*Causality versus Serial Correlation: an Asymmetric Portmanteau Test**

Amedeo Andriollo †

JOB MARKET PAPER
(Preliminary version)

Latest version: February 2025

Abstract

I study the problem of testing for noncausality in mean (one-sided conditional mean independence) between two multivariate time series within the class of testing procedures based on serial cross-correlation. Existing tests in this class either require parametrization of the joint process or are characterized under the null hypothesis of mutual independence. As a result, these tests may suffer from size distortions when misspecifying inverse causality, i.e., dependencies in the causal direction opposite to the one being tested. I propose a modified Portmanteau test statistic that incorporates a correction term to offset the influence of inverse causality, thereby eliminating the need to fully model the joint dynamics. I demonstrate that the proposed test statistic converges asymptotically to a standard normal distribution under the null hypothesis of noncausality in mean, resulting in correct asymptotic size. As an empirical application, I explore the statistical properties of my proposed test by studying three widely used measures of macroeconomic structural shocks, showing that the proposed test provides more reliable inference than the benchmark test.

Keywords: Martingale difference hypothesis; Cross-correlation; Nonfundamentalness.

JEL Codes: C12; C32; E30.

*I sincerely appreciate the supervision of Eric Renault throughout this work. I am also thankful especially to Giovanni Ricco, Luis Candelaria, Kenichi Nagasawa, Leonardo Melosi, and Cesare Robotti for their valuable suggestions. I appreciate the comments from Giulio Rossetti and Emanuele Savini. Any and all errors are my own.

†Department of Economics, University of Warwick, United Kingdom.
E-mail: amedeo.andriollo@warwick.ac.uk; *Website*: amandri.github.io.

1 Introduction

In this paper I study the problem of testing for noncausality in mean between multivariate time series within the class of testing procedures based on serial cross-correlation.

Consider two zero-mean multivariate time series $\{X_t, Z_t; t = 1, .., T\}$. Noncausality in mean, from Z to X, is understood as the property of one-sided conditional mean independence of the present X, with respect to the past X and the past and present of Z.
The null hypothesis of interest is:

$$\mathcal{H}_0: \ \mathbb{E}[X_t | \{X_s\}_{s<t}, \{Z_s\}_{s\leq t}] = 0. \tag{1}$$

Testing for this form of conditional mean independence in time series models plays an important role in contexts such as economic policy analysis (see White and Pettenuzzo (2014) for the connection with Granger causality), forecasting evaluation (e.g., the one-step conditional predictive ability test in Giacomini and White (2006)), and the study of business cycle fluctuations via impulse response analysis (see Ramey (2016), and in particular Chen et al. (2017)).
In these instances, many key questions can often be recast as problems of model specification testing in the presence of omitted variables. After specifying the conditional mean for some variables of interest, practitioners typically have interest in jointly testing two conditions: i) whether the residuals of the model, $\{X_t\}$, are conditionally mean independent with respect to their past (i.e., if the dynamic model is correctly specified), and ii) whether the variables omitted from the model, $\{Z_t\}$, do not influence the mean of the variables of interest. The null $\mathcal{H}_0$ of eq.(1) implies both conditions.

As an illustrative example, consider the variable X as a series of economic policy uncertainty shocks (i.e., unanticipated movements/news associated to the level of uncertainty about economic policy, see Baker et al. (2016)), estimated as residuals from a Structural Vector AutoRegressive (SVAR) model. The null hypothesis of interest $\mathcal{H}_0$ then jointly examines two conditions: i) whether the model is specified such that the shocks, $\{X_t\}$, are (weakly) exogenous with respect to the other present and past of the endogenous variables in the model (Ramey (2016)); ii) whether other macroeconomic variables omitted from the SVAR model, $\{Z_t\}$, do not influence the exogeneity of the shock series. If either condition is violated, the shock series is not an exogenous fluctuation of uncertainty, calling into question the validity of the impulse response analysis.

Existing tests based on serial cross-correlation either require modelling the conditional mean of the joint process (e.g., Wald tests, Hong (1996a); a similar argument extends to Hong and Lee (2005) and Escanciano and Velasco (2006)), or are characterized under the null hypothesis of mutual independence (e.g., Portmanteau tests, Haugh (1976), Hong (1996b)), being a stricter condition than noncausality in mean, as it imposes bidirectional noncausality in the entire distributions (not restricted to the mean). In both cases, inference may be jeopardized because of size distortions in the presence of misspecification when modelling inverse causality, i.e., dependencies from past and present X to present Z. This is particularly crucial in the example discussed, where it is desirable that inference about the exogeneity of the economic policy uncertainty shocks is not distorted by inverse causality, given that dependencies from past shocks, X, to the present of omitted macroeconomic variables, Z, are likely to be exist.

Using the Portmanteau test (i.e., a weighted sum of squared cross-correlation at different lags) as benchmark statistic, I address the limitations of the existing tests by proposing a modification of the test statistic, that differs by introducing an easy-to-compute correction term which offsets the influence of inverse causality. Intuitively, the correction term accounts for asymmetry in the sum of cross-correlation: while the squares (or quadratic norms) of cross-correlation symmetrically incorporate both directions of causality in their variance, the correction term isolates the component related to inverse causality, retaining only the one associated with the tested direction. As a result, my methodology avoids the need to fully model the joint dynamics, making it robust to misspecifications of the dependencies between present Z and the past and present of X. This is especially useful when there is no prior knowledge about how external variables interact with the dynamic system, allowing for a more agnostic approach to the modelling of joint dependence structure of the time series.

I show that the corrected version of the benchmark statistic achieves asymptotic normality under the null hypothesis $\mathcal{H}_0$ of interest, which ensures that the test has the correct size in large samples. This represents a clear improvement over the existing testing strategies whose asymptotic properties are studied under null hypotheses that imply the one of interest. These methodologies, usually tailored to test for mutual independence, may not be suitable for conducting inference about the specific direction of causality being tested. In fact, such procedures are likely to suffer from size distortions when the data generating process (DGP) exhibits mean independence of present X with respect to the past of the joint process, but present Z is not independent from the past and present of X. Because the variance of these test statistics accounts for dependencies in both directions -from Z

to X and from X to Z- it may be inflated relative to that of a standard normal distribution, leading to incorrect inference due to over-rejection. A similar issue arises with testing strategies that require modelling the conditional mean of the joint process. In the presence of misspecified inverse causality, size distortions may occur when present X is mean independent with respect to the past of the joint process, but present Z is not mean independent with respect to the past and present of X.

My testing strategy builds on Hong (1996b), whose test statistic is the weighted sum of squared cross-correlation between univariate time series at positive and negative lags, with weights determined by a kernel function. Given the one-sided nature of the hypothesis of interest, and following Hong (2001) and Bouhaddioui and Roy (2006), I consider its multivariate formulation restricted to positive lags as the benchmark statistic, which I regard as representative of the class of tests based on serial cross-correlation. Specifically, the benchmark is defined as the weighted sum of quadratic forms (the ℓ_2 norm) of the cross-correlation between the two multivariate processes at positive lags. My methodology differs from the benchmark by introducing a correction term that removes the influence of dependencies from X to Z, which enters in the statistic due the choice of the quadratic norm. Specifically, the correction term differentiates out the cross-product terms in the weighted sum associated to high-order moments of the joint process which capture the direction of causality inverse to the one being tested.

Pertaining to the assumptions underlying the asymptotic theory, I prioritize restrictions on the process X rather than Z, motivated by two reasons: i) the practitioner has no prior knowledge on how the omitted variables are entering the dynamic system; ii) the practitioner might have interest in placing additional restrictions on the variables of interest, as those translate in sharper identifying restrictions on the error terms $\{X_t\}$ (e.g., restrictions on the second moments for the conditional variance of the process). In particular, the main theorem requires the process $\{X_t\}$ to be a martingale in the second and fourth moments, i.e. to be conditionally homoskedastic and homokurtic.[1] While the latter condition is needed to achieve the convergence to normality, the conditional homoskedasticity is essential to isolate the effects on the quadratic forms when testing for the conditional mean.[2] Put in technical terms, due the choice of the norm, the former condition is

[1]Note that the null hypothesis of eq.(1) can be read as testing for the martingale difference property of the process $\{X_t\}$, with respect to a conditioning set that includes the past of the process itself, $\{X_s; s < t\}$, enlarged by the present and past of an additional time series, $\{Z_s; s \leq t\}$. Thus, it implies that the process $\{X_t\}$ is martingale in the first moment.

[2]When relaxing the assumption of independence, Hong (2001)'s footnote 8 briefly discussed the condition of conditional homokurtosis for establishing the asymptotic normality of his testing procedure.

necessary to correctly center the distribution of the test statistic under the null hypothesis of interest. Concerning the construction of the test statistic, the main result about the asymptotic normality is initially established under the assumption that the processes are observed. I then relax this assumption, accommodating scenarios where the processes are estimated. Such an instance, my methodology allows for flexible specifications of the conditional mean of the variables of interest, thus potentially covering a large class of DGPs.

Regarding the properties of the proposed statistic under the alternative hypotheses (i.e., its power), I consider the class of fixed alternatives characterised by nonzero cross-correlation between present X and present and past Z, as in Hong (1996a) and Hong (1996b). Under such class of alternatives, I show that, asymptotically, the corrected statistic has power equal to the benchmark testing procedures, as the correction term becomes asymptotic negligible, due the finiteness of the fourth-order cumulants. Similarly to the existing procedures based on serial cross-correlation, the test has no power against alternatives that, either i) have strong auto-correlation in the process X; ii) have non-zero causality in mean from Z to X, i.e. $\mathbb{E}[X_t|\{X_s\}_{s<t}, \{Z_s\}_{s\leq t}] \neq 0$, that is not reflected in any linear association between $\{X_t\}$ and $\{Z_s; s \leq t\}$ or, in other words, when the time series are uncorrelated but the process X is not mean-independent (non-martingale).

To validate the proposed testing strategy, I provide some simulation evidence. I study the finite sample properties of the testing procedures in a set of Monte Carlo experiments, with the purpose of examining the impact of the correction term. The Monte Carlo simulations validate the asymptotic theory: when the magnitude of the inverse causality is relevant, the corrected test statistic has good finite sample properties under the null hypothesis, with empirical size close to the nominal one, whereas the benchmark test exhibits rejection rates above the nominal level. This corroborates the theoretical discussion about the size distortions of the benchmark statistic. Regarding the empirical frequencies under the alternatives, both testing strategies generally have similar empirical power when there is nonzero cross-correlation between present X and present and past Z.

As empirical application, I study the property of fundamentalness, also known as invertibility, of popular measures of macroeconomic structural shocks. A structural shock is considered fundamental when it can be expressed as a linear function of the present and past values of the endogenous variables in the model. Intuitively, fundamentalness is related to the property of the shock being an exogenous fluctuation that depends only on the present and past of the endogenous variables in the model, rather than external variables omitted from the dynamic system or future realizations of the internal variables (see

Giannone and Reichlin (2006)). Common practice in the Structural VectorAutoRegressive (SVAR) literature is to assume this condition to hold, despite numerous empirical findings suggesting otherwise (see the survey of Alessi et al. (2011)). Failure to account for this can result in misleading inference because of model misspecification issues. Testing for fundamentalness can be reformulated into testing for the null hypothesis in eq.(1) under general conditions (see Theorem 1 in Chen et al. (2017)). With this in mind, I use my testing strategy to investigate the fundamentalness of three popular measures of macroeconomic structural shocks: Baker et al. (2016)'s economic policy uncertainty (EPU) shock, Jarociński and Karadi (2020)'s monetary policy information shock, and Känzig (2023)'s carbon policy shock. For all three series of shocks, the two testing procedures –the benchmark and its corrected version– deliver different results. By inspecting the test statistics, we notice that, in all the scenarios, the benchmark testing procedure fails to produce reliable conclusions because the inverse causality channel is non negligible. This in turn underscores the validity of the proposed approach, as the corrected test statistic is robust to misspecification of such direction of causality.

I conclude that the EPU structural shocks are not fundamental, when testing for the null hypothesis with respect to the McCracken and Ng (2016)'s macroeconomic factors. Given this additional set of controls, which alleviates the problem of non-fundamentalness, I revisit Diercks et al. (2024): when controlling for these additional macroeconomic factors previously omitted, I find that the response of inflation to the EPU shock becomes drastically positive. In light of this evidence, together with the contractionary responses of the other variables to the EPU shock, Baker et al. (2016)'s EPU structural shock can be regarded as a supply-side negative shock (similar to an 'expectational' shock, see Ascari et al. (2023)). The other two shocks, Jarociński and Karadi (2020)'s and Känzig (2023)'s, narrate a cautionary tale, by highlighting again the fragility of the benchmark test statistic. I present two scenarios where the practitioner might conclude the non-fundamentalness of the shocks when relying on the benchmark procedure -a conclusion mainly driven by the presence of inverse causality between the structural shocks and the set of omitted variables.

RELATED LITERATURE. In this paper, I contribute to three strands of the literature.

First, I contribute to the literature related to specification testing in dynamic (linear) models whose tests are typically characterized under the null hypothesis of mutual independence. Early examples are the methodologies of Hosking (1980) and Li and McLeod

(1981), which are essentially Portmanteau (or Ljung and Box (1978)'s) test statistics, i.e., statistics based on the sum of squared auto-correlation of residuals with respect to a fixed number of lags. Both papers study the asymptotic distribution of the residuals of a fitted multivariate ARMA model under the innovations being independent and identically distributed. Opposite to fitting a multivariate model, Haugh (1976) proposes a two-step procedures to test for independence between two sets of time series, which involves i) fitting univariate models to each of the series, and ii) studying the sum of the cross-correlation at different lags between the two residual series. In particular, note that Haugh (1976) does not require the modelling of the joint process. A series of papers, Hong (1996a) and Hong (1996b), generalises the Box and Pierce (1970)'s and Haugh (1976)'s tests, by proposing a test statistic based on the sum of weighted squares of residual cross-correlation, with weights depending on a kernel function. Bouhaddioui and Roy (2006) extend the testing strategy of Hong (1996b) to multivariate processes. In this paper, I adopt this last testing strategy, restricted to positive lags, as the benchmark. Thus, I improve on the literature about testing procedures that follows from Haugh (1976), by introducing a correction term which makes them more suitable for conducting inference about a specific direction of causality.

Second, I contribute to the literature concerning the testing for the martingale difference hypothesis (MDH). When considered for specification testing, such testing procedure usually require modelling the conditional mean of the joint process. An early example is Durlauf (1991), which proposes a test statistic based on estimates of the spectral density (or auto-correlation function), at different lags. Stemming from Hong (1999), two influential papers, Hong and Lee (2005) and Escanciano and Velasco (2006), generalize the previous approach to linear and nonlinear serial dependencies through the so-called generalized spectral density, thus building on the sum of auto-correlation between empirical characteristic functions at different lags. Similarly to Hong (1996a), Hong and Lee (2005) proposes a kernel-based spectral test designed explicitly to study the problem of specification testing for the conditional mean models in time series. Based on Taylor expansions, Escanciano and Velasco (2006) examines the problem of MDH for an observed time series. Under additional assumptions, this last testing procedure can be applied in the context of specification testing (e.g., Wang et al. (2022)).[3] These two testing strategies should be viewed as extensions of Ljung and Box (1978)'s approach rather than Haugh (1976)'s. Both

[3]The main difference between the two approaches is that, while the estimation error does not impact the limit distribution in Hong and Lee (2005), it does in Escanciano and Velasco (2006) as no longer asymptotically negligible.

methodologies require to specify the joint conditional mean of the multivariate time series, $\{X_t, Z_t\}$, in order to test the null hypothesis of eq. (1), which, after modelling the joint mean, is interpreted as testing for the MDH in the estimated residuals. Thus, an argument similar to that used for the class of tests following Haugh (1976) applies here as well. In short, I improve on this second strand of literature by allowing the bypassing of modelling both directions of causality in the mean, from Z to X and from X to Z, when testing the null hypothesis of interest.

Similarly to the aforementioned literature, I presume the weighting function associated to my statistic to be integrable. Székely et al. (2007) introduce a new measure of dependence between random vectors which allows for non-integrable weighting functions. A series of papers, Shao and Zhang (2014) and Lee and Shao (2018), extends the Székely et al. (2007)'s approach to the context of martingale difference hypothesis. Two recent papers develop new testing strategies based on Escanciano and Velasco (2006)'s approach and Lee and Shao (2018)'s statistic: Wang et al. (2022) and Wang (2024). Since both testing procedures fall within this previous strand of literature, my contribution also applies to this recent literature. However, extending my approach to non-integrable weighting functions is left for future research.

Third and finally, I contribute to the literature on testing the fundamentalness of structural shocks. Specifically, I propose a new testing strategy that integrates two approaches into one: i) those using Granger causality tests (e.g., Giannone and Reichlin (2006); Forni and Gambetti (2014)), and ii) those associated with the testing for conditional lagged exogeneity (e.g., Chen et al. (2017), Miranda-Agrippino and Ricco (2023)). Moreover, I provide new insights about the Economic Policy Uncertainty (EPU) shock series of Baker et al. (2016). By controlling for additional macroeconomic factors, I reveal that the EPU shock exhibits characteristics of a superadditive negative supply shock. This finding not only strengthens the results of Diercks et al. (2024) but suggests that the EPU structural shock behaves similarly to an 'expectational' shock as in Ascari et al. (2023).

The paper is structured as follows: Section 2 introduces the benchmark and the proposed test statistics, by discussing the application of the correction term. In particular, the definition of the asymmetric Portmanteau statistic is found in Section 2.2. Section **??** presents the asymptotic theory for the corrected test statistic. Section 3 offers a set of Monte Carlo simulations to investigate the finite-sample properties of the testing procedure. Section 4 presents the empirical application. Section 5 concludes.

NOTATION. Throughout the paper, I use the following standard notation. vec$(\cdot)$ stands for the vectorization operator, diag$(\cdot)$ is the main diagonal of the matrix. I denote $\xrightarrow{d}$ as convergence in distribution, $\xrightarrow{p}$ as convergence in probability. $\perp$ stands for orthogonality, and $\perp\!\!\!\perp$ for mutual independence. The notation $\sim$ reads as 'distributed as'. Given two vectors a and b, I denote the scalar/inner product $\langle a, b\rangle = a'b$, $a \otimes b$ the Kronecker product between them, and $||a|| = \sqrt{\langle a, a\rangle}$ stands for the Euclidean (or ℓ_2) norm of the vector a. For a real positive semidefinite matrix A, we denote its square-root the matrix $B = (A)^{1/2}$ such that $A = BB = BB'$, and $||A||_F = \sqrt{\text{tr}(A'A)}$ its Frobenius norm, where tr$(\cdot)$ is the trace operator.

2 The test statistics based on ℓ_2-norm

Section 2.1 covers the primary framework and provides a formal discussion on the class of Portmanteau-type test statistics. I specifically elucidate the link between the variance of existing test statistics and inverse causality. Building on this, in Section 2.2, I propose a new test statistic for testing for the null hypothesis of interest.

2.1 Preliminaries and a discussion on a class of test statistics

Let $\{X_t, Z_t; t = 1, .., T\}$ denote two zero-mean multivariate squared integrable stationary processes of respective dimensions $d_1, d_2 \in \mathbb{N}^+$. Let $\mathcal{I}(t-1)$ be the information set available at period $t-1$ of the joint time series $\{Z_t, X_s, Z_s; s < t\}$, where, together with the past of the process X, the past and present of the process Z are included. If not explicitly mentioned, I suppose these processes to be standardized.[4]

Before presenting my proposed new test statistic (eq. (5)) in Section 2.2, I provide a discussion of the testing strategies based on serial cross-correlation *à la* Hong (1996b). As a benchmark of this class of tests, I consider the following one-sided test statistic based on the sum of weighted quadratic forms:

$$\mathcal{T}_\omega = \sum_{j=0}^{T-1} \omega(j) Q(j) \tag{2}$$

$$Q(j) = ||\widehat{\Gamma}_{XZ}(j)||_F^2 = \text{tr}\left[\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)\right] = \left|\left|\text{vec}\left[\widehat{\Gamma}_{XZ}(j)\right]\right|\right|^2 \tag{3}$$

[4]In other words, Var$[X_t] = \mathbb{E}[X_t X_t'] = I_{d_1}$ and Var$[Z_t] = \mathbb{E}[Z_t Z_t'] = I_{d_2}$. This assumption is relaxed in Section **??**.

for some nonrandom non-negative weights $\{\omega(j)\}$, where $\widehat{\Gamma}_{XZ}(j)$ is the sample cross-correlation between the processes:

$$\widehat{\Gamma}_{XZ}(j) = \frac{1}{T}\sum_{t=j+1}^{T} X_t Z'_{t-j}, \quad \Gamma_{XZ}(j) = \mathbb{E}[X_t Z'_{t-j}], \qquad j = 0, 1, ..., T-1$$

The test statistic is one-sided ($j \geq 0$) as our focus is on a particular direction of causality that is tested by $\mathcal{H}_0$, i.e., the influence of past and present of Z, $\{Z_s; s \leq t\}$, on the present X, $\{X_t\}$.[5] The quadratic forms correspond to the squared ℓ_2-norms of the vectorized sample cross-correlation matrices, or alternatively, their squared Frobenius norms.[6]
Given its formulation, the test statistic can be interpreted as belonging to the class of Portmanteau tests. The statistic has an equivalent formulation in terms of spectral domain (see Lemma A.1 in Appendix A.1), where it is explicit that the weights $\{\omega(j)\}$ are determined by the chosen kernel estimator of the cross-spectrum.

By a perturbation argument, it can be shown that the quadratic form is connected with the Kullback–Leibler divergence between $\{X_t\}$ and $\{Z_{t-j}\}$ as if jointly normally distributed.[7] This result is shown in Appendix A.2. Note that the Kullback–Leibler divergence has been proposed as measure of causality between time series in Geweke (1982), Gourieroux et al. (1987) and Dufour and Taamouti (2010).

The quadratic form, based on the ℓ_2 norm, treats the direction of causality symmetrically. This becomes apparent when decomposing the test statistic: the inner product of cross-correlation matrices generates cross-product terms where variables X and Z enter symmetrically across different time lags. To show it formally, by means of some algebra (Lemma A.2):

[5]See also Hong (1996a) and Hong (2001).

[6]This natural way of generalizing the covariance-variance analysis from univariate to the multivariate dates back to Li and McLeod (1981). Further discussion is in Bouhaddioui and Roy (2006). For the connection of the one-sided test statistic to Bouhaddioui and Roy (2006)'s statistic, please refer to Lemma A.3 in Appendix A. For the connection between the Euclidean norm and the trace, please refer to the results about the trace in chapter 4 of Lütkepohl (1997).

[7]The KL divergence, D_{KL}, betweeen two mean-zero k-multivariate normal with covariance-variance matrices Σ and $\Sigma_0 = I_k$ is defined as: $D_{KL} = \frac{1}{2}\left(\ln(\det(I_k)/\det(\Sigma)) + \text{tr}[\Sigma] - k\right)$

$$
\begin{aligned}
\mathcal{T}_\omega &= \mathcal{T}_{1\omega} + \mathcal{T}_{2\omega} \\
\mathcal{T}_{1\omega} &= \frac{1}{T^2}\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}||X_t||^2||Z_{t-j}||^2 \\
\mathcal{T}_{2\omega} &= \frac{1}{T^2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t}^{T} < X_t, X_s >< Z_{t-j}, Z_{s-j} >
\end{aligned} \tag{4}
$$

The test statistic $\mathcal{T}_\omega$ consists of two components, $\mathcal{T}_{1\omega}$ and $\mathcal{T}_{2\omega}$. The former is the sum of the squared products and the latter is the sum of the cross-products. The sum of cross-products, $\mathcal{T}_{2\omega}$, incorporates the symmetric interaction between time indexes s and t, and thus generates cross-product terms that blend information about both directions of causality. This symmetry hints that the quadratic form, when used for testing the null hypothesis, may not effectively distinguish between opposing directions of causality.

A remark is needed. The distinction between components becomes relevant in understanding the asymptotic properties of the testing strategies based on Hong (1996b) and subsequent work. Intuitively, the latter sum, $\mathcal{T}_{2\omega}$, dominates the former under the null hypothesis, therefore it controls for the size of the test; conversely, the first sum, $\mathcal{T}_{1\omega}$, dominates the other under (a class of) the alternatives, and so regulates the power of the test.[8]

With the decomposition of eq.(4) in mind, I now provide a basic example to clarify the problem of testing the null in eq.(1) with the test statistic based on quadratic forms, $\mathcal{T}_\omega$. In particular, this shows how the cross-product terms may incorporate information about both directions of causality. In the example, we adopt two conditions: i) the marginal independence of the joint process $\{X_t, Z_t\}$, i.e, $X_t \perp\!\!\!\perp X_k, Z_t \perp\!\!\!\perp Z_k, t \neq k$; ii) the process X is conditionally homoskedastic with respect to the joint information set $\mathcal{I}(t-1)$, i.e., $\mathbb{E}[||X_t||^2|\mathcal{I}(t-1)] = \mathbb{E}[||X_t||^2]$.

Proposition 1. INVERSE CAUSALITY IN THE VARIANCE.
Let $\{X_t, Z_t\}$ be marginally i.i.d. processes with finite fourth moments, such that the process $\{X_t\}$ be conditionally homoskedastic with respect to the joint information set, $\mathcal{I}(t-1)$. Under the null in eq.(1), the variance of test statistic, $\mathcal{T}_\omega$, depends on the inverse causality, from X to Z, in the

[8] In Proposition 2, I describe in detail the asymptotic properties of the test under the null. For the power of the test, refer to Theorem 3 and the discussion that follows.

second moments. Indeed, we have that the variance of the cross-products of $\mathcal{T}_{2\omega}$ is:

$$\mathbb{E}[(< X_t, X_s >< Z_{t-j}, Z_{s-j} >)^2] = \begin{cases} (d_1 d_2)^2, & s \geq t-j \\ d_1 \mathbb{E}[||X_s||^2 < Z_{t-j}, Z_{s-j} >^2], & s < t-j \end{cases}$$

If we assume the two processes are independent, $X_t \perp\!\!\!\perp Z_s, \forall s, t$, we have:

$$\mathbb{E}[(< X_t, X_s >< Z_{t-j}, Z_{s-j} >)^2] = (d_1 d_2)^2$$

Proof. To show that the variance of test statistic, $\mathcal{T}_\omega$, depends on the inverse causality in the second moments, it is sufficient to study the first two moments of the two components of the quadratic forms, $Q(j)$, as described in eq.(4) (Lemma A.2).
By the conditional homoskedasticity of $\{X_t\}$ and the law of iterated expectations, we have that the first two moments of the first component, i.e. the squared products, $||X_t||^2||Z_{t-j}||^2$, are:

$$\mathbb{E}[||X_t||^2||Z_{t-j}||^2] = d_1 d_2, \quad \mathbb{E}[(||X_t||^2||Z_{t-j}||^2)^2] = \mathbb{E}[||X_t||^4]\mathbb{E}[||Z_t||^4] = \kappa_1 \kappa_2$$

which means that, for a fixed $j \geq 0$, the squared products that characterizes the first term of the test statistic, $\mathcal{T}_{1\omega}$, are independent with respect to the inverse causality (i.e., the causality from X to Z).
To study the moments of the cross-products, without loss of generality, assume $t > s$.
We have:

$$\begin{aligned} &\mathbb{E}[(< X_t, X_s >< Z_{t-j}, Z_{s-j} >)] = \mathbb{E}[X_t']\mathbb{E}[X_s < Z_{t-j}, Z_{s-j} >)] = 0 \\ &\mathbb{E}[(< X_t, X_s >< Z_{t-j}, Z_{s-j} >)^2] = d_1 \mathbb{E}[||X_s||^2 < Z_{t-j}, Z_{s-j} >^2] \end{aligned}$$

where the first equality is due the null in eq.(1), and the third equality is because of the conditional homoskedasticity of the process X.
While we cannot go further in the case $s < t-j$ without additional assumptions on the dependence structure, in the case of $s \geq t-j$ we can further say that:

$$\mathbb{E}[||X_s||^2 < Z_{t-j}, Z_{s-j} >^2] = \mathbb{E}[||X_s||^2]\mathbb{E}[< Z_{t-j}, Z_{s-j} >^2] = d_1(d_2)^2$$

This highlights that, for a fixed $j \geq 0$, the cross-products that characterizes the second term of the test statistic, $\mathcal{T}_{2\omega}$, are independent with respect to the inverse causality only

when $s \geq t-j$.
If the two processes are independent, then regardless of the time indexes:

$$\mathbb{E}[||X_s||^2 < Z_{t-j}, Z_{s-j} >^2] = \mathbb{E}[||X_s||^2]\mathbb{E}[||Z_{t-j}||^2]\mathbb{E}[||Z_{s-j}||^2] = d_1(d_2)^2$$

which is aligned to the case when $s \geq t-j$. $\square$

Proposition 1 demonstrates that, under the null hypothesis, the test statistic $\mathcal{T}_\omega$ incorporates, through the cross-products in $\mathcal{T}_{2\omega}$, the inverse causality, from X to Z, in the second moments. In particular, we see that, under the additional condition of conditional homoskedasticity, the variance of the component $\mathcal{T}_{2\omega}$ depends on the cross-moments of the joint process $\{X_t, Z_t\}$.
This no longer holds in two cases: i) when the processes are independent, $X_t \perp\!\!\!\perp Z_s, \forall t, s$, which means that the (strict) exogeneity holds both from Z to X and from X to Z, at any lags/leads; ii) when a specific ordering of the time indexes is met, $s \geq t-j$.
Note that, first, the result of Proposition 1 also follows from assuming the independence of present X with respect to the present Z and the past of X and Z, i.e., $X_t \perp\!\!\!\perp Z_{s_1}, X_{s_2}$, with $s_1 \leq t, s_2 < t$ (similar instances of such assumption are found in Hong et al. (2009), Candelon and Tokpavi (2016));[9] second, the presence of such dependencies in the variances of the cross-moments occurs even under the restrictive assumptions on the individual behavior of the sequences, that is the multivariate time series being marginally i.i.d.

Lemma 1. TOY EXAMPLES.

a) Let $\{X_t, Z_t\}$ be a bivariate mean-zero and marginally i.i.d. process, with:

$$Z_t^2 = f(X_{t-1}) + u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2), \qquad X_t \perp\!\!\!\perp u_s, \forall s, t$$

for some measurable function $f(\cdot) : \mathbb{R} \to \mathbb{R}$.
Denote: $\sigma_z^2 = Var[Z_t]$ and $\varphi_f = (\mathbb{E}[X_t^2 f(X_t)] + 1)/\sigma_z^2$.
We have:

$$Var[\mathcal{T}_{2\omega}] = \frac{1}{T^4} \sum_{j=0}^{T-2} \omega^2(j)\big(\Sigma_1(j) + \Delta_1(j)\big)$$

[9] This condition, coupled with the marginal i.i.d., is weaker than statistical independence, $X_t \perp\!\!\!\perp Z_s, \forall s, t$, where the absence of any dependence structure translates into an absence of directionality in terms of causality or, in other words, a symmetric/bidirectional noncausality for the time series, from X to Z and from Z to X.

with:

$$\Sigma_1(j) = \sigma_z^4 (T-j)(T-j-1)$$
$$\Delta_1(j) = \sigma_z^4 \sum_{s,t=j+1, s\neq t}^{T} (\varphi_f - 1)\, \mathbf{1}\{s = t-j-1\}$$

An identical result holds for the scenarios where the DGPs is such:

$$Z_t = g(X_{t-1}) + \epsilon_t, \quad \epsilon_t \sim i.i.d.(0,1), \quad X_t \perp\!\!\!\perp \epsilon_s, \forall s,t$$

for some measurable function $g(\cdot): \mathbb{R} \to \mathbb{R}$,
when we denote: $\varphi_g = (\mathbb{E}[X_t^2 g(X_t)^2] + 1)/\sigma_z^2$ (replacing the term: φ_f).

b) *Let $\{X_t, Z_t\}$ be a bivariate mean-zero process characterized by:*

$$Z_t^2 = \alpha Z_{t-1}^2 + h(X_{t-1}) + u_t, \quad X_t \sim i.i.d.(0,1), \quad u_t \sim i.i.d.(1, \sigma_u^2), \quad X_t \perp\!\!\!\perp u_s, \forall s,t$$

for some measurable function $h(\cdot): \mathbb{R} \to \mathbb{R}$, with $|\alpha| \in (0,1)$.
Denote: $\varrho_h = \mathbb{E}[h(X_s)X_s^2]$, $\mu_e = \mathbb{E}[h(X_{t-1}) + u_t]$, and $\sigma_e^2 = \mathbb{E}[(h(X_{t-1}) + u_t)^2]$.
Finally denote: $\mu_Z = \mathbb{E}[Z_t^2] = \frac{\mu_e^2}{1-\alpha}$, and $\sigma_Z^2 = \mathbb{E}[Z_t^4] = \frac{\sigma_e^2 + 2\alpha\mu_e^2/(1-\alpha)}{1-\alpha^2}$.
We have:

$$Var[\mathcal{T}_{2\omega}] = \frac{1}{T^4} \sum_{j=0}^{T-2} \omega^2(j) \big(\Sigma_2(j) + \Delta_2(j)\big)$$

with:

$$\Sigma_2(j) = \sum_{\upsilon_1=0}^{\tau(j)} \left(\alpha^{|j-\upsilon_1|}\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-\upsilon_1|-1} \alpha^l \right) + \sum_{\upsilon_2=1}^{\Upsilon(j)-\tau(j)} \left(\alpha^{j+\nu_2}\sigma_Z^2 + \mu_Z^2 \sum_{m_1=\nu_2}^{j+\nu_2-1} \alpha^{m_1} \right)$$
$$\Delta_2(j) = \sum_{\upsilon_2=1}^{\Upsilon(j)-\tau(j)} \left(\mu_Z(1+\varrho_h)\alpha^{\nu_2-1} + \mu_Z\mu_e \sum_{m_2=0}^{\nu_2-2} \alpha^{m_2} \right)$$

where: i) $\tau(j)$ is the number of times that $s \geq t-j$ up to T, at a given $j \geq 0$, such that $s \neq t$, and $s,t = j+1$; ii) $\Upsilon(j)$ is the number of times that $s < t-j$, with respect to the same conditions.

In particular, we have:

$$\begin{aligned}\tau(j) &= \frac{(T-j)(T-j-1)}{2} + \mathbb{1}(2j-T-1<0)\cdot j(T-3j/2-1/2) \\ &\quad + \mathbb{1}(2j-T-1\geq 0)\cdot (T-j)(T-j-1)/2 \\ \Upsilon(j) &= (T-j)(T-j-1)-\tau(j)\end{aligned}$$

Proof. The proofs are in Appendix A.3. $\square$

Lemma 1 derives explicitly how the dependencies from X to Z affect the variance of the second term of the test statistic, $\mathcal{T}_{2\omega}$, under two examples of general conditional heteroskedastic DGPs. Both examples pertain to DGPs where X influences Z at the immediate horizon, encompassing dynamics in the first and second-order moments, both linear and non-linear. Unlike the first set of DGPs, the second scenario includes DGPs where the time series Z exhibits autoregressive second-order moments.

In both scenarios, the variance of statistic, $\mathcal{T}_{2\omega}$, consists of two components, $\{\Sigma_i\}_{i=1,2}$ and $\{\Delta_i\}_{i=1,2}$. The first components, $\{\Sigma_i\}_{i=1,2}$, represent the variance of the statistic as if there is no causality from present and past X to present Z. Conversely, the second components, $\{\Delta_i\}_{i=1,2}$, capture the dependence from $\{X_{t-1}\}$ to $\{Z_t\}$, which occurs through the conditional variance (or the conditional mean) of the process Z. In other words, the variance of statistic, $\mathcal{T}_{2\omega}$, is augmented by an additional term due to the presence of inverse causality. As already stated by Proposition 1, these additional terms shows up, at different lags $j \geq 0$, when the time indexes are such that: $s < t-j$.

