working paper
2108

Tests for random coefficient variation in vector autoregressive models

Dante Amengual
Gabriele Fiorentini
Enrique Sentana

September 2021

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2108
September 2021

Tests for random coefficient variation in vector autoregressive models

Abstract

We propose the information matrix test to assess the constancy of mean and variance parameters in vector autoregressions. We additively decompose it into several orthogonal components: conditional heteroskedasticity and asymmetry of the innovations, and their unconditional skewness and kurtosis. Our Monte Carlo simulations explore both its finite size properties and its power against i.i.d. coefficients, persistent but stationary ones, and regime switching. Our procedures detect variation in the autoregressive coefficients and residual covariance matrix of a VAR for the US GDP growth rate and the statistical discrepancy, but they fail to detect any covariation between those two sets of coefficients.

JEL Codes: C32, C52, E01.

Keywords: GDP, GDI, Hessian matrix, information matrix test, outer product of the score.

Dante Amengual
CEMFI
amengual@cemfi.es

Gabriele Fiorentini
Università di Firenze and RCEA
gabriele.fiorentini@unifi.it

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

We would like to thank Pablo Cabrales and Jaime De la Vega for able research assistance. The first and third authors also acknowledge financial support from the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair.

1 Introduction

Following the path-breaking work by Sims (1980), vector autoregressions (VARs) have become one of the most commonly employed tools by empirical macroeconomists in academia, central banks and research departments of financial institutions. Their ability to capture dynamic relationships between multiple variables makes them particularly apt for short- and medium-run economic analysis, either on their own or in combination with structural macroeconomic models (see Canova (2007) for a textbook treatment). In fact, VARs have become the forecasting benchmark to beat, thereby replacing the univariate ARIMA models in vogue in the 1970's and 80's.

Unlike those univariate models, though, VARs are hardly ever subject to a battery of specification tests. Part of the reason is that specification testing does not fit well with the Bayesian approach to inference predominant among macroeconometricians. But the scarcity of specification tests for those models also plays an important role. The purpose of our paper is precisely to apply the information matrix test of White (1982) to VARs.

Our choice of specification test is far from random. The neglected heterogeneity interpretation of the information matrix test in Chesher (1984) provides a very relevant justification in macroeconomic applications, in which changes in the structure of the economy are a first-order concern (see e.g. Perron (1989)). There is, in fact, a long tradition of autoregressive models with time-varying parameters, which are sometimes called Random Coefficient Autoregressions (RCAs) in the time series literature (see e.g. Nicholls and Quinn (1982) for an earlier treatment, and Regis, Serra and van den Heuvel (2021) for a recent survey). Moreover, in recent years the macroeconometric literature has paid considerable attention to models in which not only the parameters governing the conditional mean change over time, but also the parameters corresponding to the variances and covariances of the innovations may also time-vary (see e.g. Primiceri (2005) or D'Agostino, Gambetti and Giannone (2013)).

Several tests for constant versus random coefficients in autoregressive models already exist in the literature, mostly in the univariate case (see e.g. Lee (1998), Akharif and Hallin (2003), Horváth and Trapani (2019), or Chen et al (2020)). Some of them use a likelihood framework, but they tend to focus on the classical triad of Wald, Likelihood Ratio and Lagrange Multiplier (LM) tests, which have a somewhat non-standard distribution under the null. There is also a huge literature on structural break tests, as well as on testing for recurrent regime switches (see, respectively, Hansen (2001) and Carrasco, Hu and Ploberger (2014), and the references therein).

Our information matrix test, though, has two main advantages: (i) unlike many of those tests, the test statistic has an asymptotic chi-square distributions under the null when the autoregressive process is covariance stationary, and (ii) it can be additively decomposed into four easily interpretable orthogonal components: a) a test for conditional heteroskedasticity of the innovations, b) a test for conditional asymmetry of those innovations, c) a test for their unconditional skewness, and d) a test for their unconditional kurtosis. These four tests can be combined into other easily interpretable tests. For example, the sum of b) and c) assesses the null hypothesis of zero covariance between the mean and variance parameters. Similarly, the sum of c) and d) checks the multivariate normality of the innovations.

An additional advantage of the information matrix test is that it can capture multiple types of deviations from constant coefficients even though it is not specifically designed for them. For that reason, we conduct an extensive Monte Carlo exercise in which we study the power of the test against three types of random coefficient variation: *i.i.d.* coefficients, as in RCAs, persistent but stationary coefficients, and finally regime switching models. In all cases, we calibrate the designs so that the unconditional variances of the coefficients is the same across these three alternatives, and compute critical values under the null using the bootstrap.

Finally, we apply our procedures to test the parameter constancy of a VAR in an important empirical context. Specifically, we study the dynamic relationship between the equally weighted average of the growth rates of the expenditure and income measures of US Gross Domestic Product (GDP) produced by the Bureau of Economic Analysis, and the statistical discrepancy, which is the difference between the (log) levels of those two measurements. Thus, we follow Almuzara, Amengual and Sentana (2019) and Almuzara, Fiorentini and Sentana (2021) in imposing cointegration between the Gross Domestic Expenditure (GDE) and Gross Domestic Income (GDI) measures. The empirical results that we obtain with our proposed information matrix tests confirm time-variation in both the autoregressive coefficients and the residual covariance matrix of the innovations, but they fail to detect any covariance between those two groups of coefficients.

The rest of the paper is organised as follows. In section 2, we derive the information matrix test for a multivariate linear regression model, while in section 3 we specialise it to VARs. We then present the results of our simulation experiments on its size and power properties in section 4, and apply it to US GDE and GDI in section 5. Our conclusions appear in section 6, followed by appendices that contain proofs together with some additional material.

2 Information matrix tests of the multivariate regression model

Consider the following multivariate normal regression model

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{\Omega}^{1/2}\varepsilon_t^*, \tag{1}$$

where the vector of dependent variables $\mathbf{y}_t$ is $N \times 1$, the vector of regressors $\mathbf{x}_t$, which often includes a constant, is $M \times 1$, the matrix of regression coefficients $\mathbf{B}$ is $N \times M$, the residual covariance matrix $\mathbf{\Omega}$ is $N \times N$ symmetric and positive definite, and the vector of standardised innovations $\boldsymbol{\varepsilon}_t^*$ follows a spherical normal distribution conditional on the regressors and the past values of the observed variables. Thus, the conditional mean vector and covariance matrix of $\mathbf{y}_t$ will be $\boldsymbol{\mu}_t(\boldsymbol{\theta}) = \mathbf{B}\mathbf{x}_t$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \mathbf{\Omega}$, respectively, where $\boldsymbol{\theta} = (\mathbf{b}', \boldsymbol{\omega}')'$, $\mathbf{b} = vec(\mathbf{B})$ and $\boldsymbol{\omega} = vech(\mathbf{\Omega})$.

Given these assumptions, the contribution from a single observation to the log-likelihood function is

$$-\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\mathbf{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}\varsigma_t(\boldsymbol{\theta}),$$

where $\varsigma_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) = \mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$.

The maximum likelihood estimators of the model parameters are known in closed form without the need to conduct any numerical optimisation. Specifically,

$$\hat{\mathbf{B}} = \left(\sum_{t=1}^{T} \mathbf{y}_t\mathbf{x}_t'\right)\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}$$

and

$$\hat{\mathbf{\Omega}} = \frac{1}{T}\left[\sum_{t=1}^{T}(\mathbf{y}_t - \hat{\mathbf{B}}\mathbf{x}_t)(\mathbf{y}_t - \hat{\mathbf{B}}\mathbf{x}_t)'\right].$$

Nevertheless, we need expressions for the score and Hessian matrix to be able to derive the information matrix test.

To compute the score, we first differentiate $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ with respect to the $q = MN + N(N+1)/2$ model parameters in $\boldsymbol{\theta}$. Specifically, the first derivatives are given by

$$\begin{aligned}\frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \mathbf{b}'} &= \mathbf{x}_t' \otimes \mathbf{I}_N \\ \frac{\partial vec[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]}{\partial \boldsymbol{\omega}'} &= \mathbf{D}_N,\end{aligned}$$

where $\mathbf{D}_N$ is the duplication matrix of order N (see Magnus and Neudecker (2019)). Thus, the log-likelihood score is

$$\mathbf{s}_t(\boldsymbol{\theta}) = \mathbf{Z}_{lt}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) + \mathbf{Z}_{st}(\boldsymbol{\theta})vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N],$$

where

$$\begin{aligned}\mathbf{Z}_{lt}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{x}_t \otimes \mathbf{\Omega}^{-1/2\prime} \\ \mathbf{0} \end{bmatrix}, \\ \mathbf{Z}_{st}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{0} \\ \frac{1}{2}\mathbf{D}_N'(\mathbf{\Omega}^{-1/2\prime} \otimes \mathbf{\Omega}^{-1/2\prime}) \end{bmatrix}.\end{aligned}$$

As a result, the scores will be

$$
\begin{aligned}
\mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= [\mathbf{x}_t \otimes \boldsymbol{\Omega}^{-1/2\prime}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] = [\mathbf{x}_t \otimes \boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)] \\
&= vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)\mathbf{x}_t'] \qquad (2)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N] \\
&= \frac{1}{2}\mathbf{D}_N' vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}]. \qquad (3)
\end{aligned}
$$

Consequently, the outer product of the scores will be

$$
\begin{aligned}
\mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta})\mathbf{s}_{\mathbf{b}t}'(\boldsymbol{\theta}) &= [\mathbf{x}_t\mathbf{x}_t' \otimes \boldsymbol{\Omega}^{-1/2\prime}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1/2}] \\
&= [\mathbf{x}_t\mathbf{x}_t' \otimes \boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}],
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta})\mathbf{s}_{\mathbf{b}t}'(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N][\mathbf{x}_t' \otimes \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1/2}] \\
&= \frac{1}{2}\mathbf{D}_N' vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}][\mathbf{x}_t' \otimes (\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta})\mathbf{s}_{\boldsymbol{\omega}t}'(\boldsymbol{\theta}) &= \frac{1}{4}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})vec[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N] \\
&\quad \times vec'[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_N](\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N \\
&= \frac{1}{4}\mathbf{D}_N' vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}] \\
&\quad \times vec'[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}]\mathbf{D}_N.
\end{aligned}
$$

To compute the Hessian, it is convenient to use the general expressions for elliptical distributions in Supplementary Appendix C of Fiorentini and Sentana (2021), namely

$$
\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) = \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},
$$

where

$$
\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\mathbf{Z}_{st}(\boldsymbol{\theta})\mathbf{Z}_{st}'(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \otimes \mathbf{I}_q\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}'
$$

and

$$
\begin{aligned}
\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' &= 2\mathbf{Z}_{lt}(\boldsymbol{\theta})\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + 8\mathbf{Z}_{st}(\boldsymbol{\theta})[\mathbf{I}_N \otimes \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})]\mathbf{Z}_{st}'(\boldsymbol{\theta}) \\
&\quad + 4\mathbf{Z}_{lt}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{st}'(\boldsymbol{\theta}) + 4\mathbf{Z}_{st}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}) \otimes \mathbf{I}_N]\mathbf{Z}_{lt}'(\boldsymbol{\theta}) \\
&\quad - 2[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-\frac{1}{2}\prime}(\boldsymbol{\theta}) \otimes \mathbf{I}_q]\partial vec[\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]\partial\boldsymbol{\theta}' \\
&\quad - \{vec'[\boldsymbol{\Sigma}_t^{-\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-\frac{1}{2}\prime}(\boldsymbol{\theta})] \otimes \mathbf{I}_q\}\partial vec\{\partial vec'[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}/\partial\boldsymbol{\theta}'.
\end{aligned}
$$

In the case of model (1), $d_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{\Omega}|$ and

$$\partial^2 d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N'(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\mathbf{D}_N \end{bmatrix}.$$

Similarly, we have that $g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}] = -\frac{1}{2}\varsigma_t(\boldsymbol{\theta})$ under normality, so that $\partial g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/\partial\varsigma = -\frac{1}{2}$ and $\partial^2 g[\varsigma_t(\boldsymbol{\theta}),\boldsymbol{\eta}]/(\partial\varsigma)^2 = 0$. Finally,

$$\partial^2\varsigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\begin{pmatrix} \mathbf{x}_t\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$+2\begin{Bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime}\otimes\boldsymbol{\Omega}^{-1/2\prime})[\mathbf{I}_N\otimes\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})](\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\mathbf{D}_N \end{Bmatrix}$$

$$+2\begin{Bmatrix} \mathbf{0} & (\mathbf{x}_t\otimes\boldsymbol{\Omega}^{-1/2\prime})[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_N](\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\mathbf{D}_N \\ \mathbf{0} & \mathbf{0} \end{Bmatrix}$$

$$+2\begin{Bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime}\otimes\boldsymbol{\Omega}^{-1/2\prime})[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\otimes\mathbf{I}_N](\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1/2}) & \mathbf{0} \end{Bmatrix}$$

$$= 2\begin{Bmatrix} (\mathbf{x}_t\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}) & [\mathbf{x}_t(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}]\mathbf{D}_N \\ \mathbf{D}_N'[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}] & \mathbf{D}_N'[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}]\mathbf{D}_N \end{Bmatrix},$$

where we have exploited the fact that the second derivatives of the conditional mean and covariance functions with respect to the model parameters are all 0.

Therefore, we can write the Hessian matrix as

$$-\begin{Bmatrix} (\mathbf{x}_t\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}) & [\mathbf{x}_t(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}]\mathbf{D}_N \\ \mathbf{D}_N'[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}] & \mathbf{D}_N'\{\boldsymbol{\Omega}^{-1}\otimes[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}-\frac{1}{2}\boldsymbol{\Omega}^{-1}]\}\mathbf{D}_N \end{Bmatrix}$$

The sum of the outer product of the score and the Hessian yields the following three terms:

$$\mathbf{bb}: [\mathbf{x}_t\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}] - (\mathbf{x}_t\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}), \tag{4}$$

$$\begin{aligned} \boldsymbol{\omega}\mathbf{b} \quad : \quad & \frac{1}{2}\mathbf{D}_N' vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}][\mathbf{x}_t'\otimes(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}] \\ & -\mathbf{D}_N'[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)\mathbf{x}_t'\otimes\boldsymbol{\Omega}^{-1}], \end{aligned} \tag{5}$$

and

$$\begin{aligned} \boldsymbol{\omega}\boldsymbol{\omega}\text{:}\ & \frac{1}{4}\mathbf{D}_N' vec[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}] \\ & \times vec'[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}]\mathbf{D}_N \\ & -\mathbf{D}_N'\{\boldsymbol{\Omega}^{-1}\otimes[\boldsymbol{\Omega}^{-1}(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)(\mathbf{y}_t-\mathbf{B}\mathbf{x}_t)'\boldsymbol{\Omega}^{-1}-\frac{1}{2}\boldsymbol{\Omega}^{-1}]\}\mathbf{D}_N. \end{aligned} \tag{6}$$

When $x_t = 1$, these formulas coincide with those in Amengual, Fiorentini and Sentana (2021), who re-write the $\boldsymbol{\omega}\mathbf{b}$ and $\boldsymbol{\omega}\boldsymbol{\omega}$ expressions in terms of multivariate Hermite polynomials of orders 3 and 4, respectively.[1] We can generalise their results for any $\mathbf{x}_t$ as follows. As in

[1] At the end of this section, we explicitly relate the test we propose to this multivariate normality test.

