Are the shocks obtained from SVAR fundamental?

Hamidi Sahneh, Mehdi

Universidad Carlos III de Madrid

15 June 2015

Are the shocks obtained from SVAR fundamental?

Mehdi Hamidi Sahneh *

Abstract

This paper provides new conditions under which the shocks recovered from the estimates of structural vector autoregressions are fundamental. I prove that the Wold innovations are unpredictable if and only if the model is fundamental. I propose a test based on a generalized spectral density to check the unpredictability of the Wold innovations. The test is applied to study the dynamic effects of government spending on economic activity. I find that standard SVAR models commonly employed in the literature are non-fundamental. Moreover, I formally show that introduction of a narrative variable that measures anticipation restores fundamentalness.

Keywords: Fundamentalness; Identification; Invertible Moving Average; Vector Autoregressive.

JEL classification: C5, C32, E62.

*Departamento de Economia, Universidad Carlos III de Madrid, Getafe, 28903, Spain. Email address: *mhamidis@eco.uc3m.es*.

1 Introduction

Since Sims's (1980) seminal paper, Structural Vector Autoregressive (SVAR) models have been used extensively for economic analysis. The underlying assumption of SVAR, known as fundamentalness, is that one is able to recover the structural shocks driving the process from linear combinations of observed present and past values of the process. Non-fundamentalness arises when observed variables do not contain enough information to recover the structural shocks and the impulse response functions. Once the representation is non-fundamental, all identification schemes, such as long-run or sign restrictions, fail to recover the true structural shocks. In this paper, I propose a test to empirically detect whether the shocks recovered from the estimation of a VAR are truly fundamental.

Although many economic models generate non-fundamental representations, little is known how to test if a model is non-fundamental. Permanent income models (Fernández-Villaverde et al., 2007), news shocks (Blanchard et al., 2013; Forni et al., 2014), and fiscal foresight (Leeper et al., 2013) are some examples that can generate equilibrium solutions with non-fundamental representation. For a comprehensive survey of this literature see Alessi et al. (2011).

The key contribution of this paper is to provide new conditions under which the shocks obtained from the estimates of the SVAR are truly fundamental. I prove that the Wold innovations from fitting a VAR to a non-fundamental model are martingale difference and therefore unpredictable (in the mean), even if one includes an infinite past of the observable variables. Consequently, to test whether the model is fundamental, one must check if the Wold innovations are unpredictable.

There are some proposals to test for the unpredictability of the Wold innovations (see Hong (1999), Domínguez and Lobato (2003), Hong and Lee (2005), Escanciano and Velasco (2006), among others). To the best of my knowledge, none of these tests are applicable to the multivariate setting of this paper. Alternatively, it is possible to apply a sequence of univariate test to each series. However, using a multivariate procedure will avoid the multiple testing problem and is more powerful, since it is possible that a single series is unpredictable, but the collection of several series is predictable. To test for the unpredictability of the Wold innovations, I extend Hong and Lee's (2005) test from univariate to multivariate setting. I show that the proposed test statistic has a convenient asymptotic standard normal distribution and diverges to infinity under the alternative hypothesis. The proposed test is simple to apply since it only needs reduced form VAR residuals as input. Therefore, my proposed test does not require any identification assumption or estimating non-fundamental models. Simulations show that the test has good size control and has power against general alternatives.

This paper is related to the literature that attempts to test if a Vector Moving Average (VMA) model is fundamental. Giannone and Reichlin (2006) prove that if a model is fundamental, then extra information should not Granger cause the variables included in the model. Similarly, Forni and Gambetti (2014) exploit the factors of a large system to propose necessary and sufficient conditions under which a VAR contains sufficient information to estimate the structural shocks, which under some assumptions could be applied to detect fundamentalness. However, these procedures are based on the untestable assumption that the extra information -such as sectoral data or factors of a large data set- that one uses to test for fundamentalness is itself fundamental.

From a methodological point of view, my proposal is similar to the proposal of Chen et al. (2012). By converting testing for fundamentalness to testing for serial independence of the Wold innovations, these authors proposed a test for fundamental VMA representation. However, their test critically depends on the *iid* assumption of the true unobserved errors, which is often rejected in macroeconomic and financial time series. Failure to accommodate these features will lead to rejection of the null of fundamentalness by mistake. In contrast, my proposal is robust to the failure of the *iid* assumption.

To illustrate the application of the proposed test, I focus on the dynamic effects of government spending shocks on economic activity in the United States in the post-war period. I find that the baseline VAR models normally considered in the empirical literature to identify these effects are non-fundamental, and therefore, the impulse responses and variance decompositions from SVAR approach appears not to be reliable. In case of rejection of the null of fundamentalness, it has been conjectured that expanding the econometrician's information set might solve the non-fundamentalness problem.[1] The proposed test of this paper can be used to formally test if adding more information solves the non-fundamentalness problem. Specifically, I show that augmenting the baseline VAR model with a narrative variable that measure *news* about future government spending restores fundamentalness. Consequently, an econometrician can proceed with the identification strategy that she finds reasonable to recover the structural shocks.

The rest of the paper is organized as follows: Section 2 provides a formal statement of the fundamental representation and the testing problem. Section 3 introduces formally the test statistic based on the generalized spectrum. Section 4 examines the finite-sample performance of the test through some Monte Carlo simulation based on a DSGE model and an empirical application to the identification of government spending shocks. Section 5 concludes. The MATLAB code for implementing the test is available from the author upon request.

[1] See for example, Giannone and Reichlin (2006) and Forni and Gambetti (2014).

2 Characterization of non-fundamental VARMA representations

Let $\{x_t\}$ be a d-dimensional stationary solution of a VARMA(p,q) model satisfying the difference equation:

$$\Phi(L)x_t = \Theta(L)\xi_t \ , \qquad t = 0, \pm 1, \pm 2, \cdots \tag{2.1}$$

where $\{\xi_t\}$ is an unpredictable process (also known as martingale difference)[2] with covariance matrix Σ_ξ and

$$\Phi(L) := I_d - \Phi_1 L - \cdots - \Phi_p L^p$$
$$\Theta(L) := I_d + \Theta_1 L + \cdots + \Theta_q L^q$$

are the AR and MA polynomials, respectively. Henceforth, $\mathbf{I}_d$ is the $d \times d$ identity matrix, $\Phi_p \neq 0$ and $\Theta_q \neq 0$ and L is the lag operator, i.e., $Lx_t = x_{t-1}$. The polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$ have no common roots, neither of the roots is on the unit circle, nor equal to zero.

To begin, lets define fundamentalness, also known as invertibility.[3]

Definition 2.1: An uncorrelated process $\{\xi_t\}$ is x_t-fundamental if $\mathcal{H}_t^\xi = \mathcal{H}_t^x$ for all $t \in \mathbb{Z}$, where $\mathcal{H}_t^\xi$ is the closed linear span of $\{\xi_s : s \leq t\}$. The process $\{\xi_t\}$ is non-fundamental if $\mathcal{H}_t^\xi \in \mathcal{H}_t^x$ and $\mathcal{H}_t^\xi \neq \mathcal{H}_t^x$, for at least one $t \in \mathbb{Z}$.

A VARMA process defined by (2.1) is said to be fundamental if and only if all the roots of $\Theta(z)$ lie outside the unit circle in the complex plane. Similarly (2.1) is

[2]A real-valued stationary time series $\{Y_t\}_{t=-\infty}^{\infty}$ is a martingale difference (MD) process if $E[Y_t|Y_{t-1}, Y_{t-2}, \cdots] = 0$. A MD process is unpredictable in the mean.