Note that the additional terms appear in the variances despite there is no causality from past and present Z to present X. Indeed, in both scenarios, we have: $X_t \perp\!\!\!\perp \{X_s, Z_{s+1}\}_{s<t}$.

In conclusions, Lemma 1 displays that the augmentation of the variance of the test statistic depends essentially on three features:

a) the timing of the inverse causality from X to Z;
(in the examples, the functions $\{f(\cdot), g(\cdot), h(\cdot)\}$);

b) the functional forms of the conditional mean and variance of the process Z;
(e.g., whether the first/second moments are autoregressive);

c) the higher order moments of the process X;
(in the examples, the moments denoted by $\{\varphi, \varrho_h\}$).

2.2 An asymmetric Portmanteau test statistic

In light of the considerations following from Proposition 1, my approach entails proposing a new test statistic, which is a modified version of the one-sided test statistic in eq.(2):

$$\begin{aligned}
\mathcal{T}^c_\omega &= \mathcal{T}_\omega - \frac{1}{T^2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T} < X_t, X_s >< Z_{t-j}, Z_{s-j} > \\
&= \mathcal{T}_\omega - \mathcal{C}_\omega \\
&= \mathcal{T}_{1\omega} + \frac{1}{T^2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s\geq t-j}^{T} < X_t, X_s >< Z_{t-j}, Z_{s-j} > \\
&= \mathcal{T}_{1\omega} + \mathcal{T}^c_{2\omega}
\end{aligned} \tag{5}$$

The purpose of the correction term, $\mathcal{C}_\omega$, applied to the test statistic $\mathcal{T}_\omega$, is to eliminate influence of the causality that is inverse of the tested one, as discussed after Proposition 1. Since the introduction of the correction term "breaks" the symmetry in the quadratic forms used in the Portmanteau test, the proposed test statistic of eq.(5) is designated as an *asymmetric* Portmanteau test. The benefits of considering the corrected test statistic will be delineated formally in the following Section, particularly in Proposition 2 and thereafter.

For understanding the rationale of the correction term, another viewpoint emerges when looking at the first two moments of the cross-products as the outcomes of specific predictive regressions.
In the following argument, I suppose the joint process, $\{X_t, Z_t\}$, to be bivariate, and additionally the considered time indexes are such: $t > s$.
Once applied the correction term to $\mathcal{T}_\omega$, the first moment of the remaining cross-products, $\mathcal{T}^c_{2\omega}$, are proportional to the coefficients of the regressions of the form: (for a fixed $s \neq 0$)

$$X_t X_s = \sum_{j=t-s}^{s-1} \phi_j Z_{t-j} Z_{s-j} + e_t$$

with e_t being a mean-zero error term, where $Z_{t-j}, Z_{s-j} \in \mathcal{I}(t-j)$ and $X_s \notin \mathcal{I}(t-j)$.

Indeed, we have:

$$\begin{aligned}\phi_j &= \text{Cov}[(X_tX_s),(Z_{t-j}Z_{s-j})]/\text{Var}[Z_{t-j}Z_{s-j}], \quad s \geq t-j \\ &\propto \mathbb{E}[X_tX_sZ_{t-j}Z_{s-j}]\end{aligned}$$

Conversely, the first moment of the cross-products which constitutes the correction term, $\mathcal{C}_\omega$, are proportional to the coefficients of these auto-regressions: (for a fixed $j \neq 0$)

$$X_tZ_{t-j} = \sum_{l=1}^{s} \varphi_l X_l Z_{l-j} + \epsilon_t$$

with ϵ_t being a mean-zero error term, and such that $X_l, Z_{l-j} \in \mathcal{I}(s)$ and $Z_{t-j} \notin \mathcal{I}(s)$, where:

$$\begin{aligned}\varphi_l &= \text{Cov}[(X_tZ_{t-j}),(X_lZ_{l-j})]/\text{Var}[X_lZ_{l-j}], \quad l < t-j \\ &\propto \mathbb{E}[X_tZ_{t-j}X_lZ_{l-j}]\end{aligned}$$

Two remarks are needed. First, in the latter case the scale of the coefficients, $\{\varphi_l\}$, is influenced by the higher moments of the joint process $\{X_t, Z_t\}$ whereas, in the former case, it is solely determined by the higher moments of the process Z.
Second, the coefficients assess two different implications of the null of interest, by application of the law of iterated expectations with respect to the X or Z processes. To see this, note that, under the null hypothesis of eq.(1), we have $\{\phi_j = 0\}$ and $\{\varphi_i = 0\}$. These conditions channel two distinct implications: i) the former case, $\{\phi_j = 0\}$, implies the following: $\mathbb{E}[X_tX_s] = 0, \ \forall s < t$; ii) the latter case, $\{\varphi_i = 0\}$, implies the following: $\mathbb{E}[X_tZ_{t-j}] = 0, \ \forall j \geq 0$. Simply put, under the null hypothesis $\mathbb{E}[X_t|\mathcal{I}(t-1)] = 0$, the first set of coefficients linearly captures the condition $\mathbb{E}[X_t|Z_t, X_k, k < t] = 0$, whereas the second set of coefficients linearly captures the condition $\mathbb{E}[X_t|Z_l, l < t] = 0$.
Despite the null of interest has similar impacts on the first moment of the cross-products (and thus the first moment of the test statistic), it becomes more evident that is no longer true when considering their second moments.
For simplicity, suppose the null hypothesis of interest holds and further assume X to be conditional homoskedastic: $\mathbb{E}[X_t^2|\mathcal{I}(t-1)] = \mathbb{E}[X_t^2]$. By a parallel argument, the second moment of the remaining cross-products, $\mathcal{T}_{2\omega}^c$, are proportional to the coefficients of the

regressions of the form: (for a fixed $s \neq 0$)

$$X_t^2 X_s^2 = \sum_j \phi_j^{(2)} Z_{t-j}^2 Z_{s-j}^2 + e_t^{(2)}, \quad \phi_j^{(2)} \propto \mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2], \;\; s \geq t-j$$

Parallel to the previous logic, the second moment of the cross-products which constitutes the correction term, $\mathcal{C}_\omega$, are then proportional to the coefficients of these auto-regressions: (for a fixed $j \neq 0$)

$$X_t^2 Z_{t-j}^2 = \sum_{l=1}^{s} \varphi_l^{(2)} X_l^2 Z_{l-j}^2 + \epsilon_t^{(2)}, \quad \varphi_l^{(2)} \propto \mathbb{E}[X_t^2 Z_{t-j}^2 X_l^2 Z_{l-j}^2], \quad l < t-j$$

The condition of conditional homoskedasticity implies:

$$\phi_j^{(2)} = \mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2]/\text{Var}[Z_{t-j}^2 Z_{s-j}^2] = d_1^2 \mathbb{E}[Z_{t-j}^2 Z_{s-j}^2]/\text{Var}[Z_{t-j}^2 Z_{s-j}^2]$$

so that the coefficients of the former regressions, $\{\phi_j^{(2)}\}$, are uniquely functions of the higher moments of the (marginal) process $\{Z_t\}$, mirroring the same property of the scale of the coefficients $\{\phi_j\}$.
Under the same conditions, however, this is not true for the latter regressions as:

$$\varphi_l^{(2)} \propto d_1 \mathbb{E}[Z_{t-j}^2 X_l^2 Z_{l-j}^2] = d_1 \mathbb{E}\Big[\big(\mathbb{E}[Z_{t-j}^2 | \mathcal{I}(t-j-1)]\big) X_l^2 Z_{l-j}^2\Big]$$

where it is evident that these last coefficients are proportional rather to higher order moments of the joint process $\{X_t, Z_t\}$ and are probably mainly shaped by the conditional variance of the process Z. Reflecting the main results of Proposition 1, this elucidates the reason of applying a correction term, so to differentiate between the two types of measure, which translates in making sure that the testing procedure is robust to (higher order) dependencies through the opposite direction of the tested one.

In a baseline scenario where the processes are stationary and directly observable, Section 2.3 outlines the asymptotic properties of the proposed test statistic of Eq.(5) under the null hypothesis of interest. Section 2.4 extends the results of Section 2.3 to a more realistic scenario where the processes are no longer observed but rather estimated from fitting a causal semi-parametric location to some variables of interest. Finally, Section 2.5 studies the asymptotic properties of the proposed test statistic under a general class of alternatives.

2.3 Asymptotics of the Statistic under the Null

Proposition 1, and the discussion which follows in Section **??**, explore the properties of the test statistic based on the ℓ_2-norm of the cross-correlation, by quantifying explicitly the trade-off between modeling the joint multivariate time series $\{X_t, Z_t\}$ and maintaining the directionality (i.e., the asymmetry) of the testing procedure. In other words, to study the asymptotic behaviour of the proposed test statistic T_ω under the null of Eq.(1), the main challenge is to carefully examine the restrictions on the joint and marginal processes, keeping in mind the tested direction of exogeneity.

Facing these considerations, this work prioritizes an agnostic approach to inverse causality, from X to Z, with placing minimal assumptions on the joint process, further formalized in Theorem 1.

Regarding the minimal restrictions on the processes, I give precedence to restrictions on X rather than Z. This is motivated by two concurring reasons.

On one hand, after choosing a set of variables of interest (i.e., the internal variables), practitioners typically define the dynamics of the system in question and, if the model is correctly specified, then its innovations, $\{X_t\}$, are correctly identified. Consequently, it is likely that the practitioner has interest in placing additional restrictions on the internal variables rather than external, since those translate in sharper identifying restrictions and so a better picture of the causality (e.g., restrictions on its variance).

On the other hand, the specification of the dynamic system often excludes a set of variables (the omitted variables). Despite this prior selection, the practitioner could still be interested in testing if the omitted variables, $\{Z_t\}$, should be included in the system. Since this question likely arises from not knowing how or if the external variables cause the internal ones, the best interest is to be as agnostic as possible on the dependence structure between the two sets of variables (internal and external). Conversely, if some prior knowledge about how to parametrize the dynamics of both sets of variables is available, the practitioner can conduct inference by means of the existing testing procedures, as testing for causality in mean would translate directly in testing for the correct specification of the parametric model (as discusses in Section **??**).[10]

Before studying the asymptotic properties of the proposed statistic, I first place some assumptions on the set of weights defined in Eq.(**??**).

[10]Clearly, if the parametrization of the causality from X to Z is incorrect, then the (parametric) model of the joint process $\{X_t, Z_t\}$ would be misspecified. Thus, testing for correct specification would lead to reject the null hypothesis of interest even if it is true. In other words, misspecifying the causality from X to Z would lead to size distortions for such class of testing procedures.

Assumption 1. *The Weighting Function: Let the sequence of weights $\{\omega(j)\}$ be a function of some sequence of integers $M = M(T)$ for which there exists an appropriate square-integrable kernel $k(\cdot) : \mathbb{R} \to [-1, 1]$, continuous at 0 and at all points except for a finite number of points, such that: $\omega(j) = k^2(j/M)$, $k(0) = 1$.*

This assumption is standard in the literature of nonparametric estimation of the spectral density via kernel functions (Hong, 2001). In this literature, the sequence of integers M characterizes the window of the kernel estimation: the larger the M, the larger the window of the kernel (or the larger the magnitude of the weights), the more importance is attached to cross-correlation at distant time lags.[11]. Define the following quantities:

$$
\begin{aligned}
\mu_{\omega,T} &= \mu[\{\omega(j)\}, T] = d_1 d_2 \sum_{j=0}^{T-1} \left(1 - \frac{j}{T}\right) \omega(j) \\
D^{(Hete)}_{\omega,T} &= D^{(Hete)}[\{\omega(j)\}, T] = \frac{d_1^2}{T^2} \sum_{j=1}^{T-2} \omega^2(j) \sum_{s,t=j+1, s\neq t, s\geq t-j}^{T} \gamma_{t-j,s-j} \\
D_{\omega,T} &= D[\{\omega(j)\}, T] = d_1^2 d_2^2 \sum_{j=0}^{T-2} \left(1 - \frac{j}{T}\right)\left(1 - \frac{j+1}{T}\right) \omega^2(j)
\end{aligned}
\tag{6}
$$

with $\gamma_{t,s} = \mathbb{E}[\langle Z_t, Z_s\rangle^2]$, when these quantities exist finite. The first two quantities are approximately the mean and the variance of the correct test statistic, $\mathcal{T}^c_\omega$, under the null hypothesis (see Proposition 2).[12] In particular, the last one corresponds to the asymptotic variance of the test statistic of Eq.(**??**) scaled by the number of observations, $(T \cdot \mathcal{T}_\omega)$, when X and Z are mutually independent (see pg.191-2 in Hong (2001)).
Lemma 1 illustrates how the last quantity, $D_{\omega,T}$, relates to the variance of the test statistic, $\mathcal{T}_\omega$. From the first scenario, we have:

$$
\begin{aligned}
\text{Var}[T \cdot \mathcal{T}_{2\omega}] &= \frac{1}{T^2} \sum_{j=0}^{T-2} \omega^2(j) \Sigma_1(j) + \frac{1}{T^2} \sum_{j=0}^{T-2} \omega^2(j) \Delta_1(j) \\
&= \sigma_Z^4 D_\omega + \frac{1}{T^2} \sum_{j=0}^{T-2} \omega^2(j) \Delta_1(j)
\end{aligned}
$$

where is explicit that the variance of the sum of the cross-products, $\mathcal{T}_{2\omega}$, consists of two

[11]Regarding the smoothing parameter M, please refer also to the discussion at the end of Section 2.3

[12]For the univariate version, please refer to Eq.(22) and the discussion thereafter in Hong (2001). For the multivariate version, please refer to pg.511-4 of Bouhaddioui and Roy (2006).

components: the variance of the statistic as if X and Z are independent, plus the augmented term due the presence of inverse causality.
To understand their asymptotic order, note that, under Assumption 1, those quantities are proportional to M, or in other words, we have $\mu_{\omega,T} = O(M)$, $D_{\omega,T} = O(M)$, and $D^{(Hete)}_{\omega,T} = O(M)$, provided that $\gamma_{t,s} = O(1), \ \forall t, s$. This is formally shown in the proof of Proposition 2, in Appendix B.1.

Similar to Proposition 1, the following proposition studies the first two moments of the corrected test statistic, $\mathcal{T}^c_\omega$.

Proposition 2. *Suppose Z has finite fourth moments. Suppose Assumption 1 holds, with $\frac{M}{T} \to 0$, as $T, M \to \infty$. We have the following:*

i) If: $\mathbb{E}[X_t X_t' | \mathcal{I}(t-1)] = \mathbb{E}[X_t X_t']$, then: $\mathbb{E}[T \cdot \mathcal{T}_{1\omega}] = \mu_{\omega,T}$.
In addition, if: $\mathbb{E}[(X_t X_t') \otimes (X_t X_t') | \mathcal{I}(t-1)] = \mathbb{E}[(X_t X_t') \otimes (X_t X_t')]$,
then: $Var[T \cdot \mathcal{T}_{1\omega}] = O(M^2/T)$.
Thus, in mean-squared error (ℓ_2-convergence):

$$\lim_{T\to\infty} \frac{T \cdot \mathcal{T}_{1\omega} - \mu_\omega}{\sqrt{D^{(Hete)}_{\omega,T}}} = 0$$

ii) Under the null hypothesis $\mathcal{H}_0$ of Eq.(1), we have: $\mathbb{E}[T \cdot \mathcal{T}^c_{2\omega}] = 0$.
In addition, if: $\mathbb{E}[X_t X_t' | \mathcal{I}(t-1)] = \mathbb{E}[X_t X_t']$, then:

$$Var[T \cdot \mathcal{T}^c_{2\omega}] = D^{(Hete)}_{\omega,T}$$

When: $\mathbb{E}[Z_t Z_t' | \{Z_s; s < t\}] = \mathbb{E}[Z_t Z_t']$, then: $Var[T \cdot \mathcal{T}^c_{2\omega}] = \sum_{j=0}^{T-2} \omega^2(j) \sum_{l=1}^{\tau(j)} \frac{d_1^2 d_2^2}{T^2}$, where $\tau(j)$ is defined as in Lemma 1.

Proof. The proofs are in Appendix B.1. □

The proposition highlights the advantages of the corrected version of the statistics. Under the conditional homoskedasticity and conditional homokurtisis of the process X with respect to the joint information set, we observe that the first two moments of the test statistics do not incorporate the inverse causality from X to Z. Specifically, the moments are solely determined by either the set of weights $\{\omega(j)\}$ or, at most, by a particular set of higher moments of the process Z alone, $\{\gamma_{t,s}\}$. Note that, when the process Z is fourth-order stationary, we have: $\gamma_{t-j,s-j} = \gamma_{t,s}$.

Requiring the process X to be a martingale with respect to the joint information set in the higher moments should be interpreted as pivotal restrictions to distinguish and isolate various causality channels, and so testing for the null hypothesis of interest. In particular, i) the conditional homoskedasticity is essential to isolate the effect of the null hypothesis to the center of the test statistic, or equivalently, to associate the mean of the corrected test statistic, $\mathcal{T}_\omega^c$, to the mean of the sum of the cross-products, $\mathcal{T}_2^c$; ii) the conditional homokurtosis serves to bound the variance of the sum of the squared products, $\mathcal{T}_{1\omega}$, or in other words, to let the sum of cross-products dominate stochastically the former under the null hypothesis.

Two additional remarks are needed. First, despite the stringency of these assumptions, they uphold the directionality of exogeneity, without placing restrictions on the inverse causal effect (i.e., from X to Z). Second, these assumptions could be considered weaker than the usual conditions under which this class of test statistics are studied, such as independence (Hong, 1996b) and past independence (Candelon and Tokpavi, 2016), or possibly comparable to some other conditions, such as approximately $q-$dependence (Hong and Lee, 2005).

Define the following quantity: $\Lambda_{s,t}^{(1)} = \sum_{j=t-s}^{s-1} \omega(j)||X_s||\langle Z_{t-j}, Z_{s-j}\rangle$. Building on the preceding findings, I now state the first main result.

Theorem 1. *Suppose the process $\{X_t\}$ is such that:*

$$\mathbb{E}[X_t X_t' | \mathcal{I}(t-1)] = \mathbb{E}[X_t X_t'], \quad \mathbb{E}[(X_t X_t') \otimes (X_t X_t') | \mathcal{I}(t-1)] = \mathbb{E}[(X_t X_t') \otimes (X_t X_t')]$$

Further, suppose the time series $\{Z_t\}$ has finite eight-order moments, the joint process $\{X_t, Z_t\}$ is strictly stationary, and Assumption 1 holds with $\frac{M^2}{T} \to 0$, as both $T, M \to \infty$. Under the null hypothesis in Eq.(1):

$$\mathcal{H}_0 : \ \mathbb{E}[X_t | \{X_s\}_{s<t}, \{Z_s\}_{s \le t}] = 0$$

we have:

$$\frac{T \cdot \mathcal{T}_\omega^c - \mu_{\omega,T}}{\sqrt{D_{\omega,T}^{(Hete)}}} \xrightarrow{d} \mathcal{N}(0,1)$$

If additionally the joint process $\{X_t, Z_t\}$ satisfies:

$$|\mathbb{E}\langle \Lambda^{(1)}_{1,t}, \Lambda^{(1)}_{1+i_1,t}\rangle| = O(i_1^{-2}), \quad |Cov[||Z_1||^2, ||Z_{1+i_2}||^2]| = O(i_2^{-2}), \quad i_1, i_2 \to +\infty$$

the asymptotic normality of the test statistic, $\mathcal{T}^c_\omega$, holds with $\frac{M}{T} \to 0$, as $T, M \to \infty$.

Proof. The proof consists of two parts.
The first part is by direct application of Proposition 2, as it is shown that the statistic $\mathcal{T}^c_\omega$ is correctly centered and standardized. The second part of the proof is in Appendix B.2. It relies on appealing to Brown (1971)'s martingale central limit theorem, which guarantees the desired asymptotic properties. As previously announced in Proposition 2, due the ℓ_2-convergence to zero of the first term of the statistic, $\mathcal{T}_{1\omega}$, the asymptotic normality of $\mathcal{T}^c_\omega$ $(= \mathcal{T}_{1\omega} + \mathcal{T}^c_{2\omega})$ under the null of interest is driven by the second term, $\mathcal{T}^c_{2\omega}$. $\square$

There are two notable improvement to the testing strategies following Hong (1996b).

First, the class of tests stemming from Hong (1996a) and Hong (1996b) is usually studied under the null hypothesis of statistical independence, which is strictly contained in the null hypothesis of Eq.(1): testing the null of statistical independence is sufficient but not necessary for testing the null of interest. This in turn means that the benchmark testing strategies based on the squared cross-correlation (i.e., the quadratic forms in Eq.(3)), might fail to have the desirable asymptotic properties under the null hypothesis of interest, as they include the inverse causality in their higher moments (see Proposition 1). In other words, under some scenarios, the inference based on the Portmanteau tests *à la* Hong (1996b) might be jeopardized because of potential size distortions. In robustifying the test against such distortions, I study a corrected version of this statistic, i.e., $\mathcal{T}^c_\omega$, which accounts for the directionality of the hypothesis of interest, by a correction term that "breaks" the symmetry in the quadratic forms. Theorem 1 shows that the asymptotic normality of the corrected test statistic, $\mathcal{T}^c_\omega$, is driven uniquely by the martingale properties of the process X with respect to the joint information set.

Second, a potentially valid statistic for testing the null of interest can be thought to belong to the class of tests designed to test the martingale difference property of a process, e.g. Escanciano and Velasco (2006) and Hong and Lee (2005). However, testing the null of Eq.(1) following their testing strategies would require either i) the joint process $\{X_t, Z_t\}$ to be a martingale difference sequence, or ii) modelling the joint process $\{X_t, Z_t\}$ such that it is possible to isolate the conditional mean of X. The former case is strictly included in the null of interest, thus the aforementioned discussion holds as well. In the latter case, before conducting inference, one would possibly require to specify the conditional mean of X, together with placing high-level assumptions (e.g., Assumption A2-A3 in Hong and Lee, 2005) that might less transparent than the ones of Theorem 1, where the restrictions on the conditional moments are rather explicit.

Aligned with the discussion of Proposition 1, these latter conditions can potentially be relaxed at the cost of constraining either the marginal behaviour of the process Z or the dependencies of the joint process (e.g. imposing mixing conditions, see Dedecker et al., 2007).

The main difference with respect to the literature stemming from Hong (1996b) is the imposition of a more stringent asymptotic rate for M, which is required to diverge at a slower rate than $\sqrt{T}$. This slower divergences comes with a cost in terms of the power of the statistic (see Theorem 3). However, note that the first part of Theorem 1 comes with the very minimal restrictions on the conditioning variable Z. By placing some additional mild restrictions on the dependence structure of the joint process, the second part of Theorem 1 highlights that the asymptotic properties of the test are still preserved with respect to a faster asymptotic rate for M, that is the standard in the literature.

In practice, the smoothing parameter M dictates the convergence of the weighted sum of independent chi-squared variables to a normal distribution, under the null hypothesis. When M is finite, as the Portmanteau-type test statistic sums over the first M squared covariances, standard result in the literature is its limiting distribution being a weighted sum of independent chi-squared random variables (Box and Pierce, 1970; Francq and Raïssi, 2007). As M increases with the sample size, more non-zero weights are spread to more covariances, meaning the sum of such independent chi-squared distributed tends to be normally distributed, by a classic central limit argument. Under the alternatives, Theorem 3 formalizes the connection between the smoothing parameter M and the power of the test statistic.

The trade-off is summarized as follows: on one hand, the rate for M needs to be fast enough to guarantee the approximating asymptotic normality of the statistic under the null, on the other hand, it needs to be slow enough to assure a good asymptotic power of the statistic.

2.4 Estimated Processes

When analyzing the variables of interest for a broad understanding of the dynamics, usual practice is to fit models to the time series processes, so that the information about the se-

rial dependence is summarized by the fitted residuals.
In our context, it translates in viewing the processes X and Z as estimated innovations, based on a particular model applied to the observed data $\{W_{1,t}, W_{2,t}; t = 1, ..., T\}$ of dimensions d_1 and d_2, respectively. I suppose both observed processes to be causal or, to put it differently, I assume that $\{W_{1,t}\}$ (and $\{W_{2,t}\}$) is represented by a causal function of $\{X_t\}$ (and $\{Z_t\}$). Formally, I consider the observed processes to have the following general class of causal conditional mean models:

$$\begin{aligned} W_{1,t} &= \mu_X(\theta_1^0, \{X_s; s < t\}) + X_t \\ W_{2,t} &= \mu_Z(\theta_2^0, \{Z_s; s < t\}) + Z_t \end{aligned} \tag{7}$$

where, $\mu_X(\theta_1^0, \cdot) \in \mathbb{R}^{d_1}$, $\mu_Z(\theta_2^0, \cdot) \in \mathbb{R}^{d_2}$ are vectors whose entries are time-varying measurable known functions with respect to two finite time-invariant parameters, θ_1^0 and θ_2^0.
The innovation process $\{X_t\}$ is defined such that:

$$\mathbb{E}[X_t | \{X_s; s < t\}] = 0$$

so that, by design, $\{\mu_X(\cdot, \cdot)\}$ is the multivariate conditional mean of W_1 with respect its own past. Put differently, under the specification of Eq.(7), the time series $\{X_t\}$ are the innovations of the conditional mean of W_1, and so a martingale difference sequence with respect to the marginal filtration, $\{X_s; s < t\}$.
A brief remark is needed. The restriction on the process W_2 should be interpreted in the context of the two-step approach in Hong (1996b), where noncausality is tested after fitting separate models. Furthermore, while it is possible to account for $W_{2,t}$ as a function of the joint past $\{W_{1,t}, W_{2,t}\}$, doing so would implicitly model inverse causality, which might not be desirable since W_2 are to be considered the omitted variables from the conditional mean model of W_1.
Clearly, many of the existing multivariate time series models that capture the first conditional moment of stationary processes are encompassed by the specifications of Eq.(7). For a process to fall within this category, the only key condition is that, by using a time-invariant and finite-dimensional set of parameters, the innovations of the first moment are correctly captured, as this last aspect is particularly crucial when testing for the hypothesis of interest.

Suppose the practitioner has $\sqrt{T}-$consistent estimators, $\{\widehat{\theta}_i\}_{i=1,2}$, of the true parameters for the prespecified functional forms in Eq.(7) (e.g., Least Squares or Maximum

Likelihood estimators for ARMA models). Given the sample data, the estimated innovations are functions of the limited past, $\widehat{\mathcal{I}}(t-1)$, and such estimators: $\widehat{X}_t(\widehat{\theta}_1), \widehat{Z}_t(\widehat{\theta}_2)$. These are the estimated pseudo-version of the innovations with arbitrary starting values, since we do not observe the infinite past of the time series, i.e., $\mathcal{I}(t-1)$.
Define the standardized innovations of the models in Eq.(7) as follows:

$$U_t = (\Gamma_X)^{-1/2} X_t, \quad V_t = (\Gamma_Z)^{-1/2} Z_t, \quad \text{s.t.} \quad \mathbb{E}[||U_t||^2] = d_1, \ \mathbb{E}[||V_t||^2] = d_2 \tag{8}$$

where the covariance-variance matrices are defined as: $\Gamma_X = \mathbb{E}[X_t X_t'], \ \Gamma_Z = \mathbb{E}[Z_t Z_t']$, with their feasible empirical counterpart being $\widehat{\Gamma}_X$ and $\widehat{\Gamma}_Y$, respectively.
Following the definitions in Eq.(8), the standardized residuals are:[13]

$$\widehat{U}_t = \left(\widehat{\Gamma}_X\right)^{-1/2} \widehat{X}_t, \quad \widehat{V}_t = \left(\widehat{\Gamma}_Z\right)^{-1/2} \widehat{Z}_t, \quad t = 1, ..., T \tag{9}$$

Parallel to Eq.(4)-(5), here below I consider the test statistic with respect to the standardized estimated processes X and Z:

$$\begin{aligned}
\widehat{\mathcal{T}}_\omega^c &= \widehat{\mathcal{T}}_\omega - \frac{1}{T^2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s<t-j}^{T} < \widehat{U}_t, \widehat{U}_s >< \widehat{V}_{t-j}, \widehat{V}_{s-j} > \\
&= \widehat{\mathcal{T}}_\omega - \widehat{\mathcal{C}}_\omega = \widehat{\mathcal{T}}_{1\omega} + \widehat{\mathcal{T}}_{\omega 2}^c \\
\widehat{\mathcal{T}}_\omega &= \sum_{j=0}^{T-1}\omega(j)||\widehat{\Gamma}_{UV}(j)||_F^2
\end{aligned} \tag{10}$$

For the formal connection with Eq.(3), refer to Lemma A.3. Similar to Eq.(6), define also:

$$\widehat{D}_{\omega,T}^{(Hete)} = \frac{d_1^2}{T^2}\sum_{j=1}^{T-2}\omega^2(j)\sum_{s,t=j+1,s\neq t,s\geq t-j}^{T}\widehat{\gamma}_{t-j,s-j}, \qquad \widehat{\gamma}_{t-j,s-j} = \frac{1}{T}\sum_{t=1}^{T} < \widehat{V}_{t-j}, \widehat{V}_{s-j} >^2$$

Theorem 2. *Suppose the processes $\{W_{i,t}\}_{t=1,...,T}^{i=1,2}$ admit the causal representation of Eq.(7). Suppose the assumptions of Theorem 1, and the regularity conditions of Eq.(14)-(15) in Appendix B.3 hold true. Let $\{\widehat{\theta}_i\}_{i=1,2}$ be $\sqrt{T}-$consistent estimators of the true parameters $\{\theta_i^0\}_{i=1,2}$.*

[13] Because of the functional dependence of the processes $\widehat{X}$ and $\widehat{Z}$ on the estimators $\{\hat{\theta}_i\}_{i=1,2}$ is presumed, I suppress the redundant notation for a cleaner exposition. See Appendix B.3 for further details.

Under the null hypothesis in Eq.(1) we have:

$$\frac{T \cdot \widehat{\mathcal{T}}_{\omega}^{c} - \mu_{\omega,T}}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Proof. The proofs are in Appendix B.3. □

Theorem 2 show that, if the processes are estimated rather than directly observed, the results of Theorem 1 still applies, up to some additional mild assumptions. In other words, if the practitioner uses a plug-in in Eq.(10)), where the standardized innovations are estimated with respect to a causal parametric specification, the limiting null distribution is a standard normal.

The additional conditions under which the estimation effect is asymptotically negligible are twofold: i) the parametric model is correct, meaning that the innovations of interest are estimated at standard rates, i.e. $\sqrt{T}$-consistency; ii) appropriate smoothness of the model specifications, which guarantees uniform ℓ_2-convergence and boundedness of the derivatives with respect to the parameters (refer to Eq.(14)-(15) in Appendix B.3). Both types of conditions are standard in the conditional mean specification literature (see the discussion in Wang et al., 2022) and the literature stemming from Hong's work (Hong and Lee, 2005; Leong and Urga, 2023).

To illustrate the broad applicability of Theorem 2, I present one example demonstrating how the proposed testing strategy can be easily adapted to test for second-order noncausality, or Granger noncausality in the conditional variance (Granger, 1969; Comte and Lieberman, 2000). Suppose the condition of Eq.(1) holds. In such context, under the assumptions about the model specifications of Eq.(7), the null hypothesis of second-order noncausality from W_2 to W_1 is:

$$\mathcal{H}_0^{VNC} : \mathbb{E}[X_t X_t' | \{X_s\}_{s<t}, \{Z_s\}_{s<t}] = \mathbb{E}[X_t X_t' | \{X_s ; s < t\}]$$

Additionally, if the model specifications of Eq.(7) is such that:

$$\mathbb{E}[X_t X_t' | \{X_s ; s < t\}] = \Gamma_X$$

once defined:

$$X_t^{\dagger} = \text{vech}[X_t X_t'] - \text{vech}[\Gamma_X]$$

the null hypothesis of second order noncausality is equivalent to testing for the following null hypothesis:

$$\widetilde{\mathcal{H}}_0^{VNC} : \mathbb{E}[X_t^\dagger | \{X_s^\dagger\}_{s<t}, \{Z_s\}_{s\leq t}] = 0$$

where is evident the connection with the null hypothesis of Eq.(1).
This example is associated with Cheung and Ng (1996) and Hong (2001): both papers study a Portmanteau-type statistic that sums, at different lags, the squared cross-correlation of the standardized (centered) squared residuals. Other instances include Tchahou and Duchesne (2013), Aguilar and Hill (2015) and Leong and Urga (2023).

2.5 Consistency under the Alternatives

In this section, we discuss the asymptotic properties of the test under a class of alternatives. Define $\kappa_{mrmr,XY}(t,j,k,l)$ being the fourth-order cumulant of the time series $\{X_{m,t}, Z_{r,t-j}, X_{m,t-k}, Z_{r,t-l}\}$, where $X_{i,t}, Z_{i,t}$ are the i^{th} entries of X_t, Z_t, respectively. I refer to the following condition as the absolute summability of the fourth-order cumulants: $\sum_{m,r=1}^{d_1,d_2} \sum_{j,k,l=-\infty}^{\infty} \kappa_{mrmr,XZ}(0,j,k,l) < \infty$.

Theorem 3. *Suppose the assumptions of Theorem 2 hold. Suppose $\{X_t, Z_t\}$ is a jointly fourth-order stationary process with absolute summability of the fourth-order cumulants. Suppose further: $\exists j > 0$, such that $||\Gamma_{XZ}(j)|| \neq 0$, with $\sum_{j=1}^{\infty} ||\Gamma_{XZ}(j)||^2 < \infty$,*
We have:

$$\frac{M^{1/2}}{T}\left(\frac{T\widehat{\mathcal{T}}_\omega^c - \mu_{\omega,T}}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}}\right) \xrightarrow{p} \Delta \sum_{j=1}^{\infty} \left|\left| vec\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2$$

for a finite $\Delta > 0$. By the asymptotic rates, equivalently:

$$\lim_{T,M\to\infty} Pr\left(\left|\frac{T\widehat{\mathcal{T}}_\omega^c - \mu_{\omega,T}}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}}\right| > K\right) \to 1, \quad \forall K \in \mathbb{R}$$

Proof. The proofs are in Appendix B.4. □

Based on the ℓ_2-convergence of the covariance estimator, Theorem 3 shows the consistency of the proposed test statistic.

Once scaled by a rate that decays to zero ($M^{1/2}/T \to 0$, as $M, T \to \infty$), the statistic converges in probability to the sum of ℓ_2-norm of the cross-correlation at different lags (up to a positive finite scale). This means that, under these fixed alternatives, the test statistic has the desired asymptotic power, with explosive rate, $T/M^{1/2}$: the slower M grows, the faster the statistic, $\widehat{\mathcal{T}}_{\omega}^{c}$, will go to infinity, and the test will be more powerful.[14] In this regard, Theorem 3 sets an upper bounds on the rate at which M should grow asymptotically, as explained at the conclusion of Section 2.3.
The test, however, has no power against the alternatives with zero cross-correlation when $\mathbb{E}[X_t|\{X_s, Y_s; s < t\}] \neq 0$, i.e. uncorrelated and non-martingales processes. To put it simply, the test may have low power against some types of nonlinear effects from past Z to present X that are not reflected in any linear association. This limitation in terms of the test's power is common in the testing strategies involving a Portmanteau-type statistics (e.g., Hong, 2001; Bouhaddioui and Roy, 2006). As well, the assumptions of fourth-order stationarity and the absolute summability of the fourth order joint cumulants are standard in the tests following Hong (1996b), as being not much restrictive.[15]

Two technical remarks about Theorem 3 are in order.
First, as mentioned in Section **??**, the proof of Theorem 3 hinges on establishing that the sum of the squared products, $\widehat{\mathcal{T}}_{1\omega}$, regulates the power of the test. In details, under the considered class of alternatives, the sum of the squared products stochastically dominates the corrected sum of the cross-products, $\widehat{\mathcal{T}}_{2\omega}^{c}$. This result is reached by appealing to Theorem 6 in Hannan (1970) (pg.210), which follows from a Isserlis (1918)-type argument.[16] Thus, to prove the consistency, two assumptions on the joint process are essential: the fourth-order stationarity and, particularly, the absolute summability of the fourth-order cumulants. With the joint process being sufficiently close to a multivariate normal distribution, these assumptions ensure that the ℓ_2-norm difference between the covariance estimator and its population counterpart goes asymptotically to zero, as mean-square convergence takes effect *"before squaring"* the covariances.
It is important to realize that, albeit the fourth-order cumulants are asymptotically negligible under the conditions of Theorem 3, they play a determinant role in Theorem 1-2, due to Proposition 2: if the process X is a martingale in the higher moments, the cumulants drive the asymptotic behavior of the test statistic under the null hypothesis of interest. In

[14] In particular, note that the explosive rate is identical to the one in Bouhaddioui and Roy (2006)'s Theorem 2. Thus, the discussion of pg. 517 in Bouhaddioui and Roy (2006) can be still applied.

