Rank-robust Regularized Wald-type tests *

Jean-Marie Dufour †
McGill University

Pascale Valéry ‡
HEC Montréal

First version: January 2005
Revised: September 2009, August 2010, October 2011, October 2013, October 2014, August 2015, May 2016
This version: May 2016
Compiled: June 8, 2020, 11:15

*We are grateful to Hafedh Bouakez, Marine Carrasco, Russell Davidson, John Galbraith, Ronald Gallant, Gloria Gonzalez-Rivera, Patrick Guggenberger, James Hamilton, Sebastian Jaimungal, Michael Jansson, Raymond Kan, Ivana Komunjer, Tae-Hwy Lee, Joris Pinkse, James Powell, Enrique Sentana, Anna Simoni, Yixiao Sun, Aman Ullah for helpful comments as well as seminar participants at PennState University, UC Berkeley, UC San Diego, UC Riverside, Université de Montréal and to seminar participants in Actuarial Sciences and Mathematical Finance seminar series at the Fields Institute, University of Toronto. This work has been presented at the 2008 NBER-NSF conference in Aarhus, the 26th Canadian Econometric Study Group in Ottawa in 2009, the 2010 Econometric Society World Congress in Shanghai as well as at the Gallant conference in Toulouse in May 2011. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

‡ Centre interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi (CIRPÉE), Service de l'enseignement de la finance, École des Hautes Études Commerciales de Montréal (HEC Montréal), 3000 chemin de la Côte-Sainte-Catherine, Montréal, Québec), Canada H3T 2A7. TEL: 1 (514) 340-7004, FAX: (514) 340-5632. email: pascale.valery@hec.ca; webpage: http://neumann.hec.ca/pages/pascale.valery/pascalevalery.htm.

ABSTRACT

This paper studies Wald-type tests in the presence of possibly *rank-deficient* covariance matrices, allowing for singular covariance matrices, either in finite samples or asymptotically. Such difficulties occur in many statistical and econometric problems, such as causality and cointegration analysis in time series, (locally) redundant restrictions, (locally) redundant moment equations in GMM, tests on the determinant of a coefficient matrix (reduced rank hypotheses), tests of linear restrictions on Average Treatment Effects in regression discontinuity designs, etc. Two different types of singularity are considered. *First*, the estimated covariance matrix has *full rank* but converges to a *singular* covariance matrix, so the Wald statistic can be computed as usual, but regularity conditions for the standard asymptotic chi-square distribution do not hold. *Second*, the estimated covariance matrix does not have full rank but converges to a population matrix whose rank may differ from the finite-sample rank. The proposed procedure works in all cases regardless of the finite-sample and asymptotic ranks. To address such difficulties, we introduce a novel mathematical object: the *regularized inverse* which is related to *generalized* inverses, although different. The regularized inverse exploits the spectral decomposition of the covariance matrix; its *unique* representation follows from the Spectral Theorem, see Eaton (2007, Theorem 1.2a, p.53). Results on total eigenprojections (that is the sum of eigenprojections over a subset of the spectral set) are combined with a *variance regularizing function*; the latter modifies small eigenvalues (using a threshold). The continuity property of the total eigenprojection technique ensures a valid asymptotic theory for the regularized inverse; it always exists and is unique. The proposed class of regularized inverse matrices includes both continuous and discontinuous regularized inverses; the Tikhonov-type inverse is continuous, the spectral cut-off regularized inverse as proposed by LB(1997) is discontinuous, and the full-rank regularized inverse we propose is continuous. Under general regularity conditions, we show that sample regularized inverse matrices converge to their regularized asymptotic counterparts. *Regularized Wald statistics* are then obtained through replacement of the usual inverse of the estimated covariance matrix (or the generalized inverse) by a regularized inverse. Both Gaussian and non-Gaussian distributions are allowed for the parameter estimates. Two classes of regularized Wald statistics are studied in relative detail. The first one admits a nonstandard asymptotic distribution, which corresponds to a linear combination of chi-square variables when the estimator used is asymptotically Gaussian. In this case, we show that the asymptotic distribution is *bounded* by the usual (full-rank) chi-square distribution, so standard critical values yield valid tests. In more general cases, we show that the asymptotic distribution can be simulated or bounded by simulation. The second class allows the threshold to vary with the sample size, but additional information is needed. This class of test statistics includes the spectral cut-off statistic proposed by Lütkepohl and Burda (1997, *J. Econometrics*) as a special case. The regularized statistics are consistent against global alternatives, with a loss of power (in certain directions) for the spectral cut-off Wald statistic. An application to U.S. data illustrates how the procedure works when testing for noncausality between saving, investment, growth and foreign direct investment.

Key words: Asymptotic singularity; Regularized Wald test; Moore-Penrose inverse; spectral cut-off and Tikhonov regularizations; Bounded distribution; Monte Carlo tests; Redundant restrictions; Noncausality tests.

Journal of Economic Literature classification: C1, C13, C12, C15, C32.

RÉSUMÉ

Mots-clefs:

Classification du Journal of Economic Literature:

Contents

List of Tables

List of Definitions, Assumptions, Propositions and Theorems

1. Introduction

This paper examines Wald-type tests in the presence of possibly *rank-deficient* covariance matrices, either in finite samples or asymptotically, so the usual regularity conditions that lead to a chi-square asymptotic distribution (with possibly reduced rank) may not hold. The method we propose consists in regularizing the relevant covariance matrix, so the latter has full rank both in finite samples and asymptotically. Our approach is "rank-robust" in the sense that the rank of the covariance matrix is arbitrary in finite samples and can converge to a matrix of any rank (which may differ from the finite-sample rank). In particular, our method allows for a sequence of statistics for which the rank of the covariance matrix varies with the sample size, *i.e.* a drifting sequence of ranks. More generally, a sequence of matrices that converges to a certain limit but whose rank alternates could also be considered. This rules out the cumbersome task of determining the asymptotic rank. Furthermore, we obtain valid inference under both types of singularity, in finite-sample or asymptotic, even though we emphasize the case of asymptotic singularity in our distributional results. The regularization method, which is simple and not computationally intensive, is valid even in the extreme case where the covariance matrix converges to a zero matrix.