Barndorff-Nielsen and Petersen (1979), define the (centred) multivariate Hermite polynomials of $\boldsymbol{\varepsilon}$ of order $k = k_1 + \ldots + k_N \geq 0$ as

$$H_{1\overset{k_1}{\ldots}1\ldots N\overset{k_N}{\ldots}N}(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}) \cdot e^{-\frac{1}{2}\boldsymbol{\varepsilon}'\boldsymbol{\Delta}\boldsymbol{\varepsilon}} = (-1)^k \frac{\partial^k}{(\partial\varepsilon_1)^{k_1}\ldots(\partial\varepsilon_N)^{k_N}} \left(e^{-\frac{1}{2}\boldsymbol{\varepsilon}'\boldsymbol{\Delta}\boldsymbol{\varepsilon}}\right), \tag{7}$$

where $\boldsymbol{\Delta} = \boldsymbol{\Omega}^{-1}$ is the inverse covariance matrix of $\boldsymbol{\varepsilon}$. As is well known, when model (1) is correctly specified: i) the expected value of any multivariate Hermite polynomial of positive degree k conditional on the regressors and the past values of the observed variables is 0; and ii) the conditional and unconditional covariance matrices of those polynomials coincide.

Let

$$\mathbf{H}_k(\boldsymbol{\varepsilon}; \boldsymbol{\Delta}) = \begin{bmatrix} H_{k,0,\cdots,0}(\boldsymbol{\varepsilon}; \boldsymbol{\Delta}) \\ H_{k-1,1,\cdots,0}(\boldsymbol{\varepsilon}; \boldsymbol{\Delta}) \\ \vdots \\ H_{0,\cdots,0,k}(\boldsymbol{\varepsilon}; \boldsymbol{\Delta}) \end{bmatrix}$$

denote the $\binom{N+k-1}{k} \times 1$ vector that contains all the non-redundant multivariate Hermite polynomials of order k, which we will simply denote by $\mathbf{H}_k(\boldsymbol{\varepsilon}_t^*)$ for the special case of $\boldsymbol{\Delta} = \mathbf{I}_N$.

Similarly, let

$$\mathbf{m}_{ht}(\boldsymbol{\theta}) = \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \otimes vech(\mathbf{x}_t\mathbf{x}_t'), \tag{8}$$

$$\mathbf{m}_{at}(\boldsymbol{\theta}) = \mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \otimes \mathbf{x}_t, \tag{9}$$

$$\mathbf{m}_{kt}(\boldsymbol{\theta}) = \mathbf{H}_4[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})], \tag{10}$$

which effectively span (4), (5) and (6), respectively. Finally, let

$$\bar{\mathbf{m}}_{lT}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{m}_{lt}(\boldsymbol{\theta}) \text{ for } l = h, a, k.$$

We can then state our main result:

Proposition 1 *Assume $\mathbf{x}_t$ has finite fourth moments. Then, the information matrix test that compares the outer product of the score with the Hessian of the multivariate regression model (1) evaluated at the Gaussian maximum likelihood estimators $\hat{\boldsymbol{\theta}}_T = (\hat{\mathbf{b}}_T', \hat{\boldsymbol{\omega}}_T')'$ is asymptotically equivalent under the null hypothesis of correct specification to the sum of the following three moment tests:*

$$h_{hT} = T \cdot \bar{\mathbf{m}}_{hT}'(\hat{\boldsymbol{\theta}}_T)\hat{V}^{+}[\mathbf{m}_{ht}(\hat{\boldsymbol{\theta}}_T)]\bar{\mathbf{m}}_{hT}(\hat{\boldsymbol{\theta}}_T), \tag{11}$$

$$h_{aT} = T \cdot \bar{\mathbf{m}}_{aT}'(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{at}(\hat{\boldsymbol{\theta}}_T)]\bar{\mathbf{m}}_{aT}(\hat{\boldsymbol{\theta}}_T), \tag{12}$$

and

$$h_{kT} = T \cdot \bar{\mathbf{m}}_{kT}'(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{kt}(\hat{\boldsymbol{\theta}}_T)]\bar{\mathbf{m}}_{kT}(\hat{\boldsymbol{\theta}}_T), \tag{13}$$

where + denotes the Moore-Penrose generalised inverse,

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{m}}_{hT}(\hat{\boldsymbol{\theta}}_T)] = V\{\mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\} \otimes V\left[vech(\mathbf{x}_t\mathbf{x}_t')\right], \tag{14}$$

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{m}}_{aT}(\hat{\boldsymbol{\theta}}_T)] = V\{\mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\} \otimes E(\mathbf{x}_t\mathbf{x}_t'),\ \textit{and} \tag{15}$$

$$\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{m}}_{kT}(\hat{\boldsymbol{\theta}}_T)] = V\{\mathbf{H}_4[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\},$$

which converge in distribution to three independent chi-square random variables whose degrees of freedom are $\binom{N+1}{2} rank\{V\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]\}$, $\binom{N+2}{3}M$ *and* $\binom{N+3}{4}$, *respectively.*

Note that if $\mathbf{x}_t$ contains either a constant or a set of dummy variables that linearly span a constant term, then $V[vech(\mathbf{x}_t\mathbf{x}_t')]$ will be singular with nullity 1, which explains the generalised inverse in (11). Given that the diagonal covariance matrices of $\mathbf{H}_k(\boldsymbol{\varepsilon}_t^*)$ for $k = 2, 3, 4$ do not depend on any unknown quantities under the null of correct specification,[2] under standard regularity conditions we can consistently estimate (14) and (15) by simply replacing $V\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]$ and $E(\mathbf{x}_t\mathbf{x}_t')$ by their sample counterparts.

The relationship between information matrix tests and the general class of moment tests of Newey (1985) and Tauchen (1985) is well known (see White (1994)). What is perhaps less known among macroeconometricians is that Chesher (1984) proved that the information matrix test can be viewed as a Lagrange multiplier test against parameter variation. Thus, we can interpret the moment test statistic (13) that looks at the unconditional mean of the fourth-order multivariate Hermite polynomials as a test of neglected heterogeneity in $\boldsymbol{\omega}$, which are the parameters that characterise the covariance matrix of the innovations. Similarly, the test statistic (12) that looks at the conditional mean of the third-order polynomials effectively assesses dependence in the neglected heterogeneity of the mean and covariance parameters $\mathbf{b}$ and $\boldsymbol{\omega}$. Finally, the test statistic (11) that looks at the conditional mean of the second-order multivariate Hermite polynomials can be understood as a test of neglected heterogeneity in the $\mathbf{b}$ parameters that determine the conditional mean of the observations. In this respect, the additive decomposition in Proposition 1 provides a multivariate generalisation of Hall (1987).[3] In particular, (11) can be regarded as the multivariate counterpart to White's (1980) heteroskedasticity test, while (12) is a multivariate version of what Bera and Lee (1993) called a test for "heterocliticity", and (13) the multivariate analogue to the Kiefer and Salmon (1983) version of the kurtosis component of the Jarque and Bera (1980) test.

[2] Specifically, the diagonal elements of $V[\mathbf{H}_2(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*2}\text{-}1) = 2$ and $V(\varepsilon_i^*\varepsilon_{i'}^*) = 1$, for $i' \neq i$, while those of $V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*3}\text{-}3\varepsilon_i^*) = 6$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^*\text{-}\varepsilon_{i'}^*) = 2$ for $i' \neq i$ and $V(\varepsilon_i^*\varepsilon_{i'}^*\varepsilon_{i''}^*) = 1$ for $i'' \neq i' \neq i$. Finally, the diagonal elements of $V[\mathbf{H}_4(\boldsymbol{\varepsilon}^*)]$ are $V[(\varepsilon_i^{*2}\text{-}3\varepsilon_i^*)^2\text{-}6] = 24$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^{*2}\text{-}\varepsilon_i^{*2}\text{-}\varepsilon_{i'}^{*2}+1) = 4$ for $i' \neq i$, $V(\varepsilon_i^{*3}\varepsilon_{i'}^*\text{-}3\varepsilon_i^*\varepsilon_{i'}^*) = 6$ for $i' \neq i$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^*\varepsilon_{i''}^*\text{-}\varepsilon_{i'}^*\varepsilon_{i''}^*) = 2$ for $i'' \neq i' \neq i$, and $V(\varepsilon_i^*\varepsilon_{i'}^*\varepsilon_{i''}^*\varepsilon_{i'''}^*) = 1$ for $i''' \neq i'' \neq i' \neq i$ (see Amengual, Fiorentini and Sentana (2021) for further details).

[3] See Bera and Lee (1993) for a related result in univariate regression models with serially correlated residuals.

Importantly, if we consider the full-rank affine transformation of the dependent variables $\mathbf{z}_t = \mathbf{c}+\mathbf{D}\mathbf{y}_t$ with $|\mathbf{D}| \neq 0$, we can show that the information matrix test statistic is always numerically invariant to the value of $\mathbf{D}$, as well as to the value of $\mathbf{c}$ when the regressors include a constant term. This numerical invariance provides a very fast numerical procedure for computing the test statistics in Proposition 1 since we can work with the standardised innovations $\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}})$ without loss of generality, whose sample mean and covariance matrix will be $\mathbf{0}$ and $\mathbf{I}_N$, respectively, when the regressors include a constant term.

In fact, when $\mathbf{x}_t' = (1, \mathbf{z}_t')$, with $\dim(\mathbf{z}_t) = m$ so that $M = m + 1$, $E(\mathbf{z}_t) = \boldsymbol{\mu}_z$ and $V(\mathbf{z}_t) = \boldsymbol{\Sigma}_{zz} > 0$ in the positive definite sense, we can further additively decompose (12) as follows:

Proposition 2 *Assume $\mathbf{z}_t$ has finite second moments. Then, the component of the information matrix test that compares the off-diagonal block of the outer product of the score with the Hessian of the multivariate regression model (1) evaluated at the Gaussian maximum likelihood estimators $\hat{\boldsymbol{\theta}}_T = (\hat{\mathbf{b}}_T', \hat{\boldsymbol{\omega}}_T')'$ is asymptotically equivalent under the null hypothesis of correct specification to the sum of the following two moment tests:*

$$h_{asT} = T \cdot \bar{\mathbf{m}}_{asT}'(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{ast}(\hat{\boldsymbol{\theta}}_T)]\bar{\mathbf{m}}_{asT}(\hat{\boldsymbol{\theta}}_T), \tag{16}$$

$$h_{adT} = T \cdot \bar{\mathbf{m}}_{adT}'(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{adt}(\hat{\boldsymbol{\theta}}_T)]\bar{\mathbf{m}}_{adT}(\hat{\boldsymbol{\theta}}_T), \tag{17}$$

where

$$\begin{aligned}
\mathbf{m}_{ast}(\boldsymbol{\theta}) &= \mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})], & (18)\\
\mathbf{m}_{adt}(\boldsymbol{\theta}) &= \mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})] \otimes (\mathbf{z}_t - \boldsymbol{\mu}_z), & (19)\\
\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{m}}_{asT}(\hat{\boldsymbol{\theta}}_T)] &= V\{\mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\}, &\\
\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{m}}_{adT}(\hat{\boldsymbol{\theta}}_T)] &= V\{\mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]\} \otimes \boldsymbol{\Sigma}_{zz}, & (20)
\end{aligned}$$

which converge in distribution to two independent chi-square random variables whose degrees of freedom are $\binom{N+2}{3}$ and $\binom{N+2}{3}m$, respectively.

In practice, we must simply replace $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}_{zz}$ by their sample counterparts to evaluate (16) and consistently estimate (20).

It is easy to see that (16) assesses the asymmetry of the unconditional distribution of the regression residuals $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\alpha} - \mathbf{B}_z\mathbf{z}_t$, where $\boldsymbol{\alpha}$ and $\mathbf{B}_z$ are the regression intercepts and slopes, respectively. Not surprisingly, it coincides with the skewness component of the aforementioned multivariate normality test in Amengual, Fiorentini and Sentana (2021) applied to $\boldsymbol{\varepsilon}_t$. In turn, (17) can be regarded as a test against pure conditional "heterocliticity". If we re-write the multivariate regression model (1) in deviation from the means form as

$$\mathbf{y}_t = \boldsymbol{\mu}_{\mathbf{y}} + \mathbf{B}_z(\mathbf{z}_t - \boldsymbol{\mu}_{\mathbf{z}}) + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^*,$$

then (16) is simply testing for dependence between random coefficient variation in the unconditional mean of the regressands $\boldsymbol{\mu}_{\mathbf{y}}$ and the covariance matrix of the residuals $\boldsymbol{\Omega}$, while (17) checks the dependence between the elements of this covariance matrix and the slope coefficients $\mathbf{B}_z$.

Despite appearances, an analogous decomposition is not available for the test statistic (11) because the first-order conditions of the estimators of $\hat{\boldsymbol{\theta}}_T$ ensure that the sample mean of $\mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)]$ is 0 when the regression contains an intercept, as explained after Proposition 1. As a result, the conditional homoskedasticity test will have no power to detect time-variation in the constant terms of the multivariate regression which is uncorrelated to the variation in any other of the model parameters.

We can further exploit the additive decompositions in Propositions 1 and 2 to obtain additional tests. For example, the sum of (11) and (17) may be interpreted as a test of random coefficient variation and covariation in the slope coefficients $\mathbf{B}_z$. Similarly, the sum of (16) and (13) coincides with the multivariate Hermite-based normality test in Amengual, Fiorentini and Sentana (2021) applied to the regression residuals $\boldsymbol{\varepsilon}_t$.