[3]Fundamentalness is slightly different from invertibility, since invertibility requires that no roots of the MA component be on or inside the unit circle. In this framework, they are equivalent since unit root in the MA polynomial is ruled out.

said to be causal if and only if all the roots of $\Phi(z)$ lie outside the unit circle in the complex plane.[4] Throughout, I assume that the model is causal.

One can show that if non-fundamental representation is excluded by mistake, the true unobserved shocks will be related to the Wold innovations through *Blaschke* matrices.[5] The following example illustrates the main ideas.

Example 2.1: Leeper et al. (2013) introduce foresight into a simple growth model. Assuming two-quarter fiscal foresight, the log-linearized equilibrium condition for capital is

$$(1-\alpha L)k_t = -\kappa(L+\theta)\xi_{\tau,t} \tag{2.2}$$

where κ is a functions of the deep parameters of the model and $0 < \alpha < 1$ and $0 < \theta < 1$. However, fundamentalness is satisfied only if $|\theta| > 1$. The fact that more recent tax news are discounted heavier than older news makes model (2.2) non-fundamental. Imposing fundamentalness, the less informed econometrician incorrectly estimates the model

$$(1-\alpha L)k_t = -\kappa(1+\theta L)\epsilon_{\tau,t} \qquad |\theta| < 1$$

or in the autoregressive form

$$\frac{(1-\alpha L)}{-\kappa(1+\theta L)}k_t = \sum_{j=0}^{\infty}\gamma_j k_{t-j} = \epsilon_{\tau,t} \qquad |\theta| < 1$$

where γ_j is a function of deep parameters and $\epsilon_{\tau,t}$ is the Wold innovation[6], related to the true unobserved errors through *Blaschke* factor, $\epsilon_{\tau,t} = \left[\frac{L+\theta}{1+\theta L}\right]\xi_{\tau,t}$.□

In practice, it is common to estimate a VAR instead of a VARMA, which makes

[4]See Brockwell and Davis (1991), Theorems 3.1.1 and 3.1.2.

[5]*Blaschke* matrices are complex-valued filters which take the roots from inside to outside the unit disc, thus generates a fundamental representation from a non-fundamental one (Lippi and Reichlin, 1994).

[6]i.e., $\epsilon_t = k_t - \mathbb{L}[k_t|\mathcal{H}_t^k]$ where, $\mathbb{L}[k_t|\mathcal{H}_t^k]$ denotes the optimal linear predictor of k_t given its past.

detecting non-fundamentalness more complicated since the DGP has undergone a further approximation. To see this, suppose the true process is a non-fundamental ARMA process (2.1), but an econometrician incorrectly imposes fundamentalness assumption. One can show that the resulting process has a representation given by

$$\Phi(L)x_t = \tilde{\Theta}(L)\epsilon_t \tag{2.3}$$

where $\{\epsilon_t\}$ are the Wold innovations related to the original innovations, $\{\xi_t\}$, through filter

$$\epsilon_t = \tilde{\Theta}^{-1}(L)\Theta(L)\xi_t \tag{2.4}$$

and $\tilde{\Theta}(L)$ has the same order as $\Theta(L)$ but all its roots are outside the unit circle. Therefore, (2.3) can be written as a VAR(∞) form:

$$\tilde{\Theta}(L)^{-1}\Phi(L)x_t = \sum_{j=0}^{\infty} \gamma_j x_{t-j} = \epsilon_t \tag{2.5}$$

For estimation of such models it is necessary to approximate the infinite order lag structure by finite order VAR(p). In practice, the order p is often selected so that the residuals are white noise. One can prove that if fundamentalness is imposed incorrectly, the Wold innovations (2.4) are still uncorrelated. Therefore, estimation methods based on second-order moment techniques do not identify non-fundamentalness. In order to deal with this identification problem the literature imposes fundamentalness by assumption.

In the non-Gaussian case, however, fundamental and non-fundamental models are distinguishable based on higher order cumulants (Lii and Rosenblatt, 1982). Using time-reversibility argument, Breidt and Davis (1992) proved that the Wold innovations from fitting an invertible ARMA model to a non-invertible one are *iid*, if and only if the error is non-Gaussian. Chen et al. (2012) extended this result to the multivariate case and proposed to test for serial dependence to detect non-fundamentalness. However, testing for serial dependence of the Wold innovations is a restrictive and may lead to rejection of the null of fundamentalness by mistake. The following is an example intended to highlight this point.

Example 2.2: Consider the ARCH process

$$x_t = \xi_t$$
$$\xi_t = h_t^{1/2} z_t \quad h_t = 0.43 + 0.57 z_{t-1}^2$$
$$z_t \sim iid\ N(0,1)$$

Definition 2.1 trivially holds and therefore ξ_t is x_t-fundamental. However, ξ_t is an ARCH process and therefore serial dependence test can incorrectly reject the null of fundamentalness.□

In this paper, I use the information available in the *Blaschke* matrix to propose a new test which is robust to the failure of the *iid* assumption. Under some mild conditions stated in Assumption 1, I prove that if the model is non-fundamental, the Wold innovations are non-MD, i.e., non-linearly predictable despite being white noise.

Assumption 1. Let ξ_{jt} denote the jth element of the true unobserved shocks $\{\xi_t\}$. There exists a $j \in 1, \cdots, d$ such that ξ_{jt} is (a) independent, and (b) continuously distributed with a non-Gaussian distribution such that $(a+1)th$ moment finite for some $a \geq 2$ and $\mathrm{Var}(\xi_{jt}) > 0$.

Proposition 2.1: Let Assumption 1 hold. The non-Gaussian VARMA model (2.1) is invertible if and only if the Wold innovations $\{\epsilon_t\}$ are MD.

For the proof see Appendix A. Assumption 1.(a) is commonly used in the empirical studies. It can be further relaxed to allow for the true unobserved shocks to be dependent.[7] Moreover, independence is a more restrictive assumption than MD.

[7] The proof holds under sub-independence assumption. Two random variables are said to

Therefore, Proposition 2.1 states that *even if* the true unobserved errors are *independent*, the Wold innovations from fitting an invertible model to a non-invertible one are *non-MD*. Intuitively, by introducing *some dependence* structure on the true shocks (for example, a GARCH process), one still expects the Wold innovations from fitting the *wrong* model to stay non-MD.

Non-Gausianity is needed to achieve identification. In fact, there are many studies that emphasize considering non-Gaussian distributions and other higher order time-varying moments (see e.g., Harvey and Siddique, 1999, 2000; Jondeau and Rockinger, 2003). Note that, no specific distributional assumption is needed. The continuity assumption is also mild and could be dropped in the univariate case or if there is only one root of the $\det \Theta(L)$ that is inside the unit circle. This is stated in the following corollary.

Corollary 2.1: If there is only one root of the determinant of the MA polynomial inside the unit circle, then the continuity assumption is not needed for the Proposition 2.1 to hold.

3 Testing for non-fundamental representations

Under the null of fundamentalness $\xi_t(\theta_0) = \epsilon_t(\theta_0)$, which following Proposition 2.1 can be restated as

$$\mathbb{H}_0 : \epsilon_t(\theta_0) \text{ are MD (unpredictable) for some } \theta_0 \in \Xi \tag{3.1}$$

where $\theta_0 = vec\{\Phi_1, \cdots, \Phi_p, \Theta_1, \cdots, \Theta_q, \Sigma_\epsilon\}$, and $vec(.)$ denote an operator on a matrix which cascades the columns of the matrix from the left to the right and

be sub-independent if the characteristic function of their sum is equal to the product of their marginal characteristic functions, i.e., $\phi_{x+y}(t) = \phi_x(t)\phi_y(t)$. This is a generalization of the concept of independence of random variables, i.e., if two random variables are independent then they are sub-independent, but not conversely, see Hamedani (2013). Unfortunately, the connection between sub-independence and MD is not clear in the literature, and I do not attempt to justify it here.

forms a column vector.