In regular setups, the regularized statistic asymptotically behaves like the standard one – though its asymptotic distribution is modified, unless the threshold varies with the sample size – while it is robust to rank deficiencies in problematic cases.[1] Asymptotically valid tests can thus be performed regardless of the asymptotic rank. More specifically, a bound is easily obtained for the full-rank regularized statistic. The bound is appealing, because it relies on usual critical points for the full rank case and is invariant to the degree of rank deficiency. These results only require information on the distribution of the estimated restricted parameters, not the data generating process (DGP). The distribution of the estimator need not be Gaussian. Should the test based on the bound be conservative, this feature can be alleviated through simulations as soon as some information on the DGP is available.

If the covariance matrix estimator of an asymptotically normal random vector converges to a singular matrix, using its generalized inverse (g-inverse) – rather than the g-inverse of the limit in the corresponding normalized quadratic form that defines a Wald-type statistic – yields a statistic whose asymptotic distribution is chi-square with a reduced number of degrees of freedom, provided the rank of the estimated covariance matrix converges to the rank of the limit matrix with probability one; see Andrews (1987).[2] Otherwise, the asymptotic distribution of the quadratic form is typically modified.

Problems of this type are quite varied in econometrics, such as many test problems in time series, tests involving (locally) redundant restrictions or redundant moment equations in GMM, tests on the determinant of a coefficient matrix (for reduced rank hypotheses), etc. Situations that lead to asymptotic rank deficiencies include: tests on impulse response coefficients in VAR models, tests of Granger non-causality in VARMA models [Boudjellaba, Dufour and Roy (1992, 1994)], tests of noncausality at multiple horizons [Dufour and Renault (1998), Dufour, Pelletier and Renault (2006)], tests of dynamic specification in time series [Sargan (1980), Gouriéroux, Monfort and Renault (1989), Galbraith and Zinde-Walsh (1997)], tests for indirect effects and mediation analysis [Sobel (1982, 1986) , MacKinnon, Lockwood, Hoffman, West and Sheets (2002), MacKinnon, Lockwood and Williams (2006), MacKinnon (2008), Tofighi and MacKinnon (2016)], tests

[1] This paper does not deal with deficient ranks due to (first-order) underidentification. For those interested in such issues, see Dovonon and Renault (2009), and Pötscher (1985). More generally, for those interested in weak identification issues in IV/GMM, see Dufour (1997), Stock and Wright (2000), Stock, Wright and Yogo (2002), Dufour and Taamouti (2005, 2007) , Antoine and Renault (2009). Nevertheless, we allow for situations of *weak* identification of θ only to the extent that the transformation $\psi(\theta)$ is identified.

[2] As pointed out by Andrews (1987, p. 349), the rank condition fails when the estimator of the covariance matrix has greater rank than the limiting matrix. This can occur when sums of outer products are used to estimate the limiting matrix in the presence of polynomial time trends, cointegrated variables, I(1) variables with innovations whose means are nonzero. Violation of the rank condition could also occur when testing nonlinear restrictions with the Wald test for some isolated values of the parameter; the derivative matrix of the restrictions when evaluated at a consistent unrestricted estimator will generally exceed that of the derivative matrix at the limit when the isolated value of the parameter is true.

of collapsibility and unions of conditional independence hypotheses in contingency tables [Glonek (1993)], tests on tetrads and nonlinear hypotheses in graphical causal modelling and factor analysis [Drton, Sturmfels and Sullivant (2007), Drton, Massam and Olkin (2008), Zwiernik and Smith (2012)], diagnostic tests on the residuals of autoregressive models [Box and Pierce (1970),Li and McLeod (1981), Ljung (1986)], tests on the coefficients of cointegrated VAR processes [Sims, Stock and Watson (1990)], tests of long-run relationships in cointegrated systems [Gonzalo and Lee (1998)], stochastic discount factor specification tests in a GMM framework [Marin (1996), Kan and Robotti (2009), Peñaranda and Sentana (2012)], etc.[3]

Finite-sample and asymptotic singularities arise naturally with redundant constraints. When dealing with nonlinear conditional moment restrictions as in Gallant and Tauchen (1989) for the I-CAPM model, many parametric restrictions turn out to be redundant; this creates collinearity problems for the Jacobian matrix. Redundant moment restrictions also arise in a dynamic panel GMM setting, when linear moment conditions imply nonlinear moment conditions under additional initial conditions on the dependent variable [Arellano and Bond (1991), Ahn and Schmidt (1995), Blundell, Bond and Windmeijer (2000), Doran and Schmidt (2006)] or when the number of parameters exceed the number of observations [Satchachai and Schmidt (2008)]. In view of such difficulties, Carrasco and Florens (2000), Carrasco, Chernov, Florens and Ghysels (2007), Carrasco, Florens and Renault (2007), and Carrasco (2012) regularize estimators when a continuum of moments is used in a GMM/IV framework. General results on regularized estimators for high dimensional covariance matrices can be found in Bickel and Levina (2004), Bickel and Levina (2008*b*, 2008*a*), Ledoit and Wolf (2004b).[4] On the estimation of high-dimensional covariance matrices for portfolio allocation and risk management, see also Ledoit and Wolf (2004a), Fan, Fan and Lv (2008), Fan, Liao and Mincheva (2011), and Carrasco and Noumon (2011).

Alternatively, in the context of polynomial restrictions, Dufour, Renault and Zinde-Walsh (2017) provide a general characterization of the asymptotic distribution of the Wald statistic under asymptotic singularity. They derive a wide array of asymptotic distributions for the original Wald statistic (without modification) possibly involving nuisance parameters for a given null hypothesis; bounds are also derived. Although very general, the characterization of the Wald statistic in irregular setups is very complicated. Indeed, as soon as the regularity conditions fail (*e.g.* the Jacobian matrix of the restriction is rank deficient), one may expect deviations from the standard asymptotic distribution. The limit distribution depends upon whether its evaluation point is regular or singular. More specifically, the limit distribution under singularity is no longer pivotal; the degree of singularity and the features of the singular point severely impact the limit distribution which is no longer unique. In contrast, our regularization method is robust to such features, in particular does not rely on any rank assumption (on the Jacobian matrix). In our case, the full-rank distribution still yields valid critical values for the regularized statistic. Moreover, testing several polynomial restrictions is even worse, that is the Wald-type statistic can diverge at a singular point even under the null hypothesis. Early detection of divergence by examining the restrictions is a difficult and cumbersome task. Alternatively, regularizing the statistic in the first place provides an easy remedy that rehabilitates the Wald statistic in terms of level

[3] Kan and Robotti (2009) note in a footnote on page 3461:

> "that we should not perform a Wald test of $H_0: \eta_1 = \beta_1,\ \psi = 0_{K_2+K_3}$. This is because the asymptotic variance of $\sqrt{n}[\hat{\eta}_1' - \hat{\beta}_1', \hat{\psi}']'$ is singular under H_0, and the Wald test statistic does not have the standard asymptotic $\chi^2_{K_1+K_2+K_3+1}$ distribution. The proof is available upon request."