In the next section, we study in detail VARs, which are such that $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$. In particular, we derive theoretical expressions for $V[vech(\mathbf{x}_t\mathbf{x}'_t)]$ and $E(\mathbf{x}_t\mathbf{x}'_t)$ that can be evaluated at the Gaussian maximum likelihood estimators as an alternative way of implementing our tests under the maintained assumption of covariance stationarity of $\mathbf{y}_t$.

3 Testing parameter constancy in vector autoregressions

Let us now focus on the important special case of the multivariate regression model (1) given by the following N-variate VAR(p) process with drift:

$$\mathbf{y}_t = \boldsymbol{\tau} + \textstyle\sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\Omega}^{1/2}\boldsymbol{\varepsilon}_t^* \text{ with } \boldsymbol{\varepsilon}_t^*|I_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Omega}). \tag{21}$$

In the next subsections we shall look at the additive components of the moment tests in Proposition 1 and 2, providing a simple regression interpretation for each of them. Like in the case of standard LM tests, these interpretations may prove particularly useful for the purposes of indicating in which specific directions our modelling efforts to enrich the Gaussian VAR model in (21) should focus.[4]

3.1 Interpretation of the influence functions

3.1.1 Conditional heteroskedasticity

Consider the multivariate regression of $\mathbf{H}_2(\boldsymbol{\varepsilon}_t^*)$ onto $1, y_{1t-1}, ..., y_{Nt-p}, y_{1t-1}^2, y_{1t-1}y_{2t-1}, ..., y_{Nt-p}^2$. Given that (8) effectively contains the normal equations of this regression evaluated under

[4] In Appendix B we illustrate the results in this section with a simple AR(1) model, which provides simpler and more intuitive expressions.

the null, it is straightforward to see that the test statistic (11) numerically coincides with the LM test of zero slopes in the aforementioned auxiliary regression (see Hall (1987) for an analogous result in the univariate case). As a consequence, if (21) generates a covariance stationary process, then the quadratic form in (11) will be asymptotically distributed as a chi-square random variable with $\left[N + \binom{N+1}{2}\right] p\binom{N+1}{2}$ degrees of freedom. Importantly, the regression intercepts do not add any degrees of freedom because $\mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\hat{\boldsymbol{\theta}}_T)] = \mathbf{0}$ from the first order conditions for $\boldsymbol{\omega}$, so they are only included to purge the remaining normal equations from the effects of sampling uncertainty resulting from the estimation of the covariance parameters. On the other hand, rejections of the null hypothesis clearly suggest the need for models with time-variation in the autoregressive coefficients, which in turn give rise to either Arch- or Qarch-type conditional heteroskedasticity depending on whether or not they are correlated with the intercepts, as explained by Hall (1987) and Bera and Lee (1993) in the univariate case, and Sentana (1995) in the multivariate one.

3.1.2 Conditional and unconditional asymmetry

Consider now the multivariate regression of $\mathbf{H}_3(\boldsymbol{\varepsilon}_t^*)$ onto $1, y_{1t-1}, ..., y_{Nt-p}$. Given that (9) effectively contains the normal equations of this auxiliary multivariate regression evaluated under the null, it is again easy to notice that the test statistic (12) numerically coincides with the LM test of zero intercepts and slopes in the aforementioned auxiliary regression. Therefore, if (21) generates a covariance stationary process, then the quadratic form in (12) will be asymptotically distributed as a chi-square random variable with $\binom{N+2}{3}(1+Np)$ degrees of freedom. Unlike in the previous subsection, though, the intercepts provide additional degrees of freedom in this case. On the other hand, rejections of the null hypothesis clearly indicate models with correlation between the random coefficients in the conditional mean vector and those in the covariance matrix, which in turn generate what Bera and Lee (1993) called heterocliticity in the univariate case.

Consider now the closely related auxiliary multivariate regressions of $\mathbf{H}_3(\boldsymbol{\varepsilon}_t^*)$ on a constant on the one hand, and the demeaned values of $y_{1t-1}, ..., y_{Nt-p}$ on the other. Given that (18) and (19) effectively provide the normal equations of these two regressions evaluated under the null, it is straightforward to see that (16) and (17) numerically coincide with the LM test of zero means and zero slopes, respectively, in these auxiliary regressions. In this respect, (16) converges in distribution to a chi-square random variable with $\binom{N+2}{3}$ degrees of freedom, while (17) will converge to an independent chi-square with $\binom{N+2}{3}Np$ degrees of freedom when (21) generates a covariance stationary process.

3.1.3 Unconditional kurtosis

Finally, consider the multivariate regression of $\mathbf{H}_4(\boldsymbol{\varepsilon}_t^*)$ on a constant. Given that (10) effectively contains the normal equations of this regression evaluated under the null, it is once more straightforward to prove that the quadratic form (13) numerically coincides with the LM test of zero intercepts in this auxiliary regression. Therefore, this test statistic will be asymptotically distributed as a chi-square random variable with $\binom{N+3}{4}$ degrees of freedom under the null. On the other hand, rejections of the null hypothesis clearly suggest the need for models with time-variation in the parameters of the residual covariance matrix, which yield excess kurtosis, as explained by Hall (1987) in the univariate case.

3.2 Covariance matrices of the influence functions

We have already explained in section 2 that under standard regularity conditions we can consistently estimate (14) and (15) by simply replacing $V\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]$ and $E(\mathbf{x}_t\mathbf{x}_t')$ by their sample counterparts. The purpose of this section is to explain how to fully exploit the structure of (21) to compute those expressions. Let us start with the VAR(1) case. We can then prove the following result:

Proposition 3 *Assume $\mathbf{y}_t$ follows a covariance stationary Gaussian* VAR(1) *with $E(\mathbf{y}_t) = \boldsymbol{\mu} \equiv (\mathbf{I}_N - \mathbf{A})^{-1}\boldsymbol{\tau}$ and $vec[V(\mathbf{y}_t)] = vec(\boldsymbol{\Upsilon}) \equiv (\mathbf{I}_{N^2} - \mathbf{A}\otimes\mathbf{A})^{-1}vec(\boldsymbol{\Omega})$. Then, the relevant population quantities required to obtain (14) and (15) are*

$$E(\mathbf{y}_{t-1}\mathbf{y}_{t-1}') = \boldsymbol{\Upsilon} + \boldsymbol{\mu}\boldsymbol{\mu}', \tag{22}$$

$$E[(\mathbf{y}_{t-1}\otimes\mathbf{y}_{t-1})\mathbf{y}_{t-1}'] = \boldsymbol{\mu}\otimes\boldsymbol{\Upsilon} + vec(\boldsymbol{\Upsilon})\boldsymbol{\mu}' + (\boldsymbol{\mu}\otimes\boldsymbol{\mu})\boldsymbol{\mu}' + (\boldsymbol{\mu}\otimes\mathbf{I}_N)\boldsymbol{\Upsilon} \tag{23}$$

and

$$\begin{aligned} E[(\mathbf{y}_{t-1}\otimes\mathbf{y}_{t-1})(\mathbf{y}_{t-1}'\otimes\mathbf{y}_{t-1}')] \;=\; & (\mathbf{I}_{N^2} + \mathbf{K}_N)\left(\boldsymbol{\Upsilon}\otimes\boldsymbol{\Upsilon} + \boldsymbol{\Upsilon}\otimes\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}'\otimes\boldsymbol{\Upsilon}\right) \\ & + vec(\boldsymbol{\Upsilon})(\boldsymbol{\mu}\otimes\boldsymbol{\mu})' + (\boldsymbol{\mu}\otimes\boldsymbol{\mu})vec'(\boldsymbol{\Upsilon}) \\ & + vec(\boldsymbol{\Upsilon})vec'(\boldsymbol{\Upsilon}) + (\boldsymbol{\mu}\otimes\boldsymbol{\mu})(\boldsymbol{\mu}\otimes\boldsymbol{\mu})'. \end{aligned} \tag{24}$$

As is well known, we can always write the N-variate VAR(p) model for $\mathbf{y}_t$ in (21) in companion form as the following VAR(1) for an augmented vector process of dimension pN:

$$\mathbf{Y}_t = \boldsymbol{\nu} + \boldsymbol{\Phi}\mathbf{Y}_{t-1} + \mathbf{U}_t$$

with $\mathbf{Y}_t' = (\mathbf{y}_t', \ldots, \mathbf{y}_{t-p+1}')$,

$$\boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\tau} \\ \mathbf{0}_{N(p-1)} \end{pmatrix}, \; \boldsymbol{\Phi} = \begin{pmatrix} \mathbf{A}_1, \; \mathbf{A}_2 & \cdots \; \mathbf{A}_p \\ \mathbf{I}_{N(p-1)} & \mathbf{0}_{N(p-1)\times N} \end{pmatrix} \text{ and } V(\mathbf{U}_t|I_{t-1}) = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Therefore, we can easily obtain the VAR(p) analogues to (22), (23) and (24) by selecting the relevant elements of those expressions with

$$E(\mathbf{Y}_{t-1}) \equiv \boldsymbol{\zeta} = [\mathbf{I}_{2N} - \boldsymbol{\Phi}]^{-1} \begin{pmatrix} \boldsymbol{\tau} \\ \mathbf{0} \end{pmatrix}$$

and

$$vec[V(\mathbf{Y}_{t-1})] \equiv \boldsymbol{\Psi} = [\mathbf{I}_{4N^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}]^{-1} vec \left[\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right].$$

3.3 A recursive-design bootstrap procedure

The theoretical results in Beran (1988) imply that if the usual Gaussian asymptotic approximation provides a reliable guide to the finite sample distribution of the sample version of the moments being tested, then the bootstrapped critical values should not only be valid, but also their errors should be of a lower order of magnitude under additional regularity conditions that guarantee the validity of a higher-order Edgeworth expansion. For that reason, we also analyse the performance of applying the bootstrap to the testing procedures we have described in sections 2 and 3.

Specifically, given an observed sample on $\mathbf{y}_t$ for $t = 1-p, \ldots, 0, 1, \ldots, T$, we can follow Bose (1988)'s recursive-design procedure and generate N_{boot} bootstrap samples by simulating in each of them T independent draws of ϵ_t^* $(t = 1, \ldots, T)$ from an N-variate spherical normal distribution, which we then use to construct

$$\tilde{\mathbf{y}}_s = \hat{\boldsymbol{\tau}}_T + \hat{\mathbf{A}}_{1T}\tilde{\mathbf{y}}_{s-1} + \cdots + \hat{\mathbf{A}}_{pT}\tilde{\mathbf{y}}_{s-p} + \hat{\boldsymbol{\Omega}}_T^{\frac{1}{2}}\boldsymbol{\epsilon}_{is}^*, \;\; s = 1, \ldots T,$$

with $\tilde{\mathbf{y}}_s = \mathbf{y}_s$ for $s = 1-p, \ldots, 0$.[5]

4 Monte Carlo analysis

In this section, we assess the finite sample size and power of the testing procedures discussed in section 3 by means of an extensive Monte Carlo analysis. Given that standard bootstrap procedures within such a simulation exercise are extremely time consuming, we focus on bivariate and trivariate models with either one or two lags. Nevertheless, our results provide relevant insights that can be extrapolated to more general designs.

[5]Under the assumption of covariance stationary of (21), in principle we could use the estimated values of $\hat{\boldsymbol{\tau}}_T$,$\hat{\mathbf{A}}_{1T}$,..., $\hat{\mathbf{A}}_{pT}$ and $\hat{\boldsymbol{\Omega}}_T$ to simulate the initial conditions $\tilde{\mathbf{y}}_s$ for $s = 1-p, \ldots, 0$ from their joint stationary distribution, which might work better for short sample periods. Given the sample sizes involved, and the fact that we use OLS to estimate the model parameters, we do not explore this possibility in section 4.

4.1 Design

We generate samples of size T from various special cases of the following general time-varying coefficients VAR process

$$\mathbf{y}_t = \boldsymbol{\tau}_t + \mathbf{A}_{1t}\mathbf{y}_{t-1} + \mathbf{A}_{2t}\mathbf{y}_{t-2} + \boldsymbol{\Omega}_t^{\frac{1}{2}}\boldsymbol{\varepsilon}_t^*, \tag{25}$$

where $\boldsymbol{\varepsilon}_t^*$ is drawn from a serially independent multivariate standard normal, $\boldsymbol{\Omega}_t = \mathbf{J}_t\boldsymbol{\Psi}_t\mathbf{J}_t'$, $\boldsymbol{\Psi}_t$ is a diagonal matrix whose elements contain the scale of the shocks, and $\mathbf{J}_t$ is a unit lower triangular matrix, so that in effect we introduce time-variation in the covariance matrix of the residuals through the LDL version of its Cholesky decomposition, as in Primiceri (2005). To simplify the calibration of the different data generating processes (DGPs), we also assume that the true $\mathbf{A}_{jt}$'s are lower triangular, although we do not impose this restriction in estimation.

For those designs that satisfy the null hypothesis of correct specification, we consider both $T = 250$, which is a realistic sample size in most macro applications with monthly or quarterly data, and $T = 1,000$, which is representative of financial applications with daily data. To avoid too many rejection rates close to 1, though, we focus on samples of length $T = 250$ to study power. The precise DGPs that we consider under the alternative hypothesis are described in section 4.1.2.

To gauge the finite sample size and power of our proposed independence tests, we generate $10,000$ samples for the DGPs that satisfy the null hypothesis and $2,500$ for the rest. Finally, we also compute non-asymptotic critical values by implementing the recursive-design bootstrap procedure described in section 3.3 with 999 samples.

4.1.1 DGPs and estimation under the null

For the bivariate processes, we generate samples of size T from (25) with

$$\mathbf{A}_{1t} = \mathbf{A}_1 = \begin{pmatrix} 1/2 & 0 \\ 1/4 & 1/3 \end{pmatrix}, \ \boldsymbol{\Omega}_t = \boldsymbol{\Omega} = \mathbf{I}_2 \ \forall t, \tag{26}$$

and $\mathbf{A}_{2t}$ either $\mathbf{0}$ or $\frac{1}{2}\mathbf{A}_1$ depending on the number of lags we consider. Similarly, for the trivariate ones we consider

$$\mathbf{A}_{1t} = \mathbf{A}_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/3 & 0 \\ 1/6 & 1/8 & 1/4 \end{pmatrix}, \ \boldsymbol{\Omega}_t = \boldsymbol{\Omega} = \mathbf{I}_3 \ \forall t, \tag{27}$$

and again $\mathbf{A}_{2t}$ either $\mathbf{0}$ or $\frac{1}{2}\mathbf{A}_1$ depending on the specification.