[15] They allows for a wide class of processes. See the discussion at pg.846 in Hong (1996a).

[16] Equivalently, one can appeal to Eq.(5.3.20) of Priestley (1981) and the discussion thereafter.

fact, it should be noted that the proposed correction term in Eq.(5) specifically targets a subset of those cumulants associated with inverse causality.
Second, Escanciano and Velasco (2006)'s approach relies on an asymptotic theory which complies to Hannan (1970)'s rationale. They propose a testing strategy based on a measure of deviations from the zero cross-spectrum in terms of quadratic (Cramér–von Mises) norm, a statistic resembling Hong and Lee (2005)'s. Under the null hypothesis of m.d.s., they show that the sum of sample auto-covariances weakly converges in ℓ_2-norm to a Gaussian process (Escanciano and Velasco (2006)'s Theorem 1).[17] This convergence takes effect *"before squaring"* the covariances: as the covariances are entering squared in their proposed test statistic, it then converges in distribution to a weighted sum of independent χ^2_1 random variables. Conversely, Hong (1996a)'s asymptotic theory focuses on the quadratic forms (Eq.(4)). *"After squaring"* the covariances, Hong (1996a)'s main result revolves around showing that the sum of cumulants converges to a normal distribution, eventually driving the desired asymptotic properties of his test under the null hypothesis. In summary, the fundamental distinction between the two asymptotic theories is that the former approach, *à la* Escanciano and Velasco, is based on the convergence driven by the squared products (i.e., *"before squaring"*), whereas the latter approach, *à la* Hong, depends on convergence driven by the cross-products (i.e., *"after squaring"*). Taking these considerations into account, my proposed testing strategy underscores the importance of fourth-order cumulants in cases where there is no available information on modeling the omitted variables Z, or put differently, where there is no knowledge of the causality channel opposite to the one being tested.

3 Simulation study

The purpose of this section is to offer Monte Carlo evidences on the impact of the proposed correction term, $\mathcal{C}_\omega$, defined in eq.(5). This entails studying the finite sample characteristics of two types of test statistics: the benchmark one *à la* Hong (1996a), $\mathcal{T}_\omega$, and its corrected

[17] They consider a particular form of sample auto-covariances, as they are a *"generalization of the usual autocovariances to measure the conditional mean dependence in a nonlinear time series framework"* (pg.155 in Escanciano and Velasco, 2006). The convergence of the process is on the Hilbert space of all square integrable functions equipped with a proper inner product. For further details, refer to pg.158-159 in Escanciano and Velasco (2006).

version, $\mathcal{T}_{\omega}^{c}$. In details, the benchmark testing procedure is defined as:

$$\frac{T \cdot \widehat{\mathcal{T}}_{\omega} - \mu_{\omega,T}}{\sqrt{D_{\omega,T}}} \tag{11}$$

where the sample quantities are specified in eq.(6) and eq.(10), with their population counterparts in eq.(2). As the corrected test, I refer to the test statistic that have been discussed in Theorem 2-3.

Since the asymptotic theory emphasizes the correction term's importance under the null hypothesis of interest, the focus of the simulations primarily lies on studying the empirical rejection rates for DGPs where the null of eq.(1) holds true. This set of results is found in Section 3.1. Regarding the alternatives, I provide the empirical rejection rates for a limited range of DGPs, since Theorem 3 shows that the two test statistics should share similar properties in terms of power. The simulations for this scenario are presented in Section 3.2.

3.1 Under the null

In this set of experiments, I consider three families of DGPs where the process X is defined as a standardized strong white noise (SWN):

$$X_t = \epsilon_x, \quad \epsilon_x \sim i.i.d.(0,1)$$

meaning that, for all the DGPs, the null hypothesis of interest holds true, as there is no causality in mean from past and present Z to present X. In line with the conclusions drawn after Lemma 1, the differences between the three scenarios can be distilled into two key aspects: i) the form of the causality from past X to present Z, i.e. the inverse causality channel, and ii) the properties of the process Z, in terms of its conditional mean and variance. Hence, I consider three representative classes of DGPs for the univariate processes X and Z:

a) DGP 1A: LINEAR-IN-MEAN

$$Z_t = \alpha Z_{t-1} + \beta X_{t-1} + \epsilon_z, \quad \epsilon_z \sim i.i.d.(0,1)$$

b) DGP 2A: SQUARED-IN-MEAN

$$Z_t = \alpha Z_{t-1} + \beta X_{t-1}^2 + \epsilon_z, \quad \epsilon_z \sim i.i.d.(0,1)$$

c) DGP 3A: SQUARED-IN-VARIANCE

$$Z_t = \alpha Z_{t-1} + \sigma_z \epsilon_z, \quad \epsilon_z \sim i.i.d.(0,1), \quad \sigma_z^2 = 0.4 + \beta X_{t-1}^2$$

with the parameters ranging: $\alpha = \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, $\beta = \{0, 0.2, 0.4, 0.6, 1, 2\}$.
Two remarks are needed. First, when $\beta = 0$, there is no causality in both directions, i.e. independence between processes X and Z. Second, in all DGPs, the process Z is conditionally homoskedastic. This last choice is dictated by the parsimony of parametrizing the DGPs. Indeed, as a result of Proposition 2, the variance of the test statistic is: $D_{\omega,T}^{(Hete)} = (d_1 d_2/T)^2 \sum_{j=0}^{T-2} \omega^2(j) \sum_{l=1}^{\tau(j)}$, where $\tau(j)$ is defined as in Lemma 1, where is clear the parallel with the benchmark test statistic's variance, D_ω (see eq.(6)).

Table 1: Rejection frequencies for DGP1A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and benchmark ($\mathcal{T}_\omega$), when the time series are generated by DGP1A; sample size, $T = 700$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12$ $\approx 2(10T)^{1/5}$ $\approx 2\ln T$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.026	0.015	0.01	0.015	0.023	0.019	0.053	0.048	0.039	0.05	0.049	0.046
$\alpha = 0.3$	0.019	0.018	0.019	0.033	0.022	0.025	0.039	0.043	0.068	0.053	0.04	0.05
$\alpha = 0.4$	0.03	0.03	0.019	0.033	0.032	0.019	0.065	0.058	0.056	0.062	0.062	0.045
$\alpha = 0.5$	0.048	0.04	0.023	0.043	0.035	0.036	0.08	0.075	0.055	0.078	0.073	0.063
$\alpha = 0.6$	0.045	0.04	0.052	0.039	0.042	0.038	0.076	0.075	0.089	0.076	0.073	0.083
$\alpha = 0.7$	0.053	0.059	0.055	0.058	0.055	0.063	0.073	0.09	0.09	0.079	0.089	0.1

$M = 30$ $\approx 5(10T)^{1/5}$ $\approx \sqrt{T}$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.016	0.015	0.009	0.009	0.023	0.016	0.055	0.045	0.033	0.045	0.053	0.053
$\alpha = 0.3$	0.02	0.018	0.019	0.028	0.023	0.019	0.052	0.056	0.059	0.063	0.053	0.056
$\alpha = 0.4$	0.028	0.029	0.019	0.039	0.025	0.018	0.072	0.072	0.056	0.075	0.066	0.053
$\alpha = 0.5$	0.04	0.038	0.028	0.043	0.043	0.023	0.085	0.088	0.058	0.093	0.093	0.068
$\alpha = 0.6$	0.048	0.039	0.056	0.052	0.045	0.045	0.095	0.089	0.103	0.108	0.098	0.096
$\alpha = 0.7$	0.053	0.078	0.065	0.068	0.07	0.069	0.13	0.129	0.12	0.123	0.132	0.128

$M = 100$ $\approx 4\sqrt{T}$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.02	0.018	0.018	0.02	0.023	0.023	0.063	0.063	0.058	0.052	0.075	0.068
$\alpha = 0.3$	0.023	0.015	0.03	0.023	0.036	0.02	0.059	0.066	0.085	0.065	0.098	0.066
$\alpha = 0.4$	0.038	0.026	0.025	0.036	0.029	0.042	0.086	0.093	0.082	0.088	0.076	0.082
$\alpha = 0.5$	0.053	0.062	0.059	0.043	0.076	0.05	0.126	0.128	0.115	0.106	0.138	0.126
$\alpha = 0.6$	0.086	0.082	0.073	0.055	0.085	0.086	0.175	0.153	0.139	0.139	0.17	0.165
$\alpha = 0.7$	0.095	0.126	0.129	0.12	0.123	0.128	0.176	0.21	0.212	0.196	0.223	0.219

In light of the discussion about the higher-order moments of the joint process (see Lemma 1 and Section 2.2), I consider the SWNs to be generated by a multivariate t-distribution with 6 degrees of freedom: $(\epsilon_x, \epsilon_z) \sim t_6(0, I_2)$.[18] The degrees of freedom for the multivariate t-distribution are selected to be appropriate for macroeconomic time series analysis, especially having in mind the Structural VAR (SVAR) literature, and so aligning the design of the experiments with the empirical application in Section 4. My primary example is Brunnermeier et al. (2021), where the structural shocks, identified through heteroskedasticity, are estimated as a scaled t-variate with 5.7 degrees of freedom.
All the results are with respect to the 5% nominal significance level.

Table 2: Rejection frequencies for DGP2A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}^c_\omega$) and benchmark ($\mathcal{T}_\omega$), when the time series are generated by DGP2A; sample size, $T = 700$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12$ $\approx 2(10T)^{1/5}$ $\approx 2\ln T$

	Corrected						Benchmark					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.026	0.015	0.033	0.026	0.03	0.029	0.053	0.039	0.05	0.055	0.049	0.042
$\alpha=0.3$	0.019	0.023	0.032	0.022	0.025	0.038	0.039	0.049	0.055	0.045	0.046	0.058
$\alpha=0.4$	0.03	0.026	0.039	0.026	0.045	0.039	0.065	0.053	0.063	0.053	0.056	0.045
$\alpha=0.5$	0.048	0.042	0.046	0.045	0.038	0.03	0.08	0.072	0.063	0.07	0.068	0.052
$\alpha=0.6$	0.045	0.058	0.056	0.042	0.078	0.045	0.076	0.093	0.085	0.055	0.098	0.065
$\alpha=0.7$	0.053	0.05	0.043	0.058	0.072	0.062	0.073	0.09	0.08	0.093	0.108	0.069

$M = 30$ $\approx 5(10T)^{1/5}$ $\approx \sqrt{T}$

	Corrected						Benchmark					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.016	0.013	0.026	0.029	0.028	0.023	0.055	0.033	0.045	0.05	0.046	0.04
$\alpha=0.3$	0.02	0.019	0.026	0.025	0.038	0.033	0.052	0.055	0.056	0.053	0.05	0.05
$\alpha=0.4$	0.028	0.029	0.049	0.036	0.04	0.039	0.072	0.069	0.073	0.062	0.06	0.055
$\alpha=0.5$	0.04	0.039	0.058	0.043	0.036	0.038	0.085	0.086	0.092	0.078	0.078	0.073
$\alpha=0.6$	0.048	0.049	0.062	0.053	0.078	0.07	0.095	0.098	0.102	0.082	0.128	0.099
$\alpha=0.7$	0.053	0.078	0.063	0.075	0.089	0.093	0.13	0.129	0.11	0.129	0.133	0.132

$M = 100$ $\approx 4\sqrt{T}$

	Corrected						Benchmark					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.02	0.022	0.019	0.029	0.035	0.05	0.063	0.069	0.055	0.062	0.068	0.053
$\alpha=0.3$	0.023	0.026	0.035	0.035	0.046	0.056	0.059	0.065	0.08	0.066	0.056	0.09
$\alpha=0.4$	0.038	0.029	0.04	0.056	0.052	0.062	0.086	0.092	0.078	0.1	0.096	0.089
$\alpha=0.5$	0.053	0.059	0.07	0.085	0.083	0.086	0.126	0.11	0.118	0.12	0.126	0.115
$\alpha=0.6$	0.086	0.073	0.102	0.102	0.1	0.11	0.175	0.175	0.17	0.169	0.166	0.148
$\alpha=0.7$	0.095	0.132	0.115	0.146	0.163	0.165	0.176	0.219	0.225	0.225	0.212	0.228

[18] In another set of experiments, I consider also the scenario in which the SWNs are generated by a multivariate normal distribution: $(\epsilon_x, \epsilon_z) \sim \mathcal{N}(0, I_2)$. As the results are similar to the presented one, this set of simulations are not reported in Section 3.1 and are available upon request.

Table 3: Rejection frequencies for DGP3A: This table presents the Rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and benchmark ($\mathcal{T}_\omega$), when the time series are generated by DGP3A; sample size, $T = 700$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12 \quad \approx 2(10T)^{1/5} \quad \approx 2\ln T$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.026	0.012	0.022	0.023	0.019	0.018	0.053	0.043	0.043	0.056	0.043	0.039
$\alpha = 0.3$	0.019	0.02	0.029	0.03	0.028	0.025	0.039	0.042	0.052	0.052	0.058	0.06
$\alpha = 0.4$	0.03	0.029	0.02	0.03	0.033	0.033	0.065	0.06	0.058	0.059	0.066	0.063
$\alpha = 0.5$	0.048	0.038	0.029	0.035	0.04	0.04	0.08	0.076	0.055	0.069	0.07	0.063
$\alpha = 0.6$	0.045	0.04	0.06	0.039	0.058	0.046	0.076	0.075	0.089	0.078	0.102	0.075
$\alpha = 0.7$	0.053	0.063	0.058	0.052	0.062	0.066	0.073	0.09	0.089	0.082	0.092	0.099

$M = 30 \quad \approx 5(10T)^{1/5} \quad \approx \sqrt{T}$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.016	0.013	0.01	0.023	0.018	0.013	0.055	0.04	0.046	0.05	0.045	0.038
$\alpha = 0.3$	0.02	0.012	0.023	0.02	0.019	0.023	0.052	0.05	0.066	0.059	0.053	0.063
$\alpha = 0.4$	0.028	0.03	0.026	0.022	0.033	0.035	0.072	0.076	0.065	0.059	0.073	0.079
$\alpha = 0.5$	0.04	0.039	0.028	0.043	0.038	0.035	0.085	0.086	0.07	0.079	0.108	0.062
$\alpha = 0.6$	0.048	0.033	0.059	0.055	0.056	0.042	0.095	0.105	0.093	0.11	0.11	0.08
$\alpha = 0.7$	0.053	0.066	0.063	0.075	0.08	0.082	0.13	0.122	0.125	0.128	0.129	0.152

$M = 100 \quad \approx 4\sqrt{T}$

	Corrected						Benchmark					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.02	0.013	0.016	0.016	0.016	0.019	0.063	0.055	0.066	0.06	0.066	0.049
$\alpha = 0.3$	0.023	0.019	0.032	0.018	0.022	0.035	0.059	0.062	0.086	0.08	0.075	0.069
$\alpha = 0.4$	0.038	0.033	0.036	0.03	0.04	0.038	0.086	0.098	0.085	0.069	0.098	0.073
$\alpha = 0.5$	0.053	0.062	0.039	0.049	0.049	0.046	0.126	0.12	0.108	0.098	0.126	0.11
$\alpha = 0.6$	0.086	0.069	0.072	0.066	0.08	0.07	0.175	0.158	0.156	0.142	0.155	0.135
$\alpha = 0.7$	0.095	0.128	0.113	0.115	0.126	0.132	0.176	0.216	0.198	0.186	0.213	0.235

Section 3 displays the results with respect to 700 independent realizations for the time series length $T = 700$. For other series lengths ($T = 300$ and $T = 100$), please refer to the Appendix C.1. Regarding the choice of the weighting function, $\omega(\cdot)$, all test statistics are calculated with respect to the Bartlett kernel:[19]

$$\omega(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

In each simulation set, the smoothing parameters considered are: $M = \{12, 30, 100\}$. The first two values adhere to the bandwidth rule specified in Hong and Lee (2005), whereas the last value is included to assess the robustness of asymptotic theory when the ratio, M/T, starts to be no longer negligible. While the first is proportional to $\ln T$, the others

[19]In additional Monte Carlo simulations, I find that the empirical rates do not change significantly with respect to the choice of the kernel function (e.g. truncated, quadratic-spectral, Daniell, Parzen). Such results are available upon request.

are proportional to the usual parametric rate, $\sqrt{T}$.

Under the null of interest, the corrected test typically performs better than the benchmark one, as it tends to over-reject less frequently. This makes the new testing strategy more robust against potential dependencies from X to Z, or equivalently, less sensitive with respect to the inverse causality channel. As Proposition 1 illustrates, these dependencies, which enter the second moments of the test statistic, can increase its variance and therefore leading to incorrect conclusions due to size distortions (i.e. over-rejections). This holds especially true when the correction term becomes relevant or, with respect to the considered DGPs, when β and α tend to be larger, aligned with Lemma 1.

Generally speaking, the correction term starts to play a significant role when the process Z is mildly autoregressive and the causality from X to Z is nonzero. Indeed, we have that, the stronger either the inverse causality (i.e., the magnitude of β) or the persistence of the process Z (i.e., the magnitude of α), the greater the size distortions in the benchmark test statistics. The results for $M = 30$ hints the following: i) Table 1 displays that, for $\alpha > 0.4$, the rejection rates for the benchmark procedure are on average 10.18%, while for the corrected version are on average 5.02%; ii) Table 2 displays that, for $0.2 < \alpha < 0.7$, the rejection rates for the benchmark procedure are on average 8.26%, while for the corrected version are on average 4.64%; iii) Table 3 displays that, for $\alpha > 0.4$, the rejection rates for the benchmark procedure are on average 5.19%, while for the corrected version are on average 10.40%.

Across smoothing parameters M, the empirical rates are roughly stable, except when the process Z is very persistent ($\alpha = 0.7$). Yet, there is a trade-off when tuning the smoothing parameter, because it needs to be large enough such that: i) M is growing asypmtotically, so that the statistic sums over a considerable number of covariances (see the discussion in Section **??**); ii) M is negligible compared to T, so that Brown (1971)'s central limit theorem can still apply because of appropriate boundedness of the higher-order moments of the statistic.

The application of the correction term, however, comes with a cost: when the process Z is not persistent enough, the corrected test typically tends to under-reject. This poor performance was announced in Section 2.2, where the distinction between the differenced-out cross-products (i.e., associated with $\mathcal{C}_\omega$), and the remaining cross-products (i.e., associated with $\mathcal{T}^c_{2\omega}$) is explained in terms of predictive regressions. When looking at the second moments of the test statistic, the latter set of cross-products is associated with regression coefficients proportional to the auto-covariance of the squared process Z, whereas the

former set of cross-products is associated with regression coefficients proportional to the conditional variance of Z with respect to the joint past of both processes X and Z. Consequently, when the process Z is weakly persistent, the sum of the remaining cross-products has potentially smaller variance and does not match the statistical information entailed in the differenced-out cross-products.
To outline key guidelines for the practitioner about my proposed testing procedure, I offer two suggestions:

1. Because of the problem of under-rejection, it is suggested to rely on the corrected test statistic when the omitted variables Z have some temporal dependence.[20]

2. Since the correction term plays a role when Z is mildly persistence, it is suggested to prioritize a smoothing parameter proportional to the parametric rate $\sqrt{T}$.

3.2 Under the alternatives

In this second set of experiments, a smaller range of DGPs is considered, as the contribution of this work is mainly related to the asymptotic properties of the test statistic under the null of interest, rather than under the alternatives. Similar to the previous ones, I define the following three classes of DGPs for the univariate processes X and Z:

a) DGP 1B: LINEAR-IN-MEAN

$$X_t = \gamma_1 Z_{t-1} + \epsilon_x, \quad Z_t = 0.4 Z_{t-1} + \beta X_{t-1} + \epsilon_z, \quad (\epsilon_x, \epsilon_z)' \sim \mathcal{N}(0, I_2)$$

b) DGP 2B: SQUARED-IN-MEAN

$$X_t = \gamma_1 Z_{t-1}^2/4 + \epsilon_x, \quad Z_t = 0.4 Z_{t-1} - \beta X_{t-1} + \epsilon_z, \quad (\epsilon_x, \epsilon_z)' \sim \mathcal{N}(0, I_2)$$

c) DGP 3B: SQUARED-IN-VARIANCE

$$X_t = \beta Z_{t-1} + \epsilon_x, \quad Z_t = 0.4 Z_{t-1} + \epsilon_z, \quad (\epsilon_x, \epsilon_z)' \sim \mathcal{N}(0, \Sigma)$$
$$[\Sigma]_{1,1} = 0.4 + \gamma_2 Z_{t-1}^2, \quad [\Sigma]_{2,2} = 1, \quad [\Sigma]_{1,2} = [\Sigma]_{2,1} = 0.5\sqrt{[\Sigma]_{1,1}[\Sigma]_{2,2}}$$

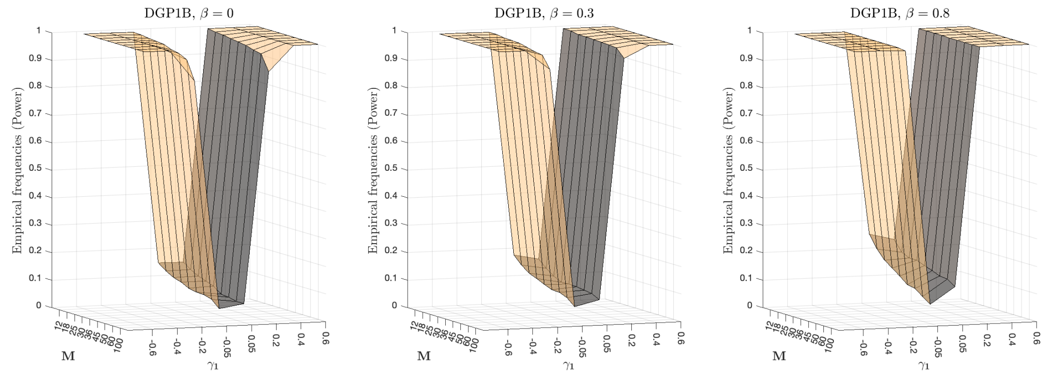
[20]This condition can be easily checked by a preliminary analysis on the magnitude (eigenvalues) of the coefficients of a VAR(p) model fitted to the multivariate process Z.

with the parameters ranging: $\gamma_1 = \{-0.6, -0.4, -0.2, -0.05, 0.05, 0.2, 0.4, 0.6\}$, $\gamma_2 = \{0.05, 0.1, 0.2, 0.4, 0.6, 0.8\}$, and $\beta = \{0, 0.3, 0.8\}$. Parallel to the previous section, the results are with respect to 700 random draws for the series length $T = 700$. Similarly, the test statistic are calculated with respect to the Bartlett kernel. In this set of simulations, the smoothing parameter ranges as: $M = \{12, 18, 25, 30, 36, 45, 50, 60, 100\}$.

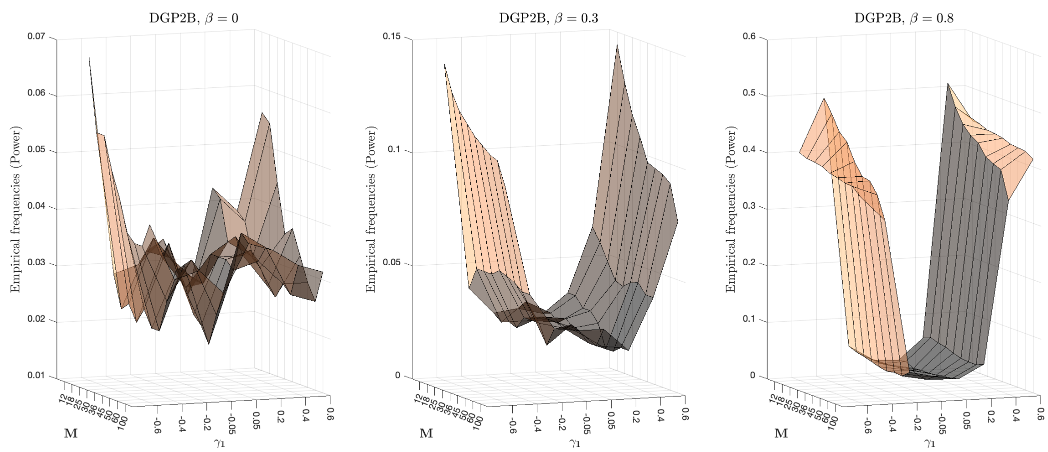
Under the alternatives, the corrected test has desired large sample properties. The first panel of Figure 1 shows that, when the $\{\gamma_i\}_{i=1,2}$ are non-negligible, the testing procedure tends to have unit power for large samples, mirroring that the probability of failing to reject the null of interest when is not true (i.e., the type II errors) asymptotically decreases to zero. For an exhaustive comparison of the two testing procedures, I provide the power curves associated to the benchmark test statistic in Appendix C.2.

The other panels regard nonlinear dynamics. The second panel of Figure 1 shows that as β increases, the power curves for DGP2B tend to resemble those of DGP1B qualitatively, although they are quantitatively lower. For $\beta = 0.8$, when $|\gamma_1| > 0.4$, the empirical rejection rates are approximately 0.5. As expected, the last panel of Figure 1 indicates that the corrected test statistic exhibits low power when the dependence between the time series is not related to any linear association, despite the strong variance dependence (i.e., large values of γ_2). When $\beta = 0$, empirical rejection rates range from 0.14 to 0.28. However, as β increases, the rejection rates also rise. Interestingly, the corrected test statistic appears more powerful than the benchmark, see in Appendix C.2.

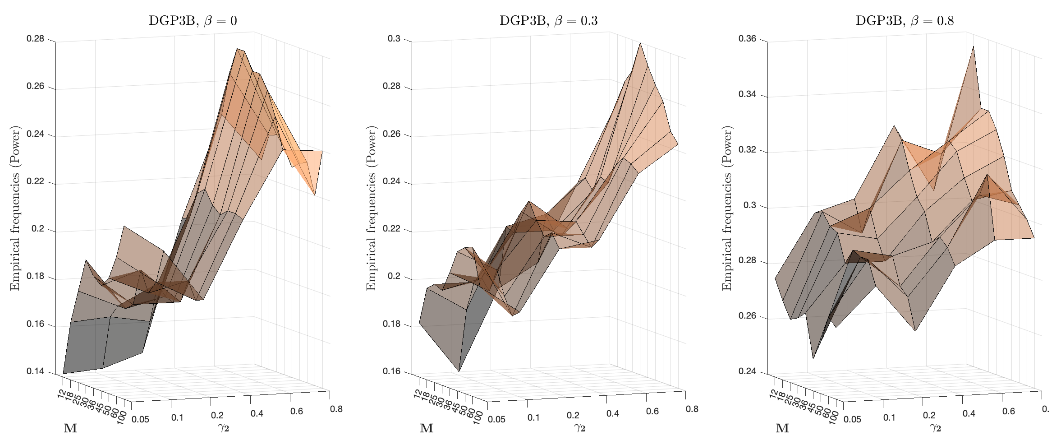
In support of the discussion in the previous section, Figure 1 illustrates how the magnitude of the smoothing parameters determines the large sample properties of the test statistic, since it regulates the approximation quality of the asymptotic theory. When the $\{\gamma_i\}_{i=1,2}$ are close to zero, thus getting to a scenario closer to the null hypothesis of interest, the power curves are more pronounced the larger the M. Indeed, as the smoothing parameter grows, the bottom of the V-shaped curves gradually stabilizes around the nominal significance level of 5%. This holds true as well for the benchmark testing procedure, as it is shown in in Appendix C.2.



(a) DGP 1B, with $\beta = \{0, 0.3, 0.8\}$



(b) DGP 2B, with $\beta = \{0, 0.3, 0.8\}$



(c) DGP 3B, with $\beta = \{0, 0.3, 0.8\}$

Figure 1: Power curves of the proposed test: These figures present the rejection rates of the testing procedure associated to the corrected test statistic ($\mathcal{T}^c_\omega$), under the alternatives (empirical power); sample size, $T = 700$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 18, 25, 30, 36, 45, 50, 60, 100\}$; nominal significance level is 5%.

4 Empirical application

This section presents an empirical application of the proposed testing procedure, with a focus on the comparison between the corrected and the benchmark testing procedures. I tackle the question of fundamentalness, also known as invertibility, of fashionable measures of structural shocks with respect to commonly used factors that summarize the state of the economy. While Section 4.1 introduces the concept of fundamentalness of structural shocks, Section 4.2 translates testing for fundamentalness to testing for the null hypothesis of interest. In Section 4.3, I investigate the fundamentalness of the uncertainty shock associated to Baker et al. (2016)'s Economic Policy Uncertainty, and Section 4.4 presents a discussion of two proxies of structural shocks: Jarociński and Karadi (2020)'s and Känzig (2023)'s.

4.1 Fundamentalness of structural shocks

Starting from the pioneering work of Sims (1980), the empirical analysis of macroeconomic time series data has progressively hinged on structural vector autoregressive (SVAR) models.

Common practice in the SVAR literature is to assume that the macroeconomic multivariate time series, $\{W_t\}$, has a vector Moving Average (MA) representation with respect to some mutually orthogonal structural shocks, $\{\epsilon_t\}$, such as:

$$W_t = B(L)\epsilon_t \tag{12}$$

where the matrices $B(L) = B_0 + B_1 L + B_2 L^2 + ...$ are moving average filters, that capture the propagation of the structural disturbances.

This rationale is supported by a twofold motivation: i) by the Wold Representation theorem, if the time series is covariance-stationary then it admits a MA(∞) representation (Brockwell and Davis (1987)), with the Wold innovations being the reduced-form residuals of the linear projection of W onto its infinite past; ii) the linear (or linearized) dynamic stochastic economic model, based on the variables W, usually admits a VARMA solution, whose structural shocks are assumed to be mutually orthogonal (Fernández-Villaverde et al. (2007)).

If the process W is causal and invertible, then there exists a linear map between structural shocks and residuals of the linear projections. This in turn means that the structural shocks can be recovered from the linear space of current and lagged values of the process W, up to a rotation matrix which governs the instantaneous relationships among the components of W.

When dealing with macroeconomic time series, the invertibility condition, however, might not always hold and, in such cases, the MA representation is said to be non-fundamental (Lippi and Reichlin (1994)). This occurs when the shocks are not solely a linear function of the past and present of the process W. Examples of non-fundamentalness might be time series associated to rational expectation models, where the economic agents forms beliefs about the future of the economy (e.g., Hansen and Sargent (2019)), and likely the agents' information space usually differs from the econometrician's. In practice, the issue of non-fundamentalness spells out as a problem of VAR misspecification, due to omitted variables or insufficient set of lagged controls (Chen et al. (2017), Miranda-Agrippino and Ricco (2023)).

For a better understanding, I use the example of Giannone and Reichlin (2006). Suppose the process W defined in eq.(12) consists of two blocks of variables, W_1 and W_2 of dimensions d_1 and d_2, respectively, such that:

$$\begin{pmatrix} W_{1,t} \\ W_{2,t} \end{pmatrix} = \begin{pmatrix} A_{1,1}(L) & A_{1,2}(L) \\ A_{2,1}(L) & A_{2,2}(L) \end{pmatrix} \begin{pmatrix} X_t \\ Z_t \end{pmatrix}, \qquad \begin{pmatrix} X_t \\ Z_t \end{pmatrix} = A_0^{-1} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

$$(X_t, Z_t)' \sim WN(0, \Sigma_w), \quad (\epsilon_{1,t}, \epsilon_{2,t})' \sim WN(0, I_{d_1+d_2}), \quad \Sigma_w = A_0^{-1}(A_0^{-1})'$$

where the reduced-form residuals $(X_t, Z_t)'$ and the structural shocks $(\epsilon_{1,t}, \epsilon_{2,t})'$ are White Noise (WN) processes. The Vector MA filters $A_{1,1}(L)$, $A_{2,2}(L)$, $A_{1,2}(L)$ are matrices of dimensions, rispectively, $d_1 \times d_1$, $d_1 \times d_2$, and $d_2 \times d_2$. The rotation matrix between reduced-form residuals and structural shocks, A_0, is of dimension $(d_1+d_2) \times (d_1+d_2)$. From eq.(12), note that: $B(L) = A(L) \cdot (A_0)^{-1}$.

Assume the $\det B(L)$ with the roots outside the unit circle, and the linear map, A_0, to be an identity matrix. As a consequence, the structural shocks coincide with the reduced-form residuals. In particular, we have: $X_t = \epsilon_{1,t}$.

Suppose the econometrician is interested in recovering the structural shocks, $\{\epsilon_{1,t}\}$, but omits the block of variables W_2 from the empirical analysis.

In such scenario, the shocks ϵ_1 are invertible, or the MA representation is fundamental, only if the restriction $A_{1,2}(L) = 0$ holds, since it implies that the shocks depend only on

the present and past values of the process W_1. The restriction on the coefficients can be read equivalently as Granger noncausality from the set of omitted variables, W_2, to the innovation process X (as it coincides with ϵ_1).
Vice versa, if the restriction $A_{1,2}(L) = 0$ does not hold, the MA representation is nonfundamental. To recover the structural shocks of interest, $\{\epsilon_{1,t}\}$, the present and past values of the process W_1 is no longer sufficient, and the econometrician needs to consider the enlarged information set, $\{W_{1,t}, W_{2,t}\}$, because of her information space being smaller than the correct one.

4.2 Testing for fundamentalness

As highlighted by the previous example, the problem of testing for fundamentalness of the structural shocks can be translated into testing for Granger causality (e.g., Giannone and Reichlin (2006), Forni and Gambetti (2014)) or, equivalently, testing for conditional lagged exogeneity (e.g., Miranda-Agrippino and Ricco (2023)).[21] Complementary to those works, Chen et al. (2017) propose a test for fundamentalness of the structural shocks by checking for the martingale difference property of the reduced-form residuals. In particular, their Theorem 1 shows that, when the DGP is a VARMA process generated by non-Gaussian i.i.d. (structural) shocks, the structural shocks are fundamental if and only if the reduced-form innovations are m.d.s. In my empirical analysis, I maintain this assumption about the underlying DGP in all three scenarios.
Thus, continuing with the illustrative VAR from the previous section, the innovations X are fundamental if: (i) they are invertible with respect to W_1, or rather, under the conditions of Theorem 1 in Chen et al. (2017), the process is an m.d.s. with respect to its own past; (ii) they are not Granger-caused in mean by the process W_2 (or equivalently, by its innovations, Z).

In light of these considerations, the two approaches to testing for invertibility of the shocks coalesce into testing for the martingale difference property of the structural shocks with respect to its own past and the past of the additional omitted variables, i.e., the following null hypothesis:

$$\mathcal{H}_0 : \ \mathbb{E}[X_t | \{X_s\}_{s<t}, \{Z_s\}_{s<t}] = 0 \tag{13}$$

[21] In particular, Giannone and Reichlin (2006) use a benchmark testing procedure based on the F-stat of the augmented regressions; Forni and Gambetti (2014) use the multivariate out-of-sample tests proposed by Gelper and Croux (2007); Miranda-Agrippino and Ricco (2023) propose a Hausman type test of lagged conditional exogeneity along the lines of Lu and White (2014).

which is equivalent to the null in eq.(1) up to properly redefining the time index for the variable Z.[22] A parallel conclusion can be drawn in the case of redefining the time index for the variable X, when the practitioner considers the following null hypothesis:

$$\mathcal{H}_0 : \ \mathbb{E}[X_{t+h}|\{X_s\}_{s<t+h}, \{Z_s\}_{s<t}] = 0, \quad h = 0, 1, ..., H$$

which is understood as the condition of the process X being a martingale difference sequence with respect its past enlarged by the h-lagged past of the process Z. Note that, this last condition implies: $\mathbb{E}[X_{t+h}|\{X_s\}_{s<t}, \{Z_s\}_{s<t}] = 0$, which might be related to the concept of non-causality at horizon h (see, for linear models, Dufour and Renault (1998), for nonlinear models or nonlinear regressors, Koop et al. (1996) and Gonçalves et al. (2021)).[23]

In the next sections, I will investigate the invertibility of estimated structural shocks and popular proxy of structural shocks. In particular, I will revisit the findings of Baker et al. (2016) (jointly with Diercks et al. (2024)), Jarociński and Karadi (2020) and Känzig (2023).