Testing (3.1) is not an easy task. Portmanteau test proposed by Box and Pierce (1970) and Ljung and Box (1978) are not suitable to reflect the non-linear dependence structure. Moreover, $\{\epsilon_t\}$ is unobserved and residuals depend on a $\sqrt{T}$-consistent estimator for θ_0, which may cause the loss of the nuisance parameter-free property of the asymptotic distribution of the test statistics.

To overcome these problems and checking for unpredictability at all lags in the sample, I extend the generalized spectral test of Hong and Lee (2005) to the multivariate setting. Compared with the existing tests in the literature, this test has some advantages: first, with the frequency domain approach, one can allow infinite number of lags as the sample size increases; second, the test has a standard normal limiting distribution and parameter estimation uncertainty has no impact on the asymptotic distribution of the test statistics. The proposed test can also be used to test the martingale hypothesis in the multivariate setting for observed raw data without any modification.

My proposal for testing the MD property of the Wold innovations is based upon the generalized spectrum of Hong (1999):

$$f(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}, \tag{3.2}$$

where $\omega \in [-\pi, \pi]$ is the frequency, $i \equiv \sqrt{-1}$, $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$, and

$$\sigma_j(u, v) = \text{cov}(e^{iu'\epsilon_t}, e^{iv'\epsilon_{t-|j|}}), \quad j = 0, \pm 1, ...$$

where $\epsilon_t \equiv \epsilon_t(\theta)$. Note that $f(\omega, u, v)$ is a complex-valued scalar function, although ϵ_t is a $d \times 1$ vector. The function $f(\omega, u, v)$ captures any type of pairwise serial dependence in $\{\epsilon_t\}$, including that with zero autocorrelation function.

The generalized spectrum $f(\omega, u, v)$ is not suitable for testing (3.1), because it also captures the serial dependence in higher order moments. For example, $f(\omega, u, v)$

captures GARCH dependence, although the process could be a MD. However, just as the characteristic function can be differentiated to generate various moments of ϵ_t, $f(\omega,u,v)$ can be differentiated to capture the serial dependence in various moments. To capture (and only capture) the serial dependence in the conditional mean, one can use

$$f^{(0,1,0)}(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j^{(1,0)}(0,v)e^{-ij\omega}, \quad \omega\in[-\pi,\pi]$$

where

$$\sigma_j^{(1,0)}(0,v) \equiv \frac{\partial}{\partial u}\sigma_j(u,v)\big|_{u=0} = \mathrm{cov}(i\epsilon_t, e^{iv'\epsilon_{t-|j|}})$$

is a $d\times 1$ vector. The measure $\sigma_j^{(1,0)}(0,v)$ checks whether the autoregression function $E(\epsilon_t|\epsilon_{t-j})=0$ at lag j is zero.[8]

In the present context, ϵ_t is not observed. Suppose we have T observations $\{x_t\}_{t=1}^T$ which is used to estimate the model and to obtain the estimated model residual

$$\hat{\epsilon}_t \equiv \hat{\Theta}^{-1}(L)\hat{\Phi}(L)x_t \tag{3.3}$$

where $\hat{\theta}$ is a $\sqrt{T}$-consistent estimator for θ_0. Examples of $\hat{\theta}$ are conditional least squares and quasi-maximum likelihood estimator. We can estimate $f^{(0,1,0)}(\omega,0,v)$ by a smoothed kernel estimator

$$\hat{f}^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=T-1}^{T-1}(1-\frac{|j|}{T})^{1/2}k(j/h)\hat{\sigma}_j^{(1,0)}(0,v)e^{-ij\omega}, \quad \omega\in[-\pi,\pi] \tag{3.4}$$

where $\hat{\sigma}_j^{(1,0)}(0,v) = \frac{\partial}{\partial u}\hat{\sigma}_j(u,v)\big|_{u=0}$, $\hat{\sigma}_j(u,v) = \hat{\varphi}_j(u,v) - \hat{\varphi}_j(u,0)\hat{\varphi}_j(0,v)$, and

$$\hat{\varphi}_j(u,v) = \frac{1}{T-|j|}\sum_{t=j+1}^{T} e^{iu'\hat{\epsilon}_t + iv'\hat{\epsilon}_{t-|j|}}$$

[8] The hypothesis of $E(\epsilon_t|I_{t-j}^{\epsilon})=0$ *a.s.* is not the same as the hypothesis of $E(\epsilon_t|\epsilon_{t-j})=0$ *a.s.* for all $j>0$. The former checks all type of dependencies, whereas the latter one only captures pairwise dependencies. See Hong (1999) for more discussion on this.

where $h \equiv h(T)$ is a bandwidth, and $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen and Quadratic spectral kernels. The factor $(1-\frac{|j|}{T})^{1/2}$ is a finite-sample correction. The effect of this correction factor is to put less weight on very large lags, for which we have less sample information. It could be replaced by unity.

Under $\mathbb{H}_0$, the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ becomes a flat spectrum:

$$f_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\sigma_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi]$$

which can be consistently estimated by

$$\hat{f}_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\hat{\sigma}_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi]$$

The estimators $\hat{f}^{(0,1,0)}(\omega, 0, v)$ and $\hat{f}_0^{(0,1,0)}(\omega, 0, v)$ converge to the same limit under $\mathbb{H}_0$, and generally converge to different limits under $\mathbb{H}_1$. Thus, any significant divergence between them can be interpret as evidence of the violation of the MD property, and hence, of the non-fundamentalness of the process.

The test statistic, that is robust to conditional heteroscedasticity and other time-varying higher order conditional moments of unknown form, is given as follows:

$$\hat{M} \equiv \left[\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left\|\hat{\sigma}_j^{(1,0)}(0, v)\right\|^2 d\mathcal{W}(v) - \hat{C}\right] \bigg/ \sqrt{\hat{D}} \tag{3.5}$$

where $T_j = T - j$, $\mathcal{W}(v) = \prod_{c=1}^{d} W(v_c)$, $W : \mathbb{R} \to \mathbb{R}^+$ is a nondecreasing function that weighs sets symmetric about zero equally, and the unspecified integrals are taken over the support of $\mathcal{W}(\cdot)$. Examples of $W(\cdot)$ include the CDF of any symmetric probability distribution, either discrete or continuous. $\hat{C}$ and $\hat{D}$ are estimate of the mean and the variance of $T \iint_{-\pi}^{\pi} \|\hat{f}^{(0,1,0)}(\omega, 0, v) - \hat{f}_0^{(0,1,0)}(\omega, 0, v)\|^2 d\omega d\mathcal{W}(v)$,

$$\hat{C} = \sum_{j=1}^{T-1} k^2(j/h) T_j^{-1} \sum_{t=j+1}^{T-1} \|\hat{\epsilon}_t\|^2 \int \left|\hat{\psi}_{t-j}(v)\right|^2 d\mathcal{W}(v)$$

$$\hat{D} = 2\hat{s}^4 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/h)k^2(l/h) \iint \left|\hat{\sigma}_{j-l}(u,v)\right|^2 d\mathcal{W}(u)d\mathcal{W}(v)$$

where $\hat{\psi}_t(v) = e^{iv'\hat{\epsilon}_t} - T^{-1}\sum_{t=1}^T e^{iv'\hat{\epsilon}_t}$, and $\hat{s}^4 = \sum_{a,b=1}^d \left(T^{-1}\sum_{t=1}^T \hat{\epsilon}_{at}\hat{\epsilon}_{bt}\right)^2$.