4.1.2 DGPs under the alternatives

As mentioned in the introduction, we study the power of our testing procedures against three types of random coefficient variation in VAR(1) models: *i.i.d.* coefficients, persistent but stationary coefficients, and recurring regime switching models. For the first two types of alternative hypotheses, it is convenient to introduce some additional notation.

Let $\boldsymbol{\theta}_t = [\boldsymbol{\tau}', vech'(\mathbf{A}_t), vecd'(\boldsymbol{\Psi}_t), vecl'(\mathbf{J}_t)]' = (\boldsymbol{\tau}_t', \mathbf{a}_t', \mathbf{j}_t', \boldsymbol{\psi}_t')$ denote the parameters characterising the stochastic process for $\mathbf{y}_t$ in (25). We generate variation in $\boldsymbol{\tau}_t$, $\mathbf{a}_t$, $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ as follows:

$$\begin{pmatrix} \boldsymbol{\tau}_t \\ \mathbf{a}_t \\ \boldsymbol{\psi}_t \\ \mathbf{j}_t \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{\tau}} \\ \bar{\mathbf{a}} \\ \bar{\boldsymbol{\psi}} \\ \bar{\mathbf{j}} \end{pmatrix} (1-\rho) + \rho \begin{pmatrix} \boldsymbol{\tau}_{t-1} \\ \mathbf{a}_{t-1} \\ \boldsymbol{\psi}_{t-1} \\ \mathbf{j}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{\tau}t} \\ \boldsymbol{\eta}_{\mathbf{a}t} \\ \boldsymbol{\eta}_{\boldsymbol{\psi}t} \\ \boldsymbol{\eta}_{\mathbf{j}t} \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{\tau}t} \\ \boldsymbol{\eta}_{\mathbf{a}t} \\ \boldsymbol{\eta}_{\boldsymbol{\psi}t} \\ \boldsymbol{\eta}_{\mathbf{j}t} \end{pmatrix} \sim iid\ N(\mathbf{0}, \boldsymbol{\Lambda}),$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{\boldsymbol{\tau}}\mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda_{\mathbf{a}}\mathbf{I}_{N(N+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \lambda_{\boldsymbol{\psi}}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \lambda_{\mathbf{j}}\mathbf{I}_{N(N-1)/2} \end{pmatrix} + \ell_{2N+N^2}\ell_{2N+N^2}'\lambda_{cov},$$

where ℓ_K is a $K \times 1$ vector of ones, so that the hyperparameters are $\boldsymbol{\upsilon} = (\bar{\mathbf{a}}', \bar{\boldsymbol{\tau}}', \bar{\mathbf{j}}', \bar{\boldsymbol{\psi}}', \boldsymbol{\lambda}', \rho)'$, with $\boldsymbol{\lambda} = (\lambda_{\boldsymbol{\tau}}, \lambda_{\mathbf{a}}, \lambda_{\boldsymbol{\psi}}, \lambda_{\mathbf{j}}, \lambda_{cov})'$.

In this context, we consider:

DGP 1: *i.i.d. time variation for the parameters.*

(a) Time variation in $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ only, with $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ constant for all t. In particular, we set $\lambda_{\boldsymbol{\tau}} = .105$, $\lambda_{\mathbf{a}} = .026$ and $\lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = \rho = 0$.

(b) Time variation in $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ only, with $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ constant for all t. Specifically, we set $\lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = .105$ and $\lambda_{\boldsymbol{\tau}} = \lambda_{\mathbf{a}} = \rho = 0$.

(c) Time variation in all parameters. In this case, we set $\lambda_{\boldsymbol{\tau}} = \lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = .105$, $\lambda_{\mathbf{a}} = .026$, $\lambda_{cov} = -.001$ and $\rho = 0$.

DGP 2: AR(1) *dynamics for the model parameters.*

(a) Time variation in $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ only, with $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ constant for all t. In particular, we set $\lambda_{\boldsymbol{\tau}} = .02$, $\lambda_{\mathbf{a}} = .005$, $\lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = 0$ and $\rho = .9$.

(b) Time variation in $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ only, with $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ constant for all t. Specifically, we set $\lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = .02$, $\lambda_{\boldsymbol{\tau}} = \lambda_{\mathbf{a}} = 0$ and $\rho = .9$.

(c) Time variation in all parameters. In this case, we set $\lambda_{\boldsymbol{\tau}} = \lambda_{\boldsymbol{\psi}} = \lambda_{\mathbf{j}} = .02$, $\lambda_{\mathbf{a}} = .005$, $\lambda_{cov} = -.001$ and $\rho = .9$.

Importantly, we have calibrated the designs (a), (b) and (c) in DGP 2 so that the unconditional variances of the VAR coefficients are the same as in the corresponding designs of DGP 1.

Finally, we model persistent but recurrent time variation in the model parameters through the following:

DGP 3: *Regime-switching model for the parameters.*

We assume that there exists a latent variable $s_t \in \{0,1\}$ that follows a (homogeneous and irreducible) first-order Markov chain whose temporal evolution is fully characterised by $\Pr(s_t = 0|s_{t-1} = 0) = p_{00}$ and $\Pr(s_t = 1|s_{t-1} = 1) = p_{11}$, with the usual constraints on p_{00} and p_{11}. This state variable drives the time variation in the parameters of the model as

$$\boldsymbol{\tau}_t = \begin{cases} \boldsymbol{\tau}_0 \text{ if } s_t = 0 \\ \boldsymbol{\tau}_1 \text{ if } s_t = 1 \end{cases} \quad \mathbf{a}_t = \begin{cases} \mathbf{a}_0 \text{ if } s_t = 0 \\ \mathbf{a}_1 \text{ if } s_t = 1 \end{cases} \quad \boldsymbol{\psi}_t = \begin{cases} \boldsymbol{\psi}_0 \text{ if } s_t = 0 \\ \boldsymbol{\psi}_1 \text{ if } s_t = 1 \end{cases} \quad \text{and } \mathbf{j}_t = \begin{cases} \mathbf{j}_0 \text{ if } s_t = 0 \\ \mathbf{j}_1 \text{ if } s_t = 1 \end{cases}.$$

As in the previous two designs, we consider the following three possibilities:

(a) Time variation in $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ only, with $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ constant for all t, so that $\boldsymbol{\upsilon} = (\boldsymbol{\tau}_0', \boldsymbol{\tau}_1', \mathbf{a}_0', \mathbf{a}_1', \bar{\boldsymbol{\psi}}', \bar{\mathbf{j}}', p_{00}, p_{11})'$.

(b) Time variation in $\mathbf{j}_t$ and $\boldsymbol{\psi}_t$ only, with $\boldsymbol{\tau}_t$ and $\mathbf{a}_t$ constant for all t, so that $\boldsymbol{\upsilon} = (\bar{\boldsymbol{\tau}}', \bar{\mathbf{a}}', \boldsymbol{\psi}_0', \boldsymbol{\psi}_1', \mathbf{j}_0', \mathbf{j}_1', p_{00}, p_{11})'$.

(c) Time variation in all parameters, which is the combination of the previous two cases, so that $\boldsymbol{\upsilon} = (\boldsymbol{\tau}_0', \boldsymbol{\tau}_1', \mathbf{a}_0', \mathbf{a}_1', \boldsymbol{\psi}_0', \boldsymbol{\psi}_1', \mathbf{j}_0', \mathbf{j}_1', p_{00}, p_{11})'$.

We set $p_{00} = p_{11} = .95$ to generate the same autocorrelation for the random coefficients as in DGP 2. Moreover, we set the remaining hyperparameters so that the unconditional variances of the VAR coefficients in designs (a), (b) and (c) coincide once again with the corresponding ones in DGP 1. In particular, we choose

$$\bar{\boldsymbol{\tau}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \boldsymbol{\tau}_0 = \begin{pmatrix} 1.3245 \\ -1.3245 \end{pmatrix}, \boldsymbol{\tau}_1 = \begin{pmatrix} .6755 \\ -.6755 \end{pmatrix},$$

$$\bar{\mathbf{a}} = \begin{pmatrix} 1/2 \\ 1/4 \\ 1/3 \end{pmatrix}, \mathbf{a}_0 = \begin{pmatrix} .662 \\ .412 \\ .496 \end{pmatrix}, \mathbf{a}_1 = \begin{pmatrix} .338 \\ .088 \\ .171 \end{pmatrix},$$

$\bar{\boldsymbol{\psi}}' = \ell_2$, $\boldsymbol{\psi}_0 = 1.325\ell_2$, $\boldsymbol{\psi}_1 = .675\ell_2$, $\bar{\mathbf{j}}' = 0$, $\mathbf{j}_0 = .325$ and $\mathbf{j}_1 = -.325$.

4.2 Simulation results

4.2.1 Size properties

Table 1 reports Monte Carlo rejection rates under the null of the information matrix tests proposed in section 3 for VAR models of orders one and two. Panels A and B present results for sample sizes of 250 and 1,000, respectively. In turn, left panels rely on asymptotic critical values while right panels on bootstrap-based ones. As can be seen, asymptotic critical values do not seem to be reliable for samples of size $T = 250$, while the bootstrap offers improvements in the right direction. Interestingly, all the models we consider do well in terms of bootstrap-based size figures even in samples of $T = 250$, despite the VAR(2) tests involving a significantly larger number of degrees of freedom.[6] On the other hand, rejection rates based on asymptotic critical values become quite reliable for moderately large samples ($T = 1,000$), with the possible exception of the tests at the 1% level, which are somewhat oversized. Once again, the bootstrap tends to correct most of the remaining size distortions in those cases.

Interestingly, the comparison of the rejection rates of the tests that use the sample counterpart of the second and fourth unconditional moments of $\mathbf{y}_t$ with those that rely on the theoretical expression in Proposition 3 shows no clear winner. It depends on whether one compares the conditional heteroskedasticity tests or the conditional asymmetry ones, as well as on the significance level chosen.

In turn, Table 2 reports the corresponding figures for trivariate VAR models. As can be observed, the same qualitative results apply. In particular, the comparison of asymptotic with bootstrapped critical values leads to analogous conclusions. Similarly, the finite sample size of the tests does not noticeably deteriorate when we move from a specification with one lag to another with two. These findings are remarkable because the number of degrees of freedom for the different chi-square limiting distributions under the null is substantially larger in trivariate models than in bivariate ones, as can be seen from the second columns of Tables 1 and 2.

4.2.2 Power of the tests

Next, we assess the power of our tests in samples of $T = 250$ in Tables 3 (bivariate) and 4 (trivariate). Panels A in those tables clearly indicate that the power of our tests against DGP 1 is quite good, which is not very surprising given Chesher's (1984) re-interpretation of the information matrix test as an LM test against this type of parameter variation.

Nevertheless, our tests also have substantial power to detect both persistent but stationary AR coefficients, as can be seen from panels B of Tables 3 and 4, and regime switching parameters,

[6] Given the number of Monte Carlo replications, the 95% asymptotic confidence intervals for the rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels, respectively.

as shown in panels C. In fact, in several instances the power to detect these other types of time-variation in the model parameters is larger than in the *i.i.d.* case despite the different DGPs generating the same unconditional variances for the coefficients.

Finally, a comparison of Tables 3 and 4 indicates that the same conclusions hold for the two cross-sectional dimensions that we consider.

5 An application to aggregate output measures

In theory, the expenditure (GDE) and income (GDI) measures of output should be equal, but in practice they differ because they are calculated from different sources. Traditionally, the difference between them, officially known as the "statistical discrepancy" (see Grimm (2007)), was regarded by many academic economists as a curiosity in the US National Input and Product Accounts (NIPA) elaborated by the Bureau of Economic Analysis (BEA) of the Department of Commerce. However, the Great Recession substantially renewed interest in this discrepancy (see e.g. Nalewaik (2010, 2011)), and in fact nowadays the BEA routinely reports an equally weighted average of the growth rates of GDE and GDI as an improved measurement.

In this section, we first fit a VAR to a suitable stationary transformation of these two aggregate output measures, and then study the constancy of its coefficients by means of the tests we have developed in section 3. In this respect, we follow Almuzara, Amengual and Sentana (2019) and Almuzara, Fiorentini and Sentana (2021) in imposing cointegration (in logs) between GDE and GDI, with cointegrating vector (-1,1). Given the non-stationary nature of those two output measures, the absence of cointegration between them would imply an implausible diverging statistical discrepancy.

Specifically, we consider the following dynamic bivariate model:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\tau} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t | I_{t-1} &\sim i.i.d.\ N(\mathbf{0}, \boldsymbol{\Omega}), \end{aligned} \tag{28}$$

where y_{1t} is the difference between the logs of GDE and GDI, while y_{2t} is the (arithmetic) average of the (geometric) quarterly rates of growth of these two output measures.

Following Almuzara, Amengual and Sentana (2019), our sample period starts in 1952Q1, soon after the Treasury–Federal Reserve Accord whereby the Fed stopped its wartime pegging of interest rates, and finishes at 2015Q2, for a total of 254 observations. This sample contains the so-called Great Moderation, a longer period prior to it which includes the turbulences in the late 1970s and early 1980s, as well as the Great Recession. Figure 1a displays the temporal evolution of the statistical discrepancy, which shows a persistent but stationary pattern whose movements are unrelated to the business cycle. In turn, Figure 1b reports the average of the growth rates of GDE and GDI, which is also noticeably stationary but serially correlated.

To determine the number of lags p in (28), we initially rely on the BIC-Schwarz criterion, which selects a single lag. Nevertheless, when inspecting the residuals of the estimated $\text{VAR}(1)$, we find evidence of residual autocorrelation, which we formally assess through a Lagrange multiplier test of multivariate white noise against a $\text{VAR}(1)$ for $\boldsymbol{\varepsilon}_t$ (see Fiorentini and Sentana (2015)). In contrast, the analogous test applied to the residuals of a $\text{VAR}(2)$ fails to reject at the usual 5% level, although it has a relatively small p–value, and the same is true of the likelihood ratio test of $\text{VAR}(2)$ versus $\text{VAR}(3)$.