Regarding the choice of the potentially omitted variables, I follow Forni and Gambetti (2014) by considering the first principal components of datasets that entail the relevant macroeconomic or financial information. The intuition is that, by the economic system admitting a state space representation, those estimated factors should capture the state variables of the system. Thus, as set of omitted variables, $\{Z_t\}$, I consider three sets of factors that should control for the state of the economy:[24] i) 8 macroeconomic factors from McCracken and Ng (2016); ii) 8 macroeconomic factors from Rapach and Zhou (2021); iii) 5 financial factors from Giglio and Xiu (2021). As FRED-MD is a macroeconomic database

[22] Note that, without invoking Chen et al. (2017)'s Theorem 1, the conclusions about fundalmentalness might still depend on the contemporaneous identification restrictions, despite the null hypothesis of eq.(13) does not include the present value of the process Z.

[23] Interestingly, for multivariate autoregressive processes, Dufour and Renault (1998) provides a decomposition of the linear causality channel at horizon h with respect to the causality at previous horizons. In terms of conditional mean expectations, this does not hold in general. However, I conjecture that, by assuming some primitive conditions on the separability of the conditioning information sets (e.g. Renault and Triacca (2015)), then comparable conclusions might be drawn. This exercise is left for future research.

[24] Regarding the first set of factors, using the FRED-MD dataset spanning until June 2021, I estimate 8 static factors by principal component analysis (PCA) adapted to allow for missing values (Stock and Watson (2002)'s EM algorithm). Regarding the second set of factors, using the same dataset, I estimate 8 static factors by sparse PCA following Rapach and Zhou (2021). Regarding the last one, I estimate 5 static factors by PCA, using the dataset in their replication package. Covering from Jan. 1976 to Nov. 2009, their dataset consists in a large panel of 647 portfolios that include US equities as well as treasury bonds, corporate bonds, and currencies. For what concerns the other vintages of McCracken and Ng (2016), I run the empirical analysis with respect to other vintages (e.g., Jan. 2020) but I have not found substantial changes with respect to the presented conclusions. Such additional results are available upon request.

of 134 monthly U.S. indicators, the first two sets of factors control for the state of macroeconomic variables. Vice versa, the last set of factors should principally control for the state of the financial markets, as being extracted by a panel of 647 portfolios, spanning from equities to corporate bonds.

The simulation results underscore that the good properties of the proposed testing procedure depend on the autoregressive properties of the process Z. Regarding this aspect, all sets of factors seem to have nonnegligible persistence. In a preliminary analysis, it emerges that, by looking at the information criteria after fitting VAR models: i) for the first two sets of factors, the AIC (BIC) suggests the optimal choice of lag controls to be around 4 (2); ii) for the last set of factors, the AIC (BIC) suggests the optimal number of lag controls to be 3 (1).

4.3 Baker et al. (2016)'s Economic Policy Uncertainty Shock

Baker et al. (2016) propose a measure of economic policy uncertainty (EPU) based on newspaper coverage frequency. In particular, for major U.S. newspapers, the index measures the frequency of articles containing terms that are associated with uncertainty, the economy, and policy.

Following Baker et al. (2016) (Section IV.D), I estimate the uncertainty structural shock at monthly frequency by: i) fitting a VAR(6) to 5 monthly US time series from Jan. 1985 to Dec. 2019; ii) imposing the following ordering restriction via Cholesky decomposition: first, the EPU index, then the log of the S&P 500 index, the federal funds rate, log employment, and lastly log industrial production.[25] As discussed in Diercks et al. (2024), the estimated structural shock is serially uncorrelated.

The small VAR system, however, arguably controls for all relevant macroeconomic conditions. This turns to be critical when questioning the fundamentalness of the estimated shock, especially since the measure should also capture *"uncertainties related to the economic ramifications of "noneconomic" policy matters [...] both near-term concerns [...] and longer term concerns"* (Baker et al. (2016)).

In response to these concerns about fundamentalness, I test for the null hypothesis of eq.(13), by employing both Portmanteau-type testing procedures: the corrected one (i.e, based on $\widehat{\mathcal{T}}^c_\omega$, see Theorem 2) and the benchmark one (i.e., based on $\widehat{\mathcal{T}}_\omega$, see eq.(11)). With respect to the times series $\{X_t, Z_t\}$, the EPU uncertainty shock is to be considered as the process X, while the McCracken and Ng (2016)'s factors as the process Z.

[25] The data is available in the replication package of Diercks et al. (2024).

The left panel of Figure 2 displays that, for both testing procedures, we reject the null

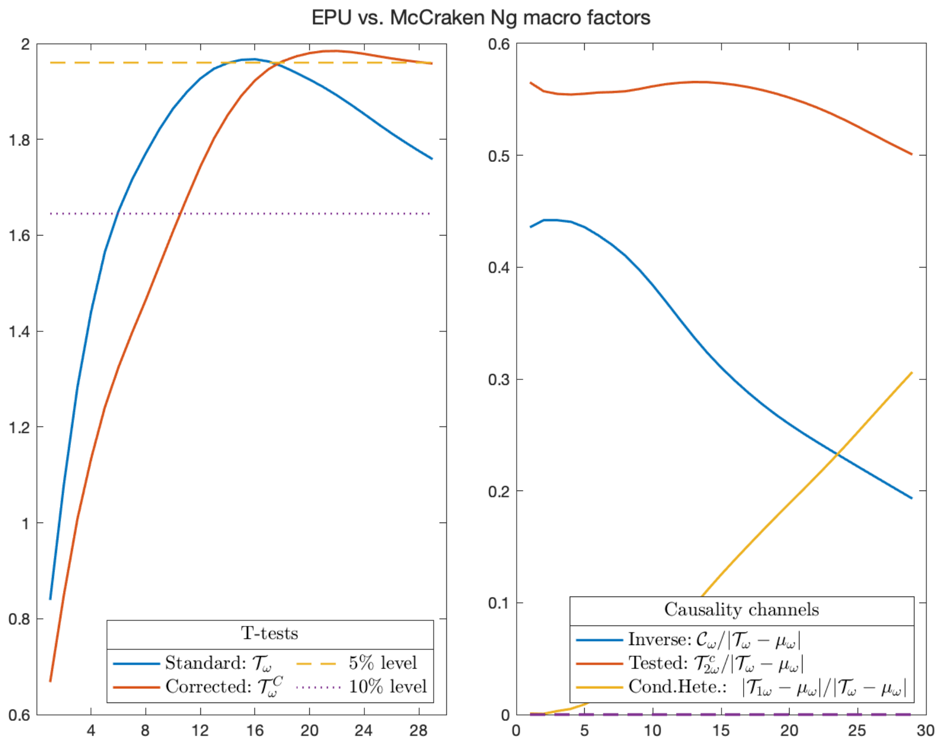


Figure 2: Baker et al. (2016). Comparison between the two testing strategies: LEFT PANEL: on the y-axes, the level of the standard/benchmark test statistic (blue solid) and corrected test statistic (orange solid) at different horizons M; on the x-axis, the smoothing parameter range from 1 to 30; the weighting function is the Bartlett kernel; nominal significance levels are 5% (dashed yellow), and 10% (dashed purple). RIGHT PANEL: on the y-axes, the contributions of the causality channels relative to the standard/benchmark test statistic (in absolute value), as decomposed in eq. (5): the inverse $\mathcal{C}_\omega$ (blue solid), the tested (orange solid) $\mathcal{T}_{2\omega}^c$, and the one associated to the conditional homoskedasticity of process X (yellow solid) $T_{1\omega}$, after being centered.

hypothesis at 5% significance level. The corrected test statistic conveys the importance of the past cross-correlation, from moderate (18 months) to long horizons (30 months).[26] On the contrary, the benchmark test statistic, indicates that the null hypothesis should be rejected for a narrow window of moderate horizons (14-16 months), thus failing to reject the null for longer horizons.

Despite seemingly puzzling, we understand the difference between the two testing strategies once looking at the right panel of Figure 2, which dissects the benchmark test statistic as in eq.(5). Albeit the relative importance of the tested causality channel (i.e., $\mathcal{T}_{2\omega}^c$) remains

[26]These values of the smoothing parameter are in line with the guideline at the end of Section 3.1. Note that, since the sample size is $T = 408$, we have: $M = 20 \approx \sqrt{T}$, and $M = 27 \approx 5(10T)^{1/5}$ (see the preliminary bandwidth in Hong and Lee (2005)).

similar across horizons, the relative importance of the inverse causality channel (i.e, C_ω) is larger at shorter horizons but fading away at longer horizons. This causes the ambiguous inference drawn by the benchmark test, which visually appears as the inverse U-shaped curve on the left panel. On the opposite, the corrected test statistic, being robust to the inverse causality channel, does not suffer such issue, and so produces a more coherent picture. Moreover, as the relative importance of the term $\mathcal{T}_{1\omega}$ tends to small, the shock series can be regarded approximately to be conditionally homoskedastic and homokurtic. Please recall that, as described in the first part of Proposition 2, the statistic $\mathcal{T}_{1\omega}$ is centered at zero when the conditional homoskedasticity holds, and has negligible variance when the conditional homokurtosis holds.

In light of these considerations, I am prone to conclude that the shock is not fundamental to the system when including the McCracken and Ng (2016)'s macroeconomic factors. Regarding the other macroeconomic factors, for both testing procedures, we fail to reject the null hypothesis.

Causal inference about the EPU uncertainty shock, thus, can beneficiate from including the McCracken and Ng (2016)'s factors. As an example, I revisit the findings of Diercks et al. (2024) about the superadditive effects of uncertainty shocks. In particular, I focus on the impulse response analysis of inflation with respect to uncertainty, measured by the EPU shocks, within the system of variables as in Baker et al. (2016).

Diercks et al. (2024) estimate the following set of state-dependent local projections:

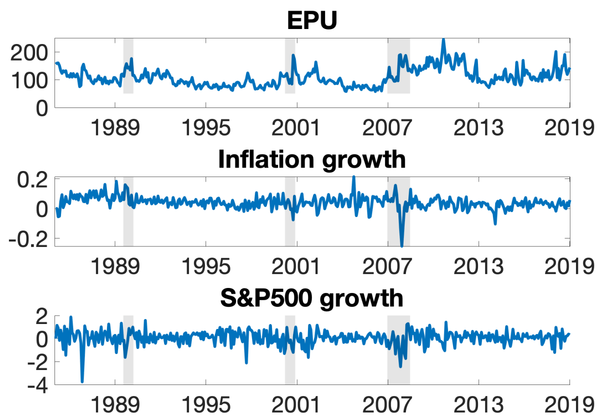
$$y_{t+h} = \alpha_h + \big(\beta_{0,h} + \beta_{1,h}\mathbb{1}\{\epsilon_{unc,t-1} > 0, .., \epsilon_{unc,t-L} > 0\}\big)\epsilon_{unc,t} + \sum_{i=1}^{p} \gamma_{i,h} w_{t-i} + u_{t+h}$$

where h sets the predictive horizon, ranging from 0 to 36 months, p is the number lags used for the control variables, $\{w_t\}$, and the indicator function takes value 1 if each one of the previous L uncertainty shocks $\{\epsilon_{unc,t-1}, .., \epsilon_{unc,t-L}\}$ has been positive. The local projections are considered as state-dependent because the indicator function measures the response to the shock conditional on having L previous consecutive positive shocks. Indeed, the $\beta_{1,h}$ coefficient is dubbed as the "state multiplier" as it captures the superaddivite effect of shocks to uncertainty: the impact of a cascade of positive uncertainty shocks is more severe than the isolated sum of them.

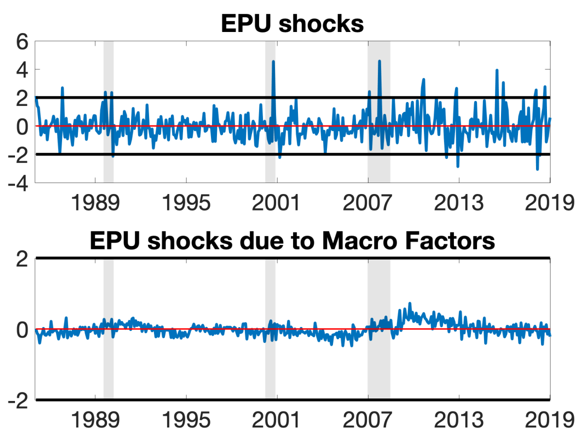
In the empirical application of their Appendix B.1 (Figure B.2), the shock of interest ϵ_{unc} is the EPU uncertainty shock, the outcome variable y is inflation (Personal Consumption Expenditures price index), the number of consecutive positive shocks is 2, $L = 1$, the number

of lags for the controls is 6, $p = 6$. The set of controls is the same as in Baker et al. (2016), augmented by the past realizations of the outcome variable. Figure 4a reproduces Diercks et al. (2024)'s results.
Once adding two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of LP controls, we notice that, while there is no much relevant change to the unconditional linear response (see Appendix D.1), the state-dependent response ($\{\beta_{1,h}\}_{h=0,\ldots,H}$) becomes significantly positive. This corroborates the superadditivity of the uncertainty shocks: a sequence of consecutive (positive) EPU structural shocks leads to a noticeable increase of inflation, substantially more severe than the case when there is no such consecutive sequence. This finding connects to two studies. First, it aligns with the impulse response analysis in Fernández-Villaverde et al. (2015), which shows that unexpected changes in fiscal policy uncertainty can lead to increased inflation within a standard New Keynesian model. Second, it relates to Ascari et al. (2023), where they demonstrate that, in a rich DSGE model with firm dynamics, a shock that raises short-term inflation expectations results in negative macroeconomic effects, causing inflation to rise while output declines (stagflationary response).
Regarding the other variables of the system, the responses to the shocks are similar to Diercks et al. (2024), except for the stock market (real S&P500 price index). In particular, the responses of industrial production and short rates to uncertainty shocks are still negative, in line with Diercks et al. (2024)'s (Figure 10-11 in Appendix D.2). Conversely, the state-dependent response of the stock market to consecutive (positive) EPU structural shocks becomes significantly more negative at longer horizons (Figure 5). This finding is consistent with Berger et al. (2020), who observe that uncertainty shocks, as associated with future volatility, are linked to declines in stock returns (refer to their discussion in Section 5.2). Upon inspecting Figure 3, we observe that EPU shocks correlate with the past state of the macroeconomy, particularly during two periods: from 2005 to 2007 and from 2009 to 2013 —i.e., the end of the Great Moderation and the aftermath of the Great Recession.

Interestingly, the response of inflation to an uncertainty shock, as measured by the one of other two proxies in Diercks et al. (2024) (namely, financial uncertainty from Ludvigson et al. (2021) and realized market volatility from Berger et al. (2020)), seems to be strongly negative. At the same time, for all the proxies, the responses of industrial production to uncertainty shocks tend to be negative, confirming the countercyclical nature of the shocks (Diercks et al. (2024)). In light of these findings, it appears that the EPU uncertainty shock could be considered a supply-side negative shock, while the other two uncertainty shocks are more accurately described as demand-side negative shocks. In conclusion, I find that, after accounting for macroeconomic factors, EPU uncertainty shocks—particularly between the end of the Great Moderation and the recovery from the Great Recession—behaves as (superadditive) 'expectational' shocks, as inspired by Ascari et al. (2023).

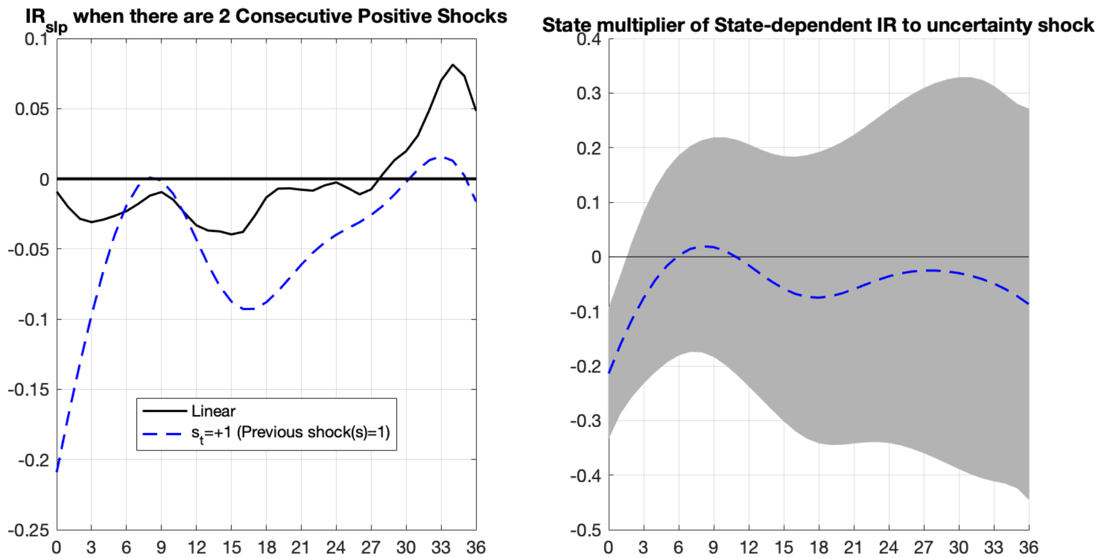


(a) Time series of EPU from Baker et al. (2016), and PCE and S&P 500 price indexes in percentage growth.

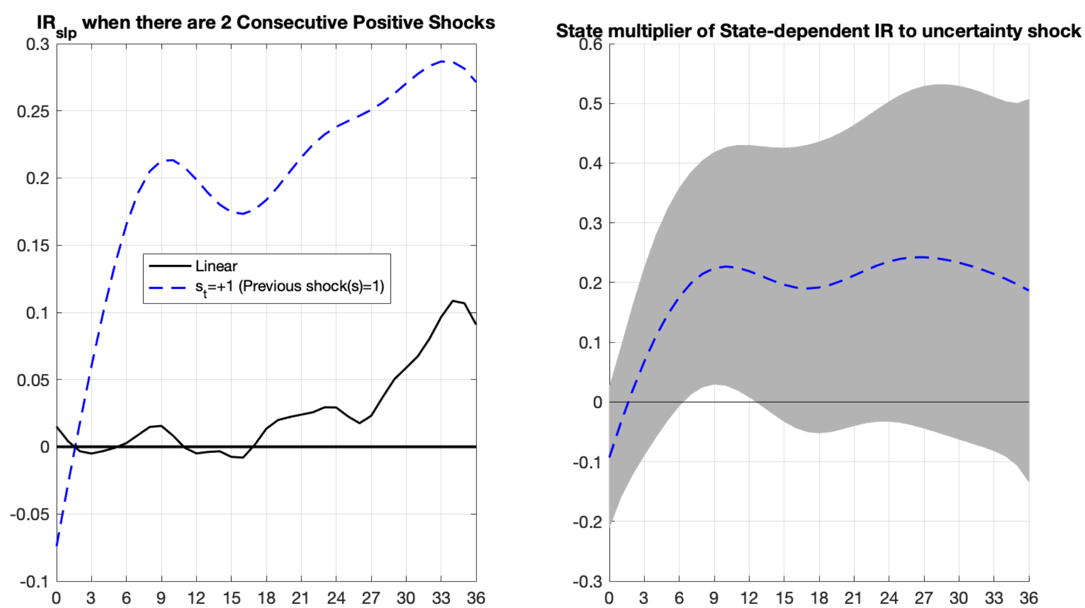


(b) Time series of the estimated EPU shock and the estimated projection of the shock series on the two lags of McCracken and Ng (2016)'s macroeconomic factors.

Figure 3: Time series and shocks:
Parallel to Figure 1 in Diercks et al. (2024), the upper panel displays the time series of EPU, together with the time series associated to inflation and stock market in percentage growth (i.e, (current/previous $-1) \times 100$). The lower panel displays the the estimated EPU shock series and its part that correlates with the past of the macroeconomic factors. The shaded areas represent NBER (National Bureau of Economic Research) recessions.

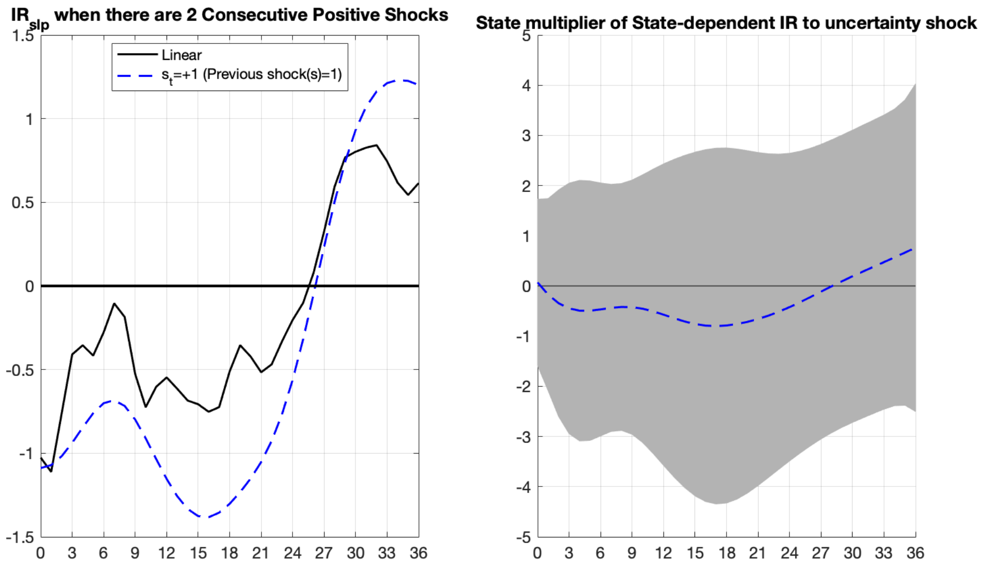


(a) Replication of Figure B.2.(e)-(f) in Diercks et al. (2024)'s Appendix B.1

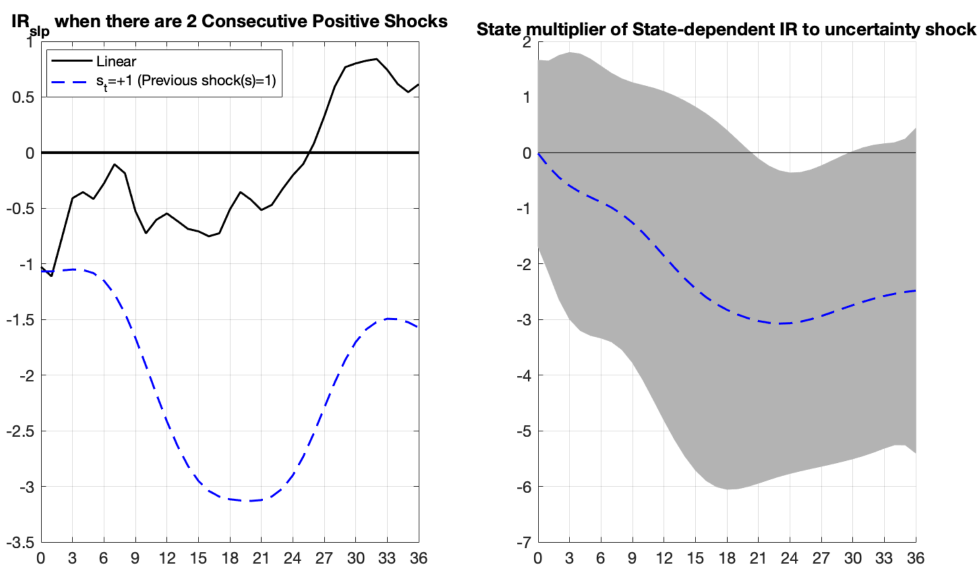


(b) Inclusion of two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of controls

Figure 4: Response of price level to consecutive positive EPU uncertainty shocks:
LEFT PANELS: the empirical state-dependent impulse responses (estimated with LPs as in Diercks et al. (2024)) to two consecutive positive uncertainty shocks (dashed blue line) and contrast it to the response to a single shock (solid black line). RIGHT PANELS: the incremental effect of the second shock, i.e. $\{\beta_{1,h}\}_{h=1,..,H}$, with 90% confidence intervals (shaded area). In both panels, on the y-axes, the level of impulse responses; on the x-axes, the horizons, h. Price level is measured by the Personal Consumption Expenditures (PCE) Price Index.



(a) Replication of Figure B.3(e)-(f) in Diercks et al. (2024)



(b) Inclusion of two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of controls

Figure 5: Response of stock market to consecutive positive EPU uncertainty shocks:
LEFT PANELS: the empirical state-dependent impulse responses (estimated with LPs as in Diercks et al. (2024)) to two consecutive positive uncertainty shocks (dashed blue line) and contrast it to the response to a single shock (solid black line). RIGHT PANELS: the incremental effect of the second shock, i.e. $\{\beta_{1,h}\}_{h=1,..,H}$, with 90% confidence intervals (shaded area). In both panels, on the y-axes, the level of impulse responses; on the x-axes, the horizons, h. Stock market is measured by real S&P 500.

4.4 Cautionary tales: Jarociński and Karadi (2020), Känzig (2023)

Contrary to the previous application, this section provides examples where the benchmark testing strategy suggests to reject the null hypothesis, in contrast with the corrected version. I consider two series of structural shocks: Jarociński and Karadi (2020)'s monetary policy information shock and Känzig (2023)'s carbon policy shock.[27]

Combining high frequency and sign restrictions identification schemes, Jarociński and Karadi (2020) propose a methodology to disentangle the monetary policy information surprises into two components: the one associated with the information about monetary policy (i.e., the monetary policy information shock) and the one associated with the central bank's assessment of the economic outlook (i.e., the central bank information shocks). In this application, I consider the monetary policy information shock identified by the "Poor Man's" sign restrictions (refer to their Section III.C).

Suppose a practitioner investigates the fundamentalness of the Jarociński and Karadi (2020)'s monetary policy information shock while controlling for the state of the economy, by testing the null hypothesis of eq.(13). With respect to the times series $\{X_t, Z_t\}$, in this example, the monetary policy shock is to be considered as the process X, while the Rapach and Zhou (2021)'s sparse factors as the process Z.

The left panel of Figure 6 shows that the benchmark testing procedure suggest to reject the null hypothesis at 5% significance level. In particular, it hints the importance of the past cross-correlation from moderate (16 months) horizons. Vice versa, by considering the corrected test statistic, we fail to reject the null at all horizons. Once appealing to the decomposition on the right panel of Figure 6, we understand that the relative magnitude of the inverse causality channel is non-negligible, therefore leading to contrasting results.

Adopting an event study approach that exploits high-frequency data and the institutional features of the European carbon market, Känzig (2023) isolates a series of carbon policy surprises. Measured around the regulatory news, these surprises are defined as changes in the carbon futures price (relative to wholesale electricity price). Using the surprise series as instrument for the energy price (i.e., Proxy-VAR approach), Känzig (2023) then estimates the carbon policy shock from a Structural VAR with 8 variables (refer to his Section 3).

[27]Both series are downloaded from the authors' github repositories: `https://github.com/marekjarocinski/jkshocks_update_fed`; `https://github.com/dkaenzig/carbonpolicyshocks`; I thank the authors, Marek Jarocinski and Diego Känzig, for publicly sharing the data.

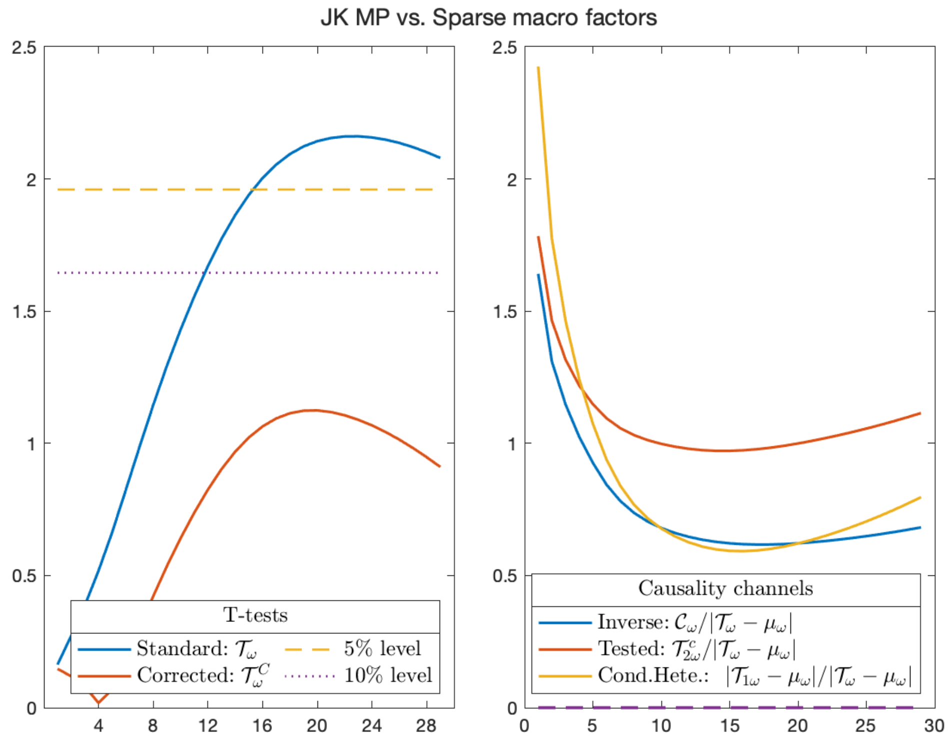


Figure 6: Jarociński and Karadi (2020). Comparison between the two testing strategies: LEFT PANEL: on the y-axes, the level of the standard/benchmark test statistic (blue solid) and corrected test statistic (orange solid) at different horizons M; on the x-axis, the smoothing parameter range from 1 to 30; the weighting function is the Bartlett kernel; nominal significance levels are 5% (dashed yellow), and 10% (dashed purple). RIGHT PANEL: on the y-axes, the contributions of the causality channels relative to the standard/benchmark test statistic (in absolute value), as decomposed in eq. (5): the inverse $\mathcal{C}_\omega$ (blue solid), the tested (orange solid) $\mathcal{T}_{2\omega}^c$, and the one associated to the conditional homoskedasticity of process X (yellow solid) $T_{1\omega}$, after being centered.

Now, suppose a practitioner investigates the fundamentalness of the Känzig (2023)'s carbon policy surprises while controlling for the state of the US financial markets, by testing the null hypothesis of eq.(13).[28] With respect to the times series $\{X_t, Z_t\}$, in this example, the carbon policy shock is to be considered as the process X, while the Giglio and Xiu (2021)'s financial factors as the process Z.

Similarly to the previous case, in the left panel of Figure 7, the benchmark testing procedure suggest to reject the null hypothesis at 10% significance level, Vice versa, by considering the corrected test statistic, we fail to reject the null at all horizons. The benchmark test statistic hints the importance of the past cross-correlation at very short horizons (3-5 months), but fading away right after. Again, by dissecting the statistic on the right panel

[28] Note that Känzig (2023) does not include any variable associated to the state of the financial markets in his Structural VAR model (e.g., S&P 500).

of Figure 7, we deduce that the notable impact of the inverse causality channel drives the conclusions about rejecting the null.

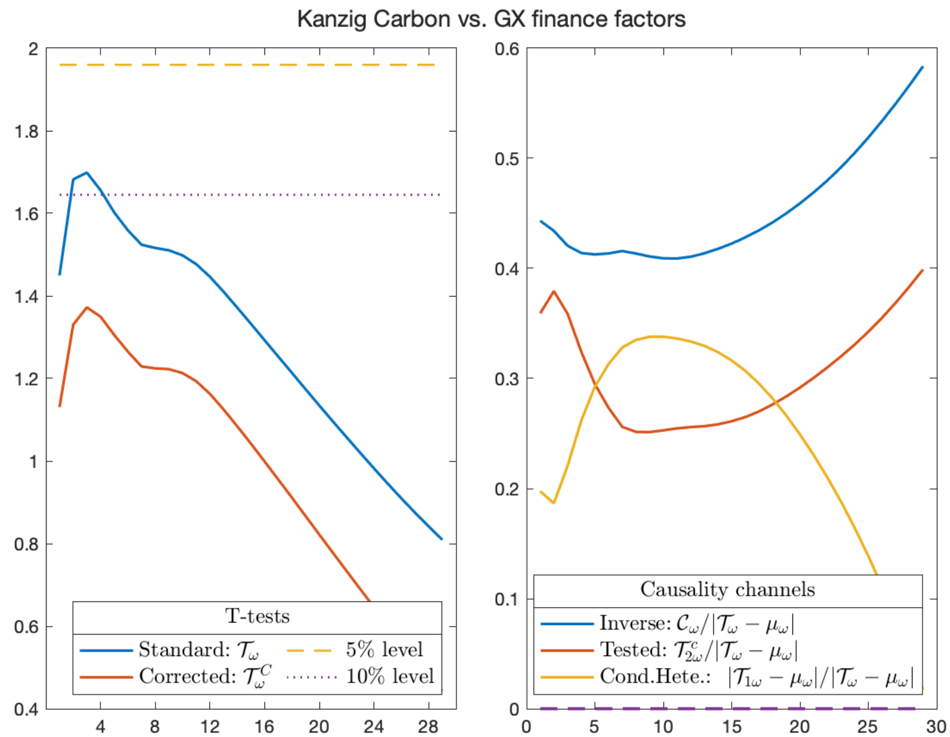


Figure 7: Känzig (2023). Comparison between the two testing strategies: LEFT PANEL: on the y-axes, the level of the standard/benchmark test statistic (blue solid) and corrected test statistic (orange solid) at different horizons M; on the x-axis, the smoothing parameter range from 1 to 30; the weighting function is the Bartlett kernel; nominal significance levels are 5% (dashed yellow), and 10% (dashed purple). RIGHT PANEL: on the y-axes, the contributions of the causality channels relative to the standard/benchmark test statistic (in absolute value), as decomposed in eq. (5): the inverse $\mathcal{C}_\omega$ (blue solid), the tested (orange solid) $\mathcal{T}_{2\omega}^c$, and the one associated to the conditional homoskedasticity of process X (yellow solid) $T_{1\omega}$, after being centered.

5 Conclusion

In this paper, I offer a strategy to test for causality in mean between multivariate time series based on serial cross-correlation. In particular, building on Hong (1996a), I propose a test statistic based on the weighted sum of ℓ_2 norm of the cross-correlation between the two multivariate processes at positive lags.

My methodology differs from the standard Portmanteau-type testing procedure by the introduction of a correction term that removes the influence of the causal dependencies whose direction is inverse to the tested one. Hence, the proposed approach is suited for scenarios where the practitioner wants to preserve an agnostic approach to the modelling of the joint conditional moments of the variables of interests and the omitted variables.

With respect to mild assumptions on the joint process, I show that the corrected version of the benchmark test statistic has desired asymptotic normality under the null hypothesis of (one-sided) conditional mean independence, for two cases: when the processes are directly observed, and when the processes are estimated from a conditional mean model. Under a class of fixed alternatives, the corrected statistic has asymptotic power similar to the benchmark statistic.

A set of Monte Carlo experiments validates the results from the asymptotic theory: when the inverse causality matters, the rejection frequencies of the proposed testing procedure are close to the nominal size, while this is not true for the benchmark procedure, which tends to over-reject. Both testing strategies generally have similar empirical power under the alternatives.