To derive the limit distribution of the test, I need to impose some regularity conditions. Throughout, I use C to denote a generic bounded constant, $\|.\|$ the Euclidean norm, and A^* the complex conjugate of A.

Assumption A1. $\{x_t\}$ is a $d \times 1$ strictly stationary time series process, and ϵ_t are MD with $\mathrm{E}\|\epsilon_t^4\| \le C$, where ϵ_t is Wold innovation from estimating an invertible model.

Assumption A2. For q sufficiently large, there exists a strictly stationary process $\{\epsilon_{q,t}\}$ measurable with respect to the sigma field generated by $\{\epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_{t-q}\}$ s.t. as $q \to \infty$, $\epsilon_{q,t}$ is independent of $\{\epsilon_{t-q-1}, \epsilon_{t-q-2}, \cdots\}$ for each t, $\mathrm{E}[\epsilon_{q,t}|I_{t-1}] = 0$ a.s., $\mathrm{E}\|\epsilon_t - \epsilon_{q,t}\|^2 \le Cq^{-\kappa}$ for some constant $\kappa \ge 1$, and $\mathrm{E}\|\epsilon_{q,t}\|^4 \le C$ for all large q.

Assumption A3. The estimator $\hat{\theta}$ is such that $\sqrt{T}(\hat{\theta} - \theta^*) = O_P(1)$, where $\theta^* \equiv plim_{T\to\infty}\hat{\theta}$. Under $\mathbb{H}_0$, $\theta^* = \theta_0$.

Assumption A4. Let $\bar{x}_0 = (x_0; \cdots; x_{1-p}; \epsilon_0; \cdots; \epsilon_{1-q})$ be some assumed initial values. Then $\mathrm{E}\|\bar{x}_0^2\| < \infty$.

Assumption A5. $k : \mathbb{R} \to [-1, 1]$ is symmetric about 0, and is continuous at 0 and all points except a finite number of points, with $k(0) = 1$ and $|k(z)| \le C|z|^{-b}$ as $z \to \infty$ for some $b > 1$.

Assumption A6. $W : \mathbb{R} \to \mathbb{R}^+$ is nondecreasing and weights sets symmetric about zero equally, with $\int \|v\|^4 dW(v) \le C$.

Assumption A7. Define $\psi_t(v) \equiv e^{iv\epsilon_t} - T^{-1}\sum_{t=1}^T e^{iv\epsilon_t}$ and $\Sigma \equiv \mathrm{E}(\epsilon_t\epsilon_t')$. Then, $\{\frac{\partial \epsilon_t}{\partial \theta}, \epsilon_t\}$ is a strictly stationary process such that

(a) $\sum_{j=1}^{\infty} \|\mathrm{cov}[\frac{\partial \epsilon_t}{\partial \theta}, \psi_{t-j}(v)]\| \leq C$;

(b) $\sum_{j=1}^{\infty} \sup_{(u,v)\in\mathbb{R}^2} |\sigma_j(u,v)| \leq C$;

(c) $\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \sup_{(u,v)\in\mathbb{R}^2} \left\|\mathrm{E}[(\epsilon_t \epsilon_t' - \Sigma)\psi_{t-j}(u)\psi_{t-l}(v)]\right\| \leq C$;

(d) $\sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \sup_{v\in\mathbb{R}} \|\kappa_{j,l,\tau}(v)\| \leq C$, where $\kappa_{j,l,\tau}(v)$ is the fourth order cumulant of the joint distribution of the process $\{\frac{\partial \epsilon_t}{\partial \theta}, \psi_{t-j}(v), \frac{\partial \epsilon_{t-l}}{\partial \theta}, \psi_{t-\tau}^*(v)\}$.

Assumption A8. $\sum_{j=1}^{\infty} \sup_{v\in\mathbb{R}} \|\sigma_j^{(1,0)}(0,v)\| \leq C$.

Assumption A1 is a regularity condition on the data generating process (DGP) $\{x_t\}$. Assumption A2 is required only under $\mathbb{H}_0$, which states that the MD $\{\epsilon_t\}$ can be approximated by a q-dependent MD process $\{\epsilon_t\}$ arbitrarily well when q is sufficiently large. Because $\{\epsilon_t\}$ is a MD, Assumption A2 essentially imposes restrictions on the serial dependence in higher order moments of $\{\epsilon_t\}$. It covers GARCH and stochastic volatility processes as special cases; see *e.g.* Hong and Lee (2005). Assumption A3 requires a $\sqrt{T}$-consistent estimator $\hat{\theta}$, such as conditional least squares estimator or a conditional quasi-maximum likelihood estimator.

Assumption A4 is a start-up value condition. It ensures that the impact of initial values assumed in the observed information set is asymptotically negligible. Assumption A5 is a regularity condition on the kernel $k(.)$. It includes all commonly used kernels in practice. 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. Assumption A6 is a condition on the weighting function $W(.)$ for the transform parameter v. It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment. Assumption A7 provides some covariance and fourth order cumulant conditions on $\{\frac{\partial \epsilon_{t-1}}{\partial \theta}, \epsilon_t\}$, which restricts the degree of serial dependence in $\{\frac{\partial \epsilon_{t-1}}{\partial \theta}, \epsilon_t\}$. Finally, Assumption A8 impose a condition on the serial dependence in $\{\epsilon_t\}$. The asymptotic properties of the test statistic is stated in the following theorem. The proof is similar to the univariate case of Hong and Lee (2005), and for the sake of space is not provided.

Proposition 4.1: Let $h = cT^{\lambda}$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ and $0 < c < \infty$. Then:

(a) Under Assumptions A1-A7 and $\mathbb{H}_0$, $\hat{M} \xrightarrow{d} N(0,1)$.

(b) Under Assumptions A1-A8 and $\mathbb{H}_1$, $\lim_{T\to\infty} P[\hat{M} > C(T)] = 1$ for any sequence $C(T) = o(T/h^{1/2})$.

Under the null, $\hat{M}$ has a simple standard normal distribution. Under the alternative hypothesis, $E(\epsilon_t|\epsilon_{t-j}) \neq 0$ a.s., at some lag $j > 0$. Then we have $\int \|\sigma_j^{(1,0)}(0,v)\|^2 d\mathcal{W}(v) > 0$ for any weighting function $\mathcal{W}(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. Therefore, $\hat{M}$ has asymptotic unit power at any given significance level.

An important feature of $\hat{M}$ is that the use of the estimated residuals $\{\hat{\epsilon}_t\}$ in place of the true errors $\{\epsilon_t\}$ has no impact on the limit distribution of $\hat{M}$. The reason is that the convergence rate of the parametric parameter estimator $\hat{\theta}$ to θ_0 is faster than that of the nonparametric kernel estimator $\hat{f}^{(0,1,0)}(w,0,v)$ to $f^{(0,1,0)}(w,0,v)$. Consequently, the limit distribution of $\hat{M}$ is solely determined by $\hat{f}^{(0,1,0)}(w,0,v)$, and replacing θ_0 by $\hat{\theta}$ has no impact asymptotically.

4 Monte Carlo evidence and empirical application

4.1 Simulation study

In this section I examine the finite sample performance of the proposed test based on artificial data generated from the DSGE model with fiscal foresight of Leeper et al.