[4] Although Bickel and Levina develop a battery of tools for regularizing high-dimensional covariance matrices in order to produce a consistent estimator for the population covariance matrix, those regularization techniques – that also employ some thresholding tools but in a different fashion– crucially rely on the sparsity assumption for the limiting covariance matrix as well as on the multivariate Gaussian (or sub-Gaussian) distribution of the variables. In contrast, our approach does not restrict to Gaussian distributions neither imposes any sparsity feature on the population matrix. Further unlike banding-type or tapering-type regularization tools that exploits natural ordering among variables or notions of distance between variables, our regularization technique does not hinge on such assumptions. Additionally, our regularization approach, which nests ridge-type regularization scheme, does not restrict to shrinkage estimators of covariance matrices; for shrinkage types of well-conditioned covariance estimators see Ledoit and Wolf (2004b).

control for testing multiple nonlinear restrictions. Thus modifying the Wald statistic as we propose, yields a wide-range, rank-free simple solution to intricate general problems (e.g., alternate ranks, drifting sequences of ranks). In other words, the prime appeal of our regularization approach is its universality. When further combined with simulations (*e.g.,* Monte Carlo tests), the bound can be enhanced, that is less conservative.

Violation of Andrews's rank condition also appear in the estimation of a lower bound on the number of mixture components in finite mixture models [see Kasahara and Shimotsu (2014),[5] and testing linear restrictions on Average Treatment Effects (ATE) where the $J \times J$ covariance matrix of the ATE estimates [a sum of two rank-one matrices] is singular [see Xu (2016)]. Within the Regression Discontinuity (RD) design, the RD estimator is the difference of two estimators with singular asymptotic covariance matrices; thus in this RD setting, the rank of the covariance matrix of the ATE's depends upon whether the Jacobian of the outcome probabilities of the two groups – treated and control groups – are linearly independent, which is usually unknown to the econometrician. In this context, Xu (2016, section 2.4, p.10-12) points out the case where the rank of the sample variance matrix happens to be larger than that of the population matrix, thereby violating Andrews's rank condition. Like Lütkepohl and Burda (1997), Duchesne and Francq (2015) study modified Moore-Penrose inverses in order to satisfy Andrews's rank condition. Further, asymptotic singularity also arises when some components of the ATE estimates enjoy different convergence rates, with some converging faster than others. Numerous situations may further occur in the econometric literature: when testing nonlinear restrictions, this can happen if an asymptotic rank deficiency obtains on sets of Lebesgue measure zero (*e.g.*, at isolated points) in the parameter space.

In this paper, we focus on testing issues. We propose a general approach to deal with deficient rank covariance matrices in order to conduct valid Wald-type tests in two different ways: (1) relatively simple asymptotic bounds, and (2) a simulation-based approach that can handle non-standard distributions in the context we consider. To overcome the problem of asymptotic singularity, Lütkepohl and Burda (1997) propose to reduce the rank of the matrix estimator in order to satisfy Andrews's rank condition. To do so, they set to zero the small problematic eigenvalues to produce a consistent estimator for the rank of the asymptotic covariance matrix. More generally within the rank testing literature Gill and Lewbel (1992), Cragg and Donald (1996, 1997), Robin and Smith (2000) and Kleibergen and Paap (2006) focus on tests for the rank of a matrix that is unobserved, but for which a $\sqrt{n}$ consistent estimator is available. Unlike Cragg and Donald (1996, 1997) , Robin and Smith (2000) and Kleibergen and Paap (2006) who assume Gaussianity for the limiting distribution of the covariance matrix estimator, our methodology [based on the theory developed by Eaton and Tyler (1994)] is more general, since the availability of a $\sqrt{n}$ asymptotically Gaussian estimator is not required for the asymptotic covariance matrix.[6] Al-Sadoon (2017) describes a general structure of rank test statistics; those are shown to be functions of implicit estimators of the null spaces of the matrix of interest. See also Doran and Schmidt (2006) for a reduced-rank weighting matrix estimate in highly-overidentified GMM setups; like Lütkepohl and Burda (1997), they discard the smallest eigenvalues to improve finite-sample properties of the estimate. Further, Gouriéroux and Jasiak (2009) have shown that the asymptotic distribution of the Wald statistic for testing the noninvertibility of a matrix A based upon the estimated determinant is seriously affected when $A = 0$. Moreover, the asymptotic distribution of a reduced-rank estimator of A is different depending upon whether $A = 0$ or $A \neq 0$; size distortions may result from using quantiles of the standard asymptotic distribution (*i.e.* those from $A \neq 0$).

When dealing with singular covariance matrices, usual inverses are discarded and replaced with g-inverses [see Moore (1977), Andrews (1987) for the generalized Wald tests] or modified inverses proposed by Lütke-

[5] In order to satisfy Andrews's rank condition, the authors follow the approach of Lütkepohl and Burda (1997), which employs a reduced-rank modified Moore-Penrose pseudo-inverse in the *average-rk statistic* defined in Kasahara and Shimotsu (2014, Eq. 10 section 3.4, p.105).

[6] Estimating the rank as Lütkepohl and Burda (1997), Robin and Smith (2000) do may not be the right thing to do when it comes to assess the finite sample distribution of such estimators. Our results somehow validate the intuition of Leeb and Pötscher (2003, 2005) who are very critical of post-model selection estimators.

pohl and Burda (1997) and Duchesne and Francq (2015). However, when using g-inverses, it is important to remain aware of two difficulties. *First*, the continuous mapping theorem so widely used in econometrics to derive asymptotic distributional results does not apply anymore because g-inverses are not (necessarily) continuous [see Andrews (1987)]. Unlike eigenvalues, eigenvectors are not continuous functions of the elements of the matrix. *Second*, when using the singular value decomposition of a matrix, the eigenvectors corresponding to the eigenvalues with multiplicity larger than one, are not uniquely defined, which rules out convergence in the usual sense. Ignoring these difficulties can lead to distributional results which are *stricto sensu* wrong.