For the empirical application, I study the property of fundamentalness of three popular measures of macroeconomic structural shocks: Baker et al. (2016)'s economic policy uncertainty (EPU) shock, Jarociński and Karadi (2020)'s monetary policy shock, and Känzig (2023)'s carbon policy shock. All three scenarios are characterized by nonnegligible inverse causality, and so the two testing procedures, the benchmark (Hong (1996a)) and its corrected version, draw different conclusions in terms of inference. Given the robustness of the proposed procedure, I deduce that the EPU structural shocks are not fundamental and, by revisiting Diercks et al. (2024), I find that the response of inflation to the EPU shock is positive, suggesting it can be dubbed as supply-side negative shock. Regading the other two shock series, Jarociński and Karadi (2020)'s and Känzig (2023)'s, while the standard procedure suggests, in both cases, to reject the null hypothesis of fundamentalness, the corrected test statistic fails to reject the null. The decomposition of the benchmark test statistic confirms that these conclusions are largely driven by the influence of the causal-

ity whose direction is inverse to the tested one.
Two interesting venues for future research might be: i) generalizing my proposed testing strategy to capture also nonlinear depedences by means of the Hong (1999)'s generalized spectrum (e.g. Hong and Lee (2005), Escanciano and Velasco (2006)). Similarly, considering the case where the dimensions of the time series are growing, i.e., high dimensional time series, is a natural extension. This line of research shares a strong connection with the distance correlation literature, as shown in Han and Shen (2024) and Wang (2024); ii) to extend my proposed testing strategy to panel data (e.g. Hong and Kao (2004), Chen (2022)), and so to apply it for testing for weak/ strict exogeneity.

References

AGUILAR, M. AND J. B. HILL (2015): "Robust score and portmanteau tests of volatility spillover," *Journal of Econometrics*, 184, 37–61.

ALESSI, L., M. BARIGOZZI, AND M. CAPASSO (2011): "Non-fundamentalness in structural econometric models: A review," *International Statistical Review*, 79, 16–47.

ASCARI, G., S. FASANI, J. GRAZZINI, AND L. ROSSI (2023): "Endogenous uncertainty and the macroeconomic impact of shocks to inflation expectations," *Journal of Monetary Economics*, 140, S48–S63.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring economic policy uncertainty," *The quarterly journal of economics*, 131, 1593–1636.

BERGER, D., I. DEW-BECKER, AND S. GIGLIO (2020): "Uncertainty shocks as second-moment news shocks," *The Review of Economic Studies*, 87, 40–76.

BOUHADDIOUI, C. AND R. ROY (2006): "A generalized portmanteau test for independence of two infinite-order vector autoregressive series," *Journal of Time Series Analysis*, 27, 505–544.

BOX, G. E. AND D. A. PIERCE (1970): "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American statistical Association*, 65, 1509–1526.

BROCKWELL, P. J. AND R. A. DAVIS (1987): "Time Series: Theory and Methods," *Springer Series in Statistics*.

BROWN, B. M. (1971): "Martingale central limit theorems," *The Annals of Mathematical Statistics*, 59–66.

BRUNNERMEIER, M., D. PALIA, K. A. SASTRY, AND C. A. SIMS (2021): "Feedbacks: financial markets and economic activity," *American Economic Review*, 111, 1845–1879.

CANDELON, B. AND S. TOKPAVI (2016): "A nonparametric test for granger causality in distribution with application to financial contagion," *Journal of Business & Economic Statistics*, 34, 240–253.

CHEN, B. (2022): "A robust test for serial correlation in panel data models," *Econometric Reviews*, 41, 1095–1112.

CHEN, B., J. CHOI, AND J. C. ESCANCIANO (2017): "Testing for fundamental vector moving average representations," *Quantitative Economics*, 8, 149–180.

CHEUNG, Y.-W. AND L. K. NG (1996): "A causality-in-variance test and its application to financial market prices," *Journal of econometrics*, 72, 33–48.

COMTE, F. AND O. LIEBERMAN (2000): "Second-order noncausality in multivariate GARCH processes," *Journal of time series analysis*, 21, 535–557.

DEDECKER, J., P. DOUKHAN, G. LANG, L. R. JOSÉ RAFAEL, S. LOUHICHI, C. PRIEUR, J. DEDECKER, P. DOUKHAN, G. LANG, L. R. JOSÉ RAFAEL, ET AL. (2007): *Weak dependence*, Springer.

DIERCKS, A. M., A. HSU, AND A. TAMONI (2024): "When it rains it pours: Cascading uncertainty shocks," *Journal of Political Economy*, 132, 694–720.

DUFOUR, J.-M. AND E. RENAULT (1998): "Short run and long run causality in time series: theory," *Econometrica*, 1099–1125.

DUFOUR, J.-M. AND A. TAAMOUTI (2010): "Short and long run causality measures: Theory and inference," *Journal of Econometrics*, 154, 42–58.

DURLAUF, S. N. (1991): "Spectral based testing of the martingale hypothesis," *Journal of Econometrics*, 50, 355–376.

ESCANCIANO, J. C. AND C. VELASCO (2006): "Generalized spectral tests for the martingale difference hypothesis," *Journal of Econometrics*, 134, 151–185.

FERNÁNDEZ-VILLAVERDE, J., P. GUERRÓN-QUINTANA, K. KUESTER, AND J. RUBIO-RAMÍREZ (2015): "Fiscal volatility shocks and economic activity," *American Economic Review*, 105, 3352–3384.

FERNÁNDEZ-VILLAVERDE, J., J. F. RUBIO-RAMÍREZ, T. J. SARGENT, AND M. W. WATSON (2007): "ABCs (and Ds) of understanding VARs," *American economic review*, 97, 1021–1026.

FORNI, M. AND L. GAMBETTI (2014): "Sufficient information in structural VARs," *Journal of Monetary Economics*, 66, 124–136.

FRANCQ, C. AND H. RAÏSSI (2007): "Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors," *Journal of Time Series Analysis*, 28, 454–470.

GELPER, S. AND C. CROUX (2007): "Multivariate out-of-sample tests for Granger causality," *Computational statistics & data analysis*, 51, 3319–3329.

GEWEKE, J. (1982): "Measurement of linear dependence and feedback between multiple time series," *Journal of the American statistical association*, 77, 304–313.

GIACOMINI, R. AND H. WHITE (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

GIANNONE, D. AND L. REICHLIN (2006): "Does information help recovering structural shocks from past observations?" *Journal of the European Economic Association*, 4, 455–465.

GIGLIO, S. AND D. XIU (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129, 1947–1990.

GONÇALVES, S., A. M. HERRERA, L. KILIAN, AND E. PESAVENTO (2021): "Impulse response analysis for structural dynamic models with nonlinear regressors," *Journal of Econometrics*, 225, 107–130.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1987): "Kullback causality measures," *Annales d'Economie et de Statistique*, 369–410.

GRANGER, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, 424–438.

HAN, Q. AND Y. SHEN (2024): "Generalized kernel distance covariance in high dimensions: non-null CLTs and power universality," *Information and Inference: A Journal of the IMA*, 13.

HANNAN, E. J. (1970): *Multiple time series*, vol. 38, John Wiley & Sons.

HANSEN, L. P. AND T. J. SARGENT (2019): "Two difficulties in interpreting vector autoregressions," in *Rational expectations econometrics*, CRC Press, 77–119.

HAUGH, L. D. (1976): "Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach," *Journal of the American Statistical Association*, 71, 378–385.

HONG, Y. (1996a): "Consistent testing for serial correlation of unknown form," *Economet-*

rica: Journal of the Econometric Society, 837–864.

——— (1996b): "Testing for independence between two covariance stationary time series," *Biometrika*, 83, 615–625.

——— (1999): "Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach," *Journal of the American Statistical Association*, 94, 1201–1220.

——— (2001): "A test for volatility spillover with application to exchange rates," *Journal of Econometrics*, 103, 183–224.

HONG, Y. AND C. KAO (2004): "Wavelet-based testing for serial correlation of unknown form in panel models," *Econometrica*, 72, 1519–1563.

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

HONG, Y., Y. LIU, AND S. WANG (2009): "Granger causality in risk and detection of extreme risk spillover between financial markets," *Journal of Econometrics*, 150, 271–287.

HOSKING, J. R. (1980): "The multivariate portmanteau statistic," *Journal of the American Statistical Association*, 75, 602–608.

ISSERLIS, L. (1918): "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, 12, 134–139.

JAROCIŃSKI, M. AND P. KARADI (2020): "Deconstructing monetary policy surprises—the role of information shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

KÄNZIG, D. R. (2023): "The unequal economic consequences of carbon pricing," Tech. rep., National Bureau of Economic Research.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of econometrics*, 74, 119–147.

LEE, C. E. AND X. SHAO (2018): "Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series," *Journal of the American Statistical Association*, 113, 216–229.

LEONG, S. H. AND G. URGA (2023): "A practical multivariate approach to testing volatility spillover," *Journal of Economic Dynamics and Control*, 104694.

LI, W. K. AND A. I. MCLEOD (1981): "Distribution of the residual autocorrelations in multivariate ARMA time series models," *Journal of the Royal Statistical Society Series B:*

Statistical Methodology, 43, 231–239.

LIPPI, M. AND L. REICHLIN (1994): "VAR analysis, nonfundamental representations, Blaschke matrices," *Journal of Econometrics*, 63, 307–325.

LJUNG, G. M. AND G. E. BOX (1978): "On a measure of lack of fit in time series models," *Biometrika*, 65, 297–303.

LU, X. AND H. WHITE (2014): "Robustness checks and robustness tests in applied economics," *Journal of econometrics*, 178, 194–206.

LUDVIGSON, S. C., S. MA, AND S. NG (2021): "Uncertainty and business cycles: exogenous impulse or endogenous response?" *American Economic Journal: Macroeconomics*, 13, 369–410.

LÜTKEPOHL, H. (1997): *Handbook of matrices*, John Wiley & Sons.

MCCRACKEN, M. W. AND S. NG (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34, 574–589.

MIRANDA-AGRIPPINO, S. AND G. RICCO (2023): "Identification with external instruments in structural VARs," *Journal of Monetary Economics*, 135, 1–19.

PRIESTLEY, M. B. (1981): *Spectral analysis and time series: Univariate series*, vol. 1, Academic press.

RAMEY, V. A. (2016): "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2, 71–162.

RAPACH, D. AND G. ZHOU (2021): "Sparse macro factors," *Available at SSRN 3259447*.

RENAULT, E. AND U. TRIACCA (2015): "Causality and separability," *Statistics & Probability Letters*, 99, 1–5.

SHAO, X. AND J. ZHANG (2014): "Martingale difference correlation and its use in high-dimensional variable screening," *Journal of the American Statistical Association*, 109, 1302–1318.

SIMS, C. A. (1980): "Macroeconomics and reality," *Econometrica: journal of the Econometric Society*, 1–48.

STOCK, J. H. AND M. W. WATSON (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American statistical association*, 97, 1167–1179.

SZÉKELY, G. J., M. L. RIZZO, AND N. K. BAKIROV (2007): "Measuring and testing dependence by correlation of distances," .

TCHAHOU, H. N. AND P. DUCHESNE (2013): "On testing for causality in variance between two multivariate time series," *Journal of Statistical Computation and Simulation*, 83, 2064–2092.

WANG, G., K. ZHU, AND X. SHAO (2022): "Testing for the martingale difference hypothesis in multivariate time series models," *Journal of Business & Economic Statistics*, 40, 980–994.

WANG, X. (2024): "Generalized Spectral Tests for Multivariate Martingale Difference Hypotheses," *Journal of Business & Economic Statistics*, 1–27.

WHITE, H. AND D. PETTENUZZO (2014): "Granger causality, exogeneity, cointegration, and economic policy analysis," *Journal of Econometrics*, 178, 316–330.

A Appendix

A.1 Formulations of the statistics

Lemma A.1.

$$2\pi \int_{2\pi} vec[\overline{f(\hat{\lambda})}]' vec[f(\hat{\lambda})] = T_\omega$$

Proof. By using the property $\text{tr}(A'C) = \text{vec}(A)'\text{vec}(C)$, together with the properties of the conjugates[29] and the interchangeability of trace and integral operator, we can rewrite the former as:

$$2\pi \int_{2\pi} \text{vec}[\overline{f(\hat{\lambda})}]'\text{vec}[f(\hat{\lambda})] = 2\pi \int_{2\pi} \textbf{tr}\Big([\overline{f(\hat{\lambda})}]'[f(\hat{\lambda})]\Big) = 2\pi\textbf{tr}\left(\int_{2\pi} \overline{[f(\hat{\lambda})]'}[f(\hat{\lambda})]\right)$$

By the definition of an appropriate kernel estimator of the cross-spectrum,

$$\hat{f}(\lambda) = \frac{1}{2\pi}\sum_{j=0}^{T-1}(\omega(j))^{1/2}\widehat{\Gamma}_{XZ}(j)e^{-ij\lambda}$$

with i being the imaginary unit, and by virtue of Parseval's theorem, we rewrite the test statistic as follows:

$$\begin{aligned}
2\pi\textbf{tr}\left(\int_{2\pi} \overline{[f(\hat{\lambda})]'}[f(\hat{\lambda})]\right) &= 2\pi\textbf{tr}\left(2\pi\sum_{j=0}^{\infty}\frac{1}{4\pi^2}\omega(j)\left(\overline{\widehat{\Gamma}_{XZ}(j)'}\right)\left(\widehat{\Gamma}_{XZ}(j)\right)\right) \\
&= \sum_{j=0}^{\infty}\omega(j)\textbf{tr}\left(\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)'\right) = \sum_{j=0}^{T-1}\omega(j)Q(j)
\end{aligned}$$

where the last equality follows from the trace property and the definition in eq.(2). □

[29] Namely: the sum of the conjugate is the conjugate of the sum, the conjugate of the transpose is the transpose of the conjugate, the conjugate of a real matrix is the real matrix itself.

Lemma A.2.

$$
\begin{aligned}
\mathcal{T}_\omega &= \sum_{j=0}^{T-1} \omega(j) Q(j) = \mathcal{T}_{1\omega} + \mathcal{T}_{2\omega} \\
&= \sum_{j=0}^{T-1} \omega(j) \left(\frac{1}{T^2} \sum_{t=j+1}^{T} ||X_t||^2 ||Z_{t-j}||^2 + \frac{1}{T^2} \sum_{s,t=j+1, s\neq t}^{T} < X_t, X_s >< Z_{t-j}, Z_{s-j} > \right)
\end{aligned}
$$

Proof. It suffices to show that:

$$
\begin{aligned}
Q(j) &= \text{Tr}\left[\widehat{\Gamma}_{XZ}(j)' \widehat{\Gamma}_{XZ}(j) \right] \\
&= \frac{1}{T^2} \text{Tr}\left[\left(\sum_{t=j+1}^{T} Z_{t-j} X_t' \right) \left(\sum_{t=j+1}^{T} X_t Z_{t-j}' \right) \right] \\
&= \frac{1}{T^2} \sum_{s,t=j+1}^{T} \text{Tr}\left[Z_{t-j} X_t' X_s Z_{s-j}' \right] \\
&= \frac{1}{T^2} \sum_{s,t=j+1}^{T} < X_t, X_s >< Z_{t-j}, Z_{s-j} >
\end{aligned}
$$

□

Lemma A.3. *Define:* $\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) = \frac{1}{T} \sum_{t=j+1}^{T} \widehat{X}_t \widehat{Z}_{t-j}'$ *(see Appendix B.3). We have:*

$$
\widehat{Q}(j) = vec\left(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) \right)' \left[\left(\widehat{\Gamma}_Z \right)^{-1} \otimes \left(\widehat{\Gamma}_X \right)^{-1} \right] vec\left(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) \right)
$$

Proof. We have:

$$
\begin{aligned}
\widehat{Q}(j) &= \text{vec}\left(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) \right)' \left[\left(\widehat{\Gamma}_Z \right)^{-1} \otimes \left(\widehat{\Gamma}_X \right)^{-1} \right] \text{vec}\left(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) \right) \\
&= \text{vec}\left[(\widehat{\Gamma}_X)^{-1/2} \widehat{\Gamma}_{\hat{X}\hat{Z}}(j) (\widehat{\Gamma}_Z^{-1/2})' \right]' \text{vec}\left[(\widehat{\Gamma}_X)^{-1/2} \widehat{\Gamma}_{\hat{X}\hat{Z}}(j) (\widehat{\Gamma}_Z^{-1/2})' \right] \\
&= \left|\left| \text{vec}\left[(\widehat{\Gamma}_X)^{-1/2} \widehat{\Gamma}_{\hat{X}\hat{Z}}(j) (\widehat{\Gamma}_Z^{-1/2})' \right] \right|\right|^2 = \left|\left| \text{vec}\left[\widehat{\Gamma}_{UV}(j) \right] \right|\right|^2
\end{aligned}
$$

where the last equality is due the definition in eq.(9). Clearly, we see that if the estimated innovations are the actual population-standardized innovation, the statistic is equal to the one defined in eq.(3). □

A.2 Connection between quadratic form and KL divergence: Proposition 3

The next proposition points out a symmetry that is intrinsically linked to the quadratic forms. In particular, it shows the explicit connection between the quadratic form and the Kullback-Leibler divergence.

Proposition 3. SYMMETRIC PERTURBATION.
For a positive $\epsilon \to 0$, we have the following:

$$\ln\left(\det\left(I_{d_1} - \epsilon\widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right)\right) = \ln\left(\det\left(I_{d_2} - \epsilon\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)\right)\right) + O(\epsilon^2)$$

and:

$$\ln\left(\det\left(I_{d_1} - \epsilon\widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right)\right) \approx \ln\left(\det\left(I_{d_2} - \epsilon\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)\right)\right) \approx -\epsilon Q(j)$$

Proof. By the Jacobi's formula up to order 1:

$$\det\left(I_{d_1} - \epsilon\widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right) = 1 - \epsilon\mathrm{tr}\left[\widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right] + O(\epsilon^2) \approx 1 - \epsilon Q(j)$$

Since, for perturbation near the identity, the determinant behaves like the trace, as we generally have: $\det(1 + \epsilon A) = 1 + \epsilon\mathrm{tr}[A] + O(\epsilon^2)$, for any bounded square matrix A and infinitesimal ε.[30] Similarly:

$$\begin{aligned}\det\left(I_{d_2} - \epsilon\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)\right) &= 1 - \epsilon\mathrm{tr}\left[\widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right] + O(\epsilon^2)\\ &= 1 - \epsilon\mathrm{tr}\left[\widehat{\Gamma}_{XZ}(j)'\widehat{\Gamma}_{XZ}(j)\right] + O(\epsilon^2) \approx 1 - \epsilon Q(j)\end{aligned}$$

where the second equality is due the property of the trace operator.
The approximation is established by the fact that, for small x, we have: $\ln(1 + x) \approx x$.
For the connection to the KL divergence measure, note that:

i) If $\{X_t, Z_{t-j}\}$ are jointly distributed: $\begin{pmatrix} X_t \\ Z_{t-j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{d_1} & \epsilon^{1/2}\Gamma_{XZ}(j) \\ \cdot & I_{d_2} \end{pmatrix}\right)$,

[30] For the Jacobi's formula, a useful discussion can be found here: https://terrytao.wordpress.com/2013/01/13/matrix-identities-as-derivatives-of-determinant-identities/.

we have: $X_t|Z_{t-j} \sim \mathcal{N}(0, \Sigma)$, with:

$$\Sigma = I_{d_1} - \epsilon \Gamma_{XZ}(j)(\Gamma_{XZ}(j))'$$

The same holds for the conditional distribution of Z_{t-j} with respect to X_t, with conditional variance that amounts to: $I_{d_2} - \epsilon \Gamma_{XZ}(j)' \Gamma_{XZ}(j)$.

ii) The KL divergence, D_{KL}, betweeen two mean-zero k-multivariate normal with covariance-variance matrices Σ and $\Sigma_0 = I_k$ is defined as:

$$D_{KL} = \frac{1}{2}\left(\ln(\det(I_k)/\det(\Sigma)) + \text{tr}[\Sigma] - k\right) \propto -\ln(\det(\Sigma))$$

which hints that the previous quantities should be understood as:

$$\ln\left(\frac{\det(I_{d_1})}{\det\left(I_{d_1} - \epsilon \widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right)}\right) = -\ln\left(\det\left(I_{d_1} - \epsilon \widehat{\Gamma}_{XZ}(j)\widehat{\Gamma}_{XZ}(j)'\right)\right)$$

□

Proposition 3 proves that, when the cross-correlation between the two time series is small, the quadratic forms, $\{Q(j)\}$, are measures connected to the variances of two conditional distributions. In particular, the quadratic forms are inversely proportional to the determinant of the perturbed variance of the conditional distribution of $\{X_t\}$ with respect to $\{Z_{t-j}\}$, and of the conditional distribution of $\{Z_{t-j}\}$ with respect to $\{X_t\}$. When $\{X_t, Z_{t-j}\}$ are jointly normal distributed with covariance $\epsilon^{1/2}\Gamma_{XZ}(j)$, we have:

$$X_t|Z_{t-j} \sim \mathcal{N}\left(0, I_{d_1} - \epsilon \Gamma_{XZ}(j)(\Gamma_{XZ}(j))'\right), \quad Z_{t-j}|X_t \sim \mathcal{N}\left(0, I_{d_2} - \epsilon(\Gamma_{XZ}(j))'\Gamma_{XZ}(j)\right)$$

To see the connection with the measure of linear feedback *à la* Geweke (1982), one can think that the processes $\{X_t, Z_t\}$ as residuals of a linear projection system (pg. 305 of Geweke (1982)) with respect to a stationary nondeterministic multivariate time series

$\{W_t\} = \{W_{1,t}, W_{2,t}\}$:

$$W_{1,t} = \sum_{s=1}^{\infty} A_s W_{1,s} + X_t$$
$$W_{1,t} = \sum_{s=1}^{\infty} B_s^{(1)} W_{1,s} + \sum_{s=1}^{\infty} B_s^{(2)} W_{2,s} + Z_t$$

with

A.3 Proof of Lemma 1

a) For the first part, I consider the scenario where the DGPs are such that:

$$Z_t^2 = f(X_{t-1}) + u_t, \quad u_t \sim i.i.d.(1, \sigma_u^2), \quad X_t \perp\!\!\!\perp u_s, \forall s, t$$

where $f(\cdot) : \mathbb{R} \to \mathbb{R}$ is a measureable function.
Since $\{X_t\}$ is an i.i.d. process that is independent of u_t, we also have:

$$\begin{aligned}\mathbb{E}[X_t|\mathcal{I}(t-1)] = \mathbb{E}[X_t] = 0\\ \mathbb{E}[X_t^2|\mathcal{I}(t-1)] = \mathbb{E}[X_t^2] = 1\end{aligned}$$

By the same logic of Proposition 1, we have:

$$\begin{aligned}&\mathbb{E}[X_t X_s Z_{t-j} Z_{s-j}] = 0\\ &\mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2]\\ &\mathbb{E}[(X_t, X_s Z_{t-j}, Z_{s-j})(X_{t+h}, X_{s+h} Z_{t-j+h}, Z_{s-j+h})] = 0, \quad \forall h \neq 0\end{aligned}$$

Denote: $\sigma_f^2 = \mathbb{E}[f(X_t)]$, $\varrho_f = \mathbb{E}[X_t^2 f(X_t)]$.
Under the assumption about the DGP:

$$\begin{aligned}&\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = (\sigma_f + 1)^2 = \sigma_z^4, \quad \text{for } s \geq t-j\\ &\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = (\varrho_f + 1)(\sigma_f + 1) = (\varrho_f + 1)\sigma_z^2, \quad \text{for } s = t-j-1\\ &\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = (\sigma_f + 1)^2 = \sigma_z^4, \quad \text{for } s < t-j-1\end{aligned}$$

Thus:

$$\begin{aligned}\text{Var}[\mathcal{T}_{2\omega}] &= \frac{1}{T^4} \sum_{j=0}^{T-1} \omega^2(j) \sum_{s,t=j+1, s\neq t}^{T} \mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2]\\ &= \frac{\sigma_z^4}{T^4} \sum_{j=0}^{T-1} \omega^2(j) \sum_{s,t=j+1, s\neq t}^{T} 1 + \left(\frac{\varrho_f + 1}{\sigma_z^2} - 1\right) \mathbb{1}\{s = t-j-1\}\end{aligned}$$

Now I consider the scenario where the DGPs are such that:

$$Z_t = g(X_{t-1}) + \epsilon_t, \quad \epsilon_t \sim i.i.d.(0, 1), \quad X_t \perp\!\!\!\perp \epsilon_s, \forall s, t$$

with $g(\cdot) : \mathbb{R} \to \mathbb{R}$ being a measureable function. Denote $\sigma_g^2 = \mathbb{E}[g(X_t)^2]$, $\varrho_g = \mathbb{E}[X_t^2 g(X_t)^2]$.
First, note that, since we have: $\mathbb{E}[Z_t] = 0$, then we must have: $\mathbb{E}[g(X_{t-1})] = 0$.
Second, under this class DGPs for Z, we have:

$$\text{Var}(Z_t) = \mathbb{E}[Z_t^2] = \sigma_g^2 + 1$$

Similar to the previous case, in this class of scenarios, we have:

$$\begin{aligned}\mathbb{E}[X_t|\mathcal{I}(t-1)] &= \mathbb{E}[X_t] = 0\\ \mathbb{E}[X_t^2|\mathcal{I}(t-1)] &= \mathbb{E}[X_t^2] = 1\end{aligned}$$

Again, by direct application of Proposition 1, we have:

$$\begin{aligned}&\mathbb{E}[X_t X_s Z_{t-j} Z_{s-j}] = 0\\ &\mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2]\\ &\mathbb{E}[(X_t, X_s Z_{t-j}, Z_{s-j})(X_{t+h}, X_{s+h} Z_{t-j+h}, Z_{s-j+h})] = 0, \quad \forall h \neq 0\end{aligned}$$

Thus:

$$\begin{aligned}\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] &= (\sigma_g + 1)^2 = \sigma_z^4, \quad \text{for } s \geq t-j\\ \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] &= (\varrho_g + 1)(\sigma_g + 1) = (\varrho_g + 1)\sigma_z^2, \quad \text{for } s = t-j-1\\ \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] &= (\sigma_g + 1)^2 = \sigma_z^4, \quad \text{for } s < t-j-1\end{aligned}$$

Finally:

$$\text{Var}[\mathcal{T}_{2\omega}] = \frac{\sigma_z^4}{T^4}\sum_{j=0}^{T-1}\omega^2(j)\sum_{s,t=j+1,s\neq t}^{T}\left(\frac{\varrho_g+1}{\sigma_Z^2} - 1\right)\mathbb{1}\{s = t-j-1\}$$

b) Finally, I consider the scenario where the DGPs are such that:

$$Z_t^2 = \alpha Z_{t-1}^2 + h(X_{t-1}) + u_t, \quad X_t \sim i.i.d.(0,1), \quad u_t \sim i.i.d.(1, \sigma_u^2), \quad X_t \perp\!\!\!\perp u_s, \forall s,t$$

for some measurable function $h(\cdot) : \mathbb{R} \to \mathbb{R}$ and $|\alpha| \in (0,1)$.
Denote: $\mu_h = \mathbb{E}[h(X_{s+1})]$ and $\varrho_h = \mathbb{E}[h(X_s)X_s^2]$, $\mu_e = \mathbb{E}[h(X_{t-1}) + u_t] = \mu_h + 1$, and

$\sigma_e^2 = \mathbb{E}[(h(X_{t-1}) + u_t)^2]$, so that we finally denote:

$$\mu_Z = \mu_e^2/(1-\alpha)$$
$$\sigma_Z^2 = \frac{\sigma_e^2 + 2\alpha\mu_e^2/(1-\alpha)}{1-\alpha^2}$$

Similar to the previous case, we have:

$$\mathbb{E}[X_t|\mathcal{I}(t-1)] = \mathbb{E}[X_t] = 0$$
$$\mathbb{E}[X_t^2|\mathcal{I}(t-1)] = \mathbb{E}[X_t^2] = 1$$

By the same logic of Proposition 1, we have:

$$\mathbb{E}[X_t X_s Z_{t-j} Z_{s-j}] = 0$$
$$\mathbb{E}[X_t^2 X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2]$$
$$\mathbb{E}[(X_t, X_s Z_{t-j}, Z_{s-j})(X_{t+h}, X_{s+h} Z_{t-j+h}, Z_{s-j+h})] = 0, \quad \forall h \neq 0$$

Note that the process $\{Z_t^2\}$ is a causal AR(1) process with i.i.d. noise, since we have:

$$Z_{t-1} \perp\!\!\!\perp X_{t-1}, \quad u_t \perp\!\!\!\perp X_{t-1}$$

Consider the two scenarios:

i) When $s \geq t-j$, we have:

$$\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[X_s^2]\mathbb{E}[Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[Z_{t-j}^2 Z_{s-j}^2]$$
$$= \alpha^{|j-v_1|}\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-v_1|-1} \alpha^l$$

where $v_1 = s - t + j \geq 0$.

To see the last equality, note that we have the following:

- For $s = t-j$, that is $v_1 = 0$:

$$\begin{aligned}
\mathbb{E}[Z_{t-j}^2 Z_{s-j}^2] &= \mathbb{E}[Z_s^2 Z_{s-j}^2] \\
&= \alpha^j \mathbb{E}[Z_{s-j}^4] + \mathbb{E}[Z_{s-j}^2]\mu_e \sum_{l=0}^{j-1} \alpha^l \\
&= \alpha^j \mathbb{E}[Z_{s-j}^4] + \frac{1}{1-\alpha}\mu_e^2 \sum_{l=0}^{j-1} \alpha^l \\
&= \alpha^j \left(\frac{\sigma_e^2 + 2\alpha\mu_e^2/(1-\alpha)}{1-\alpha^2} \right) + \frac{\mu_e^2}{1-\alpha} \sum_{l=0}^{j-1} \alpha^l \\
&= \alpha^j \sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{j-1} \alpha^l
\end{aligned}$$

- For $s = t-j+1$, that is $v_1 = 1$:

$$\mathbb{E}[Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] = \alpha^{|j-1|}\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-1|-1} \alpha^l$$

since:

$$\begin{aligned}
\mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-1}^2 Z_s^2] = \alpha\sigma_Z^2 + \mu_Z^2 && \text{for } j = 0 \\
\mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-1}^2 Z_{s-1}^2] = \sigma_Z^2 && \text{for } j = 1 \\
&\cdots \\
\mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-1}^2 Z_{m+1}^2] = \alpha^m\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{m-1} \alpha^l && \text{for } j = m+1
\end{aligned}$$

- For $s = t-j+2$, that is $v_1 = 2$:

$$\mathbb{E}[Z_{t-j}^2 Z_{s-j}^2] = \mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] = \alpha^{|j-2|}\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-2|-1} \alpha^l$$

since:

$$\begin{aligned}
\mathbb{E}[Z_{s-2}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-2}^2 Z_s^2] = \alpha^2\sigma_Z^2 + \mu_Z^2(1+\alpha) && \text{for } j=0 \\
&\dots \\
\mathbb{E}[Z_{s-2}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-2}^2 Z_{s-2}^2] = \sigma_Z^2 && \text{for } j=2 \\
&\dots \\
\mathbb{E}[Z_{s-1}^2 Z_{s-j}^2] &= \mathbb{E}[Z_{s-1}^2 Z_{m+1}^2] = \alpha^m\sigma_Z^2 + \mu_Z^2\sum_{l=0}^{m-1}\alpha^l && \text{for } j=m+2
\end{aligned}$$

Summing up the terms, we have:

$$\sum_{s\geq t-j, s\neq t}^{T} \mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \sum_{v_1=0}^{\tau(j)} \left(\alpha^{|j-v_1|}\sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-v_1|-1} \alpha^l\right)$$

where $\tau(j)$ is the number of times that $s \geq t-j$ up to T, at a given $j \geq 0$, such that: $s \neq t$, and $s, t = j+1$. In particular, we have:

$$\begin{aligned}
\tau(j) = &\frac{(T-j)(T-j-1)}{2} \\
&+ \mathbb{1}(2j-T-1<0)\cdot\big((T-2j-1)j + j(j+1)/2\big) \\
&+ \mathbb{1}(2j-T-1\geq 0)\cdot(T-j)(T-j-1)/2
\end{aligned}$$

To see that, note that the summation is with respect to two indexes:

i. The first term, $\frac{(T-j)(T-j-1)}{2}$, is the number of times that $s \geq t-j$ when increasing along the s index;

ii. The second term, $(T-2j-1)j + j(j+1)/2$, is the number of times that $s \geq t-j$ while increasing along the t index when $j < \frac{T+1}{2}$;

iii. The third term, $\frac{(T-j)(T-j-1)}{2}$, is the number of times that $s \geq t-j$ while increasing along the t index when $j \geq \frac{T+1}{2}$.