(2013). The model is characterized by a representative household that maximizes expected log utility,

$$E_0 \sum_{t=0}^{\infty} \beta^t \log(C_t)$$

$$s.t. \qquad C_t + K_t + T_t \leq (1 - \tau_t) A_t K_{t-1}^{\alpha}$$

where C_t, K_t, Y_t, T_t, and τ_t denote time$-t$ consumption, capital, output, lump-sum taxes, and the income tax rate, respectively, and A_t is an exogenous technology shock. The parameters satisfy $0 < \alpha < 1$, $0 < \beta < 1$. The government sets the tax rate according to $T_t = \tau_t Y_t$, and labor is supplied inelastically. Let A and τ_k denote the steady states values of technology and the tax rate. The log-linearized equilibrium condition for the capital and the tax rate is given by the following bivariate VARMA model

$$\hat{\tau}_t = \Psi(L)\xi_{\tau,t}$$

$$k_t = \alpha k_{t-1} + \xi_{a,t} - \frac{\tau(1-\theta)}{1-\tau} \sum_{k=0}^{\infty} \theta^k E_t \hat{\tau}_{t+k+1}$$

where $\theta = \alpha\beta\frac{1-\tau_y}{1-\tau_k}$ and the lower case letters denote percentage deviations from steady state values, $k_t = \log(K_t) - \log(K)$, $a_t = \log(A_t) - \log(A)$, and $\hat{\tau}_t = \log(\tau_t) - \log(\tau)$.

To model foresight, I assume the tax rate evolves as

$$\hat{\tau}_t = \sum_{j=0}^{J} \psi_j \xi_{\tau,t-j} = \Psi(L)\xi_{\tau,t} \tag{4.1}$$

where $\sum_{j=0}^{J} \psi_j = 1$, and $\psi_j \in [0, 1]$ determines the relative weight of the shock at time j. I consider five different processes for the tax rate (Table 1), that embed many of the information flows that appear in theoretical studies of foresight (see, e.g., Leeper et al., 2013; Forni et al., 2014; Schmitt-Grohé and Uribe, 2012). DGP1 is an example of no foresight, and therefore the model is fundamental. DGP2 is an example of a fundamental model with two period foresight.[9] DGP3 is an example of a non-fundamental process with two period foresight, with weights reciprocal to the DGP2. DGP4 and DGP5 are examples of non-fundamental processes with roots zero, which are commonly used in the literature with news shocks. Although Proposition 2.1 rules out these kind of processes, it would be interesting to see how the test performs.

For the simulation exercise, I generate artificial series for the capital and the tax rate setting $\alpha = 0.36$, $\beta = 0.99$, and $\tau = 0.25$, as in Leeper et al. (2013). The structural shocks $\xi_{a,t}$ and $\xi_{\tau,t}$ are generated as centered *iid lognormal*$(0,1)$, mutually independent at all leads and lags.

Chen et al. (2012) consider the stronger null hypothesis that the errors are serially independent. However, testing for serial independence of the errors is a more restrictive condition than (3.1); in particular, one might reject a correct null model because of higher order dependence. Their proposed test statistic to check for serial dependence of the residuals is of the form

$$\hat{Q} \equiv \left[\sum_{j=1}^{T-1} k^2(j/h) T_j \iint \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W}(u) d\mathcal{W}(v) - \hat{C}_q\right] \Big/ \sqrt{\hat{D}_q}$$

where

$$\hat{C}_q = \sum_{j=1}^{T-1} k^2(j/h) \left[\int \hat{\sigma}_0(v,-v) d\mathcal{W}(v)\right]^2$$

$$\hat{D}_q = 2 \sum_{j=1}^{T-2} k^4(j/h) \left[\int |\hat{\sigma}_0(u,v)|^2 d\mathcal{W}(u) d\mathcal{W}(v)\right]^2$$

which also has an asymptotic standard normal null distribution. To examine why it is important to take into account the impact of higher order time-varying moments in testing $\mathbb{H}_0$, I also consider a GARCH process for $\xi_{a,t} = \sigma_t^{\frac{1}{2}} z_t, \quad \sigma_t^2 = 0.001 + 0.09\xi_{t-1}^2 + 0.9\sigma_{t-1}^2$ and $\xi_{\tau,t} \sim$ *iid lognorm*$(0,1)$. A similar GARCH process is used

[9]The roots of the determinant of the MA component are complex conjugate with modulus 2.82.

Table 1: Information Flow Processes

Process	Description	Coefficients
DGP1	No foresight	$\psi_0 = 1$
DGP2	2-qtr concentrated news	$\psi_0 = 0.8, \psi_1 = 0.1, \psi_2 = 0.1$
DGP3	2-qtr concentrated news	$\psi_0 = 0.1, \psi_1 = 0.1, \psi_2 = 0.8$
DGP4	2-qtr perfect foresight	$\psi_2 = 1$
DGP5	8-qtr perfect foresight	$\psi_8 = 1$

Note: Coefficient settings in tax rule (4.1).

by Escanciano and Velasco (2006).[10]

I estimate a VAR(p) based on a sample size of 250 which is about the size of most postwar data sets. The number of Monte Carlo replication is 500. I also throw away the first 1000 observations for removing initial conditions effects on the simulations. I choose the order of VAR, p, using the Akaike Information Criterion (AIC) to reduce the probability of choosing a small order VAR by mistake.[11] If the VARMA representation is non-invertible, it does not admit a VAR representation mapping economic shocks to a vector of observable variables and its lags. Therefore, I expect that VAR estimation give a reasonable approximation for DGP1 and DGP2, but a poor one for DGP3-DGP5.

Some comments are in order. First, $\hat{M}$ involves $d-$ and $2d-$ dimensional numerical integration, which can be computationally cumbersome when d is large. In practice, one may approximate the integrals by choosing a finite number of grid points symmetric about zero or generate a finite number of points drawn from the uniform distribution on $[-1, 1]^d$. Alternatively, for some weighting functions there is a closed form expression for the test statistics. In this paper, I use a closed form solution obtained by choosing $d\mathcal{W}(\cdot)$ as the $d-$dimensional Gaussian CDF.

[10] As a robustness check, I examined many combinations of alternative volatility forms and found results that are consistent with those of Table 2.

[11] The results (not reported here) are very similar when I use the BIC and HQ criteria. The finding that choosing different lag order does not solve the invertibility problem is in accordance with the fact that if a model is non-invertible, we can not recover the true shocks even if we include infinite lags.

Table 2: Empirical rejections probabilities for DGP1-DGP5

		DGP1(size)		DGP2(size)		DGP3		DGP4		DGP5	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Panel A: IID											
$\bar{h}=5$	$\hat{M}$	1.2	0.4	1.6	0.8	91.0	86.2	90.2	85.8	91.8	88.4
	$\hat{Q}$	27.4	20.6	38.0	30.2	93.4	88.6	93.8	89.2	95.8	92.6
$\bar{h}=10$	$\hat{M}$	1.0	0.4	1.6	0.8	90.6	81.4	90.4	85.2	89.8	85.6
	$\hat{Q}$	27.0	20.4	37.2	30.8	91.8	85.8	93.0	87.4	93.2	89.0
$\bar{h}=15$	$\hat{M}$	1.0	0.4	1.8	0.6	87.6	81.6	88.8	80.2	86.0	82.6
	$\hat{Q}$	27.6	20.2	37.0	30.4	90.8	84.6	91.2	83.0	89.6	84.6
Panel B: GARCH											
$\bar{h}=5$	$\hat{M}$	2.2	1.8	2.0	1.2	100	99.0	99.8	97.6	100	100
	$\hat{Q}$	57.6	51.2	59.8	51.8	100	100	100	100	100	100
$\bar{h}=10$	$\hat{M}$	2.2	1.6	1.8	1.4	100	98.4	98.8	96.0	100	99.0
	$\hat{Q}$	58.4	52.8	66.6	57.7	100	100	100	100	100	100
$\bar{h}=15$	$\hat{M}$	2.2	1.4	1.6	1.2	98.2	95.6	99.2	96.0	100	99.4
	$\hat{Q}$	58.0	52.6	66.4	57.6	100	100	100	100	100	100

Notes: (1) $\hat{M}$ is the multivariate martingale test; (2) $\hat{Q}$ is the multivariate independence test proposed by Chen et al. (2012); (3) $\bar{h}$ is the preliminary lag order used in a plug-in method to select a data-driven lag order; (4) The number of Monte Carlo replication is 500; (5) Sample size is 250.