To address such difficulties, we introduce a class of *regularized* inverses whose convergence properties exploit the technique of *total eigenprojection*, *i.e.* an eigenprojection operator taken over a subset of the spectral set. Following Kato (1966) and Tyler (1981), we work with the *total eigenprojection* to overcome the discontinuity and non-uniqueness of eigenvectors; like eigenvectors, individual eigenprojections may still be discontinuous in the elements of the matrix. A lemma given by Tyler (1981) states the continuity property for the *total eigenprojection*. As a result, the important continuity property is preserved for eigenvalues and *total eigenprojections* even though eigenvectors are *not* continuous. We further define a perturbation function of the inverse of the eigenvalues called the *variance regularizing function* (VRF). The VRF $g(\lambda;c)$ modifies the small eigenvalues that fall below a certain threshold so that their inverse is well behaved whereas the large eigenvalues remain unchanged. Depending on the choice of the variance regularizing function, we can produce regularized inverses that are continuous, like the full-rank regularized inverse, or discontinuous versions of it, like the modified Moore-Penrose inverse (see Lütkepohl and Burda (1997). Under specific regularity conditions, the regularized inverse converges to its regularized population counterpart. The distributional theory of the test statistic resulting from the total eigenprojection technique is therefore valid.

Our contributions can be summarized as follows. *First*, we introduce a novel mathematical object: a *full-rank regularized inverse* which employs a new regularization scheme [*i.e.* $g(\lambda;c)=\frac{1}{c}$ for eigenvalues smaller than the threshold c]. This full-rank regularized inverse is contrasted with g-inverses. This new class of inverses has *full rank*, and satisfies a decomposition property: a *regular* component based on large eigenvalues, and a *nonregular* component based on small eigenvalues which may be associated with small or zero eigenvalues of the asymptotic covariance matrix. This matrix decomposition determines a corresponding decomposition of the regularized Wald statistic. Under simple conditions on the VRF, we show that the regularized inverse converges to its full rank regularized counterpart; the convergence holds component by component. Besides, the class of regularized inverses is general, including as special cases the spectral cut-off type inverse and a Tikhonov-type inverse. *Second*, we define a regularized Wald statistic that relies on a fixed value of the threshold in the VRF $g(\lambda;c)$. Another version allows the threshold to vary with the sample size, but requires more information on the behavior of estimated eigenvalues. The first regularized Wald statistic admits a nonstandard asymptotic distribution in the general case, which corresponds to a linear combination of chi-square variables if the restrictions are Gaussian. A *conservative bound* is then obtained for the distribution of the regularized Wald statistic. Hence, the test is *asymptotically valid*: usual critical points (given by the chi-square variable with *full rank*) can be used, but are conservative. Interestingly, the bound is invariant to the degree of rank deficiency of the covariance matrix. When the threshold goes to zero with the sample size, we obtain the spectral cut-off modified Wald statistic proposed by Lütkepohl and Burda (1997) as a special case. Under normality, the test statistic has the chi-square asymptotic distribution whose reduced rank is given by the number of eigenvalues greater than zero. Note that Lütkepohl and Burda's (1997) result only holds for distinct eigenvalues whereas our result accounts for eigenvalues with multiplicity larger than one. *Third*, to complement our bound, we propose three alternative ways to conduct the (regularized) Wald test by simulation: *(i)* when a DGP is completely specified, the distribution of the test statistic can be simulated by simulating the DGP; *(ii)* when the DGP is not available, but the asymptotic distribution of the estimator is known (at least in large sample), the test statistic can be simulated by simulating the estimator; *(iii)* when the restrictions (evaluated at the unrestricted parameter estimate) can be simulated, this

also provides a way of simulating the test statistic. These three approaches require different amounts of information on the model and the estimator employed, so they have different reliable properties with respect to asymptotic error, nonlinearity and identification. For example, simulating under the law of the restrictions may allow one to bypass identification problems raised by the presence of unidentified parameters.

We investigate in a Monte Carlo experiment the finite and large-sample properties of the regularized test statistics. Our findings can be summarized as follows. *i)* Regarding level control, the standard Wald statistic (*i.e., W*) suffers from severe over-rejections in small samples, or from under-rejections in large samples in non-regular setups. Similarly, the reduced rank Wald statistic (*i.e.,* W_{LB}) displays the same poor, finite sample behavior as the standard statistic in non-regular setups, with critical size distortions when parameter values approach the nonstationary region. However, it exhibits good size properties asymptotically. In contrast, the full-rank regularized statistic that uses the bound is conservative. We observe that this feature can be alleviated by using simulation-based versions of the regularized statistics. If one directly simulates the DGP, one can control the level of the test for the full-rank regularized statistic even in small samples. Thus, it is very important to simulate from a well-behaved statistic to produce a reliable test. *ii)* In terms of power, the full-rank regularized test statistics do not entail a significant loss of power under the alternative compared to their oversized infeasible competitors W and W_{LB} in small samples for the asymptotic tests. Finally, the most striking result is the severe *under-performance* of the reduced rank statistic W_{LB} in a regular setup. As already mentioned by Lütkepohl and Burda (1997), by underestimating the true rank of the covariance matrix, this reduced rank statistic puts more weight on the first restriction that remains fulfilled in this case. A violation of the null hypothesis coming from the second restriction will not be detected by a statistic that underestimates the rank; a full-rank regularized statistic dominates in such a case. Thus, these results on power reinforce the better properties of the full-rank regularized statistics over the spectral cut-off one.

iii) We finally illustrate the procedure on U.S. data by conducting noncausality tests at several horizons to assess any causal relation between Saving, Investment, Growth and Foreign Direct Investment (FDI) (in the presence of (locally) redundant restrictions). While most of the procedures are not able to reject the null hypothesis that Saving does not cause Growth at all horizons, we unambiguously find that Growth causes Saving, and that Investment causes Growth in the presence of FDI on U.S. data. Our findings support the original literature by Houthakker (1961, 1965), and Modigliani (1970) at the cross-country level. Moreover, our findings confirm Dufour and Renault (1998, Proposition 4.5)'s results that in a VAR(1) model, it is sufficient to have noncausality up to horizon 2 for noncausality to hold at all horizons.

The paper is organized as follows. In Section 2 we describe a general framework with minimal assumptions. In Section 3, we provide specific examples in the presence of (asymptotic) singular covariance matrices. In Section 4, we introduce the class of *regularized* inverses. The *regularized* test statistic is presented in Section 5. In Section 6, we review and adapt some results on total eigenprojections. In Section 7, we establish the asymptotic properties of the new regularized inverse based on a fixed threshold. In Section 8, we state new asymptotic distributional results for the regularized Wald test statistic that uses a fixed threshold. We exploit the decomposition of the regularized statistic to derive an upper bound. In Section 9, we obtain, as a special case, the Lütkepohl and Burda's (1997) result in the Gaussian case. Finally, we illustrate the procedure by conducting noncausality tests at several horizons on U.S. data in Section 10. Concluding remarks follow while proofs and simulation experiments on finite and large sample properties of the tests are gathered in the appendix.