Here below, I present a table that helps to visualize the previous terms:

	$t=j+1$	$t=j+2$	$t=j+3$	$t=j+4$	$j+4<t<2j+2$	$t=2j+2$	$t=2j+3$	$t=2j+4$
$s=j+1$	0	1	1	1	1	0	0	0
$s=j+2$	1	0	1	1	1	1	0	0
$s=j+3$	1	1	0	1	1	1	1	0
$s=j+4$	1	1	1	0	1	1	1	1
$\cdots$	.	.	.	.	.	.	.	.
$s=2j+2$	1	1	1	1	1	0	1	1
$s=2j+3$	1	1	1	1	1	1	0	1
$s=2j+4$	1	1	1	1	1	1	1	0

The entries take value one if $s \geq t-j$, i.e. $\mathbb{1}(s \geq t-j)$. Because of the additional condtion that $s \neq t$, the diagonal entries are zero.
The first term, $(T-j)(T-j-1)/2$, is the summation over the lower-triangular entries. The second and third terms are the summations over the upper-triangular entries. When $j < \frac{T+1}{2}$, some entries are zeros on the top-right corner, meaning that the upper sum is j-times $T-2j-1$, plus the bottom-right corner of the upper-triangular part, which consists of $j(j+1)/2$ elements. When $j \geq \frac{T+1}{2}$, the top-right corner contains no null entry or, in other words, the entries are symmetric about the main diagonal, which means that the upper-triangular summation is identical to the lower-triangular one.

ii) when $s < t-j$, denote $\upsilon_2 = t-s-j > 0$. We have the following:

$$\mathbb{E}[X_s^2 Z_{t-j}^2 Z_{s-j}^2] = \alpha^{j+\nu_2}\sigma_Z^2 + \mu_Z \left[\mu_Z \sum_{m_1=\nu_2}^{j+\nu_2-1} \alpha^{m_1} + (1+\varrho_h)\alpha^{\nu_2-1} + (1+\mu_h)\sum_{m_2=0}^{\nu_2-2}\alpha^{m_2}\right]$$

To see it, note that:

- For $s = t-j-1$, meaning $\upsilon_2 = 1$:

$$\begin{aligned}
\mathbb{E}[Z_{s+1}^2 X_s^2 Z_{s-j}^2] &= \alpha\mathbb{E}[X_s^2 Z_s^2 Z_{s-j}^2] + \mathbb{E}[u_{s+1}X_s^2 Z_{s-j}^2] + \mathbb{E}[h(X_s)X_s^2 Z_{s-j}^2] \\
&= \alpha\mathbb{E}[Z_s^2 Z_{s-j}^2] + \mathbb{E}[Z_{s-j}^2] + \mathbb{E}[h(X_s)X_s^2 Z_{s-j}^2] \\
&= \alpha\left(\alpha^j\sigma_Z^2 + \mu_Z^2\sum_{l=0}^{j-1}\alpha^l\right) + \mu_Z + \varrho_h\mu_Z \\
&= \alpha^{j+1}\sigma_Z^2 + \mu_Z^2\sum_{l=1}^{j}\alpha^l + \mu_Z(1+\varrho_h)
\end{aligned}$$

- For $s = t - j - 2$, meaning $\upsilon_2 = 2$:

$$\begin{aligned}
\mathbb{E}[Z_{s+2}^2 X_s^2 Z_{s-j}^2] &= \alpha \mathbb{E}[Z_{s+1}^2 X_s^2 Z_{s-j}^2] + \mathbb{E}[u_{s+2} X_s^2 Z_{s-j}^2] + \mathbb{E}[h(X_{s+1}) X_s^2 Z_{s-j}^2] \\
&= \alpha \mathbb{E}[Z_{s+1}^2 X_s^2 Z_{s-j}^2] + \mathbb{E}[Z_{s-j}^2] + \mathbb{E}[h(X_{s+1})]\mathbb{E}[Z_{s-j}^2] \\
&= \alpha \left(\alpha^{j+1} \sigma_Z^2 + \mu_Z^2 \sum_{l=1}^{j} \alpha^l + \mu_Z (1 + \varrho_h) \right) + \mu_Z (1 + \mu_h) \\
&= \alpha^{j+2} \sigma_Z^2 + \mu_Z^2 \sum_{l=2}^{j+1} \alpha^l + \alpha \mu_Z (1 + \varrho_h) + \mu_Z (1 + \mu_h)
\end{aligned}$$

- For $s = t - j - 3$, meaning $\upsilon_2 = 3$:

$$\begin{aligned}
&\mathbb{E}[Z_{s+3}^2 X_s^2 Z_{s-j}^2] = \alpha \mathbb{E}[Z_{s+2}^2 X_s^2 Z_{s-j}^2] + \mu_Z (1 + \mu_h) \\
&= \alpha \left(\alpha^{j+2} \sigma_Z^2 + \mu_Z^2 \sum_{l=2}^{j+1} \alpha^l + \alpha \mu_Z (1 + \varrho_h) + \mu_Z (1 + \mu_h) \right) + \mu_Z (1 + \mu_h)
\end{aligned}$$

Thus, reconciling both results, we conclude:

$$\begin{aligned}
&T^4 \mathrm{Var}[\mathcal{T}_{2\omega}] = \sum_{j=0}^{T-2} \omega^2(j) \sum_{\upsilon_1=0}^{\tau(j)} \left(\alpha^{|j-\upsilon_1|} \sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-\upsilon_1|-1} \alpha^l \right) \\
&+ \sum_{\upsilon_2=1}^{(T-j)(T-j-1)-\tau(j)} \left(\alpha^{j+\nu_2} \sigma_Z^2 + \mu_Z \left[\mu_Z \sum_{m_1=\nu_2}^{j+\nu_2-1} \alpha^{m_1} + (1 + \varrho_h) \alpha^{\nu_2 - 1} + (1 + \mu_h) \sum_{m_2=0}^{\nu_2-2} \alpha^{m_2} \right] \right) \\
&= \sum_{j=0}^{T-2} \omega^2(j) \sum_{\upsilon_1=0}^{\tau(j)} \left(\alpha^{|j-\upsilon_1|} \sigma_Z^2 + \mu_Z^2 \sum_{l=0}^{|j-\upsilon_1|-1} \alpha^l \right) + \sum_{\upsilon_2=1}^{(T-j)(T-j-1)-\tau(j)} \left(\alpha^{j+\nu_2} \sigma_Z^2 + \mu_Z^2 \sum_{m_1=\nu_2}^{j+\nu_2-1} \alpha^{m_1} \right) \\
&+ \mu_Z \sum_{j=0}^{T-2} \omega^2(j) \sum_{\upsilon_2=1}^{(T-j)(T-j-1)-\tau(j)} \left((1 + \varrho_h) \alpha^{\nu_2 - 1} + (1 + \mu_h) \sum_{m_2=0}^{\nu_2-2} \alpha^{m_2} \right)
\end{aligned}$$

B Appendix A: Proof of Proposition 2, Theorem 1 to 3

B.1 Proof of Proposition 2

i) The proof consists of three parts:

A. From its definition in eq.(4):

$$\mathbb{E}[T \cdot \mathcal{T}_{1\omega}] = \frac{1}{T}\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}\mathbb{E}[||X_t||^2||Z_{t-j}||^2]$$

The expectations are such that, by law of iterated expectations:

$$\mathbb{E}[||X_t||^2||Z_{t-j}||^2] = \mathbb{E}[\mathbb{E}[||X_t||^2|\mathcal{I}(t-1)]||Z_{t-j}||^2] = d_1\mathbb{E}[||Z_{t-j}||^2] = d_1 d_2$$

where the second equality is because of the assumption of conditional homoskedasticity, $\mathbb{E}[X_t X_t'|\mathcal{I}(t-1)] = \mathbb{E}[X_t X_t']$, and the processes being standardized. Substituting above:

$$\mathbb{E}[T \cdot \mathcal{T}_{1\omega}] = \frac{1}{T}\sum_{j=0}^{T-1}\omega(j)(T-j)d_1 d_2 = \mu_{\omega,T}$$

B. By definition:

$$\text{Var}[T \cdot \mathcal{T}_{1\omega}] = \mathbb{E}[(T \cdot \mathcal{T}_{1\omega} - \mu_{\omega,T})^2]$$

For simplicity, denote:

$$T \cdot \mathcal{T}_{1\omega} - \mu_{\omega,T} = \frac{1}{T}\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}\left(||X_t||^2||Z_{t-j}||^2 - d_1 d_2\right) = \frac{1}{T}\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}$$

so that, using the Mean Square Error (MSE) norm (i.e., $||\cdot||_{MSE} = \mathbb{E}[\cdot]^2$)

the previous expression can be bounded as follows:

$$\begin{aligned}\text{Var}[T\mathcal{T}_{1\omega}] &= \frac{1}{T^2}\mathbb{E}\left[\left(\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right)^2\right] = \frac{1}{T^2}\left|\left|\sum_{j=0}^{T-1}\omega(j)\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right|\right|_{MSE}^2 \\ &\leq \frac{1}{T^2}\left(\sum_{j=0}^{T-1}\omega(j)\left|\left|\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right|\right|_{MSE}\right)^2\end{aligned}$$

by virtue of Minkowski inequality.

By the assumptions on the conditional moments, we have:

$$\mathbb{E}[||X_t||^4||Z_{t-j}||^4] = \mathbb{E}[\mathbb{E}[X_t'X_tX_t'X_t|\mathcal{I}(t-1)]||Z_{t-j}||^4] = \mathbb{E}[||X_t||^4||]\mathbb{E}[Z_{t-j}||^4]$$

thus:

$$\begin{aligned}\left|\left|\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right|\right|_{MSE}^2 &= \mathbb{E}\left[\left(\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right)^2\right] \\ &\leq \mathbb{E}\left[\left(\sum_{t=j+1}^{T}||X_t||^4||Z_t||^4\right) + 2d_1d_2\left(\sum_{t=j+1}^{T}||X_t||^2||Z_t||^2\right)\right. \\ &\qquad \left. + d_1^2d_2^2(T-j)\right. \\ &\leq (T-j)\max\{3d_1^2d_2^2, \mathbb{E}[||X_t||^4||]\mathbb{E}[Z_{t-j}||^4]\} = O(T)\end{aligned}$$

where the last equality is by the finiteness of the fourth moments.

Equivalently, this means:

$$\left|\left|\sum_{t=j+1}^{T}\Upsilon_{t,j}^{(1)}\right|\right|_{MSE} = O(T^{1/2})$$

Thus:

$$\text{Var}[T\mathfrak{T}_{1\omega}] \leq \frac{\Delta}{T}\left(\sum_{j=1}^{T-1}\omega(j)\right)^2 = \frac{\Delta}{T}M^2\left(M^{-1}\sum_{j=1}^{T-1}\omega(j)\right)^2 = O(M^2/T)$$

for finite $\Delta > 0$, where the last equality comes by realizing that, under

Assumption 1, the following convergence holds:

$$M^{-1}\sum_{j=1}^{T-1}\omega(j) \to \int_0^\infty k^2(z)dz < \infty$$

C. By the same reason as above, under Assumption 1, the following convergence holds:

$$M^{-1}D_{\omega,T} \to \frac{1}{2}\int_0^\infty k^4(z)dz < \infty$$

so that:

$$D_{\omega,T} = O(M)$$

Given the finiteness of the fourth moments of Z, and the result in Lemma 1, we also have that:

$$\begin{aligned}
D^{(Hete)}_{\omega,T} &= \frac{d_1^2}{T^2}\sum_{j=1}^{T-1}\omega^2(j)\sum_{s,t=j+1,s\geq t-j}^{T}\gamma_{t-j,s-j} \asymp \frac{1}{T^2}\sum_{j=1}^{T-1}\omega^2(j)\sum_{l=1}^{\tau(j)} \\
&\asymp \sum_{j=1}^{T-1}\omega^2(j)\Bigg(\frac{(T-j)(T-j-1)}{2T^2} + \mathbb{1}(2j-T-1<0) \\
&\qquad \cdot \frac{\big((T-2j-1)j + j(j+1)/2\big)}{T^2} \\
&\qquad + \mathbb{1}(2j-T-1\geq 0)\frac{(T-j)(T-j-1)}{2T^2}\Bigg) \\
&= O(M)
\end{aligned}$$

where $a \asymp b$ stands for the joint condition: $a = O(b)$ and $b = O(a)$. Together with the previous findings, we have:

$$\mathbb{E}\left[\frac{T\cdot\mathcal{T}_{1\omega} - \mu_{\omega,T}}{\sqrt{D^{(Hete)}_{\omega,T}}}\right] = 0, \quad \text{Var}\left[\frac{T\cdot\mathcal{T}_{1\omega} - \mu_{\omega,T}}{\sqrt{D^{(Hete)}_{\omega,T}}}\right] = O(M/T)$$

so that, under the asymptotics of Assumption 1, when $M/T \to 0$ as $T \to \infty$, we have ℓ_2 (or MSE) convergence to zero.

ii) The proof consists of two parts:

A. From its definition in eq.(5):

$$\begin{aligned}\mathbb{E}[T \cdot \mathcal{T}^c_{2\omega}] &= \frac{2}{T}\sum_{j=0}^{T-2}\omega(j)\sum_{t=j+2}^{T}\sum_{s=t-j}^{t-1}\mathbb{E}[< X_t, X_s >< Z_{t-j}, Z_{s-j} >] \\ &= \frac{2}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=t-s}^{s-1}\omega(j)\mathbb{E}[< X_t, X_s >< Z_{t-j}, Z_{s-j} >]\end{aligned}$$

Under the null hypothesis in eq.(1) and by law of iterated expectations:

$$\mathbb{E}[< X_t, X_s >< Z_{t-j}, Z_{s-j} >] = \mathbb{E}[\mathbb{E}[X_t'|\mathcal{I}(t-1)]X_s < Z_{t-j}, Z_{s-j} >] = 0$$

as the indexes are such that $t > s$. Thus:

$$\mathbb{E}[T \cdot \mathcal{T}^c_{2\omega}] = 0$$

B. Since, under the null hypothesis in eq.(1), the following property holds:

$$\mathbb{E}[< X_t, X_s >< Z_{t-j}, Z_{s-j} >< X_{t-h}, X_{s-h} >< Z_{t-j-h}, Z_{s-j-h} >] = 0$$

for any integers $h \neq 0$, as the indexes are always such that: $t > s$.
Given the uncorrelatedness of the process, by definition:

$$\text{Var}[T \cdot \mathcal{T}^c_{2\omega}] = \frac{4}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=t-s}^{s-1}\omega^2(j)\mathbb{E}[(\langle X_t, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle)^2]$$

where:

$$\begin{aligned}\mathbb{E}[(\langle X_t, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle)^2] &= \mathbb{E}[\mathbb{E}[X_t'X_sX_s'X_t\langle Z_{t-j}, Z_{s-j}\rangle^2|\mathcal{I}(t-1)]] \\ &= \mathbb{E}[\mathbb{E}[\text{tr}(X_t'X_t)|\mathcal{I}(t-1)]\langle X_s, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle^2] \\ &= d_1\mathbb{E}[||X_s||^2\langle Z_{t-j}, Z_{s-j}\rangle^2] \\ &= d_1\mathbb{E}[\mathbb{E}[||X_s||^2|\mathcal{I}(s-1)]\langle Z_{t-j}, Z_{s-j}\rangle^2] \\ &= d_1^2\mathbb{E}[Z_{t-j}'Z_{s-j}Z_{s-j}'Z_{t-j}] = d_1^2\gamma_{t-j,s-j}\end{aligned}$$

with $\gamma_{t,s} = \mathbb{E}[\langle Z_t, Z_s\rangle^2]$. If the process Z is fourth-order stationary, then the lag index j could be dropped, meaning that the previous expression could be written as:

$$\mathbb{E}[(\langle X_t, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle)^2] = d_1^2\gamma_{t,s}$$

In conclusions, we have:

$$\text{Var}[T\cdot\mathcal{T}_{2\omega}^c] = \frac{4d_1^2}{T^2}\sum_{j=0}^{T-2}\omega^2(j)\sum_{t=j+2}^{T}\sum_{s=t-j}^{t-1}\gamma_{t-j,s-j}$$

Under the assumption of conditional homoskedasticity of $\{Z_t\}$ with respect to the own information set, that is $\mathbb{E}[Z_tZ_t'|\mathcal{I}_Z(t-1)] = \mathbb{E}[Z_tZ_t']$, by law of iterated expectations, we have:

$$\mathbb{E}[(\langle X_t, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle)^2] = d_1^2d_2^2$$

thus:

$$\text{Var}[T\cdot\mathcal{T}_{2\omega}^c] = \sum_{j=0}^{T-2}\omega^2(j)\sum_{l=1}^{\tau(j)}\frac{d_1^2d_2^2}{T^2}$$

B.2 Proof of Theorem 1

The proof of asymptotic normality of the test statistic, $\mathcal{T}_\omega^c$, translates into proving the asymptotic normality of the dominating term $T\cdot\mathcal{T}_{2\omega}^c$, as because of Proposition 2 we have asymptotically:

$$\frac{T\mathcal{T}_{1\omega}-\mu_{\omega,T}}{\sqrt{D_{\omega,T}^{(Hete)}}} = o_p(1)$$

when $M/T\to 0$ as $M,T\to\infty$. Note that, if we have $M^2/T\to 0$, the asymptotical negligibility holds as well.

For simplicity, define:

$$T \cdot \mathcal{T}_{2\omega}^{c} = \frac{2}{T}\sum_{t=2}^{T} J_t$$
$$J_t = \sum_{s=3}^{t-1}\sum_{j=t-s}^{s-1} \omega(j)\langle X_t, X_s\rangle\langle Z_{t-j}, Z_{s-j}\rangle = \sum_{s=3}^{t-1}\langle X_t, X_s\rangle W_{ts}$$

such that: $W_{ts} = \sum_{j=t-s}^{s-1}\omega(j)\langle Z_{t-j}, Z_{s-j}\rangle$.
Note that, under Assumption 1, $\sum_{j=t-s}^{s-1}\omega(j) = O(M)$, as discussed in Proposition 2.

The process $\{(J_t, \mathcal{I}(t-1)), t \in \mathbb{Z}^+\}$ constitutes a martingale difference sequence, since: 1) we have $\mathbb{E}[J_t|\mathcal{I}(t-1)] = 0$ under the null hypothesis of eq.(1); 2) $\mathbb{E}[|J_t|] < \infty$ by the finiteness of the moments.
Finally, to invoke Brown (1971)'s theorem, two conditions need to be verified.

i) The following Lindeberg condition needs to hold:

$$T^{-2}(D_{\omega}^{(Hete)})^{-1}\sum_{t=2}^{T}\mathbb{E}\Big[(J_t)^2 \cdot \mathbf{1}\big\{|J_t| > \epsilon(D_{\omega}^{(Hete)})^{1/2}\big\}\Big] \to 0$$

To prove it, it is sufficient to show that the Lyapounov condition holds:

$$T^{-4}(D_{\omega}^{(Hete)})^{-2}\sum_{t=2}^{T}\mathbb{E}[(J_t)^4] \to 0$$

We have the following inequalities:

$$\begin{aligned}
\mathbb{E}[(J_t)^4] &\leq \mathbb{E}\left[\sum_{s=1}^{t-1}||X_t||||X_s||W_{ts}\right]^4 \\
&\leq \mathbb{E}\left[\mathbb{E}\left[||X_t||^4\middle|\mathcal{I}(t-1)\right]\left[\sum_{s=1}^{t-1}||X_s||W_{ts}\right]^4\right] \\
&\leq \mathbb{E}[||X_t||^4]\mathbb{E}\left[\sum_{s=1}^{t-1}||X_s||W_{ts}\right]^4
\end{aligned}$$

where the first inequality is by Cauchy-Schwarz, and the last one is by the

assumption on the conditional moments (conditional homokurtosis).
Denote: $\Lambda^{(1)}_{s,t} = ||X_s||W_{ts}$.
Note that, for a given t, $\{(\Lambda^{(1)}_{s,t}, \sigma_\Lambda(s-1) = \sigma(\Lambda^{(1)}_{l,t}; l < s), s \in \mathbb{Z}^+)\}$ is a martingale difference sequence, with respect to its natural $\sigma-$algebra, by the null hypothesis since the indexes are such that $s \geq t-j$, and by the aforementioned justifications for the m.d.s. property of the process J.
Thus, we define:

$$b_{i,t} = \max_{i \leq l \leq t} \left|\left| \Lambda^{(1)}_{i,t} \sum_{k=i}^{l} \mathbb{E}[\Lambda^{(1)}_{k,t} | \sigma_\Lambda(i)] \right|\right|_{MSE} = \max \left\{ 0, \left|\left| \Lambda^{(1)}_{i,t} \right|\right|^2_{MSE} \right\}$$

where the last equality is by the martingale difference property and by law of iterated expectations, since we have for $k > i$, we have $\mathbb{E}[\Lambda^{(1)}_{k,t}|\sigma_\Lambda(i)] = 0$.
The ℓ_2 norm is to be interpreted in MSE terms, as previously referred in Proof of B.1.
By definition:

$$\begin{aligned} \left|\left| \Lambda^{(1)}_{i,t} \right|\right|^2_{MSE} &= \mathbb{E}\left[\left(\Lambda^{(1)}_{i,t}\right)' \left(\Lambda^{(1)}_{i,t}\right) \right] \\ &= \mathbb{E}\left[\mathbb{E}\left[||X_s||^2 | \mathcal{I}(s-1) \right] W^2_{ts} \right] = d_1 \mathbb{E}\left[W^2_{ts}\right] \end{aligned}$$

where the last equality is because of the assumption of the conditional moments (conditional homoskedasticity). By a similar logic of part (b) of proof of Proposition 2, in ℓ_2 norm:

$$\mathbb{E}\left[W^2_{ts}\right] \leq \left(\sum_{j=t-s}^{s-1} \omega(j) \mathbb{E}[< Z_{t-j}, Z_{s-j} >^2] \right)^2 = O(M^2)$$

where the last equality is by the assumption of finiteness of the higher order cross-moments, $\forall t, s, \ |\gamma_{t,s}| < \infty$, and by Assumption 1 on the weighting function.
Note that, this last equality is also trivially satisfied under the assumption of conditional homoskedasticity of the process Z with respect to the own information set, that is: $\mathbb{E}[Z_t Z_t' | \{Z_s, s < t\}] = \mathbb{E}[Z_t Z_t']$.
By application of Proposition 5.4 of Dedecker et al. (2007) (Burkholder's in-

equality):

$$\mathbb{E}[(J_t)^4] \leq \left(4\sum_{s=1}^{t-1} b_{i,t}\right)^2 = O(t^2M^4)$$

which means that under Assumption 1:

$$T^{-4}(D_\omega^{(Hete)})^{-2}\left(\sum_{t=2}^{T}\mathbb{E}[(J_t)^4]\right) = \Delta T^{-4}M^{-2}\left(T^3M^4\right) = O(M^2/T) = o_p(1)$$

for finite $\Delta > 0$, under $M^2/T \to 0$, as $M, T \to \infty$.

To reach a sharper conclusion, the alternative line is to appeal to Marcinkiewicz-Zygmund inequalities. Under the assumptions of the time series $\{\Lambda_{i,t}^{(1)}\}$:

$$|\text{Cov}[\Lambda_{1,t}^{(1)}, \Lambda_{1+i,t}^{(1)}]| = O(i^{-4/2}) = O(i^{-2}), \quad i \to +\infty$$

by application of Theorem 4.1 of Dedecker et al. (2007), we have that:

$$\mathbb{E}[(J_t)^4] = O(t^2M^2)$$

Similar result can be obtained via Rosenthal type inequalities. For further details, please refer to Theorem 4.2 and Corollary 5.4 of Dedecker et al. (2007). For this latter scenario, we have the sharper conclusion:

$$T^{-4}(D_\omega^{(Hete)})^{-2}\left(\sum_{t=2}^{T}\mathbb{E}[(J_t)^4]\right) = \Delta T^{-4}M^{-2}\left(T^3M^2\right) = O(1/T) = o_p(1)$$

for finite $\Delta > 0$, as $T \to \infty$.

ii) The following condition needs to hold:

$$T^{-2}(D_\omega^{(Hete)})^{-1}\sum_{t=2}^{T}\mathbb{E}[(J_t)^2|\mathcal{I}(t-1)] \xrightarrow{p} 1$$

Similar to the previous point, we prove the sufficient condition:

$$T^{-4}(D_{\omega}^{(Hete)})^{-2}\text{Var}\left[\sum_{t=2}^{T}\mathbb{E}[(J_t)^2|\mathcal{I}(t-1)]\right]\to 0$$

To ease the notation, let us define $\Lambda_t' = \sum_{s=1}^{t-1} W_{ts}X_s'$.
Using the trace operator properties, we can express the conditional expectations as follows:

$$\begin{aligned}\mathbb{E}[(J_t^\star)|\mathcal{I}(t-1)] &= \text{tr}\left\{\mathbb{E}\left[\sum_{s=1}^{t-1}W_{ts}X_s'(X_tX_t')\sum_{s=1}^{t-1}X_s(W_{ts})'\middle|\mathcal{I}(t-1)\right]\right\}\\ &\leq \text{tr}\left\{\mathbb{E}\left[||X_t||^2\middle|\mathcal{I}(t-1)\right]\right\}\text{tr}\left\{\mathbb{E}\left[\Lambda_t'\Lambda_t\middle|\mathcal{I}(t-1)\right]\right\} = d_1\cdot\text{tr}\left\{\Lambda_t'\Lambda_t\right\}\end{aligned}$$

where the last equality is because of the assumption of the conditional moments (conditional homoskedasticity), and by realizing that $\Lambda_t \in \mathcal{I}(t-1)$.
Denote:

$$\begin{aligned}\text{tr}\left\{\Lambda_t'\Lambda_t\right\} &= \sum_{s=1}^{t-1}||X_s||^2(W_{ts}^\star)^2 + 2\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\langle X_{s_1}, X_{s_2}\rangle W_{ts_1}W_{ts_2}\\ &= \tilde{B}_t + 2\tilde{A}_t\end{aligned}$$

A. The second addend, $\tilde{A}_t$, is a sum of martingale difference sequences, over the index s_1 , because of the null hypothesis of eq.(1).
Define:

$$\Lambda_{s_1,t}^{(2)} = \sum_{s_2=1}^{s_1-1}\langle X_{s_1}, X_{s_2}\rangle W_{ts_1}W_{ts_2}$$

By similar arguments to the previous ones:

$$\tilde{b}_{i,t} = \max_{i\leq l\leq t}\left|\left|\Lambda_{i,t}^{(2)}\sum_{k=i}^{l}\mathbb{E}[\Lambda_{k,t}^{(2)}|\sigma_\Lambda(i)]\right|\right|_{MSE} = \max\left\{0, \left|\left|\Lambda_{i,t}^{(2)}\right|\right|_{MSE}^2\right\}$$

By the martingale difference property of the sequence, we write:

$$\left|\left|\Lambda_{i,t}^{(2)}\right|\right|_{MSE}^2 = \sum_{s_2=3}^{s_1} ||\langle X_{s_1}, X_{s_2}\rangle W_{ts_1} W_{ts_2}||_{MSE}^2$$

with:

$$\begin{aligned}
||\langle X_{s_1}, X_{s_2}\rangle W_{ts_1} W_{ts_2}||_{MSE}^2 &= \left|\left|\langle X_{s_1}, X_{s_2}\rangle \sum_{j=t-s_1+1}^{s_1-1} \omega(j)\langle Z_{t-j}, Z_{s_1-j}\rangle W_{ts_2}\right|\right|_{MSE}^2 \\
&\leq \left|\left|\langle X_{s_1}, X_{s_2}\rangle \langle Z_{h_1}, Z_{h_2}\rangle M \left(M^{-1} \sum_{j=t-s_1+1}^{s_1-1} \omega(j)\right) W_{ts_2}\right|\right|_{MSE}^2 \\
&\leq M^4 ||\langle X_{s_1}, X_{s_2}\rangle \langle Z_{h_1}, Z_{h_2}\rangle \langle Z_{h_3}, Z_{h_4}\rangle||_{MSE}^2 \\
&\leq M^4 ||\langle X_{s_1}, X_{s_2}\rangle||_{MSE}^4 \left(||\langle Z_{h_1}, Z_{h_2}\rangle||_{MSE}^4\right)^2 = O(M^4)
\end{aligned}$$

for some appropriate indexes $\{h_i\}_{i=1}^4$, where the first inequality is due stationarity, while the last inequality is by use of Cauchy-Schwarz and the assumption of the conditional moments (conditional homokurtosis). The last equality is due the finiteness of the higher moments of the process Z. Thus:

$$\left|\left|\Lambda_{i,t}^{(2)}\right|\right|_{MSE}^2 = O(tM^4)$$

By application of Proposition 5.4 of Dedecker et al. (2007) (Burkholder's inequality):

$$||\tilde{A}_t||_{MSE}^2 = \mathbb{E}[\tilde{A}_t^2] \leq \left(4 \sum_{s_1=2}^{t-1} \tilde{b}_{i,t}\right) = O(t^2 M^4)$$

which in turn means that, under Assumption 1, we have:

$$T^{-4}(D_\omega^{(Hete)})^{-2} \left(\sum_{t=2}^{T} \mathbb{E}[(\tilde{A}_t)^2]\right) = \Delta T^{-4} M^{-2} \left(T^3 M^4\right) = O(M^2/T) = o_p(1)$$

for finite $\Delta > 0$, where the last equality is under Assumption 1 and under the asymptotic rate such that $M^2/T \to 0$, as $M, T \to 0$.

Similar to the previous part i), under the assumptions of the time series $\{\Lambda_i^{(1)}\}$:

$$|\text{Cov}[\Lambda_1^{(1)}, \Lambda_{1+i}^{(1)}]| = O(i^{-4/2}) = O(i^{-2}), \quad i \to +\infty$$

since:

$$\sum_{s_1=2}^{t-1} \Lambda_{s_1}^{(2)} = \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \langle \Lambda_{s_1}^{(1)}, \Lambda_{s_2}^{(1)} \rangle$$

by application of Cauchy-Schwarz and the previous result by virtue of Theorem 4.1 of Dedecker et al. (2007), we must have that:

$$\mathbb{E}[(\tilde{A}_t)^2] = O(t^2 M^2)$$

so that, under Assumption 1:

$$T^{-4}(D_\omega^{(Hete)})^{-2} \left(\sum_{t=2}^{T} \mathbb{E}[(\tilde{A}_t)^2] \right) = \Delta T^{-4} M^{-2} \left(T^3 M^2 \right) = O(1/T) = o_p(1)$$

for finite $\Delta > 0$, as $T \to \infty$.

B. The first addend, $\tilde{B}_t$, can be expressed as:

$$\begin{aligned} \sum_{s=1}^{t-1} ||X_s||^2 (W_{ts}^\star)^2 &= \sum_{s=1}^{t-1} (||X_s||^2 - d_1)(W_{ts})^2 + d_1 \sum_{s=1}^{t-1} (W_{ts})^2 \\ &= B_t^{(1)} + B_t^{(2)} \end{aligned}$$

Note that the first term, $B_t^{(1)}$, is a sum of martingale difference sequences over indexes s, under the conditional homoskedasticity of process X. Following a similar arguments to the previous ones, by the application of Proposition 5.4 of Dedecker et al. (2007):

$$\begin{aligned} \mathbb{E}\left[(B_t^{(1)})^2 \right] &\leq \left(4 \sum_{s=1}^{t-1} ||(||X_s||^2 - d_1)(W_{ts})^2||^2 \right) \\ &\leq \left(4\Delta \sum_{s=3}^{t-1} ||(W_{ts})^2||^2 \right) = O(tM^4) \end{aligned}$$

for finite $\Delta > 0$. The second inequality is because of the assumption of the conditional moments (conditional homokurtosis), while the last equality is by the previous results on the fourth moments of the time series $\{W_{ts}\}$ and finiteness of the moments of the process Z.
By the last placed restrictions, we write:

$$\mathbb{E}\left[(B_t^{(2)})^2\right] \leq O(t^2 M^4)$$

Alternative, one can notice that the order of the last term, $B_t^{(2)}$, is determined by the magnitude of the unconditional variance, $\mathbb{E}[J_t^2]$.
Thus:

$$\sum_{t=3}^{T} \mathbb{E}[(\tilde{B}_t^{(1)})^2] = O(T^2 M^4), \quad \sum_{t=3}^{T} \mathbb{E}[(\tilde{B}_t^{(2)})^2] = O(T^3 M^4)$$

Regarding the last term, under the assumptions of the time series $\{\Lambda_i^{(1)}\}$:

$$|\text{Cov}[\Lambda_1^{(1)}, \Lambda_{1+i}^{(1)}]| = O(i^{-2}), \quad i \to +\infty$$
$$|\text{Cov}[||V_1||^2, ||V_{1+i}||^2]| = O(i^{-2}), \quad i \to +\infty$$

we must have:

$$\sum_{t=3}^{T} \mathbb{E}[(\tilde{B}_t^{(1)})^2] = O(T^3 M^2), \quad \sum_{t=3}^{T} \mathbb{E}[(\tilde{B}_t^{(2)})^2] = O(T^2 M^4) = O(T^3 M^2)$$

where the last equality is because of $M/T \to 0$.
For the first (weaker) case, in conclusions, we have:

$$T^{-4}(D_\omega^{(Hete)})^{-2} \text{Var}\left[\sum_{t=3}^{T} \mathbb{E}[J_t^2 | \mathcal{I}(t-1)]\right] \leq \Delta T^{-4} M^{-2} \left(T^3 M^4\right) = O(M^2/T) = o_p(1)$$

for finite $\Delta > 0$, where the last equality is due Assumption 1 and under the asymptotics $M^2/T \to 0$.
Conversely, for the second (sharper) case, i.e., under the additional as-

sumptions:

$$|\text{Cov}[\Lambda_1^{(1)}, \Lambda_{1+i}^{(1)}]| = O(i^{-2}), \quad i \to +\infty$$
$$|\text{Cov}[||Z_1||^2, ||Z_{1+i}||^2]| = O(i^{-2}), \quad i \to +\infty$$

we have:

$$T^{-4}(D_\omega^{(Hete)})^{-2}\text{Var}\left[\sum_{t=3}^{T} \mathbb{E}[J_t^2|\mathcal{I}(t-1)]\right] \leq \Delta T^{-4} M^{-2}\left(T^3 M^3\right) = O(M/T) = o_p(1)$$

for finite $\Delta > 0$, where the last equality is due Assumption 1 and under the asymptotics $M/T \to 0$.

B.3 Proof of Theorem 2

I introduce some additional notations and objects.
Given a parameter belonging to a parameter space, $\theta \in \Theta$, I define the gradient and Hessian operators with respect to θ to be ∇_θ and ∇_θ^2, whenever have proper meaning. Define $\{\Theta_i\}_{i=1,2}$ the parameter spaces of the real parameters $\{\theta_i^0\}_{i=1,2}$, respectively.
Following eq.(7), given $\{\theta_i \in \Theta_i\}_{i=1,2}$, I defined the processes:

$$\widetilde{X}_t(\theta_1) = W_{1,t} - \mu_X(\theta_1, \mathcal{I}_X(t-1))$$
$$\widetilde{Z}_t(\theta_2) = W_{2,t} - \mu_Z(\theta_2, \mathcal{I}_Z(t-1))$$

with corresponding standardized innovations of the unobservable infinite past:

$$\widetilde{U}_t = (\Gamma_X)^{-1/2}\,\widetilde{X}_t(\theta_1), \qquad \widetilde{V}_t = (\Gamma_Z)^{-1/2}\,\widetilde{Y}_t(\theta_2)$$

Under the parametrization of eq.(7), we have:

$$U_t = \widetilde{U}_t(\theta_1^0), \qquad V_t = \widetilde{V}_t(\theta_2^0)$$

Define further the population-standardized estimated innovations as follow:

$$\breve{U}_t = (\Gamma_X)^{-1/2}\,\widehat{X}_t, \qquad \breve{V}_t = (\Gamma_Z)^{-1/2}\,\widehat{Z}_t$$

where, using the previous notation, we have:

$$\widehat{X}_t = W_{1,t} - \mu_X(\widehat{\theta}_1, \widehat{\mathcal{I}}_X(t-1))$$
$$\widehat{Z}_t = W_{2,t} - \mu_Z(\widehat{\theta}_2, \widehat{\mathcal{I}}_Z(t-1))$$

where $\widehat{\mathcal{I}}_X(t-1)$ and $\widehat{\mathcal{I}}_Z(t-1)$ are the feasible information sets, i.e., the information sets constrained to the observable finite past of the time series, $\{W_{i,t}\}_{t=1,\ldots,T}^{i=1,2}$.
Note that, generally: $\widehat{X}_t \neq \widetilde{X}_t(\widehat{\theta}_1),\ \widehat{Z}_t \neq \widetilde{Z}_t(\widehat{\theta}_2)$.
Recall that:

$$\widehat{U}_t = \left(\widehat{\Gamma}_X\right)^{-1/2}\widehat{X}_t, \qquad \widehat{V}_t = \left(\widehat{\Gamma}_Z\right)^{-1/2}\widehat{Z}_t$$

Denote the k^{th} entry-wise element of $\breve{U}_t, \widetilde{U}_t, U_t$ with $\breve{U}_{k,t}, \widetilde{U}_{k,t}, U_{k,t}$, respectively. In a similar fashion, denote $\breve{V}_{l,t-j}, \widetilde{V}_{l,t-j}, V_{l,t-j}$ the l^{th} element of, respectively, $\breve{V}_{t-j}, \widetilde{V}_{t-j}, V_{t-j}$.
We denote:

$$C_{UV}(j) = \frac{1}{T}\sum_{t=1}^{T} U_t(V_{t-j})', \;\; C_X = \frac{1}{T}\sum_{t=1}^{T} X_t(X_t)', \;\; C_Y = \frac{1}{T}\sum_{t=1}^{T} Y_t(Y_t)'$$
$$\widehat{\Gamma}_{UV}(j) = \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_t(\widehat{V}_{t-j})', \;\; \breve{C}_{UV}(j) = \frac{1}{T}\sum_{t=1}^{T} \breve{U}_t(\breve{V}_{t-j})' = (\Gamma_X)^{-1/2}\widehat{\Gamma}_{\hat{X}\hat{Z}}(j)(\Gamma_Z)^{-1/2}$$
$$\widehat{\Gamma}_{\hat{X}\hat{Z}}(j) = \frac{1}{T}\sum_{t=j+1}^{T} \widehat{X}_t\widehat{Z}'_{t-j}$$

which are respectively: the sample covariance between standardized innovations, the sample variances of the innovations, the sample covariance between feasible standardized residuals, and the population-standardized sample covariance between estimated residuals.