In view of this mixed evidence, in Panel A of Table 5 we report the OLS estimates and their standard errors for $p = 1$ and $p = 2$. Interestingly, the estimated residuals of the two models are quite similar, as can be seen from Figure 2. Indeed, the correlation of the estimated residuals of the statistical discrepancy for the specifications with one and two lags is .982 while the corresponding figure for the average GDE-GDI growth is .988.

Panel B of Table 5 contains the results of applying our information matrix tests to assess the constancy of the mean and variance parameters of the two VAR models. As can be seen, the null hypotheses of conditional homoskedasticity and mesokurtosis are strongly rejected for both specifications. However, the null hypotheses of conditional and unconditional symmetry are not rejected. Therefore, our proposed information matrix tests indicate time-variation in both the autoregressive coefficients and the residual covariance matrix of the innovations, but they fail to detect any covariation between those two groups of coefficients.

To inspect whether there are specific periods leading to the rejection, in Figure 3 we report the contribution of observation t to each of the test statistics for the $\text{VAR}(2)$ specification. Specifically, we plot the fourth root of $T \cdot \mathbf{m}'_{lt}(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{lt}(\hat{\boldsymbol{\theta}})]\mathbf{m}_{lt}(\hat{\boldsymbol{\theta}}_T)$ for $l = h, a, k$ to "normalise" these quadratic forms, a transformation proposed by Hawkins and Wixley (1986) for chi-square distributed variables. Those plots suggest that the oil crises in the seventies and the recession of the early 2000s are events that generate high values for those quantities.

6 Conclusions and directions for further research

We propose the information matrix test to assess the constancy of mean and variance parameters in vector autoregressions. We additively decompose this test into four easily interpretable orthogonal components which check conditional heteroskedasticity and asymmetry of the innovations, as well as their unconditional skewness and kurtosis. We also conduct extensive Monte

Carlo simulations to study the finite size and power properties of these tests against *i.i.d.* coefficients, persistent but stationary ones, and regime switching. Finally, our procedures detect time-variation in the autoregressive coefficients and residual covariance matrix of a VAR for the US GDP growth rate and the statistical discrepancy, but they fail to detect any covariation between those two sets of coefficients.

If the autoregressive polynomial $(\mathbf{I}_N - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)$ had some unit roots, $\mathbf{y}_t$ would be a (co-) integrated process, and the estimators of the conditional mean parameters would have non-standard asymptotic distributions, as some of them would converge at the faster rate T. In contrast, the distribution of the usual OLS estimators of the residual variance parameters would remain standard (see, e.g., Phillips and Durlauf (1986)). Presumably, the asymptotic distribution of the sample averages of the multivariate Hermite polynomials of the regression residuals evaluated at the maximum likelihood estimators would also remain standard. Therefore, the complication would arise because the second and fourth moments of the $\mathbf{y}_t$'s that appear in Propositions 1 and 2 would diverge. The study of the properties of our proposed tests in those circumstances constitutes an interesting avenue for research.

Macroeconomists may also be interested in assessing the constancy of the mean and variance parameters of model (21) without maintaining the Gaussianity of its shocks. While it makes no sense to robustify the kurtosis test (13) when the shocks are not normal, one could easily use robust versions of (11) and (12) under the maintained assumption of conditional independence of ε_t^* given $\mathbf{x}_t$. In practice, a researcher would simply need to replace the theoretical expressions for the covariance matrix of the corresponding vector of multivariate Hermite polynomials that appear in (14) and (15) by either their sample second moment matrix or their covariance matrix, both of which are consistent under the null. In fact, one could also consider more robust versions of these moment test statistics that do not exploit the Kronecker structure in (14) and (15). However, as Gonçalves and Killian (2004) forcefully argue for the case of the mean coefficients, the bootstrap procedure discussed in section 3.3 will not work in these more general contexts, so extending it provides another promising research avenue.

References

Akharif, A. and Hallin, M. (2003): "Efficient detection of random coefficients in autoregressive models", *Annals of Statistics* 31, 675–704.

Almuzara, M., Amengual, D. and Sentana, E. (2019): "Normality tests for latent variables", *Quantitative Economics* 10, 981–1017.

Almuzara, M., Fiorentini, G. and Sentana, E. (2021): "Aggregate output measurements: a common trend approach", CEMFI Working Paper 2101.

Amengual, D., Fiorentini, G. and Sentana, E. (2021): "Multivariate Hermite polynomials and information matrix tests", CEMFI Working Paper 2103.

Barndorff-Nielsen, O. and Petersen, B.V. (1979): "The bivariate Hermite polynomials up to order six", *Scandinavian Journal of Statistics* 6, 127–128.

Bera, A. and Lee, S (1993): "Information matrix test, parameter heterogeneity and Arch: a synthesis", *Review of Economic Studies* 60, 229–240.

Beran, R. (1988): "Prepivoting test statistics: a bootstrap view of asymptotic refinements", *Journal of the American Statistical Association* 83, 687–697.

Bose, A. (1988): "Edgeworth correction by bootstrap in autoregression", *Annals of Statistics* 16, 1709–1722.

Canova, F. (2007): *Methods for applied macroeconomic research*, Princeton University Press.

Carrasco, M., Hu, L. and Ploberger, W. (2014): "Optimal test for Markov switching parameters", *Econometrica* 82, 765–784.

Chen, J., Wang, D., Li, C. and Huang, J. (2020): "Estimation and testing of multivariate random coefficient autoregressive model based on empirical likelihood", *Communications in Statistics - Simulation and Computation* https://doi.org/10.1080/03610918.2020.1855445

Chesher, A. (1984): "Testing for neglected heterogeneity", *Econometrica* 52, 865–872.

D'Agostino, A., Gambetti, L. and Giannone, D. (2013): "Macroeconomic forecasting and structural change", *Journal of Applied Econometrics* 28, 82–101.

Fiorentini, G. and Sentana, E. (2021): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved components and time series econometrics*, 118–189, Oxford University Press.

Fiorentini, G. and Sentana, E. (2021): "Specification tests for non-Gaussian maximum likelihood estimators", *Quantitative Economics* 12, 683–742.

Gonçalves, S. and Killian, L. (2004): "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form", *Journal of Econometrics* 123, 89–120.

Grimm, B. T. (2007): "The statistical discrepancy", Bureau of Economic Analysis, Washington D.C.

Hall, A. (1987): "The information matrix test for the linear model", *Review of Economic Studies* 54, 257–263.

Hansen, B.E. (2001): "The new econometrics of structural change: dating breaks in U.S. labour productivity", *Journal of Economic Perspectives* 15, 117–128.

Hawkins, D.M. and Wixley, R. A. J. (1986): "A note on the transformation of chi-squared variables to normality", *American Statistician* 40, 296–298.

Holmquist, B. (1996): "The d-variate vector Hermite polynomial of order k", *Linear Algebra and its Applications* 237/238, 155–190.

Horváth, L. and Trapani, L. (2019): "Testing for randomness in a random coefficient autoregression model", *Journal of Econometrics* 209, 338–352.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economic Letters* 6, 255–259.

Kiefer, N.M. and Salmon, M. (1983): "Testing normality in econometric models", *Economic Letters* 11, 123–127.

Magnus, J.R. and Neudecker, H. (1979): "The commutation matrix: some properties and applications", *Annals of Statistics* 7, 381–394.

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in statistics and econometrics*, 3rd edition, Wiley.

Lee, S. (1998): "Coefficient constancy test in a random coefficient autoregressive model", *Journal of Statistical Planning and Inference* 74, 93–101.

Nalewaik, J. (2010): "The income- and expenditure-side measures of output growth", *Brookings Papers on Economic Activity* 1, 71–106.

Nalewaik, J. (2011): "The income- and expenditure-side measures of output growth - An update through 2011Q2", *Brookings Papers on Economic Activity* 2, 385–402.

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047–70.

Nicholls, D.F. and Quinn, B.G. (1981): "Multiple autoregressive models with random coefficients", *Journal of Multivariate Analysis* 11, 185–198.

Nicholls, D.F. and Quinn, B.G. (1982): *Random coefficient autoregressive models: an introduction*, Springer.

Perron, P. (1989): "The great crash, the oil price shock and the unit root hypothesis", *Econometrica* 57, 1361–1401.

Phillips, P.C.B. and Durlauf, S.N. (1986): "Multiple time series regression with integrated processes", *Review of Economic Studies* 53, 473–495.

Primiceri. G.E. (2005): "Time varying structural vector autoregressions and monetary policy", *Review of Economic Studies* 72, 821–852.

Regis, M., Serra P. and van den Heuvel, E.R. (2021): "Random autoregressive models: a structured overview", *Econometric Reviews* https://doi.org/10.1080/07474938.2021.1899504

Sentana, E. (1995): "Quadratic ARCH models", *Review of Economic Studies* 62, 639–661.

Sims, C. (1980): "Macroeconomics and reality", *Econometrica* 48, 1–48.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415–443.

White, H. (1980): "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity", *Econometrica* 48, 817–838.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1–25.

White, H. (1994): *Estimation, inference and specification analysis*, Cambridge University Press.

Appendices

A Proofs

Proposition 1

Given that the random vectors $\mathbf{H}_k(\boldsymbol{\varepsilon}_t^*)$ for $k = 2,3,4$ are not only mean independent from the regressors $\mathbf{x}_t$ under the null, but also their conditional covariance matrices are constant, it is easy to see that

$$\begin{aligned} V[\mathbf{m}_{ht}(\boldsymbol{\theta})] &= V(\mathbf{H}_{2t}) \otimes E\left[vech(\mathbf{x}_t\mathbf{x}_t')\right], \\ V[\mathbf{m}_{at}(\boldsymbol{\theta})] &= V(\mathbf{H}_{3t}) \otimes E(\mathbf{x}_t\mathbf{x}_t') \end{aligned}$$

and

$$V[\mathbf{m}_{kt}(\boldsymbol{\theta})] = V(\mathbf{H}_{4t}),$$

where –to shorten the expressions– $\mathbf{H}_{kt}$ denotes $\mathbf{H}_k[\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})]$.

In practice, though, we must evaluate the influence functions at the Gaussian maximum likelihood estimators $\hat{\boldsymbol{\theta}}_T = (\hat{\mathbf{b}}_T', \hat{\boldsymbol{\omega}}_T')'$, so to correct for sampling uncertainty we need to find their residual covariance matrices after projecting them onto the linear span of the scores $\mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta})$ and $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta})$. In this respect, it is easy to see that we can re-write (2) and (3) as

$$\begin{aligned} \mathbf{s}_{\mathbf{b}t}(\boldsymbol{\theta}) &= (\mathbf{I}_M \otimes \boldsymbol{\Omega}^{-1/2\prime})(\mathbf{x}_t \otimes \mathbf{H}_{1t}), \\ \mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N\mathbf{H}_{2t}, \end{aligned}$$

which means that we only have to consider the projection of $\mathbf{m}_{ht}(\boldsymbol{\theta})$ onto the linear span of $\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta})$.

Given that

$$E\{[\mathbf{H}_{2t} \otimes vech(\mathbf{x}_t\mathbf{x}_t')]\mathbf{s}_{\boldsymbol{\omega}t}'(\boldsymbol{\theta})\} = \{V(\mathbf{H}_{2t}) \otimes E[vech(\mathbf{x}_t\mathbf{x}_t')]\}\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\mathbf{D}_N$$

and

$$V[\mathbf{s}_{\boldsymbol{\omega}t}(\boldsymbol{\theta})] = \frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N V(\mathbf{H}_{2t})\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\mathbf{D}_N,$$

the residual covariance matrix will be

$$
\begin{aligned}
& \{V(\mathbf{H}_{2t}) \otimes E\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]\} - [V(\mathbf{H}_{2t}) \otimes vech(\mathbf{x}_t\mathbf{x}_t')]\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\mathbf{D}_N \\
& \times \left[\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N V\{\mathbf{H}_{2t})\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\mathbf{D}_N\right]^{-1} \\
& \times\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N\{V(\mathbf{H}_{2t}) \otimes E[vech(\mathbf{x}_t\mathbf{x}_t')]\} \\
= \; & \{V(\mathbf{H}_{2t}) \otimes E\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]\} - [V(\mathbf{H}_{2t}) \otimes vech(\mathbf{x}_t\mathbf{x}_t')]\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\mathbf{D}_N \\
& \times 2\mathbf{D}_N^+(\boldsymbol{\Omega}^{1/2} \otimes \boldsymbol{\Omega}^{1/2})\mathbf{D}_N^{+\prime}V^{-1}(\mathbf{H}_{2t})2\mathbf{D}_N^+(\boldsymbol{\Omega}^{1/2\prime} \otimes \boldsymbol{\Omega}^{1/2\prime})\mathbf{D}_N^{+\prime} \\
& \times\frac{1}{2}\mathbf{D}_N'(\boldsymbol{\Omega}^{-1/2\prime} \otimes \boldsymbol{\Omega}^{-1/2\prime})\mathbf{D}_N\{V(\mathbf{H}_{2t}) \otimes E[vech(\mathbf{x}_t\mathbf{x}_t')]\} \\
= \; & \{V(\mathbf{H}_{2t}) \otimes E\left[vech(\mathbf{x}_t\mathbf{x}_t')vech'(\mathbf{x}_t\mathbf{x}_t')\right]\} \\
& -\{V(\mathbf{H}_{2t}) \otimes E[vech(\mathbf{x}_t\mathbf{x}_t')]\}V^{-1}(\mathbf{H}_{2t})\{V(\mathbf{H}_{2t}) \otimes E[vech'(\mathbf{x}_t\mathbf{x}_t')]\} \\
= \; & V(\mathbf{H}_{2t}) \otimes V\left[vech(\mathbf{x}_t\mathbf{x}_t')\right],
\end{aligned}
$$

where we have used Theorem 3.13 in Magnus and Neudecker (2019). This expression simply reflects the fact that the projection residual is effectively

$$\mathbf{H}_{2t} \otimes \{vech(\mathbf{x}_t\mathbf{x}_t') - E\left[vech(\mathbf{x}_t\mathbf{x}_t')\right]\}.$$

Given that multivariate Hermite polynomials of different orders are uncorrelated (see Holmquist (1996)), this adjusted influence function is orthogonal to both $\mathbf{m}_{at}(\boldsymbol{\theta})$ and $\mathbf{m}_{kt}(\boldsymbol{\theta})$, which in turn are orthogonal to each other. Therefore, the joint moment test will be sum of the three components. □

Proposition 2

Consider the unconditional moment conditions

$$E\left[\mathbf{H}_{3t} \otimes \begin{pmatrix} 1 \\ \mathbf{z}_t \end{pmatrix}\right] = \mathbf{0}, \tag{A1}$$

which result from the conditional moment restrictions $E(\mathbf{H}_{3t}|\mathbf{z}_t) = \mathbf{0}$.

