockwell, P. J. and R. A. Davis (1991), *Time Series: Theory and Methods*. Springer-Verlag, New York. [154]

Burnside, C., M. Eichenbaum, and J. D. M. Fisher (2004), "Fiscal shocks and their consequences." *Journal of Economic Theory*, 115, 89–117. [169]

Chan, K.-S., L.-H. Ho, and H. Tong (2006), "A note on time-reversibility of multivariate linear processes." *Biometrika*, 93, 221–227. [152]

Chen, B. and Y. Hong (2011), "Generalized spectral testing for multivariate continuous-time models." *Journal of Econometrics*, 164, 268–293. [161]

Cheng, Q. (1992), "On the unique representation of non-Gaussian linear processes." *The Annals of Statistics*, 20, 1143–1145. [152]

Cúrdia, V., M. Negro, and D. L. Greenwald (2014), "Rare shocks, great recessions." *Journal of Applied Econometrics*, 29, 1031–1052. [152]

Doornik, J. A. (1996), "Testing vector error autocorrelation and heteroscedasticity." Nuffield College, unpublished manuscript. [170, 176]

Durbin, J. (1973), "Weak convergence of the sample distribution function when parameters are estimated?" *The Annals of Statistics*, 1, 274–290. [157]

Edelberg, W., M. Eichenbaum, and J. D. M. Fisher (1999), "Understanding the effects of a shock to government purchases." *Review of Economic Dynamics*, 2, 166–206. [169]

Escanciano, J. C. (2006), "Goodness-of-fit tests for linear and nonlinear time series models." *Journal of the American Statistical Association*, 101, 531–541. [157]

Escanciano, J. C., I. N. Lobato, and L. Zhu (2013), "Automatic specification testing for vector autoregressions and multivariate nonlinear time series models." *Journal of Business & Economic Statistics*, 31, 426–437. [175]

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007), "ABCs (and Ds) of understanding VARs." *American Economic Review*, 97, 1021–1026. [150, 173]

Fisher, J. D. M. and R. Peters (2010), "Using stock returns to identify government spending shocks." *The Economic Journal*, 120, 414–436. [169]

Forni, M. and L. Gambetti (2014), "Sufficient information in structural VARs." *Journal of Monetary Economics*, 66, 124–136. [151, 163]

Forni, M., L. Gambetti, M. Lippi, and L. Sala (forthcoming), "Noise bubbles." *The Economic Journal.* [151]

Forni, M., L. Gambetti, and L. Sala (2014), "No news in business cycles." *The Economic Journal*, 124, 1168–1191. [150]

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009), "Opening the black box: Structural factor models with large cross sections." *Econometric Theory*, 25, 1319–1347. [173]

Francq, C. and J. M. Zakoïan (1998), "Estimating linear representations of nonlinear processes." *Journal of Statistical Planning and Inference*, 68, 145–165. [161]

Francq, C., R. Roy, and J. M. Zakoïan (2005), "Diagnostic checking in ARMA models with uncorrelated errors." *Journal of the American Statistical Association*, 100, 532–544. [155]

Galí, J., J. D. López-Salido, and J. Vallés (2007), "Understanding the effects of government spending on consumption." *Journal of the European Economic Association*, 5, 227–270. [169]

Geweke, J. F. (1993), "Bayesian treatment of the independent student-t linear model." *Journal of Applied Econometrics*, 8, S19–S40. [152]

Geweke, J. F. (1994), "Priors for macroeconomic time series and their application." *Econometric Theory*, 10, 609–632. [152]

Gospodinov, N. and S. Ng (2015), "Minimum distance estimation of possibly non-invertible moving average models." *Journal of Business & Economic Statistics*, 33, 403–417. [152]

Gourieroux, C. and A. Monfort (2014), "Revisiting identification and estimation in structural VARMA models." CREST Working paper No. 2014-30. [152]

Hamilton, J. D. and G. Lin (1996), "Stock market volatility and the business cycle." *Journal of Applied Econometrics*, 11, 573–593. [158]

Hansen, L. P. and T. J. Sargent (1980), "Formulating and estimating dynamic linear rational expectations models." *Journal of Economic Dynamics and Control*, 2, 7–46. [150, 155, 173]

Hansen, L. P. and T. J. Sargent (1991), "Two difficulties in interpreting vector autoregressions." In *Rational Expectations Econometrics* (L. P. Hansen et al., eds.), 77–120, Westview Press, Boulder. [150, 155]

Hong, Y. (1999), "Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach." *Journal of the American Statistical Association*, 94, 1201–1220. [153, 157, 173]