Additional to the assumptions listed in Theorem 2, I presume the following conditions hold true:

$$\begin{aligned}\sup_{\theta_1\in\Theta_1}\sum_{t=1}^{T}\mathbb{E}||U_t(\theta_1) - \widetilde{U}_t(\theta_1)||^2 = O(1/T)\\ \sup_{\theta_2\in\Theta_2}\sum_{t=1}^{T}\mathbb{E}||V_t(\theta_2) - \widetilde{V}_t(\theta_2)||^2 = O(1/T)\end{aligned} \tag{14}$$

and

$$\sup_{\theta_1 \in \Theta_1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}||\nabla_{\theta_1} \widetilde{U}_t(\theta_1)||^4 = O(1), \quad \sup_{\theta_2 \in \Theta_2} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}||\nabla_{\theta_2} \widetilde{V}_t(\theta_2)||^4 = O(1)$$
$$\sup_{\theta_1 \in \Theta_1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}||\nabla^2_{\theta_1} \widetilde{U}_t(\theta_1)||^4 = O(1), \quad \sup_{\theta_2 \in \Theta_2} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}||\nabla^2_{\theta_2} \widetilde{V}_t(\theta_2)||^4 = O(1) \tag{15}$$

which are essentially conditions on the uniform ℓ_2-convergence, the second-order differentiability and boundedness of the derivatives.
These conditions are relatively standard in the literature. For instance, see Assumption A3 in Hong and Lee (2005), and Assumption 2.3-2.4 in Leong and Urga (2023). In particular, with respect to the latter ones, the conditions in eq.(15) are stricter than their conditions in Assumption 2.4.

Notice that we can write the following:

$$\frac{T \cdot \widehat{\mathcal{T}}^c_\omega - \mu_{\omega,T}}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}} = \frac{T\left(\widehat{\mathcal{T}}_\omega - \widehat{\mathcal{T}}^\star_\omega\right)}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}} + \frac{T\left(\widehat{\mathcal{T}}^\star_\omega - \mathcal{T}^\star_\omega\right)}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}} - \frac{T\left(\widehat{\mathcal{C}}_\omega - \mathcal{C}^\star_\omega\right)}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}}$$
$$+ \frac{\sqrt{D^{(Hete)}_{\omega,T}}}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}} \left(\frac{T \cdot \mathcal{T}^{\star c}_\omega - \mu_{\omega,T}}{\sqrt{D^{(Hete)}_{\omega,T}}}\right)$$

with:

$$
\begin{aligned}
\widehat{\mathcal{T}}_\omega &= \sum_{j=0}^{T-1} \omega(j) ||\widehat{\Gamma}_{UV}(j)||_F^2 \\
&= \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{t=j+1}^{T} ||\widehat{U}_t||^2 ||\widehat{V}_{t-j}||^2 + \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{s,t=j+1, s\neq t}^{T} < \widehat{U}_t, \widehat{U}_s >< \widehat{V}_{t-j}, \widehat{V}_{s-j} > \\
\widehat{\mathcal{T}}_\omega^\star &= \sum_{j=0}^{T-1} \omega(j) ||\widehat{C}_{UV}(j)||_F^2 \\
&= \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{t=j+1}^{T} ||\breve{U}_t||^2 ||\breve{V}_{t-j}||^2 + \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{s,t=j+1, s\neq t}^{T} < \breve{U}_t, \breve{U}_s >< \breve{V}_{t-j}, \breve{V}_{s-j} > \\
\mathcal{T}_\omega^\star &= \sum_{j=0}^{T-1} \omega(j) ||C_{UV}(j)||_F^2 \\
&= \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{t=j+1}^{T} ||U_t||^2 ||V_{t-j}||^2 + \frac{1}{T^2} \sum_{j=0}^{T-1} \omega(j) \sum_{s,t=j+1, s\neq t}^{T} < U_t, U_s >< V_{t-j}, V_{s-j} > \\
\widehat{\mathcal{C}}_\omega &= \frac{1}{T^2} \sum_{j=0}^{T-2} \omega(j) \sum_{s,t=j+1, s\neq t, s<t-j}^{T} < \widehat{U}_t, \widehat{U}_s >< \widehat{V}_{t-j}, \widehat{V}_{s-j} > \\
\mathcal{C}_\omega^\star &= \frac{1}{T^2} \sum_{j=0}^{T-2} \omega(j) \sum_{s,t=j+1, s\neq t, s<t-j}^{T} < U_t, U_s >< V_{t-j}, V_{s-j} > \\
\widehat{D}_{\omega,T}^{(Hete)} &= \frac{d_1^2}{T^2} \sum_{j=0}^{T-2} \omega^2(j) \sum_{s,t=j+1, s\neq t, s\geq t-j}^{T} \mathbb{E}[< V_{t-j}, V_{s-j} >^2] \\
\mathcal{T}_\omega^{\star c} &= \mathcal{T}_\omega^\star - \mathcal{C}_\omega^\star
\end{aligned}
$$

where the correction term, $\mathcal{C}_\omega^\star$, is defined accordingly to eq.(5).
The proof of Theorem 2 follows from Propositions (4)-(7), and by direct application of Slutsky's theorem and Theorem 1, for which:

$$
\frac{T \cdot \mathcal{T}_\omega^{\star c} - \mu_{\omega,T}}{\sqrt{D_{\omega,T}^{(Hete)}}} \xrightarrow{d} \mathcal{N}(0,1)
$$

Proposition 4. *Under the Assumptions of Theorem 2, we have:*

$$\frac{T\left(\widehat{\mathcal{T}}_{\omega}^{\star} - \mathcal{T}_{\omega}^{\star}\right)}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}} \xrightarrow{p} 0$$

Proof. The proof is provided in Appendix B.3.1. □

Proposition 5. *Under the Assumptions of Theorem 2, we have:*

$$\frac{T\left(\widehat{\mathcal{T}}_{\omega} - \widehat{\mathcal{T}}_{\omega}^{\star}\right)}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}} \xrightarrow{p} 0$$

Proof. The proof is provided in Appendix B.3.2. □

Proposition 6. *Under the Assumptions of Theorem 2, we have:*

$$\frac{T\left(\widehat{\mathcal{C}}_{\omega} - \mathcal{C}_{\omega}^{\star}\right)}{\sqrt{\widehat{D}_{\omega,T}^{(Hete)}}} \xrightarrow{p} 0$$

Proof. The proof is provided in Appendix B.3.3. □

Proposition 7. *Under the Assumptions of Theorem 2, we have:*

$$\sqrt{\widehat{D}_{\omega,T}^{(Hete)}} \xrightarrow{p} \sqrt{D_{\omega,T}^{(Hete)}}$$

Proof. Given the consistency of the estimators, the boundedness of the moments and the conditions of eq.(14)-(15), showing that $\left(\widehat{D}_{\omega,T}^{(Hete)} - D_{\omega,T}^{(Hete)}\right) = o_p(1)$ is parallel to the first part of the proof of Proposition 5 (Appendix B.3.2). Using the continuous mapping theorem directly concludes the proof. □

B.3.1 Proof of Proposition 4

The proof is parallel to the one of Hong (2001)'s Lemma A.1-2, Bouhaddioui and Roy (2006)'s Lemma 2, and Leong and Urga (2023)'s Appendix B.

The aim is to show that:

$$T\left(\widehat{\mathcal{T}}_\omega^\star - \mathcal{T}_\omega^\star\right) = o_p(M^{1/2})$$

since $\widehat{D}_{\omega,T}^{(Hete)} = O(M)$ by Proposition 2 and Proposition 7.
The initial section of the proof studies the difference:

$$\begin{aligned}
T\left(\widehat{\mathcal{T}}_\omega^\star - \mathcal{T}_\omega^\star\right) &= T\sum_{j=1}^{T-1}\omega(j)\left(\left|\left|\text{vec}\left[\widehat{C}_{UV}(j)\right]\right|\right|^2 - ||\text{vec}\left[C_{UV}(j)\right]||^2\right) \\
&= T\sum_{j=1}^{T-1}\omega(j)\Big(||\text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))||^2 \\
&\qquad + 2\langle \text{vec}(C_{UV}(j)), \text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))\rangle\Big)
\end{aligned}$$

The proof then consists of two parts:

i) We have:

$$\begin{aligned}
&||\text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))|| \\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\frac{1}{T}\left(\sum_{t=j+1}^{T}\breve{U}_{k,t}\breve{V}_{l,t-j} - U_{k,t}V_{l,t-j}\right) \\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\Bigg(\frac{1}{T}\sum_{t=j+1}^{T}(\breve{U}_{k,t} - U_{k,t})V_{l,t-j} + U_{k,t}(\breve{V}_{l,t-j} - V_{l,t-j}) \\
&\qquad + (\breve{U}_{k,t} - U_{k,t})(\breve{V}_{l,t-j} - V_{l,t-j})\Bigg) \\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\left(F_{Tj}^{(1)} + F_{Tj}^{(2)} + F_{Tj}^{(3)}\right)
\end{aligned}$$

Thus, applying of Cauchy-Schwarz inequality:

$$||\text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))||^2 \leq \Delta\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\left((F_{Tj}^{(1)})^2 + (F_{Tj}^{(2)})^2 + (F_{Tj}^{(3)})^2\right)$$

for some finite $\Delta > 0$. I now study the three terms:

A. For the term $F_{Tj}^{(3)}$, we have:

$$
\begin{aligned}
&\sup_j (F_{Tj}^{(3)})^2 \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T}(\breve{U}_{k,t} - U_{k,t})^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{V}_{l,t-j} - V_{l,t-j})^2\right) \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t} + \widetilde{U}_{k,t} - U_{k,t})^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{V}_{l,t-j} - \widetilde{V}_{l,t-j} + \widetilde{V}_{l,t-j} - V_{l,t-j})^2\right) \\
&\leq \left\{\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t})^2\right) + \left(\frac{1}{T}\sum_{t=1}^{T}(\widetilde{U}_{k,t} - U_{k,t})^2\right)\right\} \\
&\qquad \left\{\left(\frac{1}{T}\sum_{t=1}^{T}(\breve{V}_{l,t-j} - \widetilde{V}_{l,t-j})^2\right) + \left(\frac{1}{T}\sum_{t=1}^{T}(\widetilde{V}_{l,t-j} - V_{l,t-j})^2\right)\right\} \\
&\leq \left\{F_T^{(31)} + F_T^{(32)}\right\}\left\{F_T^{(33)} + F_T^{(34)}\right\}
\end{aligned}
$$

where the first and the third inequalities are by virtue of Cauchy-Schwarz inequality.
Because of the condition eq.(14), we have:

$$
F_T^{(31)} = \frac{1}{T}\sum_{t=1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t})^2 = O_p(T^{-2}), \quad F_T^{(33)} = \frac{1}{T}\sum_{t=1}^{T}(\breve{V}_{l,t} - \widetilde{V}_{l,t})^2 = O_p(T^{-2})
$$

Now, I turn the attention on the other two terms, $F_T^{(32)}, F_T^{(34)}$.
By joint application of the mean value theorem and the Cauchy-Schwarz inequality:

$$
F_T^{(32)} \leq ||\hat{\theta}_1 - \theta_1^0||^2 \left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}||\nabla_{\theta_1}\widetilde{U}_{k,t}(\breve{\theta}_1)||^2\right\}
$$

with $\breve{\theta}_1 \in [\hat{\theta}_1, \theta_1^0]$. By the consistency of the estimator and the conditions of eq.(15): $F_T^{(32)} = O(T^{-1})$.
The same logic holds for $F_T^{(34)} = O(T^{-1})$.
In conclusions:

$$
\sup_j (F_{Tj}^{(3)})^2 = O_p(T^{-2})
$$

B. For the term $F_{Tj}^{(1)}$, we have:

$$\begin{aligned}
F_{Tj}^{(1)} &= \frac{1}{T}\sum_{t=j+1}^{T}(\breve{U}_{k,t} - U_{k,t})V_{l,t-j} \\
&= \frac{1}{T}\sum_{t=j+1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t} + \widetilde{U}_{k,t} - U_{k,t})V_{l,t-j} \\
&= \frac{1}{T}\sum_{t=j+1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t})V_{l,t-j} + \frac{1}{T}\sum_{t=j+1}^{T}(\widetilde{U}_{k,t} - U_{k,t})V_{l,t-j} \\
&= F_T^{(11)} + F_T^{(12)}
\end{aligned}$$

By Cauchy-Schwarz inequality and the boundedness of the fourth moments of the process Y:

$$(F_{Tj}^{(11)})^2 \leq \frac{1}{T^2}\sum_{t=j+1}^{T}(\breve{U}_{k,t} - \widetilde{U}_{k,t})^2(V_{l,t-j})^2 \leq F_{Tj}^{(31)}\left(\frac{1}{T}\sum_{t=1}^{T}(V_{l,t-j})^2\right) = O_p(T^{-2})$$

By the conditions of eq.(15), the term $F_T^{(12)}$ can be expressed into two terms using its Taylor expansion (up to the second order):

$$\begin{aligned}
F_T^{(11)} &= \frac{1}{T}\sum_{t=j+1}^{T}(\widetilde{U}_{k,t} - U_{k,t})V_{l,t-j} \\
&= \frac{1}{T}(\hat{\theta}_1 - \theta_1^0)'\sum_{t=1}^{T}(\nabla_{\theta_1}\widetilde{U}_{k,t}(\theta_1^0)V_{l,t-j}) \\
&\quad + \frac{1}{2T}(\hat{\theta}_1 - \theta_1^0)'\sum_{t=1}^{T}(\nabla_{\theta_1}^2\widetilde{U}_{k,t}(\breve{\theta}_1)V_{l,t-j})(\hat{\theta}_1 - \theta_1^0)
\end{aligned}$$

where $\breve{\theta}_1 \in [\hat{\theta}_1, \theta_1^0]$. Applying twice the Cauchy-Schwarz inequality:

$$\begin{aligned}
\frac{1}{T}(\hat{\theta}_1 - \theta_1^0)' \sum_{t=1}^{T} (\nabla_{\theta_1} \widetilde{U}_{k,t}(\theta_1^0) V_{l,t-j}) &= \frac{1}{T}(\hat{\theta}_1 - \theta_1^0)' \sum_{t=1}^{T} ((\nabla_{\theta_1} U_{k,t})(V_{l,t-j})) \\
&\leq \frac{1}{T^2} \mathbb{E}\left[||\hat{\theta}_1 - \theta_1^0||^2\right] \mathbb{E}\left[\left\{\sum_{t=1}^{T} ||(\nabla_{\theta_1} U_{k,t})(V_{l,t-j})||\right\}^2\right] \\
&\leq \frac{1}{T^2} \mathbb{E}\left[||\hat{\theta}_1 - \theta_1^0||^2\right] \sum_{t=1}^{T} \mathbb{E}\left[||\nabla_{\theta_1} U_{k,t}||^4\right] \mathbb{E}\left[||V_{l,t-j}||^2\right] = O_p(T^{-2})
\end{aligned}$$

where the last equality follows from the boundedness of the moments, the consistency of the estimator, and the conditions in eq.(15).
A remark is needed. To prove the last inequality, there is a trade-off between bounding the moments of the derivatives $\{\nabla_{\theta_1} U_{k,t}, \nabla^2_{\theta_1} U_{k,t}\}$ and imposing additional orthogonality conditions. In fact, instead of imposing the boundedness of the fourth moment of $\{\nabla_{\theta_1} U_{k,t}\}$, one could reach the same conclusions by imposing orthogonality between $\{\nabla_{\theta_1} U_t\}$ and $\{V_{t-j}\}$. Since the purpose of the proposed statistic is to infer about non-causality, we wish not to preclude possible causality channels by imposing extra restrictions on those.
By the same logic, the other term:

$$\begin{aligned}
&\frac{1}{2T}(\hat{\theta}_1 - \theta_1^0)' \sum_{t=1}^{T} (\nabla^2_{\theta_1} \widetilde{U}_{k,t}(\breve{\theta}_1) V_{l,t-j})(\hat{\theta}_1 - \theta_1^0) \\
&\qquad \leq \frac{1}{4T^2} \mathbb{E}\left[||\hat{\theta}_1 - \theta_1^0||^4\right] \mathbb{E}\left[\left|\left|\sum_{t=1}^{T} \nabla^2_{\theta_1} \widetilde{U}_{k,t}(\breve{\theta}_1) V_{l,t-j}\right|\right|^2\right] = O_p(T^{-2})
\end{aligned}$$

which means: $F_T^{(12)} = O_p(T^{-2})$. In conclusions:

$$(F_T^{(1)})^2 = O_p(T^{-2}), \quad (F_T^{(2)})^2 = O_p(T^{-2})$$

where the second equation is by reasoning analogue to the one above.

Finally:

$$\sum_{j=1}^{T-1} \omega(j) ||\text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))||^2$$

$$\leq \Delta \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} \left((F_{Tj}^{(1)})^2 + (F_{Tj}^{(2)})^2 + (F_{Tj}^{(3)})^2 \right) = O_p(MT^{-2}) = o_p(M^{1/2}T^{-1})$$

ii) We have:

$$\begin{aligned}
&\sum_{j=1}^{T-1} \omega(j) \langle \text{vec}(C_{UV}(j)), \text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j)) \rangle \\
&= \sum_{j=1}^{T-1} \omega(j) \text{vec}(C_{UV}(j))' (\text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j))) \\
&= \sum_{j=1}^{T-1} \omega(j) \sum_{k=1}^{d_1 d_2} \sum_{l=1}^{d_1 d_2} C_{k,UV}(j) \left(\widehat{C}_{l,UV}(j) - C_{l,UV}(j) \right) \\
&= \sum_{k=1}^{d_1 d_2} \sum_{l=1}^{d_1 d_2} \sum_{j=1}^{T-1} \omega(j) C_{k,UV}(j) \left(\widehat{C}_{l,UV}(j) - C_{l,UV}(j) \right) \\
&\leq \sum_{k=1}^{d_1 d_2} \sum_{l=1}^{d_1 d_2} \left(\sum_{j=1}^{T-1} \omega(j) (C_{k,UV}(j))^2 \right)^{1/2} \left(\sum_{j=1}^{T-1} \omega(j) \left(\widehat{C}_{l,UV}(j) - C_{l,UV}(j) \right)^2 \right)^{1/2}
\end{aligned}$$

where the last inequality is by Cauchy-Schwarz inequality.
By the results of the previous parts:

$$\left(\sum_{j=1}^{T-1} \omega(j) \left(\widehat{C}_{l,UV}(j) - C_{l,UV}(j) \right)^2 \right)^{1/2} = O_p(M^{1/2}T^{-1})$$

$$\left(\sum_{j=1}^{T-1} \omega(j) (C_{k,UV}(j))^2 \right)^{1/2} = O_p(M^{1/2}T^{-1/2})$$

where the last equality is by:

$$\sum_{j=1}^{T-1} \omega(j) (C_{k,UV}(j))^2 \leq \sum_{j=1}^{T-1} \omega(j) ||C_{k,UV}(j)||^2 = O(MT^{-1})$$

In conclusions:

$$\sum_{j=1}^{T-1} \omega(j) \langle \text{vec}(C_{UV}(j)), \text{vec}(\widehat{C}_{UV}(j)) - \text{vec}(C_{UV}(j)) \rangle = O_p(MT^{-3/2}) = o_p(M^{1/2}T^{-1})$$

As announced:

$$\widehat{\mathcal{T}}^{\star}_{\omega} - \mathcal{T}^{\star}_{\omega} = o_p(M^{1/2}T^{-1})$$

which means that the proof is concluded by Proposition 2 and Proposition 7.

B.3.2 Proof of Proposition 5

The proof is parallel to the one of Hong (2001)'s Lemma A.1, Bouhaddioui and Roy (2006)'s Proposition 3, and Leong and Urga (2023)'s Appendix B.
Similar to the proof of Proposition 4, the aim is to show that:

$$T\left(\widehat{\mathcal{T}}_{\omega} - \widehat{\mathcal{T}}^{\star}_{\omega}\right) = O_p(MT^{-1/2})$$

since $\widehat{D_{\omega,T}}^{(Hete)} = O(M)$ by Proposition 7 and Proposition 2.
By Lemma A.3, we can write:

$$\begin{aligned}
T\left(\widehat{\mathcal{T}}_{\omega} - \widehat{\mathcal{T}}^{\star}_{\omega}\right) &= T \sum_{j=1}^{T-1} \left(||\widehat{\Gamma}_{UV}(j)||_F^2 - ||\widehat{C}_{UV}(j)||_F^2 \right) \\
&= T \sum_{j=1}^{T-1} k^2\left(\frac{j}{M}\right) \left(\text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))' \left(\widehat{\Gamma}_Z^{-1} \otimes \widehat{\Gamma}_X^{-1} \right) \text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j)) \right. \\
&\qquad \left. - \text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))' \left(\Gamma_Z^{-1} \otimes \Gamma_X^{-1} \right) \text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j)) \right) \\
&= T \sum_{j=1}^{T-1} k^2\left(\frac{j}{M}\right) \text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))' \left(\widehat{\Gamma}_Z^{-1} \otimes \widehat{\Gamma}_X^{-1} - \Gamma_Z^{-1} \otimes \Gamma_X^{-1} \right) \text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))
\end{aligned}$$

Recall that $\widehat{X}_{k,t}, X_{k,t}$ are the k^{th} element of $\widehat{X}_t(\hat{\theta}_1)$, X_t, respectively.
Consider the term $(\widehat{\Gamma}_X - \Gamma_X)$:

$$||\widehat{\Gamma}_X - \Gamma_X||_F \leq \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} ||\widehat{\Gamma}_{kl,X} - \Gamma_{kl,X}||$$

By triangular equality, it follows that:

$$||\widehat{\Gamma}_{kl,X} - \Gamma_{kl,X}|| = ||\widehat{\Gamma}_{kl,X} - C_{kl,X} + C_{kl,X} - \Gamma_{kl,X}|| \leq ||\widehat{\Gamma}_{kl,X} - C_{kl,X}|| + ||C_{kl,X} - \Gamma_{kl,X}||$$

For the first term, by virtue of Cauchy-Schwarz inequality:

$$\begin{aligned}
\widehat{\Gamma}_{kl,X} - C_{kl,X} &= \frac{1}{T}\sum_{t=1}^{T} 2(\widehat{X}_{k,t} - X_{k,t})X_{k,t} + (\widehat{X}_{k,t} - X_{k,t})^2 \\
&\leq 4\left(\frac{1}{T}\sum_{t=1}^{T}(X_{k,t})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}(\widehat{X}_{k,t} - X_{k,t})^2\right)^{1/2} + \frac{1}{T}\sum_{t=1}^{T}(\widehat{X}_{k,t} - X_{k,t})^2 = O_p(T^{-1/2})
\end{aligned}$$

since we have that the term $\frac{1}{T}\sum_{t=1}^{T}(X_{k,t})^2 = O_p(1)$ because of Chebyshev inequality and boundedness of moments, while the term $\frac{1}{T}\sum_{t=1}^{T}(\widehat{X}_{k,t} - X_{k,t})^2 = O_p(T^{-1})$ as consequence of what proved in Proposition 4.
By application of Chebyshev inequality: $||C_{kl,X} - \Gamma_{kl,X}|| = O_p(T^{-1/2}),\ \forall k, l$.
This in turn implies:

$$\widehat{\Gamma}_{kl,X} - \Gamma_{kl,X} = O_p(T^{-1/2})$$

By analogue reasoning:

$$\widehat{\Gamma}_{kl,Z} - \Gamma_{kl,Z} = O_p(T^{-1/2})$$

and $\frac{1}{T}\sum_{t=1}^{T}(Z_{k,t})^2 = O_p(1)$, as well as $\Gamma_{kl,X} = O(1)$ and $\Gamma_{kl,Z} = O(1)$.
Thus, by virtue of the continuous mapping theorem, I can conclude that:

$$\widehat{\Gamma}_X^{-1} \otimes \widehat{\Gamma}_Z^{-1} - \Gamma_X^{-1} \otimes \Gamma_Z^{-1} = O_p(T^{-1/2})$$

Now, since Γ_X and Γ_Z are bounded (i.e., $\Gamma_X = O(1), \Gamma_Z = O(1)$), studying the boundedness of the term, $\widehat{\Gamma}_{\hat{X}\hat{Z}}(j)$, is equivalent to studying the boundedness of the term, $\widehat{C}_{UV}(j)$. Thus, I direct my focus to the term that follows:

$$\begin{aligned}
&\sum_{j=1}^{T-1}\omega(j)\big(\text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))'\text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))\big)\\
&\qquad\asymp\sum_{j=1}^{T-1}\omega(j)\big(\text{vec}(\widehat{C}_{UV}(j))'\text{vec}(\widehat{C}_{UV}(j))\big)\\
&\qquad=\sum_{j=1}^{T-1}\omega(j)\Big(\text{vec}(\widehat{C}_{UV}(j))'\text{vec}(\widehat{C}_{UV}(j))-\text{vec}(C_{UV}(j))'\text{vec}(C_{UV}(j))\\
&\qquad\qquad\qquad\qquad+\text{vec}(C_{UV}(j))'\text{vec}(C_{UV}(j))\Big)
\end{aligned}$$

i) We have:

$$\begin{aligned}
&\sum_{j=1}^{T-1}\omega(j)\left(\text{vec}(\widehat{C}_{UV}(j))'\text{vec}(\widehat{C}_{UV}(j))-\text{vec}(C_{UV}(j))'\text{vec}(C_{UV}(j))\right)\\
&=\sum_{j=1}^{T-1}\omega(j)\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\widehat{C}^2_{kl,UV}(j)-C^2_{kl,UV}(j)\\
&=\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\sum_{j=1}^{T-1}\omega(j)\left\{\left(\widehat{C}_{kl,UV}(j)-C_{kl,UV}(j)\right)^2+2\widehat{C}_{kl,UV}(j)\left(\widehat{C}_{kl,UV}(j)-C_{kl,UV}(j)\right)\right\}\\
&=O_p(MT^{-1})
\end{aligned}$$

where the last equality is because of Proposition 4.

ii) We have:

$$\sum_{j=1}^{T-1}\omega(j)\big(\text{vec}(C_{UV}(j))'\text{vec}(C_{UV}(j))\big)=\sum_{r=1}^{d_1d_2}\sum_{j=1}^{T-1}\omega(j)(C_{r,UV}(j))^2=O_p(MT^{-1})$$

where the last equality follows from Proposition 4, since $C_{i,UV}(j)$ as the i^{th} entry of the matrices $\text{vec}(C_{UV}(j))$.

This in turn means that:

$$\sum_{j=1}^{T-1}\omega(j)\text{vec}(\widehat{C}_{UV}(j))'\text{vec}(\widehat{C}_{UV}(j))=O_p(MT^{-1})$$

In conclusion:

$$T\left(\widehat{\mathcal{T}}_\omega - \widehat{\mathcal{T}}^\star_\omega\right) = T\sum_{j=1}^{T-1} k^2\left(\frac{j}{M}\right)\text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))'\left(\widehat{\Gamma}_Z^{-1}\otimes\widehat{\Gamma}_X^{-1} - \Gamma_Z^{-1}\otimes\Gamma_X^{-1}\right)\text{vec}(\widehat{\Gamma}_{\hat{X}\hat{Z}}(j))$$
$$= T\cdot O_p(MT^{-1})\cdot O_p(T^{-1/2}) = O_p(MT^{-1/2})$$

which concludes the proof as $\widehat{D}^{(Hete)}_{\omega,T} = O(M)$ by Proposition 7 and Proposition 2.

B.3.3 Proof of Proposition 6

The proof is parallel to Proposition 4 and Proposition 5.
As previously done, I consider the following decomposition:

$$\frac{T\left(\widehat{\mathcal{C}}_\omega - \breve{\mathcal{C}}_\omega\right)}{\sqrt{\widehat{D_\omega}^{(Hete)}}} + \frac{T\left(\breve{\mathcal{C}}_\omega - \mathcal{C}^\star_\omega\right)}{\sqrt{\widehat{D_\omega}^{(Hete)}}} \xrightarrow{p} 0$$

with:

$$\breve{\mathcal{C}}_\omega = \frac{1}{T^2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T} < \breve{U}_t, \breve{U}_s >< \breve{V}_{t-j}, \breve{V}_{s-j} >$$

By a similar argument of Proposition 5:

$$\frac{T\left(\widehat{\mathcal{C}}_\omega - \breve{\mathcal{C}}_\omega\right)}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}} \xrightarrow{p} 0$$

since the convergence is uniquely driven by the distance between $\widehat{\Gamma}_X$ and Γ_X, and between $\widehat{\Gamma}_Y$ and Γ_Y. It remains to prove:

$$T\left(\breve{\mathcal{C}}_\omega - \mathcal{C}^\star_\omega\right) = o_p(M^{1/2})$$

We have:

$$
\begin{aligned}
&T\left(\breve{\mathcal{C}}_\omega - \mathcal{C}^\star_\omega\right)\\
&= \frac{1}{T}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T}\left(\breve{U}_t'\breve{U}_s\breve{V}_{t-j}'\breve{V}_{s-j} - (U_t)'U_s(V_{t-j})'V_{s-j}\right)\\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\frac{1}{T}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T}\left(\breve{U}_{k,t}\breve{U}_{k,s}\breve{V}_{l,t-j}\breve{V}_{l,s-j} - U_{k,t}U_{k,s}V_{l,t-j}V_{l,s-j}\right)\\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\frac{1}{T}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T}\Bigg(\left[\breve{U}_{k,t}\breve{U}_{k,s} - U_{k,t}U_{k,s}\right]V_{l,t-j}V_{l,s-j}\\
&\quad + U_{k,t}U_{k,s}\left[\breve{V}_{l,t-j}\breve{V}_{l,s-j} - V_{l,t-j}V_{l,s-j}\right] + \left[\breve{U}_{k,t}\breve{U}_{k,s} - U_{k,t}U_{k,s}\right]\left[\breve{V}_{l,t-j}\breve{V}_{l,s-j} - V_{l,t-j}V_{l,s-j}\right]\Bigg)\\
&= \sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\sum_{j=0}^{T-2}\omega(j)\sum_{s,t=j+1,s\neq t,s<t-j}^{T}\Bigg(\\
&\quad T^{-1}\Big[(\breve{U}_{k,t} - U_{k,t})U_{k,s} + (\breve{U}_{k,s} - U_{k,s})U_{k,t} + (\breve{U}_{k,t} - U_{k,t})(\breve{U}_{k,s} - U_{k,s})\Big]V_{l,t-j}V_{l,s-j}\\
&\quad + T^{-1}U_{k,t}U_{k,s}\Big[(\breve{V}_{l,t-j} - V_{l,t-j})V_{l,s-j} + (\breve{V}_{l,s-j} - V_{l,s-j})V_{l,t-j} + (\breve{V}_{l,t-j} - V_{l,t-j})(\breve{V}_{l,s-j} - V_{l,s-j})\Big]\\
&\quad + T^{-1}\left[\breve{U}_{k,t}\breve{U}_{k,s} - U_{k,t}U_{k,s}\right]\left[\breve{V}_{l,t-j}\breve{V}_{l,s-j} - V_{l,t-j}V_{l,s-j}\right]\Bigg)
\end{aligned}
$$

Proposition 4 shows that all terms inside the brackets are at most $O_p(T^{-2})$, which concludes the proof as:

$$\breve{\mathcal{C}}_\omega - \mathcal{C}^\star_\omega = O_p(MT^{-2}) = o_p(M^{1/2}T^{-1})$$

B.4 Proof of Theorem 3

For the definition of the objects and the regularity conditions, please refer to the first part of the proof of Theorem 2 in Appendix B.3. The proof is parallel to the one of Hong (2001)'s Theorem 2, Bouhaddioui and Roy (2006)'s Theorem 2, and Leong and Urga (2023)'s Appendix C.

As done for Theorem 2, the test statistic is decomposed as follows:

$$\left(\frac{M^{1/2}}{T}\right)\frac{T\cdot\widehat{\mathcal{T}}^{c}_{\omega}-\mu_{\omega,T}}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}}=\frac{\left(\widehat{\mathcal{T}}_{\omega}-\widehat{\mathcal{T}}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}+\frac{\left(\widehat{\mathcal{T}}^{\star}_{\omega}-\mathcal{T}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}-\frac{\left(\widehat{\mathcal{C}}_{\omega}-\mathcal{C}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}$$
$$+\frac{\sqrt{M^{-1}D^{(Hete)}_{\omega,T}}}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}\left(\frac{\mathcal{T}^{\star c}_{\omega}-\mu_{\omega,T}}{\sqrt{M^{-1}D^{(Hete)}_{\omega,T}}}\right)$$

By Assumption 1, as $\mu_{\omega,T}$ and $D^{(Hete)}_{\omega,T}$ are of order $O(M)$, together with Proposition 7, I can conclude:

$$\left(\frac{M^{1/2}}{T}\right)\frac{T\cdot\widehat{\mathcal{T}}^{c}_{\omega}-\mu_{\omega,T}}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}}\xrightarrow{p}\frac{\left(\widehat{\mathcal{T}}_{\omega}-\widehat{\mathcal{T}}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}+\frac{\left(\widehat{\mathcal{T}}^{\star}_{\omega}-\mathcal{T}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}-\frac{\left(\widehat{\mathcal{C}}_{\omega}-\mathcal{C}^{\star}_{\omega}\right)}{\sqrt{M^{-1}\widehat{D}^{(Hete)}_{\omega,T}}}$$
$$+\left(\frac{\mathcal{T}^{\star c}_{\omega}}{\sqrt{\Delta\int_{0}^{\infty}k^{4}(z)dz}}\right)+o_p(1)$$

for some finite $\Delta>0$.

The proof of Theorem 3 follows from: i) Corollary 2, which show that the first three terms are of order $o_p(1)$, ii) Corollary 1 together with Slutsky's theorem, and iii) Lemma B.4.

In conclusions, we have the following:

$$\frac{M^{1/2}}{T}\frac{T\widehat{\mathcal{T}}^{c}_{\omega}-\mu_{\omega,T}}{\sqrt{\widehat{D}^{(Hete)}_{\omega,T}}}\xrightarrow{p}\frac{\mathcal{T}^{\star c}_{\omega}}{\sqrt{\Delta\int_{0}^{\infty}k^{4}(z)dz}}+o_p(1)$$

Corollary 1. *Suppose Assumptions of Theorem 3 hold. It follows that:*

$$M^{-1}\widehat{D}^{(Hete)}_{\omega,T}\xrightarrow{p}\Delta\int_{0}^{\infty}k^{4}(z)dz$$

for some finite $\Delta > 0$.