2. Framework

Consider a family of probability spaces $\{(\mathscr{L}, \mathscr{A}_{\mathscr{L}}, \bar{\mathbb{P}}_\theta) : \theta \in \Theta\}$, where $\mathscr{L}$ is a sample space, $\mathscr{A}_{\mathscr{L}}$ is a σ-algebra of subsets of $\mathscr{L}$, and $\bar{\mathbb{P}}_\theta$ is a probability measure on the measurable space $(\mathscr{L}, \mathscr{A}_{\mathscr{L}})$ indexed by a parameter θ in $\Theta \subset \mathbb{R}^p$. The sets $\mathscr{L}, \mathscr{A}_{\mathscr{L}}$, and Θ are all nonempty. Typically, we are interested in a

transformation $\psi : \Theta_1 \to \Psi$, defined on a nonempty subset Θ_1 of Θ on which we want to test hypotheses of the form $H_0(\psi_0) : \psi(\theta) = \psi_0$. Let Γ_0 be a nonempty subset of Ψ, $\Theta_0 = \{\theta \in \Theta_1 \subset \mathbb{R}^p : \psi(\theta) \in \Gamma_0 \subset \mathbb{R}^q\}$. A usual test statistic for testing the null hypothesis is the Wald-type statistic as soon as a consistent estimator $\hat{\psi}_n$ of the restrictions is available. We first consider a general Wald-type statistic based on an arbitrary weighting matrix A_n:

$$W_n(\hat{\psi}_n, A_n) = \left[H_n(\hat{\psi}_n - \psi_0)\right]' A_n \left[H_n(\hat{\psi}_n - \psi_0)\right], \tag{2.1}$$

where $\hat{\psi}_n$ is an unrestricted estimator of the restrictions $\psi(\theta)$. $W_n \equiv W_n(\hat{\psi}_n, A_n)$ is continuous with respect to (w.r.t) the restrictions and the weighting matrix A_n which allows fairly weak conditions. Usually A_n is the inverse of a covariance matrix estimator $\boldsymbol{\Sigma}_n$ for $\hat{\psi}_n$. However, this specification allows more general forms of the weighting matrix A_n. More generally, this setup includes as special cases either the well-known standard case whenever the estimator and its limit have full rank - in that case $A_n = \boldsymbol{\Sigma}_n^{-1}$ - or deficient ranks with $A_n = \boldsymbol{\Sigma}_n^{\dagger}$. Thus, among regularity conditions usually made when conducting tests based on quadratic forms such as Wald-type tests, is the well-known rank condition for the covariance matrix. When the population matrix $\boldsymbol{\Sigma}$ and its sample analog $\boldsymbol{\Sigma}_n$ have full ranks, we are in the regular case with the $q \times q$-weighting matrix $\boldsymbol{\Sigma}$ being nonsingular, and therefore W_n has an asymptotic $\chi^2(q)$ distribution. This is not necessarily true, however, if $\boldsymbol{\Sigma}$ is singular. In this case, $\boldsymbol{\Sigma}$ does not admit a usual inverse, but can still be inverted by means of a generalized inverse. However, when the population matrix $\boldsymbol{\Sigma}$ has a reduced rank, the rank of the sample matrix has to converge almost surely (a.s.) towards the *reduced rank* of the population matrix for the quadratic form to have a limiting chi-square distribution, with fewer degrees of freedom, when the restrictions are assumed to be asymptotically Gaussian. This is the case covered by Andrews (1987); we shall relax this assumption in this paper. In other words, the method we propose is applicable under more general assumptions: it is valid even though the finite sample (covariance) matrix $\boldsymbol{\Sigma}_n$ is not invertible (hence requiring a g-inverse), or is invertible but converges to a singular population matrix $\boldsymbol{\Sigma}$. For notations $\underset{n\to\infty}{\overset{\mathscr{L}}{\to}}$, $\overset{a.s.}{\to}$ and $\overset{p}{\to}$ denote the convergence in law, the almost sure convergence and the convergence in probability respectively, and $\mathscr{L}(X)$ denotes the law of X. Let $\hat{\psi}_n$ satisfy the following assumption, where implicitly $\psi = \psi(\theta)$.

Assumption 2.1 CONVERGENCE IN LAW OF THE RESTRICTIONS WITH DIFFERENT RATES OF CONVERGENCE. *Let X_n and X be random vectors in $\mathbb{R}^q$. $H_n = (h_{n,ij})_{i,j=1,\dots,q}$ is a sequence of real-valued $q \times q$ matrices such that $h_{n,ij} \to \infty$, and $X_n \equiv H_n(\hat{\psi}_n - \psi) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X$, where $\mathscr{L}(X)$ is known.*

Assumption 2.1 significantly enlarges the family of admissible laws for $\hat{\psi}_n$; the typical Gaussian distribution for X can easily be replaced by a chi-square distribution, or a Cauchy distribution. The assumption allows us to deviate from the typical root-n convergence rate of the estimator of the restrictions. If H_n is a diagonal matrix with distinct elements, the rates of convergence of $\hat{\psi}_n$ differ across components. Generally speaking, any distribution that can be consistently estimated by simulations is admissible. Therefore, if $\mathscr{L}(X)$ is not known, but can be simulated through bootstrap techniques, *e.g.*, then the techniques proposed in this paper can be applied to provide *valid* tests under nonregular conditions. More importantly, note that Assumption 2.1 only requires ψ to be identified; in other words, θ can be unidentified, but there exist transformations of θ, *i.e.* $\psi(\theta)$, that can be identified. In regression problems, it is frequent to encounter situations where only certain components of the parameter of interest θ are identified; in such a case, inference is limited to the identified components. Whereas Lütkepohl and Burda (1997) assume the availability of an asymptotically Gaussian estimator of θ, as in equation (2.4), that unnecessarily restricts to situations where θ is identified, we relax this assumption here. In doing so, we allow for situations of *weak* identification only to the extent that $\psi(\theta)$ is identified. Like Lütkepohl and Burda (1997), Duchesne and Francq (2015) examine the behavior of generalized Wald tests under singular normal distributions. Note that ψ will alternately equal ψ_0 under

the null hypothesis, or ψ_1 under the alternative. Of course, the distributions characterizing the null and the alternative are distinct.