Second, a practical issue in implementing the test is the choice of the bandwidth parameter $\hat{h}$. Following Hong and Lee (2005), one can choose a data-driven bandwidth $\hat{h} = \hat{c}_0 T^{\frac{1}{2q+1}}$ via the plug-in method, which lets data themselves determine an appropriate lag.[12] The data-driven bandwidth $\hat{c}_0$, involves the choice of a preliminary bandwidth $\bar{h}$, which can be fixed or grow with the sample size T. Applying the data-driven method to choose the bandwidth, while considering a wide range of the bandwidth, $\bar{h} \in \{4, \cdots, 16\}$, the simulation results show that the test is not sensitive to the choice of preliminary bandwidth. For the sake of space, I only report the results for $\bar{h} = 5, 10$ and 15, using the Bartlett kernel. Simulations suggest that the choice of $k(\cdot)$ has little impact on both the level and the power of the test.

Table 2 reports the rejection rates of the tests at the 10% and 5% levels. The simulation results show that $\hat{M}$ severely under-rejects $\mathbb{H}_0$. Similar under-rejection has been reported by Hong and Lee (2005).[13] This could be due to the fact that the asymptotic standard normal distribution only approximates the small sample distribution of the test statistic under the null hypothesis, and $T = 250$ is rather small. For example, when I increase the sample size to $T = 500$, the size for DGP1 improves to 2.6 and 6.4 at 5% and 10% level of significance, respectively.[14] The fact that the test is under-rejects the null hypothesis is not harmful. However, this might be imply that the test is also under power.

For the sake of comparison, I also report the multivariate independence test $\hat{Q}$ proposed by Chen et al. (2012). As can be seen from Table 2, $\hat{Q}$ does not control the size, even under the *iid* assumption. The rejection of the null hypothesis of serial independence can be due to the truncation error. Theoretically, the truncation error

[12] q is called the characteristic exponent of $k(.)$. For Bartlett kernel, $q = 1$; for quadratic spectral (QS) and Tukey kernels, $q = 2$.

[13] These authors argue that the under-rejection is due to the parameter estimation uncertainty in the finite-sample.

[14] Hong and Lee (2007) argue that the under-rejection might be due to the impact of parameter estimation uncertainty in small samples. Indeed, this might be the case for my simulations since using AIC, I may estimate a long VAR when it is unnecessary. For example, for the DGP1, which we know the correct order is $p = 1$, the average lag order chosen by AIC is 2.67. When I estimate a VAR with $p = 1$, the size performance improves.

associated with the estimation of a finite order VAR(p) which only approximates the exact infinite order VAR representation is expected to be small. However, it might be the case that the lag order p necessary to recover the structural shock maybe very large, and therefore the errors after truncation might be dependent even under the invertibility assumption (see, e.g., Chari et al., 2005; Ravenna, 2007).

4.2 Empirical application

As an empirical application, I focus on the dynamic effects of government spending shocks on economic activity in the United States. It has been argued that fiscal policy should be the primary tool for the economy to recover from the Great Recession and operate near potential level of output and employment. Yet there is a sharp conflict over the efficacy of discretionary fiscal policy.

Using VAR techniques, Blanchard and Perotti (2002) find moderate estimates of government spending output multipliers, an increase in consumption and the real wages (see also, Galí et al., 2007; Mountford and Uhlig, 2009). In contrast, Ramey (2011) argue that big increases in military spending are anticipated several quarters before they actually occur. Leeper et al. (2013) argue that fiscal foresight can create non-fundamentalness and therefore econometric methods using VAR models can not recover the correct structural shocks and impulse response functions.

To check whether fiscal foresight plays an important role in measuring the government spending shocks, I apply the test to the VAR specification standard in the empirical fiscal policy literature. To this end, suppose an economy is represented by a VMA model

$$x_t = \Gamma(L)\xi_t \tag{4.2}$$

where x_t consists of variables of interest and $\Gamma(L)$ is a polynomial in the lag operator. For the baseline specification, I include quarterly real per capita taxes, government spending, GDP, and the tax rate. This set of variables is similar to the ones used recently by Ramey (2011), covering the period 1948:I-2008:IV, and is available on Valerie Ramey's website.

Obtaining structural shocks from a VAR involves two steps: first, impose invertibility on (4.2) and construct a reduced form VAR model

$$\Pi(L)x_t = \xi_t \tag{4.3}$$

where $\Pi(L)$ is an autoregressive polynomial in the lag operator. Wold innovations can be recovered from estimating (4.3) with p lags. Second, structural disturbances are identified from the reduced-form errors, imposing some restrictions derived from economic theory.

To apply the test, I only need model residuals from the first step. This is consistent with what one would expect: no identification scheme is valid if the VAR is non-fundamental. Following Ramey (2011), I specify the VAR in levels, with a quadratic time trend and four lags included. Panel A of Table 3 reports the p-values of the tests applied to the residuals of this model.

Applying the tests to the residuals obtained from VAR, one observes that both $\hat{M}$ and $\hat{Q}$ reject the null of fundamentalness at the 10% level for the baseline specification. This implies that based on the results of the tests, given the data and variables selected in the baseline model, the impulse responses from SVAR approach appears not to be reliable.

Giannone and Reichlin (2006) proposed to restore the fundamentalness by expanding the econometrician's information set using extra information. Ramey (2011) argues that many shocks identified from a SVAR are anticipated changes in defense spending, which accounts for almost all of the volatility of government spending. Motivated by the importance of measuring anticipation, Ramey uses narrative evidence to construct a new variable, which measures the expected discounted value of government spending changes. Augmenting the baseline model with this

Table 3: Testing for Fundamentalness of VAR

	Panel A: Baseline specification		
	$\bar{h} = 5$	$\bar{h} = 10$	$\bar{h} = 15$
$\hat{Q}$	0.000	0.000	0.000
$\hat{M}$	0.026	0.042	0.069
	Panel B: *News*-augmented specification		
$\hat{Q}$	0.000	0.000	0.000
$\hat{M}$	0.382	0.384	0.392

Notes: (1) P-values for the null hypothesis that the structural model is fundamental; (2) $\hat{M}$ is the multivariate martingale test; (3) $\hat{Q}$ is the multivariate independence test proposed by Chen et al. (2012); (4) $\bar{h}$ is the preliminary lag order used in a plug-in method to select a data-driven lag order.

narrative variable, Ramey finds very different effects of government spending on economic activities, and conjectures that this new narrative variable might solve the non-fundamentalness problem.[15]

My proposed test can be used to formally show if adding more information solves the non-fundamentalness problem. Panel B of Table 3 reports the p-values for the null of fundamentalness for the $\hat{M}$ and $\hat{Q}$, which suggest that we fails to reject the null for the *news*-augmented model. This implies that based on the results of the tests, the SVAR model augmented with the *news* variable is fundamental, and the impulse responses appear to be reliable. In contrast, serial dependence test, $\hat{Q}$, rejects the null of fundamentalness at 5% level for the *news*-augmented model. As discussed in the simulation study, this could be due to the fact that the $\hat{Q}$ test over-reject the null hypothesis.

[15]Using the narrative tax series constructed by Romer and Romer (2010), Mertens and Ravn (2012) also find that the effects of anticipated tax changes are very different from the unanticipated ones.