Under the maintained assumption that $E(\mathbf{H}_{3t}\mathbf{H}_{3t}'|\mathbf{z}_t) = E(\mathbf{H}_{3t}\mathbf{H}_{3t}')$, we will have that

$$
\begin{aligned}
V\left[\mathbf{H}_{3t} \otimes \begin{pmatrix} 1 \\ \mathbf{z}_t \end{pmatrix}\right] &= E\left[\mathbf{H}_{3t}\mathbf{H}_{3t}' \otimes \begin{pmatrix} 1 & \mathbf{z}_t' \\ \mathbf{z}_t & \mathbf{z}_t\mathbf{z}_t' \end{pmatrix}\right] \\
= E\left[E(\mathbf{H}_{3t}\mathbf{H}_{3t}'|\mathbf{z}_t) \otimes \begin{pmatrix} 1 & \mathbf{z}_t' \\ \mathbf{z}_t & \mathbf{z}_t\mathbf{z}_t' \end{pmatrix}\right] &= \left[V(\mathbf{H}_{3t}) \otimes \begin{pmatrix} 1 & \boldsymbol{\mu}_z' \\ \boldsymbol{\mu}_z & \boldsymbol{\mu}_z\boldsymbol{\mu}_z' + \boldsymbol{\Sigma}_{zz} \end{pmatrix}\right].
\end{aligned}
$$

Therefore, the moment test that uses the asymptotic covariance matrix will be given by

$$
\begin{aligned}
&T\left(\frac{1}{T}\sum_{t=1}^{T}[\mathbf{H}_{3t}'\otimes\left(\begin{array}{cc}1 & \mathbf{z}_t'\end{array}\right)]\right)\left[V(\mathbf{H}_{3t})\otimes\left(\begin{array}{cc}1 & \boldsymbol{\mu}_z'\\ \boldsymbol{\mu}_z & \boldsymbol{\mu}_z\boldsymbol{\mu}_z'+\boldsymbol{\Sigma}_{zz}\end{array}\right)\right]^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{H}_{3t}\otimes\left(\begin{array}{c}1\\ \mathbf{z}_t\end{array}\right)\right]\right)\\
&=\frac{1}{T}\left(\sum_{t=1}^{T}[\mathbf{H}_{3t}'\otimes\left(\begin{array}{cc}1 & \mathbf{z}_t'\end{array}\right)]\right)\left[V^{-1}(\mathbf{H}_{3t})\otimes\left(\begin{array}{cc}1 & \boldsymbol{\mu}_z'\\ \boldsymbol{\mu}_z & \boldsymbol{\mu}_z\boldsymbol{\mu}_z'+\boldsymbol{\Sigma}_{zz}\end{array}\right)^{-1}\right]\left(\sum_{t=1}^{T}\left[\mathbf{H}_{3t}\otimes\left(\begin{array}{c}1\\ \mathbf{z}_t\end{array}\right)\right]\right)\\
&\qquad=T\left(\frac{1}{T}\sum_{t=1}^{T}[\mathbf{H}_{3t}'\otimes\left(\begin{array}{cc}1 & \mathbf{z}_t'\end{array}\right)]\right)\left[V^{-1}(\mathbf{H}_{3t})\otimes\left(\begin{array}{cc}1+\boldsymbol{\mu}_z'\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z & -\boldsymbol{\mu}_z'\boldsymbol{\Sigma}_{zz}^{-1}\\ -\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z & \boldsymbol{\Sigma}_{zz}^{-1}\end{array}\right)\right]\\
&\qquad\qquad\times\left(\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{H}_{3t}\otimes\left(\begin{array}{c}1\\ \mathbf{z}_t\end{array}\right)\right]\right). \qquad \text{(A2)}
\end{aligned}
$$

Let us now project $\mathbf{H}_{3t}\otimes\mathbf{z}_t$ onto the linear span of $\mathbf{H}_{3t}\otimes 1=\mathbf{H}_{3t}$. We end up with the following projection residuals

$$
\mathbf{H}_{3t}\otimes\mathbf{z}_t-[V(\mathbf{H}_{3t})\otimes\boldsymbol{\mu}_z]\left[V^{-1}(\mathbf{H}_{3t})\otimes 1\right](\mathbf{H}_{3t}\otimes 1)=\mathbf{H}_{3t}\otimes(\mathbf{z}_t-\boldsymbol{\mu}_z),
$$

whose asymptotic covariance matrix is

$$
V(\mathbf{H}_{3t})\otimes(\boldsymbol{\mu}_z\boldsymbol{\mu}_z'+\boldsymbol{\Sigma}_{zz})-[V(\mathbf{H}_{3t})\otimes\boldsymbol{\mu}_z]\left[V^{-1}(\mathbf{H}_{3t})\otimes 1\right]\left[V(\mathbf{H}_{3t})\otimes\boldsymbol{\mu}_z'\right]=V(\mathbf{H}_{3t})\otimes\boldsymbol{\Sigma}_{zz}.
$$

In addition, these projection residuals are orthogonal to $(\mathbf{H}_{3t}\otimes 1)$ by construction.

Hence, we can re-write the quadratic form (A2) as

$$
\begin{aligned}
&\frac{1}{T}\left[\sum_{t=1}^{T}\{\mathbf{H}_{3t}'\otimes\left[\begin{array}{cc}1 & \mathbf{z}_t-E(\mathbf{z}_t')\end{array}\right]\}\right]\left\{V^{-1}(\mathbf{H}_{3t})\otimes\left[\begin{array}{cc}1 & 0\\ 0 & V^{-1}(\mathbf{z}_t)\end{array}\right]\right\}\left[\sum_{t=1}^{T}\left\{\mathbf{H}_{3t}\otimes\left[\begin{array}{c}1\\ \mathbf{z}_t-E(\mathbf{z}_t)\end{array}\right]\right\}\right]\\
&\qquad=T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_{3t}'\right)V^{-1}(\mathbf{H}_{3t})\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}_{3t}\right)\\
&\qquad+T\left[\frac{1}{T}\sum_{t=1}^{T}\{\mathbf{H}_{3t}'\otimes[\mathbf{z}_t'-\boldsymbol{\mu}_z']\}\right]\left[V^{-1}(\mathbf{H}_{3t})\otimes\boldsymbol{\Sigma}_{zz}^{-1}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\{\mathbf{H}_{3t}\otimes[\mathbf{z}_t-\boldsymbol{\mu}_z)]\}\right].
\end{aligned}
$$

Effectively, what we are doing is to premultiply the original moment conditions (A1) by

$$
\left(\begin{array}{cc}1 & \mathbf{0}'\\ -\boldsymbol{\mu}_z & \mathbf{I}_N\end{array}\right)
$$

and adjust the covariance matrix accordingly. □

Proposition 3

Since $E(\mathbf{x}_{t-1})=\boldsymbol{\mu}\equiv(\mathbf{I}_N-\mathbf{A})^{-1}\boldsymbol{\tau}$ and $vec[V(\mathbf{y}_t)]=vec(\boldsymbol{\Upsilon})\equiv(\mathbf{I}_{N^2}-\mathbf{A}\otimes\mathbf{A})^{-1}vec(\boldsymbol{\Omega})$, we trivially have that $E(\mathbf{x}_{t-1}\mathbf{x}_{t-1}')=\boldsymbol{\Upsilon}+\boldsymbol{\mu}\boldsymbol{\mu}'$. Next, we can use Theorem 4.3 (i) and (i.v) in Magnus and Neudecker (1979) to compute $E[vech(\mathbf{x}_{t-1}\mathbf{x}_{t-1}')vech'(\mathbf{x}_{t-1}\mathbf{x}_{t-1}')]$. Thus, we can show that

$$V(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})=(\mathbf{I}_{N^2}+\mathbf{K}_N)\,(\boldsymbol{\Upsilon}\otimes\boldsymbol{\Upsilon}+\boldsymbol{\Upsilon}\otimes\boldsymbol{\mu}\boldsymbol{\mu}'+\boldsymbol{\mu}\boldsymbol{\mu}'\otimes\boldsymbol{\Upsilon}),$$

which together with

$$E(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})=vec(\boldsymbol{\Upsilon})+\boldsymbol{\mu}\otimes\boldsymbol{\mu}$$

and

$$E[(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})']=V(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})+E(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})E(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})'$$

imply the stated result.

Finally, we can combine the fact that

$$\begin{aligned}\mathbf{0} &= E\{[(\mathbf{x}_{t-1}-\boldsymbol{\mu})\otimes(\mathbf{x}_{t-1}-\boldsymbol{\mu})](\mathbf{x}_{t-1}-\boldsymbol{\mu})']\}\\ &= E\{[(\mathbf{x}_{t-1}-\boldsymbol{\mu})\otimes(\mathbf{x}_{t-1}-\boldsymbol{\mu})]\mathbf{x}'_{t-1}]\}-E[(\mathbf{x}_{t-1}-\boldsymbol{\mu})\otimes(\mathbf{x}_{t-1}-\boldsymbol{\mu})]\boldsymbol{\mu}'\\ &= E\{[(\mathbf{x}_{t-1}-\boldsymbol{\mu})\otimes(\mathbf{x}_{t-1}-\boldsymbol{\mu})]\mathbf{x}'_{t-1}]\}-vec(\boldsymbol{\Upsilon})\boldsymbol{\mu}'\end{aligned}$$

and

$$(\mathbf{x}_{t-1}-\boldsymbol{\mu})\otimes(\mathbf{x}_{t-1}-\boldsymbol{\mu})=\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1}-\boldsymbol{\mu}\otimes\mathbf{x}_{t-1}-\mathbf{x}_{t-1}\otimes\boldsymbol{\mu}+\boldsymbol{\mu}\otimes\boldsymbol{\mu}$$

to obtain the stated result for $E[(\mathbf{x}_{t-1}\otimes\mathbf{x}_{t-1})\mathbf{x}'_{t-1}]$ because

$$E[(\mathbf{x}_{t-1}\otimes\boldsymbol{\mu})\mathbf{x}'_{t-1}]=E[vec(\boldsymbol{\mu}\mathbf{x}'_{t-1})\mathbf{x}'_{t-1}]=(\mathbf{I}_N\otimes\boldsymbol{\mu})E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1})=(\mathbf{I}_N\otimes\boldsymbol{\mu})\boldsymbol{\Upsilon}$$

and $E[(\boldsymbol{\mu}\otimes\mathbf{x}_{t-1})\mathbf{x}'_{t-1}]=\boldsymbol{\mu}\otimes E[\mathbf{x}_{t-1}\mathbf{x}'_{t-1}]=\boldsymbol{\mu}\otimes\boldsymbol{\Upsilon}$. □

B The special case of an AR(1)

Consider the simplest possible univariate version of model (21):

$$y_t=\tau+\alpha y_{t-1}+u_t,\ \text{ where }\ u_t\sim i.i.d.\ N(0,\omega^2),\tag{B3}$$

so that under the null the conditional mean and variance functions will be $\mu_t(\boldsymbol{\theta})=\tau+\alpha y_{t-1}$ and $\sigma_t^2(\boldsymbol{\theta})=\omega^2$, respectively, with $\boldsymbol{\theta}=(\tau,\alpha,\omega^2)'$. The first derivatives of these functions are

$$\frac{\partial\mu_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=(1,y_{t-1},0)',\ \frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=(0,0,1)',$$

while the second ones are

$$\frac{\partial\mu_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}=\mathbf{0}\ \text{ and }\ \frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}=\mathbf{0}.$$

As a consequence, the outer product of the score is

$$\frac{\partial \ln f(\varepsilon_t^*)}{\partial \boldsymbol{\theta}}\frac{\partial \ln f(\varepsilon_t^*)}{\partial \boldsymbol{\theta}'} = \frac{1}{\omega^2}\left[\begin{array}{ccc} \varepsilon_t^{*2} & \varepsilon_t^{*2}y_{t-1} & \varepsilon_t^*(\varepsilon_t^{*2}-1)/(2\omega) \\ \varepsilon_t^{*2}y_{t-1} & \varepsilon_t^{*2}y_{t-1}^2 & \varepsilon_t^*(\varepsilon_t^{*2}-1)y_{t-1}/(2\omega) \\ \varepsilon_t^*(\varepsilon_t^{*2}-1)/(2\omega) & \varepsilon_t^*(\varepsilon_t^{*2}-1)y_{t-1}/(2\omega) & (\varepsilon_t^{*2}-1)^2/(4\omega^2) \end{array}\right],$$

while the Hessian is

$$\mathbf{h}_t(\boldsymbol{\theta}) = \frac{\partial^2 \ln f(\varepsilon_t^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = -\frac{1}{\omega^2}\left[\begin{array}{ccc} 1 & y_{t-1} & \varepsilon_t^*/\omega \\ y_{t-1} & y_{t-1}^2 & \varepsilon_t^* y_{t-1}/\omega \\ \varepsilon_t^*/\omega & \varepsilon_t^* y_{t-1}/\omega & (2\varepsilon_t^{*2}-1)/(2\omega^2) \end{array}\right].$$

Therefore, the expression for the sum of these two matrices reduces to

$$-\frac{1}{\omega^2}\left[\begin{array}{ccc} \varepsilon_t^{*2}-1 & (\varepsilon_t^{*2}-1)y_{t-1} & (\varepsilon_t^{*3}-3\varepsilon_t^*)/(2\omega) \\ (\varepsilon_t^{*2}-1)y_{t-1} & (\varepsilon_t^{*2}-1)y_{t-1}^2 & (\varepsilon_t^{*3}-3\varepsilon_t^*)y_{t-1}/(2\omega) \\ (\varepsilon_t^{*3}-3\varepsilon_t^*)/(2\omega) & \varepsilon_t^*(\varepsilon_t^{*2}-1)y_{t-1}/(2\omega) & (\varepsilon_t^{*4}-6\varepsilon_t^{*2}+3)^2/(4\omega^2) \end{array}\right].$$

In terms of Hermite polynomials, these expressions are proportional to

$$\left[\begin{array}{ccc} H_2(\varepsilon_t^*) & H_2(\varepsilon_t^*)y_{t-1} & H_3(\varepsilon_t^*)/(2\omega) \\ H_2(\varepsilon_t^*)y_{t-1} & H_2(\varepsilon_t^*)y_{t-1}^2 & H_3(\varepsilon_t^*)y_{t-1}/(2\omega) \\ H_3(\varepsilon_t^*)/(2\omega) & H_3(\varepsilon_t^*)y_{t-1}/(2\omega) & H_4(\varepsilon_t^*)/(4\omega^2) \end{array}\right].$$

On this basis, we can provide the following interpretation of the different asymptotically independent test statistics in Propositions 1 and 2:

Conditional heteroskedasticity: This is obtained by regressing $H_2[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)]$ onto 1, y_{t-1} and y_{t-1}^2. The associated influence functions are

$$\mathbf{m}_{ht}(\boldsymbol{\theta}) = \left[\begin{array}{c} 1 \\ H_{2t}y_{t-1} \\ H_{2t}y_{t-1}^2 \end{array}\right],$$

whose covariance matrix under the null is

$$2\left[\begin{array}{ccc} 1 & E(y_{t-1}) & E(y_{t-1}^2) \\ E(y_{t-1}) & E(y_{t-1}^2) & E(y_{t-1}^3) \\ E(y_{t-1}^2) & E(y_{t-1}^3) & E(y_{t-1}^4) \end{array}\right].$$

Since the intercept appears only to adjust the test statistic for the parameter uncertainty in estimating ω^2, the resulting test statistic will be asymptotically distributed as a chi-square random variable with two degrees of freedom. This can be seen more clearly by noticing that the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{m}}_{hT}(\hat{\boldsymbol{\theta}}_T)$ under the null will be

$$2\left[\begin{array}{ccc} 0 & 0 & 0 \\ 0 & V(y_{t-1}) & cov(y_{t-1}, y_{t-1}^2) \\ 0 & cov(y_{t-1}, y_{t-1}^2) & V(y_{t-1}^2) \end{array}\right],$$

which explains why it is not possible to detect random variation in the intercept τ uncorrelated to the random variation in α or ω^2.