Further, a general condition given by Eaton and Tyler (1994) states the convergence result for the weighting matrix A_n (or a set of parameters).

Assumption 2.2 EATON-TYLER CONDITION. *A_n is a sequence of $p \times q$ real random matrices and A is a $p \times q$ real nonstochastic matrix such that $Q_n = b_n(A_n - A) \xrightarrow[n\to\infty]{\mathscr{L}} Q$, where b_n is a sequence of real constants such that $b_n \to +\infty$, and Q a random matrix.*

Note that Assumption 2.2 is less restrictive than that of Robin and Smith (2000, Assumption 2.2, p. 154) and Kleibergen and Paap (2006, Assumption 1, p. 103). Indeed, Assumption 2.2 allows situations whose matrix estimator is not asymptotically Gaussian, *e.g.*, Wishart distributions. More specifically, b_n may differ from the conventional $\sqrt{n}$ convergence rate. The Eaton-Tyler condition is stated for rectangular matrices, but most of the time we will consider square matrices that are symmetric with real eigenvalues. Assumptions 2.1 and 2.2 will constitute the cornerstone for the validity of the distributional results developed further. In particular, we do not require the finite-sample rank to converge to the asymptotic rank for the procedure to be valid. Our approach is rank-robust even allowing the (finite-sample) rank to alternate. It is also important to note that the generality of Assumption 2.2 enables a mixture of a continuous distribution and of a Delta-Dirac distribution at an eigenvalue $\lambda = c$. Therefore, it is not superfluous to examine this case, especially for non-continuous distributions of matrices and their eigenvalues, to provide a thorough and comprehensive distributional theory. Note that Assumption 2.2 implies that $A_n \xrightarrow{p} A$. Under Assumptions 2.1 and 2.2, we can easily obtain the distribution of the Wald statistic $W_n(\hat{\psi}_n, A_n)$ given in a general form. It is worth mentioning at this stage that under those assumptions only a Wald test can be conducted to the extent that no information is available on the DGP. Only an estimator (of the restrictions) is assumed available to perform the test. As a result, a LR-type test can hardly be conducted without further information.

Lemma 2.1 *Under Assumption 2.1 and 2.2, the statistic $W_n(\hat{\psi}_n, A_n)$ defined in equation (2.1) is such that:*

$$W_n(\hat{\psi}_n, A_n) \xrightarrow[n\to\infty]{\mathscr{L}} X'AX. \tag{2.2}$$

The general form of the statistic $W_n(\hat{\psi}_n, A_n)$ in equation (2.1) based on the general weighting matrix A_n bypasses any issue related to the invertibility of the covariance matrix estimator $\boldsymbol{\Sigma}_n$. As soon as a pseudo-inverse exists, one can conduct the test, at the cost of a slightly intricate distributional theory. Most of the time, the Wald test is implemented using the inverse of the covariance matrix of the restrictions under normality. Indeed, if normality is assumed as in Assumption 2.4 below, the Wald statistic follows a chi-square distribution with the number of degrees of freedom given by the rank of the asymptotic covariance matrix. Intentionally, H_n in equation (2.1) allows situations where some components of $\hat{\psi}_n$, or linear combinations of them, do not converge at the same rate. Some components might converge faster or slower than the conventional $\sqrt{n}$ rate; for example *superconsistent* estimators arise in a unit root framework, or in a simple time trend model (see Hamilton (1994, chapter 16, page 457-460)).

While $\psi = \psi(\theta)$ in Assumption 2.1 can accommodate some identification problems on some components of θ, it might involve some discontinuity at some specific values, *e.g.*, $\{\theta = (\theta_1, \theta_2) \in \Omega : \theta_2 = 0\}$ for $\psi(\theta) = \theta_1/\theta_2$. In this case, one should rather work with θ and place oneself under the alternative assumption:

Assumption 2.3 CONVERGENCE IN LAW OF THE ESTIMATOR OF THE PARAMETER WITH DIFFERENT RATES OF CONVERGENCE. *Let $\tilde{X}_n$ and $\tilde{X}$ be random vectors in $\mathbb{R}^p$. $\tilde{H}_n = (\tilde{h}_{n,ij})_{i,j=1,\dots,q}$ is a sequence of real-valued $p \times p$ matrices such that $\tilde{h}_{n,ij} \to \infty$, and $\tilde{X}_n \equiv \tilde{H}_n(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathscr{L}} \tilde{X}$, where $\mathscr{L}(\tilde{X})$ is known.*

Finally, a data generating process (DGP) may be available in specific settings. One could exploit the DGP (or the corresponding parametric model) to derive the distribution of $\hat{\theta}_n$ or that of $\psi(\hat{\theta}_n)$. The knowledge of the parameter θ completely specifies the distribution of the data. Let us express the usual Wald statistic as a function of the parameter θ:

$$W_n(\hat{\theta}_n, A_n) = [H_n(\psi(\hat{\theta}_n) - \psi(\theta))]' A_n [H_n(\psi(\hat{\theta}_n) - \psi(\theta))]. \tag{2.3}$$

Most of the time, the weighting matrix A, as well as its sample analog A_n, is interpreted as a covariance matrix. Nevertheless, such an interpretation is very restrictive and discards too many distributions, for instance those whose moments do not exist, like the Cauchy distribution. Therefore, Assumptions 2.1 and 2.2 are purposely formulated to allow general distributions for nonlinear restrictions on the parameter θ. For instance, nonlinear transformations of a Gaussian estimator of θ such as high-order polynomial transformations will not typically be Gaussian, *e.g.,* tests on the determinant of a matrix. Also quadratic forms based on superconsistent estimators in the unit-root literature typically involve non-Gaussian random variables.

Let us now focus on the usual case where the weighting matrix A_n in Assumption 2.2 is equal to $\boldsymbol{\Sigma}_n$, *i.e.*, a consistent estimator of the limiting covariance matrix $\boldsymbol{\Sigma}$ of the restrictions.

A special case of Assumption 2.1 that is usually encountered in the econometric literature consists in specifying a Gaussian distribution for X whose parametrization hinges on $\boldsymbol{\Sigma}$ with $H_n = \sqrt{n} I_q$ as in Lütkepohl and Burda (1997).