5 Conclusions

This paper provides a new theoretical and empirical tool for testing fundamentalness assumption of macroeconomic models. I convert the fundamentalness testing problem into one of testing the unpredictability of the Wold innovations. To test the unpredictability, I extend the generalized spectral density test of Hong and Lee (2005) to the multivariate case. The proposed test is simple to apply since it only needs model residual as input and has a convenient asymptotic standard normal distribution. In addition, the test is robust to the failure of the *iid* assumption and does not need information outside of the specified model to check for fundamentalness. The Monte Carlo study based on a DSGE model with fiscal foresight exhibits a satisfactory finite-sample performance of the proposed test. Furthermore, an empirical application to the identification of government spending shocks illustrates how to use the proposed test to a variety of empirical problems.

If the null hypothesis is rejected, it has been conjectured that expanding the econometrician's information set may restore the fundamentalness. The proposed test can be used to formally check if adding more information solves the non-fundamentalness problem. In the empirical application, I show that augmenting a standard VAR model with a narrative variable that measure anticipations solves the non-fundamentalness problem.

Acknowledgments

The author is deeply indebted to Carlos Velasco for guidance and encouragement. I also benefited most from the comments of Juan Carlos Escanciano, Miguel Delgado and Jesus Gonzalo. The research was supported by the Spanish Plan Nacional de I+D+I (ECO2012-31748) and (ECO2014-57007).

Appendix

I first prove Lemma 1, which is an extension of Theorem 5.4.1 Rosenblatt (2000), by dropping the identical distribution assumption. In Lemma 2, I use Lemma 1 to prove the univariate case of Proposition 2.1, and then show that under Assumption 1 the multivariate case can be reduced to the univariate case.

Lemma 1: Consider a causal and non-invertible ARMA(p, q) model

$$\sum_{i=0}^{p} \alpha_i \epsilon_{t-i} = \sum_{i=0}^{q} \beta_i \xi_{t-i} \tag{A.1}$$

and let $\phi^t(\tau)$ denote the characteristic function of ξ_t and $\phi^t_{\tau_0}(\cdot) = \frac{\partial \phi^t(\cdot)}{\partial \tau_0}$. Then linearity of the best predictor in mean square implies that

$$\sum_{k=-\infty}^{\infty} \big(\gamma_k - \sum_{l=1}^{\infty} \mathbf{b}_l \gamma_{k-l}\big) h^{t-k} \big(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}\big) = 0 \tag{A.2}$$

where $h^t(\vartheta) = \frac{\phi^t_{\tau_0}(\vartheta)}{\phi^t(\vartheta)}$ and $\mathbf{b}_l$'s are the coefficients of the best linear predictor of ϵ_t in mean square

$$\epsilon_t^* = \sum_{l=1}^{\infty} \mathbf{b}_l \epsilon_{t-l} \tag{A.3}$$

Proof of Lemma 1: Writing (A.1) in the MA form we have:

$$\epsilon_t = \sum_{k=0}^{\infty} \gamma_k \xi_{t-k}, \quad \gamma_k = 0 \quad \forall k < 0 \tag{A.4}$$

The joint characteristic function of $\{\epsilon_{t-j}, j \geq 0\}$ is given by

$$\begin{aligned}\eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots) &= E\Big\{ \exp\big(i \sum_{l=0}^{\infty} \tau_l \xi_{t-l}\big) \Big\} \\ &= \prod_{k=-\infty}^{\infty} \phi^{t-k}\big(\sum_{l=0}^{\infty} \tau_l \gamma_{t-l}\big) \qquad \text{(A.5)}\end{aligned}$$

while the joint characteristic function of $\{\epsilon_{t-j}, j \geq 1\}$ is

$$\tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \prod_{k=-\infty}^{\infty} \phi^{t-k}\big(\sum_{l=1}^{\infty} \tau_l \gamma_{t-l}\big) \qquad \text{(A.6)}$$

Differentiating $\eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots)$ w.r.t. τ_0 we have

$$\begin{aligned}&\frac{\partial}{\partial \tau_0} \eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots)|_{\tau_0=0} = \eta^t_{\tau_0}(0, \tau_1, \cdots, \tau_p, \cdots) \\ &\quad = \int i\epsilon_t \; \exp(i \sum_{l=1}^{\infty} \tau_l \epsilon_{t-l}) \; dF^t(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots) \qquad \text{(A.7)} \\ &\quad = i \int E[\epsilon_t | \epsilon_{t-s}, s > 0] \; \exp(i \sum_{l=1}^{\infty} \tau_l \epsilon_{t-l}) \; dF^t(\epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots)\end{aligned}$$

where $F^t(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots)$ is the joint cumulative distribution function of $\epsilon_{t-j}, j \geq 0$. Also by differentiating the logarithm of (A.4) w.r.t. τ_0 we get:

$$\frac{\eta^t_{\tau_0}(0, \tau_1, \cdots, \tau_p, \cdots)}{\eta^t(0, \tau_1, \cdots, \tau_p, \cdots)} = \sum_{k=-\infty}^{\infty} \gamma_k h^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}). \qquad \text{(A.8)}$$

Similarly, differentiating the logarithm of $\tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots)$ w.r.t. τ_j, $j = 1, 2, \cdots$, we have

$$\frac{\partial}{\partial \tau_j} log \; \tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \sum_{k=-\infty}^{\infty} \gamma_{k-j} h^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}), \quad j = 1, 2, \cdots \qquad \text{(A.9)}$$

If the best predictor in mean square is linear we must have

$$\eta^t_{\tau_0}(0, \tau_1, \cdots) = \sum_{k=1}^{\infty} \mathbf{b}_k \tilde{\eta}^t_{\tau_k}(\tau_1, \tau_2, \cdots) \tag{A.10}$$

which implies

$$\sum_{k=-\infty}^{\infty} \big(\gamma_k - \sum_{l=1}^{\infty} \mathbf{b}_l \gamma_{k-l}\big) h^{t-k} \big(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}\big) = 0. \tag{A.11}$$

□

Lemma 2: Let Assumption 1 hold. The univariate non-Gaussian ARMA model (2.1) is invertible if and only if the Wold innovations $\{\epsilon_t\}$ are MDS.

Proof of Lemma 2: A standard result for ARMA processes is that any ARMA(p, q) process $\{x_t\}$ which is non-invertible with respect to the noise sequence $\{\xi_t\}$ can also be modeled as an invertible ARMA(p, q) with respect to a new noise sequence $\{\epsilon_t\}$ defined by[16]

$$\epsilon_t = \frac{\prod_{r_B < i \le q} (1 - b_i^{-1} L)}{\prod_{r_B < i \le q} (1 - b_i L)} \xi_t, \qquad |b_i| < 1. \tag{A.12}$$

which can be written as:

$$\sum_{i=0}^{q-r_B} \alpha_i \epsilon_{t-i} = \sum_{i=0}^{q-r_B} \beta_i \xi_{t-i} \tag{A.13}$$

Let $y_t = \sum_{i=0}^{q-r_\Theta} \alpha_i \epsilon_{t-i}$. Then (A.13) can be written as:

$$y_t = \sum_{i=0}^{q-r_\Theta} \beta_i \xi_{t-i}. \tag{A.14}$$

Because y_t in a non-invertible MA of order $(q - r_\Theta)$, Lemma 1 and Corollary 5.4.3 of Rosenblatt (2000) implies that the best one-step predictor of y_t is non-linear, i.e., $E[y_t | y_{t-s}, s \ge 1]$ is non-linear. On the other hand, y_t is causal since all the roots of $\prod_{r_\Theta < i \le q} (1 - b_i L)$ are outside the unit circle. Therefore, the $\sigma-$algebras $\sigma(\epsilon_{t-s}, s \ge 1)$

[16]See Brockwell and Davis (1991), page 103.