Conditional and unconditional asymmetry: This is obtained by regressing $H_3[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)]$ onto 1 and y_{t-1}. The associated influence functions are

$$\mathbf{m}_{at}(\boldsymbol{\theta}) = \begin{bmatrix} H_{3t} \\ H_{3t}y_{t-1} \end{bmatrix},$$

whose covariance variance under the null is

$$6\begin{bmatrix} 1 & E(y_{t-1}) \\ E(y_{t-1}) & E(y_{t-1}^2) \end{bmatrix}.$$

The resulting test statistic will be asymptotically distributed as a chi-square random variable with two degrees of freedom.

As explained in Proposition 2, this test can in turn be additively decomposed into two asymptotically orthogonal components: one obtained by regressing $H_3[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)]$ onto a constant, and another one which regresses this Hermite polynomial on the demeaned value of y_{t-1}. The residual variances of those regressions are 6 and $6V(y_{t-1})$, respectively.

Kurtosis: This is obtained by regressing $H_4[\varepsilon_t^*(\hat{\boldsymbol{\theta}}_T)]$ onto a constant. The relevant influence function is

$$\mathbf{m}_{kt}(\boldsymbol{\theta}) = H_{4t},$$

whose variance is 24 under the null. The resulting test statistic will be asymptotically distributed as a chi-square random variable with one degree of freedom.

Finally, if we impose the Gaussian null in full, we can replace the unconditional moments of y_{t-1} in the above expressions with the following terms

$$E(y_t) = \frac{\tau_0}{1-\alpha_0},$$

$$E(y_t^2) = \frac{\omega_0^2}{1-\alpha_0^2} + \left(\frac{\tau_0}{1-\alpha_0}\right)^2,$$

$$E(y_t^3) = 3\frac{\omega_0^2}{1-\alpha_0^2}\frac{\tau_0}{1-\alpha_0} + \left(\frac{\tau_0}{1-\alpha_0}\right)^3,$$

and

$$E(y_t^4) = 3\left(\frac{\omega_0^2}{1-\alpha_0^2}\right)^2 + 6\frac{\omega_0^2}{1-\alpha_0^2}\left(\frac{\tau_0}{1-\alpha_0}\right)^2 + \left(\frac{\tau_0}{1-\alpha_0}\right)^4.$$

Table 1: Monte Carlo size of random coefficient variation tests: Bivariate models.

	df	Asymptotic critical values			Bootstrap (999 samples) critical values		
		10%	5%	1%	10%	5%	1%
		Panel A: $\text{VAR}(1),\ T = 250$					
h_{hT}^S	15	8.82	4.94	1.55	9.74	4.71	0.96
h_{hT}^T	15	9.66	5.57	2.20	9.70	4.58	0.94
h_{aT}^S	12	10.72	6.53	2.29	10.26	5.10	0.97
h_{aT}^T	12	10.87	6.68	2.35	10.27	5.18	0.95
h_{kT}	5	9.05	5.67	2.48	10.10	4.98	0.86
		Panel B: $\text{VAR}(2),\ T = 250$					
h_{hT}^S	42	9.37	5.71	2.05	9.57	4.82	1.01
h_{hT}^T	42	10.55	7.02	3.23	9.53	4.86	0.87
h_{aT}^S	20	11.09	7.27	3.04	10.13	4.91	0.99
h_{aT}^T	20	11.38	7.42	3.12	10.02	4.85	0.98
h_{kT}	5	8.89	5.82	2.50	10.07	5.06	0.82
		Panel C: $\text{VAR}(1),\ T = 1,000$					
h_{hT}^S	15	10.09	5.30	1.49	10.34	5.11	1.02
h_{hT}^T	15	10.22	5.80	1.81	10.05	5.21	1.09
h_{aT}^S	12	10.77	5.83	1.57	10.29	5.13	1.00
h_{aT}^T	12	10.81	5.86	1.58	10.30	5.08	0.99
h_{kT}	5	9.72	5.20	1.63	10.06	4.74	0.98
		Panel D: $\text{VAR}(2),\ T = 1,000$					
h_{hT}^S	42	10.51	5.84	1.54	10.24	5.23	1.02
h_{hT}^T	42	11.11	6.48	1.93	9.98	5.16	1.05
h_{aT}^S	20	10.80	5.89	1.52	9.81	4.72	0.83
h_{aT}^T	20	10.83	5.96	1.51	9.81	4.66	0.81
h_{kT}	5	9.77	5.21	1.64	9.95	4.67	0.96

Notes: Results based on 10,000 samples of size $T = 250$ and $1,000$ from bivariate $\text{VAR}(1)$ and $\text{VAR}(2)$ models. Left panels report the rejection rates based on the asymptotic critical values while the right ones do the same but conducting the bootstrap explained in section 3.3. The row labels h_{hT}, h_{aT}, and h_{kT} refer to the moment tests in Propositions 1. In the cases of h_{hT} and h_{aT}, the additional superscript S refers to using the sample analogues of (14) and (15), while the superscript T denotes those that rely on the population covariance in Proposition 3 evaluated at the OLS estimators.

Table 2: Monte Carlo size of random coefficient variation tests: Trivariate models.

	df	Asymptotic critical values 10%	5%	1%	Bootstrap (999 samples) critical values 10%	5%	1%
				Panel A: $\text{V}\textsc{ar}(1),\ T = 250$			
h_{hT}^S	54	10.16	6.33	2.35	10.09	5.26	1.02
h_{hT}^T	54	11.13	7.31	3.39	9.75	5.11	0.88
h_{aT}^S	40	11.47	7.60	3.10	10.09	5.09	0.97
h_{aT}^T	40	11.81	7.81	3.19	10.11	5.10	0.98
h_{kT}	15	11.16	7.61	3.71	10.18	5.07	1.06
				Panel B: $\text{V}\textsc{ar}(2),\ T = 250$			
h_{hT}^S	162	10.28	6.77	2.82	9.81	5.18	1.01
h_{hT}^T	162	12.12	8.35	4.22	9.42	4.65	0.99
h_{aT}^S	70	12.77	8.97	4.20	10.30	5.22	1.08
h_{aT}^T	70	13.27	9.27	4.34	10.22	5.14	1.06
h_{kT}	15	11.31	7.80	3.73	10.32	5.25	1.04
				Panel C: $\text{V}\textsc{ar}(1),\ T = 1,000$			
h_{hT}^S	54	10.73	5.80	1.41	10.38	5.08	1.00
h_{hT}^T	54	11.61	6.67	2.00	10.50	5.32	1.10
h_{aT}^S	40	11.50	6.70	1.95	10.50	5.18	1.02
h_{aT}^T	40	11.60	6.75	1.97	10.52	5.17	1.02
h_{kT}	15	10.72	6.39	2.09	9.80	5.01	0.98
				Panel D: $\text{V}\textsc{ar}(2),\ T = 1,000$			
h_{hT}^S	162	11.12	6.02	1.94	10.25	4.92	1.02
h_{hT}^T	162	12.03	7.13	2.72	9.82	5.02	1.10
h_{aT}^S	70	12.01	6.71	2.30	9.99	4.72	1.05
h_{aT}^T	70	12.16	6.75	2.33	10.09	4.70	1.08
h_{kT}	15	10.89	6.32	2.16	10.03	5.00	0.97

Notes: Results based on 10,000 samples of size $T = 250$ and $1,000$ from trivariate $\text{V}\textsc{ar}(1)$ and $\text{V}\textsc{ar}(2)$ models. Left panels report the rejection rates based on the asymptotic critical values while the right ones do the same but conducting the bootstrap explained in section 3.3. The row labels h_{hT}, h_{aT}, and h_{kT} refer to the moment tests in Propositions 1. In the cases of h_{hT} and h_{aT}, the additional superscript S refers to using the sample analogues of (14) and (15), while the superscript T denotes those that rely on the population covariance in Proposition 3 evaluated at the OLS estimators.

Table 3: Monte Carlo power of random coefficient variation tests with bootstrap: Sample size $T = 250$, $N = 2$.

	Time-varying mean parameters only			Time-varying variance parameters only			Time-varying mean and variance parameters		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%
	Panel A: DGP 1 – Shocks: $\theta_{it} = \bar{\theta} + \eta_{it}$ with $\eta_{it} \sim iidN(0, \sigma^2_{\eta_i})$								
h^S_{hT}	74.08	64.24	42.20	38.04	25.40	9.88	95.16	91.52	80.36
h^T_{hT}	72.36	61.96	39.56	31.04	20.24	8.04	92.96	88.36	76.00
h^S_{aT}	34.00	25.04	13.36	63.60	52.00	29.76	78.60	69.44	50.16
h^T_{aT}	33.84	25.08	13.32	63.60	52.12	29.76	78.44	69.40	50.04
h_{kT}	25.16	16.48	6.96	88.68	82.68	62.04	83.32	75.28	54.48
	Panel B: DGP 2 – Shocks: $\theta_{it} = \bar{\theta}(1 - .95) + .95\theta_{it-1} + \eta_{it}$ with $\eta_{it} \sim iidN[0, \sigma^2_{\eta_i}(1 - .95^2)]$								
h^S_{hT}	68.00	61.12	46.84	70.80	61.00	38.48	97.20	96.16	90.96
h^T_{hT}	56.96	48.40	33.60	60.40	49.28	28.96	95.52	93.36	85.96
h^S_{aT}	40.48	32.84	22.44	61.84	51.80	31.68	87.64	82.76	70.56
h^T_{aT}	40.32	32.92	22.24	61.92	52.28	31.76	88.20	83.32	71.20
h_{kT}	26.68	20.76	13.44	84.16	77.76	55.52	89.92	85.36	73.24
	Panel C: DGP 3 – Shocks: $\theta_{it} = \bar{\theta}1_{\{s_t=1\}} + \underline{\theta}1_{\{s_t=0\}}$ with s_t binary Markov chain (p, q)								
h^S_{hT}	41.96	32.68	18.92	81.68	75.40	60.08	92.20	89.60	80.48
h^T_{hT}	45.68	35.60	20.80	74.84	67.60	50.36	93.92	91.40	84.28
h^S_{aT}	61.64	53.00	39.12	66.40	56.76	37.36	89.44	85.84	74.00
h^T_{aT}	61.52	53.08	39.20	66.40	56.84	37.72	89.40	85.76	74.04
h_{kT}	51.00	42.96	29.32	86.28	80.28	65.84	90.48	87.00	76.28

Notes: Results based on 2,500 samples of size $T = 250$ from bivariate VAR(1) models. Left panels report the rejection rates based on the asymptotic critical values while the right ones do the same but conducting the bootstrap explained in section 3.3. The row labels h_{hT}, h_{aT}, and h_{kT} refer to the moment tests in Propositions 1. In the cases of h_{hT} and h_{aT}, the additional superscript S refers to using the sample analogues of (14) and (15), while the superscript T denotes those that rely on the population covariance in Proposition 3 evaluated at the OLS estimators.

Table 4: Monte Carlo power of random coefficient variation tests with bootstrap: Sample size $T = 250$, $N = 3$.

	Time-varying mean parameters only			Time-varying variance parameters only			Time-varying mean and variance parameters		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%
	Panel A: DGP 1 – Shocks: $\theta_{it} = \bar{\theta} + \eta_{it}$ with $\eta_{it} \sim iidN(0, \sigma^2_{\eta_i})$								
h^S_{hT}	82.56	74.80	54.44	52.00	39.40	20.20	99.28	98.60	95.16
h^T_{hT}	80.52	71.08	50.56	41.30	29.80	15.30	98.76	97.28	92.08
h^S_{aT}	53.04	41.92	23.48	79.60	70.90	46.50	94.96	91.60	79.60
h^T_{aT}	53.32	42.04	23.24	79.40	71.00	46.60	94.92	91.40	79.72
h_{kT}	33.92	22.72	9.48	94.70	90.70	76.40	94.80	90.72	76.04
	Panel B: DGP 2 – Shocks: $\theta_{it} = \bar{\theta}(1 - .95) + .95\theta_{it-1} + \eta_{it}$ with $\eta_{it} \sim iidN[0, \sigma^2_{\eta_i}(1 - .95^2)]$								
h^S_{hT}	79.76	73.00	58.88	84.56	75.56	54.48	99.52	98.72	96.48
h^T_{hT}	69.52	61.20	44.80	72.92	62.80	40.76	98.96	98.04	94.72
h^S_{aT}	51.80	44.48	32.72	80.48	71.16	49.76	97.32	95.64	89.44
h^T_{aT}	52.20	44.56	32.96	80.64	71.00	49.80	97.44	95.60	89.56
h_{kT}	32.68	26.12	17.44	91.36	86.68	69.40	96.44	93.44	84.36
	Panel C: DGP 3 – Shocks: $\theta_{it} = \bar{\theta}1_{\{s_t=1\}} + \underline{\theta}1_{\{s_t=0\}}$ with s_t binary Markov chain (p, q)								
h^S_{hT}	78.80	72.84	60.00	90.28	86.92	78.44	98.72	98.16	95.40
h^T_{hT}	84.76	78.84	66.52	86.84	82.72	70.60	98.88	98.52	97.00
h^S_{aT}	96.48	95.96	93.28	83.00	76.76	61.12	99.08	98.88	97.92
h^T_{aT}	96.60	95.96	93.64	82.96	76.88	61.12	99.08	98.92	97.92
h_{kT}	95.96	94.92	93.08	90.12	86.32	76.48	99.32	99.20	98.72

Notes: Results based on 2,500 samples of size $T = 250$ from trivariate VAR(1) models. Left panels report the rejection rates based on the asymptotic critical values while the right ones do the same but conducting the bootstrap explained in section 3.3. The row labels h_{hT}, h_{aT}, and h_{kT} refer to the moment tests in Propositions 1. In the cases of h_{hT} and h_{aT}, the additional superscript S refers to using the sample analogues of (14) and (15), while the superscript T denotes those that rely on the population covariance in Proposition 3 evaluated at the OLS estimators.