Assumption 2.4 ROOT-n ASYMPTOTIC NORMALITY. *Let X_n and X be random vectors in $\mathbb{R}^q$. $X_n \equiv \sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta)) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X$, where $\mathscr{L}(X) = N(0, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is a fixed $q \times q$ matrix.*

Note that Assumption 2.4 allows for the most degenerate case corresponding to $\boldsymbol{\Sigma} = 0$. In this case, the population eigenvalue denoted by d_j is zero with multiplicity q, namely $d_j = 0$, with $m(d_j) = m(0) = q$. Usually, one derives the asymptotic normality of the restrictions from the root-n asymptotic normality of the estimator $\hat{\theta}_n$ of the underlying parameter θ through the delta method, *i.e.*,

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \mathrm{N}(0, \boldsymbol{\Sigma}_\theta). \tag{2.4}$$

This requires the continuously differentiability of the restrictions unlike Assumption 2.1. In doing so, econometricians unnecessarily restrict the family of admissible restrictions to those for which the delta method is applicable. Thus, when the delta method is applied to the Gaussian estimator given in equation (2.4), the covariance matrix has the typical form $\boldsymbol{\Sigma} = P(\theta)\boldsymbol{\Sigma}_\theta P(\theta)'$ which critically hinges on the differentiability of the restrictions, *i.e.* $P(\theta) = \partial\psi(\theta)/\partial\theta'$ as in Lütkepohl and Burda (1997). By contrast, Andrews (1987, Theorem 1) does not rely on the differentiability property of the restrictions, nor on the delta method, but on the Gaussian distribution of the random variable X, and on the consistency of the sample *covariance* matrix to its population counterpart. Indeed, any weighting matrix can be used in the Wald-type statistic but only the *covariance* matrix of the restrictions yields the standard chi-square distribution. If a different weighting matrix is used instead, the resulting distribution is modified.

3. Examples

In this section, we first provide examples of rank-deficient covariance matrices, emphasizing the asymptotic singularity case that affects the distribution of the Wald test statistic. Then we illustrate situations where the restrictions tested are not necessarily Gaussian, thereby altering the limiting distribution of the Wald-type test statistic.

3.1. Multistep noncausality

As already observed by Lütkepohl and Burda (1997), when testing for noncausality with a Wald test statistic, one may encounter asymptotically singular covariance matrices. For the sake of comparison, we examine the example studied by Lütkepohl and Burda (1997). For simplicity, a VAR(1) process is considered for the (3×1) vector $\mathbf{y}_t = [x_t' y_t' z_t']'$ as follows:

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = A_1 \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + u_t = \begin{bmatrix} \theta_{xx} & \theta_{xy} & \theta_{xz} \\ \theta_{yx} & \theta_{yy} & \theta_{yz} \\ \theta_{zx} & \theta_{zy} & \theta_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_{x,t} \\ u_{y,t} \\ u_{z,t} \end{bmatrix} .$$

Suppose $Y \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, $B \equiv (A_1)$, $Z_t \equiv [\mathbf{y}_t]$, $\quad Z \equiv (Z_0, \ldots, Z_{n-1})$, $U \equiv [u_t]_{t=1,\ldots,n} = (u_1, \ldots, u_n)$, where $u_t = [u_{x,t}' u_{y,t}' u_{z,t}']'$ is a white noise with a 3×3 nonsingular covariance matrix $\boldsymbol{\Sigma}_u$. Using the standard column stacking operator *vec*, let $\theta = vec(A_1) = vec(B)$, where B is (3×3) and Y, Z and U are $(3 \times n)$. Testing the null hypothesis of multi-step noncausality running from y to x, *i.e.* $H_0 : y_t \overset{(\infty)}{\not\to} x_t$, requires to test 2 restrictions on θ of the following form [see Dufour and Renault (1998)]:

$$\psi(\theta) = \begin{bmatrix} \theta_{xy} \\ \theta_{xx}\theta_{xy} + \theta_{xy}\theta_{yy} + \theta_{xz}\theta_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

These restrictions are fulfilled in the following three parameter settings:

$$\theta_{xy} = \theta_{xz} = 0, \theta_{zy} \neq 0, \quad \theta_{xy} = \theta_{zy} = 0, \theta_{xz} \neq 0, \quad \theta_{xy} = \theta_{xz} = \theta_{zy} = 0. \tag{3.1}$$

We observe that the first-order partial derivative of the restrictions leads to a singular matrix

$$\frac{\partial \psi}{\partial \theta'} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \theta_{xy} & 0 & 0 & \theta_{xx} + \theta_{yy} & \theta_{xy} & \theta_{xz} & \theta_{zy} & 0 & 0 \end{bmatrix} \tag{3.2}$$

if (3.1) holds; the rank is equal to 1 instead of 2. Under such circumstances, the Wald test statistic does not have the standard chi-square distribution under the null.

3.2. Jacobian matrix degenerate at isolated values for a stochastic volatility model

A two-step GMM-type estimator for estimating $\theta = (a_w, r_w, r_y)'$ has been proposed by Dufour and Valéry (2009) in the context of a lognormal stochastic volatility model:

$$y_t = c y_{t-1} + u_t, \quad |c| < 1, \quad u_t = [r_y \exp(w_t/2)] z_t \,, \; w_t = a_w w_{t-1} + r_w v_t, \quad |a_w| < 1 \,.$$

based on the following moment conditions: $\mu_2(\theta) = \mathsf{E}(u_t^2) = r_y^2 \exp[(1/2) r_w^2/(1 - a_w^2)]$, $\mu_4(\theta) = \mathsf{E}(u_t^4) = 3 r_y^4 \exp[2 r_w^2/(1 - a_w^2)]$, $\mu_{2,2}(1|\theta) = \mathsf{E}[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1 - a_w)]$. Testing for homoskedasticity ($a_w = r_w = 0$) in this model can be written $\psi(\theta) = 0$ with $\psi(\theta) = (a_w, r_w)'$; there are two restrictions, and the derivative matrix of the restrictions

$$P(\theta) = \frac{\partial \psi}{\partial \theta'} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

has full rank two, so it is regular. However, the Jacobian of the moment conditions does not have full rank when evaluated at a point that satisfies the null hypothesis: it is easily shown that

$$\frac{\partial \mu}{\partial \theta'} = \begin{bmatrix} 0 & 0 & 2r_y \\ 0 & 0 & 12r_y^3 \\ 0 & 0 & 4r_y^3 \end{bmatrix} \tag{3.3}$$

when $a_w = r_w = 0,$ so that the Jacobian $\partial \mu / \partial \theta'$ has at most rank one (instead of three in the full-rank case). But GMM identification requires a full-rank Jacobian; see Newey and McFadden (1994, p. 2127).