and $\sigma(y_{t-s}, s \geq 1)$ coincide, and

$$\begin{aligned}
E[y_t|y_{t-s}, s \geq 1] &= E[y_t|\epsilon_{t-s}, s \geq 1] \quad a.s. \\
&= E[\epsilon_t - \alpha_1\epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta}\epsilon_{t-q-r_\Theta}|\epsilon_{t-s}, s \geq 1] \quad a.s. \\
&= E[\epsilon_t|\epsilon_{t-s}, s \geq 1] - \alpha_1\epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta}\epsilon_{t-q-r_\Theta} \quad a.s. \qquad \text{(A.15)}
\end{aligned}$$

If ϵ_t were a MD, i.e. $E[\epsilon_t|\epsilon_{t-s}, s \geq 1] = 0$, then

$$E[y_t|y_{t-s}, s \geq 1] = -\alpha_1\epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta}\epsilon_{t-q-r_\Theta} \quad a.s. \qquad \text{(A.16)}$$

which is linear -a clear contradiction- and therefore ϵ_t can not be a MD.$\square$

Proof of Proposition 2.1: Note that without loss of generality we can assume that the first component of $\{\xi_t\}$ satisfies Assumption 1. It is clear that if $\{x_t\}$ is invertible $\{\epsilon_t\} \equiv \{\xi_t\}$ are MD. I want to prove the reciprocal, that is if $\{x_t\}$ is non-invertible then $\{\epsilon_t\}$ is non-MD. The proof in the univariate case follows from Lemma 2. I want to show under Assumption 1 we can reduce the multivariate to the univariate case. Let $\tilde{\Theta}^{-1}(L)\Theta(L) = A(L)$. Write

$$A(L) = \begin{bmatrix} A_1(L) & A_2(L) \end{bmatrix}$$

Where $A_1(L)$ is $d \times 1$ and $A_2(L)$ is $d \times (d-1)$. From (2.4) we have

$$\epsilon_t = \tilde{\Theta}^{-1}(L)\Theta(L)\xi_t$$

$$\begin{bmatrix} \epsilon_t & M \end{bmatrix} = A(L)\begin{bmatrix} \xi_t & \begin{matrix} 0_{1\times(d-1)} \\ I_{d-1} \end{matrix} \end{bmatrix}$$

where $M = A_2(L)$. Define $\tilde{\epsilon}_t = \det\begin{bmatrix} \epsilon_t & M \end{bmatrix}$, and note that by Assumption 1 and the property $A^*(1)A(1) = I_d$, $\{\tilde{\epsilon}_t\}$ is a non-zero measurable transformation of $\{\xi_t\}$.

Furthermore from the properties of determinants we have

$$\tilde{\epsilon}_t = \det(A(L))\xi_{t1}$$

where ξ_{t1} is the first component of ξ_t. Theorem 3 in Lippi and Reichlin (1994) implies that for some non-zero constant C

$$\tilde{\epsilon}_t = C\frac{\Theta^*(L)}{\Theta^f(L)}\xi_{t1}$$

where $\Theta^*(L)$ contains the non-invertible roots, i.e.,

$$\Theta^*(L) := \prod_{i=s+1}^{dq}(1 - b_i^{-1}z), \quad |b_i| < 1$$

and $\Theta^f(L)$ is the flipped-root polynomial defined as

$$\Theta^f(L) := \prod_{i=s+1}^{dq}(1 - b_i^* z), \quad |b_i^*| < 1$$

with $s \in \left[0, dq\right]$ is the number of the invertible roots of $\det(\Theta(z)) = 0$. Then by Lemma 2 $\{\tilde{\epsilon}_t\}$ and hence $\{\xi_t\}$ is non-MD.□

Proof of Corrolary 2.1: Note that by the proof of Theorem 1 in Lippi and Reichlin (1994)

$$\tilde{\Theta}^{-1}(L)\Theta(L) = R(\alpha, L)K$$

where K is an orthogonal matrix and

$$R(\alpha, L) = \begin{pmatrix} \frac{1+b^{-1}L}{1+bL} & 0 \\ 0 & \mathbf{I}_{d-1} \end{pmatrix}$$

Then, if $\{\xi_t\}$ is a martingale difference process, $\{\tilde{\xi}_t := K\xi_t\}$ is also a martingale difference process, and the results from Lemma 2 applied to

$$\epsilon_t = \frac{1 + b^{-1}L}{1 + bL}\tilde{\xi}_{1t}$$

where $\tilde{\xi}_{1t}$ is the first component of $\tilde{\xi}_t$.$\square$

References

Alessi, L., Barigozzi, M., and Capasso, M., 2011. Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1):16–47.

Blanchard, O. J. and Perotti, R., 2002. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4):1329–1368.

Blanchard, O. J., L'Huillier, J.-P., and Lorenzoni, G., 2013. News, noise, and fluctuations: An empirical exploration. *American Economic Review*, 103(7):3045–70.

Box, G. E. and Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Breidt, F. J. and Davis, R. A., 1992. Time-reversibility, identifiability and independence of innovations for stationary time series. *Journal of Time Series Analysis*, 13(5):377–390.

Brockwell, P. J. and Davis, R. A., 1991. *Time Series: Theory and Methods: Theory and Methods*. Springer series in statistics. Springer.

Chari, V., Kehoe, P. J., and McGrattan, E. R., 2005. *A critique of structural VARs using business cycle theory*. Federal Reserve Bank, Research Department.

Chen, B., Choi, J., and Escanciano, J. C., 2012. Testing for fundamental vector moving average representations.

Domínguez, M. A. and Lobato, I. N., 2003. Testing the martingale difference hypothesis. *Econometric Reviews*, 22(4):351–377.

Escanciano, J. C. and Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, 134(1):151–185.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., and Watson, M. W., 2007. ABCs (and Ds) of Understanding VARs. *American Economic Review*, 97 (3):1021–1026.

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

Forni, M., Gambetti, L., and Sala, L., 2014. No news in business cycles. *The Economic Journal*.

Galí, J., López-Salido, J. D., and Vallés, J., 2007. Understanding the effects of government spending on consumption. *Journal of the European Economic Association*, 5(1):227–270.

Giannone, D. and Reichlin, L., 2006. Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2-3):455–465.

Hamedani, G., 2013. Sub-independence: An expository perspective. *Communications in Statistics-Theory and Methods*, 42(20):3615–3638.

Harvey, C. R. and Siddique, A., 1999. Autoregressive conditional skewness. *Journal of financial and quantitative analysis*, 34(04):465–487.

Harvey, C. R. and Siddique, A., 2000. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295.

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(448):1201–1220.

Hong, Y. and Lee, Y.-J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, 72(2):499–541.

Hong, Y. and Lee, Y.-J., 2007. An improved generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Econometric Theory*, 23(01):106–154.

Jondeau, E. and Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

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

Lii, K.-S. and Rosenblatt, M., 1982. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The Annals of Statistics*, pages 1195–1208.

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

Ljung, G. M. and Box, G. E., 1978. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Mertens, K. and Ravn, M. O., 2012. Empirical evidence on the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy*, 4(2):145–81.

Mountford, A. and Uhlig, H., 2009. What are the effects of fiscal policy shocks? *Journal of applied econometrics*, 24(6):960–992.

Ramey, V. A., 2011. Identifying government spending shocks: it's all in the timing*. *The Quarterly Journal of Economics*, 126(1):1–50.

Ravenna, F., 2007. Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, 54(7):2048–2064.

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

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

Schmitt-Grohé, S. and Uribe, M., 2012. What's news in business cycles. *Econometrica*, 80(6):2733–2764.

Sims, C. A., 1980. Macroeconomics and Reality. *Econometrica*, 48(1):1–48.