Table 5: A bivariate model for GDE-GDI measures.

Panel A: OLS estimates

$\textsc{Var}(1)$

$\boldsymbol{\tau}$			$vec(\mathbf{A}_1)$		
Param.	Est.	Std.Err.	Param.	Est.	Std.Err.
τ_1	.046	.034	a_1	.872	.032
τ_2	.425	.037	a_2	−.011	.059
			a_3	−.001	.064
			a_4	.445	.056

$\textsc{Var}(2)$

$\boldsymbol{\tau}$			$vec(\mathbf{A}_1)$			$vec(\mathbf{A}_2)$		
Param.	Est.	Std.Err.	Param.	Est.	Std.Err.	Param.	Est.	Std.Err.
τ_1	.038	.036	a_1	.707	.034	a_5	.190	.115
τ_2	.397	.066	a_2	−.245	.066	a_6	.271	.059
			a_3	.012	.034	a_7	−.019	.114
			a_4	.445	.062	a_8	.014	.059

Panel B: Testing for random coefficients

		Sample covariance			Theoretical covariance		
			p-value			p-value	
Test	df	Stat.	Asym.	Boot.	Stat.	Asym.	Boot.
			$\textsc{Var}(1)$				
h_{hT}	15	27.526	.025	.034	33.474	.004	.024
h_{aT}	12	15.626	.209	.174	15.260	.228	.186
h_{saT}	4	6.408	.171	.141	6.408	.171	.141
h_{daT}	8	9.218	.324	.265	9.053	.338	.271
h_{kT}	5	34.462	.000	.002	34.462	.000	.002
h_{saT} & h_{kT}	9	40.870	.000	.003	40.870	.000	.003
			$\textsc{Var}(2)$				
h_{hT}	42	78.417	.001	.003	89.308	.000	.004
h_{aT}	20	19.074	.517	.411	18.703	.541	.425
h_{saT}	4	6.706	.152	.131	6.706	.152	.131
h_{daT}	16	12.368	.718	.588	12.159	.733	.597
h_{kT}	5	37.036	.000	.001	37.036	.000	.001
h_{saT} & h_{kT}	9	43.742	.000	.001	43.742	.000	.001

Notes: Data: Quarterly real GDE and GDI from 1952Q1 to 2015Q2. y_{1t} is the difference between the logs of GDE and GDI, while y_{2t} is the (arithmetic) average of the (geometric) quarterly rates of growth of these two output measures. The row labels h_{hT}, h_{aT}, and h_{kT} refer to the moment tests in Proposition 1, while h_{saT} and h_{daT} to those in Proposition 2. In the cases of h_{hT}, h_{aT} and h_{daT}, "Sample covariance" refers to using the sample analogues of (14) and (15), while "Theoretical covariance" refers to those that rely on the population covariance in Proposition 3 evaluated at the OLS estimators.

Figure 1: Data

Figure 1a: Statistical discrepancy

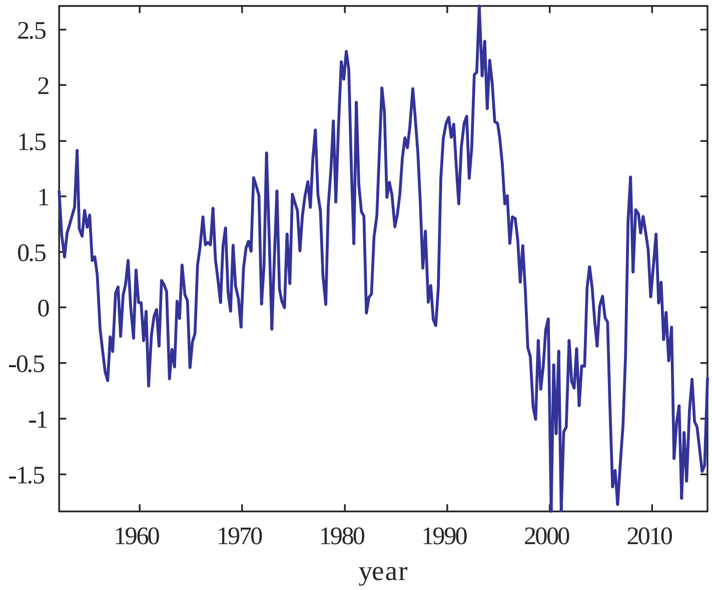
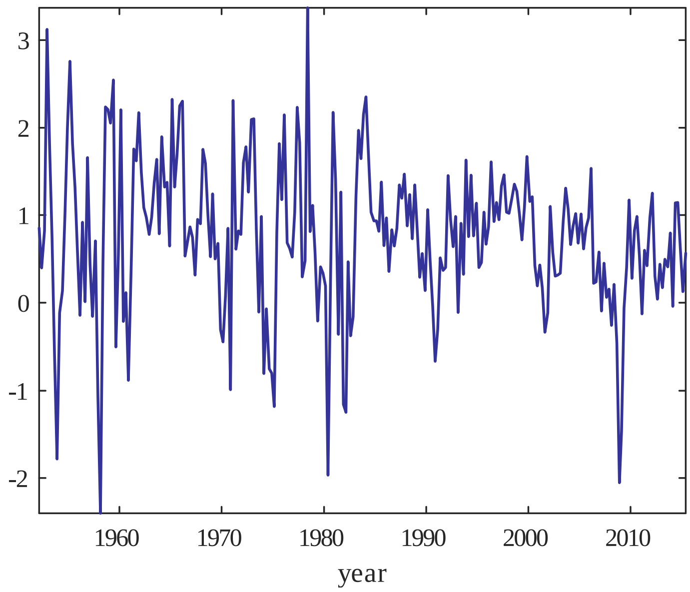


Figure 1b: Average GDE-GDI growth rate



Notes: Data: Quarterly real GDE and GDI from 1952Q1 to 2015Q2.

Figure 2: Estimated innovations: VAR(1) and VAR(2)

Figure 2a: Statistical discrepancy

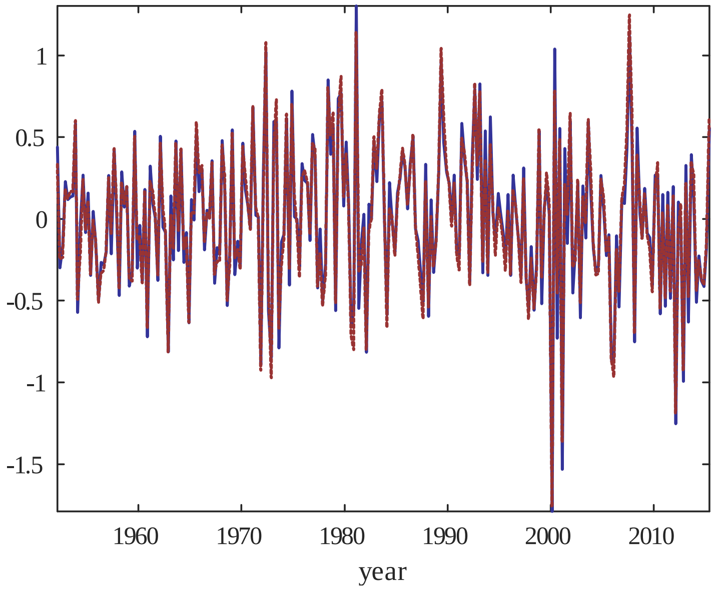
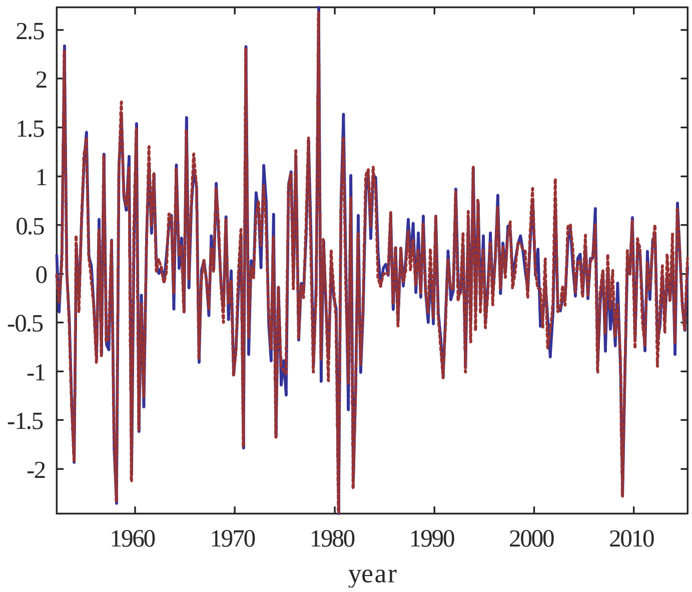


Figure 2b: Average GDE-GDI growth rate



Notes: Data: Quarterly real GDE and GDI from 1952Q1 to 2015Q2. The blue continuous line depicts the OLS residuals from the VAR(1) while the red dotted one correspond to the VAR(2) ones.

Figure 3: Contribution of observation t to the different tests: VAR(2)

Figure 3a: Conditional heteroskedasticity

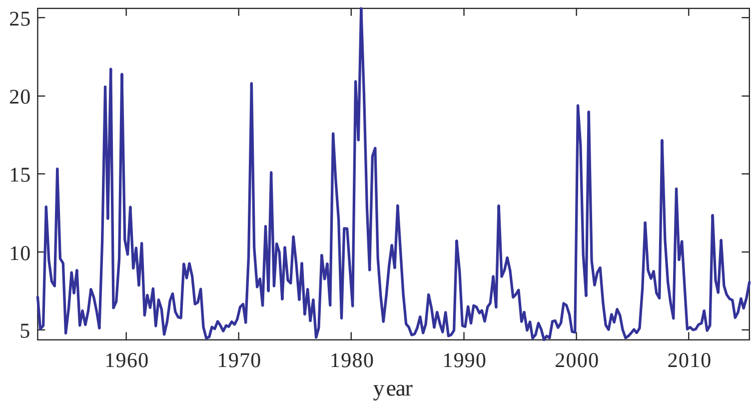


Figure 3b: Conditional skewness

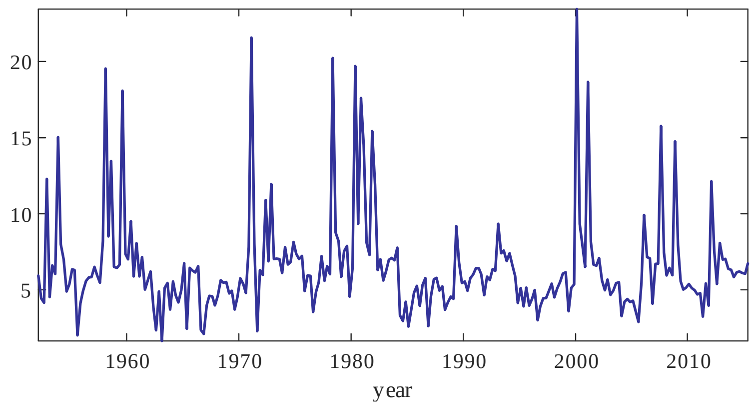
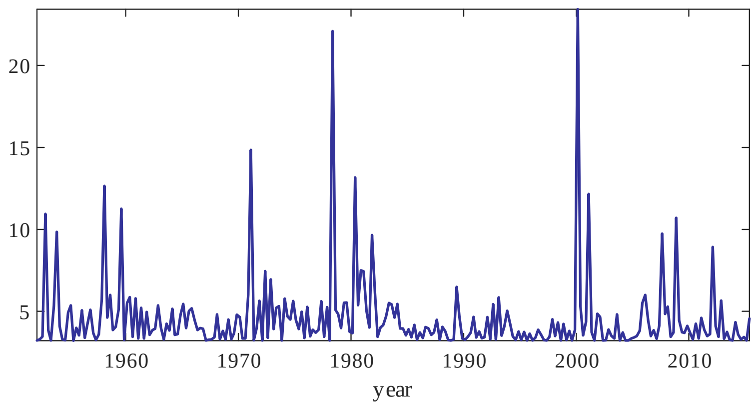


Figure 3c: Unconditional kurtosis



Notes: In this figure we report the contribution of observation t to each of the test statistics we consider for the vector autoregression of order two. For the influence functions based on the second and third multivariate Hermite polynomials, we report results from estimating the covariance matrix of the moments through its sample analogue. Specifically, the contribution of observation t to the test statistic j is simply $T \cdot \mathbf{m}'_{jt}(\hat{\boldsymbol{\theta}}_T)\hat{V}^{-1}[\mathbf{m}_{jt}(\hat{\boldsymbol{\theta}})]\mathbf{m}_{jt}(\hat{\boldsymbol{\theta}}_T)$ for $j = h, a, \dot{k}$, the expression for $\mathbf{m}_{jt}(\boldsymbol{\theta})$ and $V[\mathbf{m}_{jt}(\boldsymbol{\theta})]$ can be found in section 3. In order to "normalise" these figures, which tend to have huge spikes occasionally and almost no action the rest of the time, we follow Hawkins and Wixley (1986) so that we plot the fourth root of these quantities.