Hong, Y. and Y.-J. Lee (2005), "Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form." *Review of Economic Studies*, 72, 499–541. [153, 157, 164, 176]

Huang, J. and Y. Pawitan (2000), "Quasi-likelihood estimation of non-invertible moving average processes." *Scandinavian Journal of Statistics*, 27, 689–702. [150]

Kelejian, H. H. (1982), "An extension of a standard test for heteroskedasticity to a systems framework." *Journal of Econometrics*, 20, 325–333. [170, 176]

Kilian, L. (2001), "Impulse response analysis in vector autoregressions with unknown lag order." *Journal of Forecasting*, 20, 161–179. [164]

Lanne, M., M. Meitz, and P. Saikkonen (2013), "Testing for linear and nonlinear predictability of stock returns." *Journal of Financial Econometrics*, 11, 682–705. [152]

Leeper, E. M. (1989), "Policy rules, information, and fiscal effects in a 'Ricardian' model." Board of Governors of the Federal Reserve System, International Finance Discussion Papers 360. [150]

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013), "Fiscal foresight and information flows." *Econometrica*, 81, 1115–1145. [151, 152, 153, 154, 155, 162, 163, 166, 169, 170, 173]

Lippi, M. and L. Reichlin (1993), "The dynamic effects of aggregate demand and supply disturbances: Comment." *American Economic Review*, 83, 644–652. [150, 155]

Lippi, M. and L. Reichlin (1994), "VAR analysis, non-fundamental representations, Blaschke matrices." *Journal of Econometrics*, 63, 307–325. [150, 155]

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York. [170, 176]

Mertens, K. and M. O. Ravn (2010), "Measuring the impact of fiscal policy in the face of anticipation: A structural VAR approach." *The Economic Journal*, 120, 393–413. [169]

Mountford, A. and H. Uhlig (2009), "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, 24, 960–992. [169]

Perotti, R. (2005), "Estimating the effects of fiscal policy in OECD countries." Proceedings, Federal Reserve Bank of San Francisco. [169]

Perotti, R. (2008), "In search of the transmission mechanism of fiscal policy." In *NBER Macroeconomics Annual 2007*, Vol. 22, 169–226, University of Chicago Press, Chicago. [169]

Pham, D. T. and L. T. Tran (1985), "Some mixing properties of time series models." *Stochastic Processes and their Applications*, 19, 297–303. [161]

Priestley, M. B. (1981), *Spectral Analysis and Time Series*. Academic Press, London. [158]

Ramey, V. A. (2011), "Identifying government spending shocks: It's all in the timing." *The Quarterly Journal of Economics*, 126, 1–50. [149, 152, 153, 168, 169, 171, 172, 173]

Ramey, V. A. and M. D. Shapiro (1998), "Costly capital reallocation and the effects of government spending." *Carnegie-Rochester Conference Series on Public Policy*, 48, 145–194. [169, 171, 172]

Ramsey, J. B. and A. Montenegro (1992), "Identification and estimation of non-invertible non-Gaussian MA(q) processes." *Journal of Econometrics*, 54, 301–320. [150, 152]

Rondina, G. (2008), "Incomplete information and informative pricing." University of California, San Diego, unpublished manuscript. [150]

Rosenblatt, M. (2000), *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer-Verlag, New York. [152, 154, 155, 156]

Rozanov, Y. A. (1967), *Stationary Random Processes*. Holden Day, San Francisco. [154, 155]

Sahneh, M. H. (2015), "Are the shocks obtained from SVAR fundamental?" Working paper, Universidad Carlos III de Madrid. [152]

Schmitt-Grohe, S. and M. Uribe (2008), "What's news in business cycles." NBER Working Papers 14215. [152]

Schmitt-Grohe, S. and M. Uribe (2012), "What's news in business cycles." *Econometrica*, 80, 2733–2764. [152]

Sims, E. R. (2012), "News, non-invertibility, and structural VARs." *Advances in Econometrics*, 28, 81–136. [166]

Stinchcombe, M. B. and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative." *Econometric Theory*, 14, 295–325. [161]

Tsay, R. S. (1993), "Testing for non-invertible models with applications." *Journal of Business & Economic Statistics*, 11, 225–233. [153]

Watson, M. W. (1986), "Vector autoregressions and cointegration." In *Handbook of Econometrics*, Vol. IV (R. F. Engle and D. McFadden, eds.), 2843–2915, Elsevier, Amsterdam. [153]

White, H. (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, 48, 817–838. [170, 176]

Yang, S.-C. S. (2005), "Quantifying tax effects under policy foresight." *Journal of Monetary Economics*, 52, 1557–1568. [150]