Proof. The proof follows directly from Proposition 7 and by having:

$$M^{-1} D_{\omega,T}^{(Hete)} \xrightarrow[M\to\infty]{} \Delta \int_0^\infty k^4(z)dz$$

for some finite $\Delta > 0$, as discussed in Appendix B.1. □

Corollary 2. *Suppose the assumptions of Theorem 3 hold. It follows that:*

$$\widehat{\mathcal{T}}_\omega - \widehat{\mathcal{T}}_\omega^\star = o_p(1)$$
$$\widehat{\mathcal{T}}_\omega^\star - \mathcal{T}_\omega^\star = o_p(1)$$
$$\widehat{\mathcal{C}}_\omega - \mathcal{C}_\omega^\star = o_p(1)$$

Proof. The proof follows from direct application of Proposition 4-5-6, since:

$$\widehat{\mathcal{T}}_\omega - \widehat{\mathcal{T}}_\omega^\star = o_p(M^{1/2}T^{-1}), \quad \widehat{\mathcal{T}}_\omega^\star - \mathcal{T}_\omega^\star = o_p(M^{1/2}T^{-1}), \quad \widehat{\mathcal{C}}_\omega - \mathcal{C}_\omega^\star = o_p(M^{1/2}T^{-1})$$

and concludes by Assumption 1, as $M^{1/2}T^{-1} \to 0$. □

Lemma B.4. *Suppose Assumptions of Theorem 3 hold. We have:*

$$\mathcal{T}_\omega^{\star c} \xrightarrow{p} \sum_{j=1}^{\infty} \left|\left| vec\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 + o_p(1)$$

Proof. The proof is focused on establishing the consistency of the statistic, so that the difference between the two quantities is of order $o_p(1)$. We have:

$$\mathcal{T}_\omega^{\star c} - \sum_{j=1}^{\infty} ||\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}||^2 = \mathcal{T}_\omega^\star - \mathcal{C}_\omega^\star - \sum_{j=1}^{\infty} ||\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}||^2$$
$$= \sum_{j=1}^{T-1} \omega(j) \left(||\text{vec}\left[C_{UV}(j)\right]||^2 - \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 \right) - \mathcal{C}_\omega^\star$$
$$+ \sum_{j=1}^{T-1} (\omega(j) - 1) \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 + \sum_{T}^{\infty} \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2$$

For the first term, we have:

$$\begin{aligned}
&\sum_{j=1}^{T-1} \omega(j) \left(||\text{vec}\left[C_{UV}(j)\right]||^2 - \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 \right) - \mathcal{C}_\omega^\star \\
&= \left[\sum_{j=1}^{T-1} \omega(j) \left|\left|\text{vec}\left[C_{UV}(j)\right] - \text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 - \mathcal{C}_\omega^\star\right] + 2\sum_{j=1}^{T-1}\omega(j)\lambda^\star(j) \\
&= \left[\sum_{k,l=1}^{d_1,d_2}\sum_{j=1}^{T-1} \omega(j)\text{Var}[C_{kl,UV}(j)] - \mathcal{C}_\omega^\star\right] + 2\sum_{j=1}^{T-1}\omega(j)\lambda^\star(j)
\end{aligned}$$

with $\lambda^\star(j) = \left\langle \text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right], \text{vec}\left[C_{UV}(j) - \Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right\rangle$.
Denote $\rho_{kl}(j)$, the $(k,l)-$ entry of the matrix $\left(\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right)$.
If $\{X_t, Z_t\}$ is a fourth-order stationary process so $\{U_t, V_t\}$ is fourth-order stationary as well, by Hannan (1970) pg.209-210 or, equivalently, by Priestley (1981) pg. 325–26:

$$\begin{aligned}
\text{Var}[C_{kl,UV}(j)] = T^{-1}\sum_{j=1}^{T-1}\omega(j)\sum_{i=-T+1}^{T-1}\left(1-\frac{i}{T}\right)\rho_{kl}(i+j)\rho_{kl}(i-j) \\
+ T^{-1}\sum_{j=1}^{T-1}\omega(j)\sum_{i=-T+1}^{T-1}\left(1-\frac{|i|}{T}\right)\kappa_{klkl,XZ}(0,j,i,j+i)
\end{aligned}$$

Following a similar argument in Hong (2001)'s Lemma A.6, Bouhaddioui and Roy (2006)'s Lemma A.7, or Leong and Urga (2023)'s Lemma C.3,
as $\sum_{m,r=1}^{d_1,d_2}\sum_{j,k,l=-\infty}^{\infty}\kappa_{mrmr,XZ}(0,j,k,l) < \infty$, we have:

$$\sum_{k,l=1}^{d_1,d_2}\sum_{j=1}^{T-1}\omega(j)\text{Var}[C_{kl,UV}(j)] - \mathcal{C}_\omega^\star = O_p(1/T + M/T) = o_p(1)$$

where the last equality is because of the assumption on the asymptotic rates ($\frac{M}{T} \to 0$, as $T, M \to \infty$) and by realizing that $\mathcal{C}_\omega^\star$ represents the fourth-order cumulants of:
$\{U_{m,t}, V_{r,t-j}, U_{m,t-k}, V_{r,t-l}\}$.
By the dominated convergence theorem, the boundedness condition on the covariances, $\sum_{j=-\infty}^{\infty}||\Gamma_{XZ}(j)||^2 < \infty$, and by the assumptions on the asymptotic rates, I

can conclude that:

$$\sum_{j=1}^{T-1} (\omega(j) - 1) \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 = o_p(1)$$
$$\sum_{T}^{\infty} \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 = o_p(1)$$

This in turn implies as well:

$$\sum_{j=1}^{T-1} \omega(j) \left\langle \text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right], \text{vec}\left[C_{UV}(j) - \Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right\rangle = o_p(1)$$

Hence:

$$\mathcal{T}_\omega^{\star c} \xrightarrow{p} \sum_{j=1}^{\infty} \left|\left|\text{vec}\left[\Gamma_X^{-1/2}\Gamma_{XZ}(j)\Gamma_Z^{-1/2}\right]\right|\right|^2 + o_p(1)$$

which concludes the proof. □

C Appendix

C.1 Monte Carlo simulations: under the null

C.1.1 Small sample: $T = 300$

Table 4: Rejection frequencies for DGP2A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP2A; sample size, $T = 300$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12 \quad \approx 2(10T)^{1/5} \quad \approx 2\ln T$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.01	0.02	0.02	0.02	0.02	0.06	0.04	0.04	0.04	0.05	0.05
$\alpha = 0.3$	0.03	0.02	0.03	0.02	0.02	0.02	0.05	0.05	0.06	0.04	0.05	0.03
$\alpha = 0.4$	0.03	0.03	0.04	0.02	0.04	0.03	0.06	0.06	0.07	0.04	0.08	0.06
$\alpha = 0.5$	0.03	0.03	0.03	0.04	0.03	0.04	0.05	0.06	0.07	0.08	0.07	0.06
$\alpha = 0.6$	0.03	0.05	0.05	0.04	0.07	0.04	0.06	0.08	0.09	0.06	0.09	0.07
$\alpha = 0.7$	0.07	0.06	0.07	0.04	0.07	0.05	0.10	0.09	0.09	0.07	0.10	0.08

$M = 30 \quad \approx 5(10T)^{1/5} \quad \approx \sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.02	0.02	0.02	0.01	0.03	0.02	0.06	0.04	0.04	0.04	0.06	0.05
$\alpha = 0.3$	0.02	0.02	0.03	0.02	0.02	0.02	0.06	0.07	0.07	0.05	0.05	0.05
$\alpha = 0.4$	0.04	0.03	0.03	0.03	0.03	0.03	0.07	0.07	0.08	0.05	0.07	0.07
$\alpha = 0.5$	0.04	0.04	0.04	0.05	0.04	0.04	0.07	0.07	0.07	0.09	0.09	0.07
$\alpha = 0.6$	0.03	0.05	0.05	0.04	0.07	0.04	0.08	0.10	0.10	0.07	0.12	0.08
$\alpha = 0.7$	0.09	0.08	0.09	0.06	0.09	0.06	0.14	0.13	0.14	0.12	0.13	0.10

$M = 100 \quad \approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.02	0.02	0.02	0.03	0.05	0.06	0.05	0.06	0.07	0.09	0.09
$\alpha = 0.3$	0.04	0.04	0.05	0.04	0.04	0.03	0.06	0.08	0.10	0.08	0.08	0.09
$\alpha = 0.4$	0.06	0.04	0.05	0.04	0.04	0.06	0.11	0.08	0.09	0.08	0.10	0.11
$\alpha = 0.5$	0.08	0.07	0.08	0.07	0.09	0.08	0.13	0.10	0.12	0.14	0.13	0.11
$\alpha = 0.6$	0.09	0.10	0.10	0.12	0.12	0.12	0.12	0.16	0.15	0.17	0.16	0.17
$\alpha = 0.7$	0.18	0.16	0.18	0.16	0.15	0.17	0.23	0.21	0.22	0.22	0.22	0.21

Table 5: Rejection frequencies for DGP3A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP3A; sample size, $T = 300$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12 \quad \approx 2(10T)^{1/5} \quad \approx 2\ln T$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.03	0.03	0.03	0.03	0.04	0.06	0.05	0.04	0.04	0.06	0.05
$\alpha = 0.3$	0.03	0.03	0.04	0.04	0.03	0.03	0.06	0.05	0.06	0.05	0.05	0.04
$\alpha = 0.4$	0.03	0.03	0.05	0.03	0.04	0.05	0.07	0.05	0.08	0.05	0.06	0.07
$\alpha = 0.5$	0.03	0.03	0.04	0.05	0.05	0.04	0.05	0.06	0.06	0.07	0.07	0.06
$\alpha = 0.6$	0.04	0.06	0.05	0.04	0.06	0.04	0.06	0.09	0.07	0.07	0.09	0.07
$\alpha = 0.7$	0.08	0.07	0.05	0.06	0.08	0.07	0.11	0.1	0.08	0.08	0.1	0.1

$M = 30 \quad \approx 5(10T)^{1/5} \quad \approx \sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.02	0.02	0.04	0.04	0.04	0.06	0.04	0.04	0.05	0.07	0.05
$\alpha = 0.3$	0.02	0.03	0.04	0.04	0.05	0.05	0.06	0.05	0.06	0.05	0.06	0.06
$\alpha = 0.4$	0.04	0.04	0.04	0.06	0.04	0.04	0.08	0.06	0.07	0.07	0.07	0.07
$\alpha = 0.5$	0.04	0.03	0.04	0.06	0.06	0.06	0.08	0.06	0.07	0.08	0.1	0.06
$\alpha = 0.6$	0.03	0.07	0.05	0.06	0.08	0.08	0.08	0.11	0.09	0.09	0.12	0.1
$\alpha = 0.7$	0.09	0.08	0.08	0.09	0.12	0.1	0.15	0.13	0.11	0.13	0.14	0.13

$M = 100 \quad \approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.03	0.04	0.06	0.06	0.08	0.07	0.07	0.05	0.07	0.08	0.08
$\alpha = 0.3$	0.04	0.05	0.06	0.06	0.08	0.09	0.07	0.07	0.09	0.08	0.09	0.09
$\alpha = 0.4$	0.06	0.05	0.07	0.09	0.09	0.09	0.11	0.09	0.09	0.11	0.09	0.1
$\alpha = 0.5$	0.09	0.08	0.09	0.09	0.12	0.11	0.13	0.12	0.12	0.13	0.14	0.12
$\alpha = 0.6$	0.09	0.12	0.13	0.16	0.15	0.14	0.12	0.17	0.16	0.18	0.17	0.17
$\alpha = 0.7$	0.18	0.16	0.17	0.19	0.21	0.23	0.23	0.21	0.2	0.23	0.24	0.23

Table 6: Rejection frequencies for DGP4A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP4A; sample size, $T = 300$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12$ $\approx 2(10T)^{1/5}$ $\approx 2\ln T$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.02	0.02	0.02	0.03	0.02	0.06	0.05	0.04	0.05	0.06	0.04
$\alpha = 0.3$	0.03	0.03	0.03	0.04	0.04	0.03	0.06	0.06	0.07	0.07	0.06	0.06
$\alpha = 0.4$	0.03	0.03	0.04	0.03	0.04	0.04	0.07	0.06	0.07	0.06	0.07	0.08
$\alpha = 0.5$	0.03	0.04	0.03	0.04	0.03	0.04	0.05	0.05	0.06	0.08	0.07	0.07
$\alpha = 0.6$	0.04	0.06	0.06	0.04	0.05	0.06	0.06	0.09	0.09	0.06	0.09	0.09
$\alpha = 0.7$	0.08	0.08	0.07	0.05	0.06	0.07	0.11	0.1	0.09	0.09	0.09	0.12

$M = 30$ $\approx 5(10T)^{1/5}$ $\approx \sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.01	0.03	0.01	0.03	0.03	0.06	0.04	0.06	0.04	0.06	0.05
$\alpha = 0.3$	0.02	0.02	0.03	0.03	0.04	0.03	0.06	0.06	0.07	0.06	0.07	0.05
$\alpha = 0.4$	0.04	0.04	0.04	0.04	0.04	0.04	0.08	0.07	0.07	0.07	0.08	0.08
$\alpha = 0.5$	0.04	0.04	0.05	0.04	0.04	0.04	0.08	0.07	0.08	0.07	0.08	0.08
$\alpha = 0.6$	0.03	0.06	0.05	0.05	0.06	0.06	0.08	0.1	0.09	0.1	0.11	0.11
$\alpha = 0.7$	0.09	0.08	0.08	0.08	0.08	0.1	0.15	0.13	0.13	0.12	0.14	0.15

$M = 100$ $\approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.02	0.01	0.02	0.03	0.03	0.07	0.05	0.06	0.05	0.07	0.06
$\alpha = 0.3$	0.04	0.04	0.04	0.04	0.04	0.04	0.07	0.08	0.09	0.07	0.08	0.08
$\alpha = 0.4$	0.06	0.05	0.05	0.05	0.07	0.05	0.11	0.08	0.1	0.09	0.12	0.09
$\alpha = 0.5$	0.09	0.07	0.09	0.06	0.08	0.08	0.13	0.11	0.14	0.11	0.11	0.12
$\alpha = 0.6$	0.09	0.1	0.11	0.1	0.11	0.12	0.12	0.16	0.16	0.17	0.15	0.18
$\alpha = 0.7$	0.18	0.18	0.18	0.17	0.17	0.2	0.23	0.22	0.21	0.21	0.23	0.25

C.1.2 Small sample: $T = 100$

Table 7: Rejection frequencies for DGP2A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP2A; sample size, $T = 100$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12$ $\approx 2(10T)^{1/5}$ $\approx 2\ln T$

	Corrected						Standard					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.02	0.03	0.04	0.03	0.02	0.03	0.05	0.05	0.07	0.05	0.05	0.05
$\alpha=0.3$	0.03	0.02	0.03	0.03	0.03	0.04	0.05	0.05	0.05	0.04	0.05	0.07
$\alpha=0.4$	0.04	0.05	0.03	0.03	0.04	0.02	0.06	0.07	0.06	0.05	0.06	0.05
$\alpha=0.5$	0.04	0.05	0.04	0.05	0.05	0.03	0.05	0.05	0.06	0.07	0.08	0.06
$\alpha=0.6$	0.05	0.04	0.06	0.05	0.06	0.05	0.06	0.07	0.09	0.08	0.08	0.07
$\alpha=0.7$	0.07	0.07	0.05	0.06	0.05	0.07	0.1	0.1	0.08	0.1	0.07	0.09

$M = 30$ $\approx 5(10T)^{1/5}$ $\approx \sqrt{T}$

	Corrected						Standard					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.03	0.02	0.03	0.03	0.03	0.03	0.05	0.04	0.05	0.05	0.05	0.06
$\alpha=0.3$	0.04	0.03	0.04	0.04	0.03	0.04	0.05	0.06	0.06	0.05	0.05	0.06
$\alpha=0.4$	0.04	0.05	0.05	0.04	0.03	0.05	0.08	0.07	0.06	0.05	0.06	0.06
$\alpha=0.5$	0.05	0.06	0.05	0.05	0.06	0.05	0.07	0.08	0.07	0.07	0.09	0.08
$\alpha=0.6$	0.08	0.05	0.07	0.07	0.08	0.06	0.1	0.08	0.1	0.1	0.11	0.09
$\alpha=0.7$	0.1	0.1	0.11	0.12	0.08	0.1	0.13	0.13	0.14	0.14	0.1	0.12

$M = 100$ $\approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=1$	$\beta=2$
$\alpha=0.2$	0.07	0.06	0.07	0.05	0.06	0.07	0.07	0.07	0.09	0.06	0.08	0.1
$\alpha=0.3$	0.09	0.09	0.08	0.07	0.07	0.09	0.09	0.1	0.09	0.09	0.1	0.12
$\alpha=0.4$	0.11	0.1	0.1	0.1	0.08	0.1	0.12	0.11	0.1	0.12	0.11	0.12
$\alpha=0.5$	0.13	0.14	0.14	0.11	0.13	0.13	0.14	0.15	0.13	0.14	0.16	0.16
$\alpha=0.6$	0.18	0.17	0.18	0.17	0.16	0.15	0.17	0.15	0.17	0.17	0.17	0.16
$\alpha=0.7$	0.23	0.23	0.23	0.24	0.23	0.25	0.22	0.23	0.23	0.25	0.22	0.26

Table 8: Rejection frequencies for DGP3A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}_\omega^c$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP3A; sample size, $T = 100$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12 \quad \approx 2(10T)^{1/5} \quad \approx 2\ln T$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.02	0.03	0.03	0.04	0.02	0.04	0.05	0.04	0.06	0.05	0.04	0.03
$\alpha = 0.3$	0.03	0.03	0.03	0.04	0.03	0.04	0.05	0.04	0.05	0.05	0.04	0.06
$\alpha = 0.4$	0.04	0.04	0.03	0.04	0.02	0.03	0.06	0.05	0.05	0.06	0.03	0.04
$\alpha = 0.5$	0.04	0.04	0.05	0.04	0.03	0.05	0.05	0.06	0.06	0.06	0.06	0.06
$\alpha = 0.6$	0.05	0.05	0.05	0.05	0.04	0.04	0.06	0.07	0.07	0.08	0.06	0.05
$\alpha = 0.7$	0.07	0.07	0.06	0.05	0.06	0.05	0.1	0.1	0.09	0.07	0.08	0.07

$M = 30 \quad \approx 5(10T)^{1/5} \quad \approx \sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.03	0.03	0.04	0.05	0.04	0.05	0.04	0.05	0.05	0.05	0.03
$\alpha = 0.3$	0.04	0.03	0.04	0.05	0.04	0.06	0.05	0.05	0.05	0.06	0.04	0.05
$\alpha = 0.4$	0.04	0.03	0.05	0.05	0.05	0.07	0.08	0.06	0.06	0.06	0.04	0.07
$\alpha = 0.5$	0.05	0.05	0.04	0.06	0.07	0.09	0.07	0.07	0.07	0.06	0.06	0.08
$\alpha = 0.6$	0.08	0.07	0.07	0.08	0.08	0.09	0.1	0.07	0.1	0.09	0.07	0.09
$\alpha = 0.7$	0.1	0.11	0.09	0.13	0.12	0.13	0.13	0.13	0.11	0.14	0.12	0.11

$M = 100 \quad \approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.07	0.05	0.07	0.07	0.1	0.11	0.07	0.06	0.07	0.07	0.09	0.1
$\alpha = 0.3$	0.09	0.08	0.09	0.11	0.12	0.13	0.09	0.09	0.09	0.10	0.12	0.11
$\alpha = 0.4$	0.11	0.09	0.11	0.13	0.14	0.16	0.12	0.10	0.11	0.11	0.12	0.16
$\alpha = 0.5$	0.13	0.12	0.12	0.16	0.18	0.22	0.14	0.14	0.12	0.14	0.16	0.19
$\alpha = 0.6$	0.18	0.16	0.19	0.21	0.2	0.26	0.17	0.16	0.17	0.19	0.17	0.22
$\alpha = 0.7$	0.23	0.22	0.25	0.29	0.3	0.3	0.22	0.22	0.22	0.28	0.27	0.26

Table 9: Rejection frequencies for DGP4A: This table presents the rejection frequencies of two testing procedure, corrected ($\mathcal{T}^c_\omega$) and standard ($\mathcal{T}_\omega$), when the time series are generated by DGP4A; sample size, $T = 100$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 30, 100\}$; nominal significance level is 5%.

$M = 12 \quad \approx 2(10T)^{1/5} \quad \approx 2\ln T$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.02	0.03	0.03	0.03	0.03	0.03	0.05	0.05	0.07	0.06	0.07	0.06
$\alpha = 0.3$	0.03	0.03	0.03	0.02	0.03	0.03	0.05	0.05	0.05	0.05	0.05	0.05
$\alpha = 0.4$	0.04	0.04	0.04	0.04	0.03	0.04	0.06	0.06	0.05	0.07	0.05	0.06
$\alpha = 0.5$	0.04	0.05	0.04	0.05	0.04	0.04	0.05	0.07	0.06	0.08	0.07	0.07
$\alpha = 0.6$	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.07	0.08	0.08	0.08	0.07
$\alpha = 0.7$	0.07	0.08	0.07	0.07	0.06	0.07	0.10	0.10	0.09	0.09	0.08	0.09

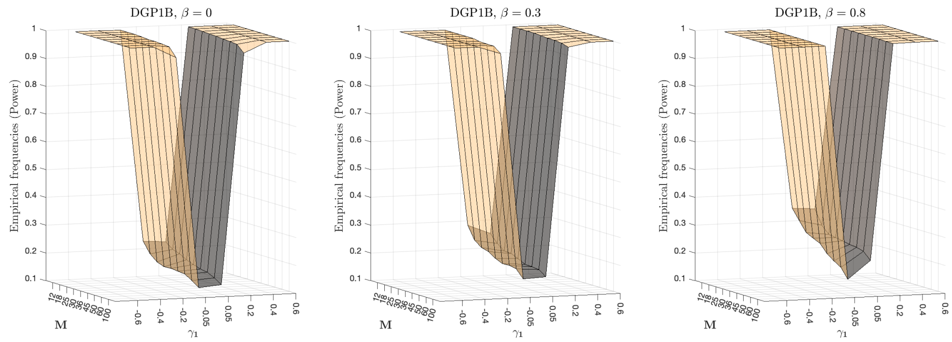
$M = 30 \quad \approx 5(10T)^{1/5} \quad \approx \sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.03	0.02	0.03	0.03	0.04	0.03	0.05	0.04	0.05	0.04	0.06	0.05
$\alpha = 0.3$	0.04	0.03	0.04	0.02	0.04	0.03	0.05	0.06	0.06	0.05	0.05	0.06
$\alpha = 0.4$	0.04	0.05	0.05	0.04	0.04	0.04	0.08	0.08	0.08	0.06	0.05	0.06
$\alpha = 0.5$	0.05	0.06	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.09	0.09
$\alpha = 0.6$	0.08	0.05	0.06	0.07	0.08	0.07	0.1	0.08	0.1	0.1	0.1	0.1
$\alpha = 0.7$	0.1	0.11	0.11	0.12	0.1	0.11	0.13	0.12	0.13	0.13	0.13	0.13

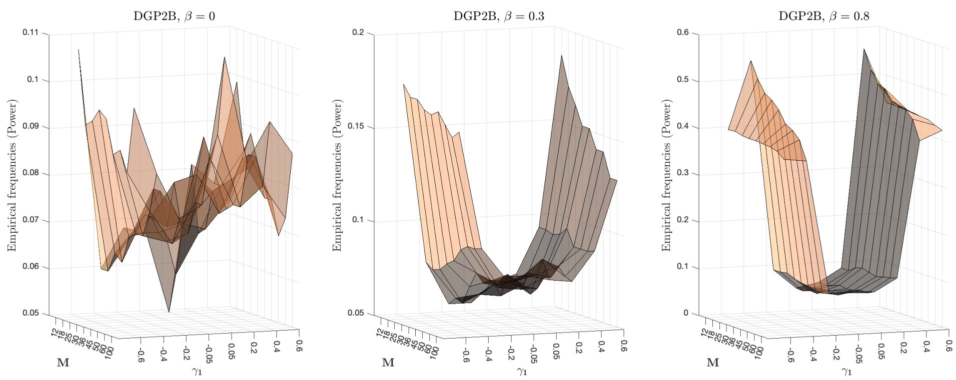
$M = 100 \quad \approx 4\sqrt{T}$

	Corrected						Standard					
	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$	$\beta = 0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 1$	$\beta = 2$
$\alpha = 0.2$	0.07	0.07	0.07	0.06	0.09	0.07	0.07	0.07	0.08	0.07	0.1	0.08
$\alpha = 0.3$	0.09	0.09	0.09	0.07	0.07	0.1	0.09	0.11	0.09	0.08	0.07	0.08
$\alpha = 0.4$	0.11	0.09	0.1	0.11	0.08	0.08	0.12	0.1	0.09	0.11	0.1	0.09
$\alpha = 0.5$	0.13	0.13	0.13	0.13	0.15	0.13	0.14	0.13	0.13	0.13	0.15	0.13
$\alpha = 0.6$	0.18	0.16	0.18	0.17	0.19	0.19	0.17	0.16	0.18	0.17	0.18	0.17
$\alpha = 0.7$	0.23	0.23	0.25	0.26	0.25	0.28	0.22	0.23	0.23	0.25	0.23	0.26

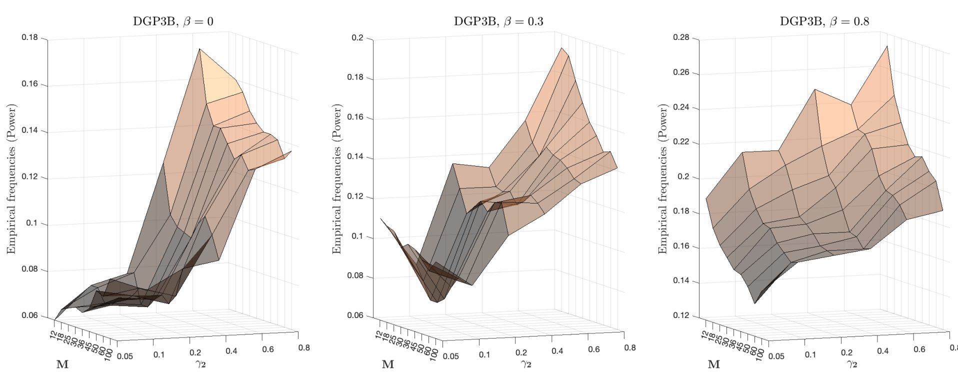
C.2 Monte Carlo simulations: under the alternatives



(a) DGP 1B, with $\beta = \{0, 0.3, 0.8\}$



(b) DGP 2B, with $\beta = \{0, 0.3, 0.8\}$

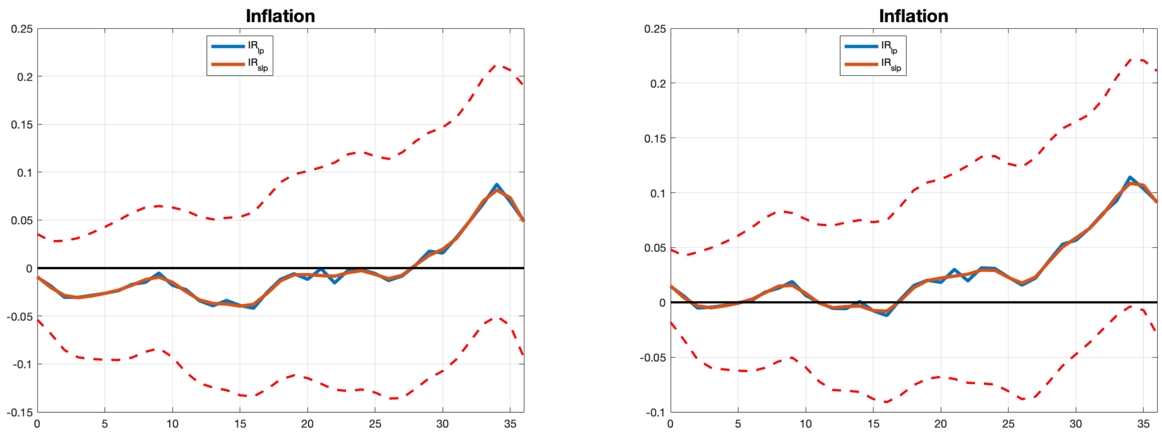


(c) DGP 3B, with $\beta = \{0, 0.3, 0.8\}$

Figure 8: Power curves of the original test: These figures present the rejection rates of the testing procedure associated to the original test statistic ($\mathcal{T}_\omega$), under the alternatives (empirical power); sample size, $T = 700$; 700 iterations; the weighting function is the Bartlett kernel; the smoothing parameter range is: $M = \{12, 18, 25, 30, 36, 45, 50, 60, 100\}$; nominal significance level is 5%.

D Appendix

D.1 Linear responses of inflation to uncertainty shocks

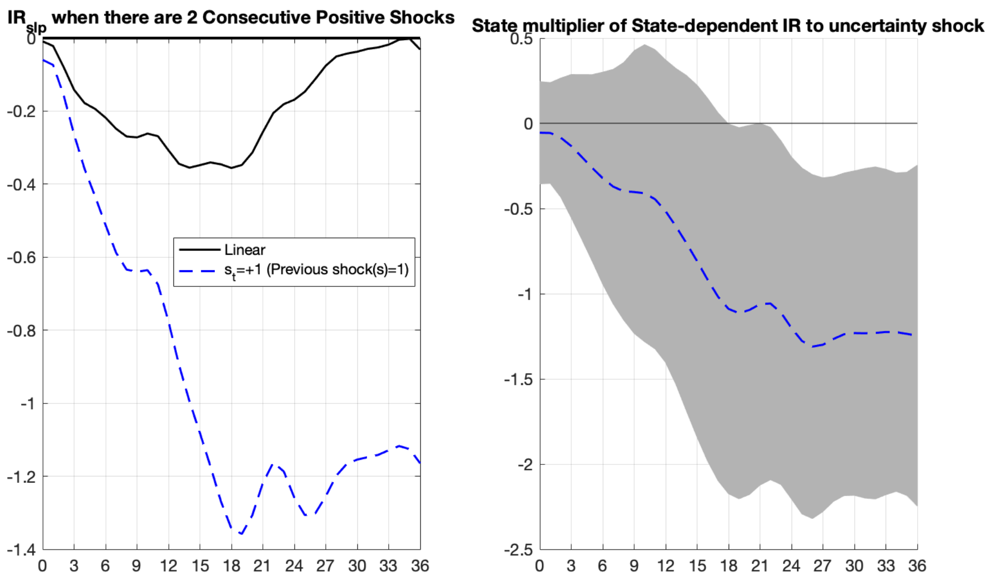


(a) Linear responses in Figure B.2.(e)-(f) in Diercks et al. (2024)'s Appendix B.1

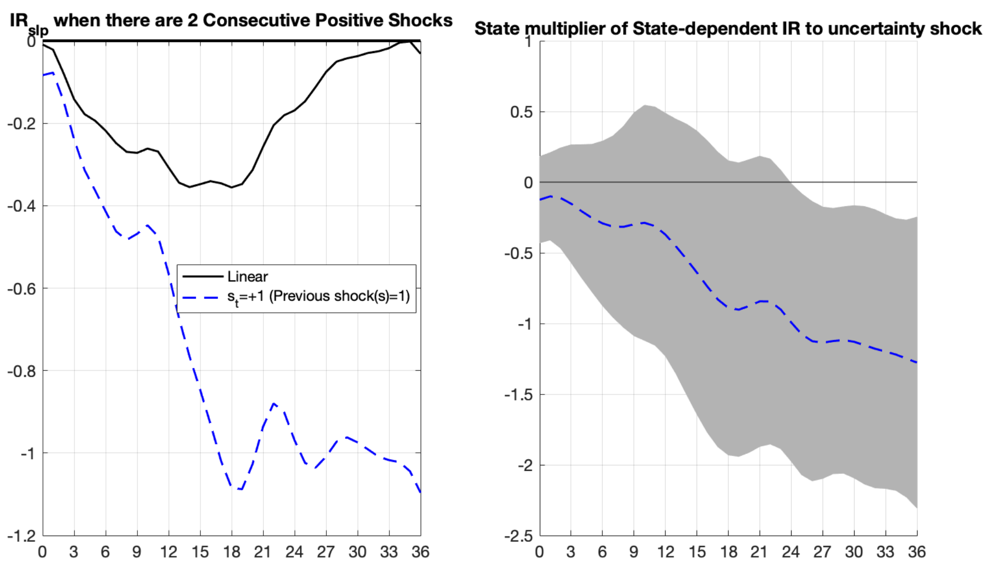
(b) Inclusion of two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of controls

Figure 9: Linear response of price level to EPU uncertainty shocks: The panels show the empirical unconditional impulse responses, i.e. $\{\beta_{0,h}\}_{h=1,\ldots,H}$. On the y-axes, the level of impulse responses; on the x-axes, the horizons, h; solid blue line represents the standard LPs and red solid line represents the Smoothed LPs; dashed red line stands for the 90% confidence intervals. LEFT PANEL: the original results in Diercks et al. (2024). RIGHT PANEL: the results once controlling for the McCracken and Ng (2016)'s macroeconomic factors.

D.2 Responses of the other variables to EPU shocks

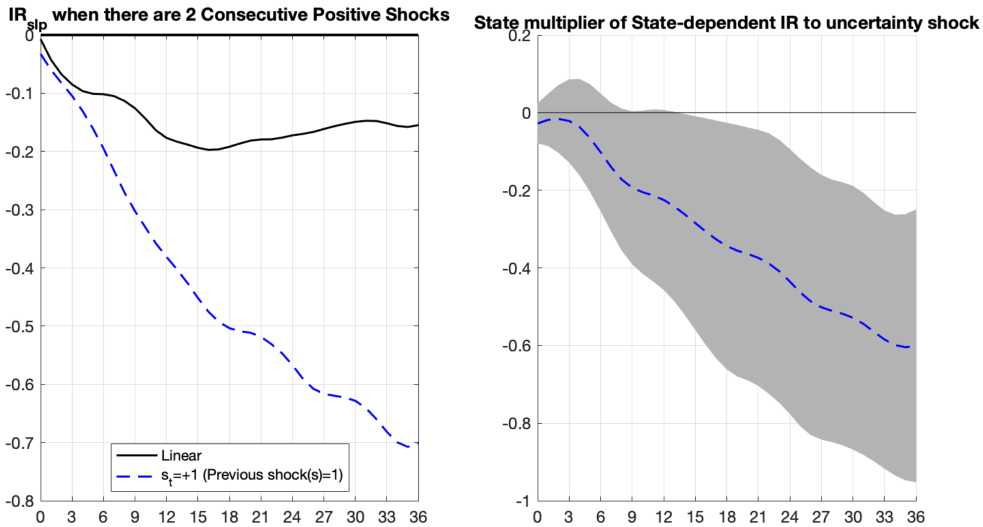


(a) Replication of Figure 2.E-F in Diercks et al. (2024)

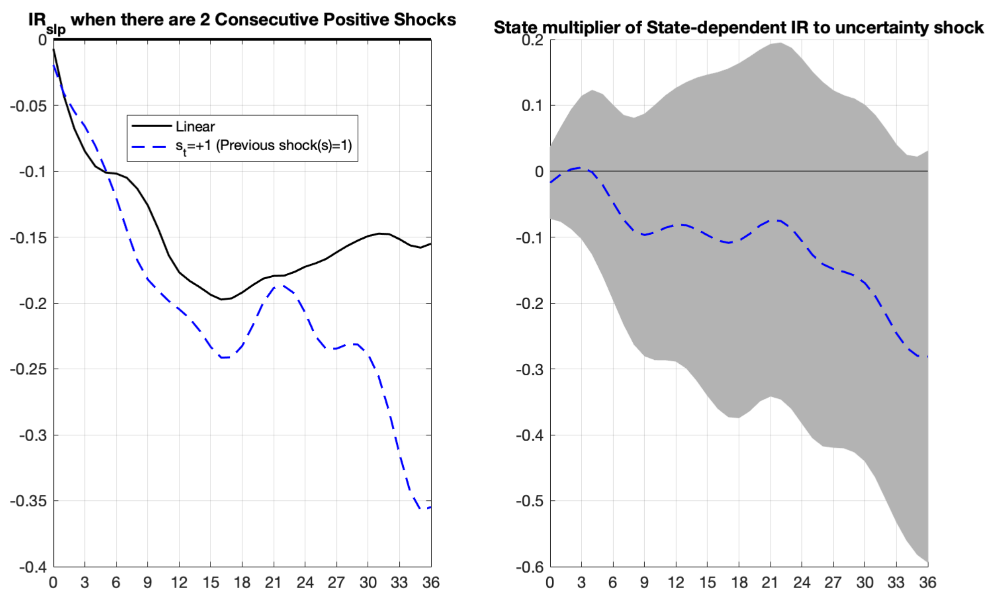


(b) Inclusion of two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of controls

Figure 10: Response of industrial production to consecutive positive EPU uncertainty shocks:
LEFT PANELS: the empirical state-dependent impulse responses (estimated with LPs as in Diercks et al. (2024)) to two consecutive positive uncertainty shocks (dashed blue line) and contrast it to the response to a single shock (solid black line). RIGHT PANELS: the incremental effect of the second shock, i.e. $\{\beta_{1,h}\}_{h=1,..,H}$, with 90% confidence intervals (shaded area). In both panels, on the y-axes, the level of impulse responses; on the x-axes, the horizons, h.



(a) Replication of Figure B.1(e)-(f) in Diercks et al. (2024)



(b) Inclusion of two lags of McCracken and Ng (2016)'s macroeconomic factors to the set of controls

Figure 11: Response of short rate to consecutive positive EPU uncertainty shocks:
LEFT PANELS: the empirical state-dependent impulse responses (estimated with LPs as in Diercks et al. (2024)) to two consecutive positive uncertainty shocks (dashed blue line) and contrast it to the response to a single shock (solid black line). RIGHT PANELS: the incremental effect of the second shock, i.e. $\{\beta_{1,h}\}_{h=1,..,H}$, with 90% confidence intervals (shaded area). In both panels, on the y-axes, the level of impulse responses; on the x-axes, the horizons, h.