Thus, $\partial \mu / \partial \theta'$ typically has full rank when it is evaluated at a point that does not satisfy the null hypothesis, for example at an unrestricted point estimate of $\theta,$ as in Wald-type statistics. Therefore, the rank of $\partial \mu / \partial \theta',$ when evaluated at an unrestricted point estimate of θ, generally exceeds the rank of $\partial \mu / \partial \theta'$ evaluated at the true θ when $a_w = r_w = 0$ holds. This again violates the standard regularity condition entailing a non-regular asymptotic distribution for the Wald statistic.

3.3. Tests on residual autocorrelations in time series models

In time series analysis, the adequacy of a model often involves testing the absence of serial autocorrelations in the residuals of autoregressive models. It is well known that such statistics are often singular normal distributions, *e.g.,* Box and Pierce (1970), Li and McLeod (1981), Ljung (1986). For instance as pointed out by Duchesne and Francq (2015), one might want to test whether the first q residual autocorrelations of an $AR(q_0)$ model, with $q_0 < q$ with i.i.d stationary normally distributed observations are zero. Thus, under the null hypothesis $H_0 : \mu_0 = (\rho(1), \ldots, \rho(q))' = 0_q$,

$$\sqrt{n}\left(\hat{\rho}_n^{(q)} - \mu_0\right) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \mathrm{N}_q(0_q, \boldsymbol{\Sigma})$$

where $\hat{\rho}_n^{(q)} \equiv (\hat{\rho}_n(1), \ldots, \hat{\rho}(q))'$ denotes the first q empirical autocorrelations of the above stationary observations and the $q \times q$ covariance matrix $\boldsymbol{\Sigma} = diag(0'_{q_0}, 1'_{q-q_0})$, where $0'_{q_0}$ and $1'_{q-q_0}$ are vectors of zeros and ones, respectively. Clearly, $\boldsymbol{\Sigma}$ is rank-deficient, *i.e.* less than q.

3.4. Deviation from Normality: the Delta method breaks down

In this section we illustrate a situation where the nonlinear transformation of a Gaussian estimator is no longer Gaussian causing the delta method to break down. Suppose the underlying parameter θ is a $q \times 1$ vector such as

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \mathrm{N}[0, I_q], \tag{3.4}$$

and suppose we want to test a null hypothesis of this form:

$$H_0(\psi_0) : \psi(\theta) = \theta'\theta = 0. \tag{3.5}$$

The data generating process corresponding to (3.4) is:

$$Y = \theta\iota + u, \quad u \sim \mathrm{N}[0, I_q],$$

where Y is $q \times n$, θ is $q \times 1$, ι is $1 \times n$ and u is $q \times n$. Using the multivariate least square estimator, we can write:

$$\hat{\theta}_n = [(\iota\iota')^{-1}\iota \otimes I_q]y = \frac{1}{n}(\iota \otimes I_q)y \tag{3.6}$$

where $y = \text{vec}(Y)$ is $qn \times 1$. Although the initial estimator is Gaussian according to (3.4), and the restrictions are differentiable at the true value of the parameters $\theta = 0$, the estimator of the restrictions are not Gaussian anymore:

$$n\psi(\hat{\theta}_n) = (\sqrt{n}\hat{\theta}_n)'(\sqrt{n}\hat{\theta}_n) \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2(q). \tag{3.7}$$

The weighting matrix used in the quadratic form is:

$$\boldsymbol{\Sigma} = P(\theta)\boldsymbol{\Sigma}_\theta P(\theta)', \quad \boldsymbol{\Sigma}_\theta = I_q, \quad \text{with } P(\theta) = \frac{\partial \psi}{\partial \theta'} = 2\theta', \quad \boldsymbol{\Sigma} = 4\theta'\theta\,. \tag{3.8}$$

Thus, the delta method breaks down because the distribution of the estimator of the restriction is not Gaussian anymore, but belongs to a new family, the χ^2 distribution. A consistent estimator of $\boldsymbol{\Sigma}$ is given by:

$$\boldsymbol{\Sigma}_n = 4\hat{\theta}_n'\hat{\theta}_n.$$

A Wald-type statistic can be built upon those restrictions, that is:

$$W_n = n\psi(\hat{\theta}_n)'\left[n\,\boldsymbol{\Sigma}_n\right]^{-1} n\,\psi(\hat{\theta}_n) = (n\,\hat{\theta}_n'\hat{\theta}_n)\left[n4\hat{\theta}_n'\hat{\theta}_n\right]^{-1}(n\,\hat{\theta}_n'\hat{\theta}_n) = \frac{1}{4}n\,\hat{\theta}_n'\hat{\theta}_n \xrightarrow[n\to\infty]{\mathscr{L}} \frac{1}{4}\,\chi^2(q)\,. \tag{3.9}$$

As a result, the asymptotic distribution of the Wald-type statistic is nonstandard, with a scale factor of $\frac{1}{4}$.

3.5. Tests on the determinant of a matrix

This example comes from Gouriéroux and Jasiak (2009). Suppose one wants to test the invertibility of a $q \times q$ matrix A based on the significance of its determinant, *i.e.*

$$H_0 : \det(A) = 0\,.$$

This null hypothesis is of economic interest as shown in Gouriéroux and Jasiak (2009), since it allows to determine the autoregressive order of a multivariate ARCH model, or the hypothesis of nonpredictability of asset returns in risk premium analysis that hinges on the condition $A = 0$. Suppose a consistent asymptotically Gaussian estimator $\hat{A}_n$ of A is available.

To test the null hypothesis based on the estimated determinant $\det(\hat{A}_n)$, one can deduce its asymptotic distribution from the distribution of $\text{vec}(\hat{A}_n)$ through the delta method, where the $\text{vec}(.)$ operator stacks the columns of the $q \times q$ matrix A into a q^2-dimensional vector. Applying the delta method onto:

$$\sqrt{n}[\text{vec}(\hat{A}_n) - \text{vec}(A)] \xrightarrow[n\to\infty]{\mathscr{L}} \text{N}(0, \Omega)\,,$$

we get:

$$\sqrt{n}[\det(\hat{A}_n) - \det(A)] \xrightarrow[n\to\infty]{\mathscr{L}} \text{N}\Big(0, \text{vec}[\text{cof}(A)]'\Omega\,\text{vec}[\text{cof}(A)]\Big)\,,$$

where $\text{vec}[\text{cof}(A)] = \frac{\partial(\det(A))}{\partial(\text{vec}(A))}$ and $\text{cof}(A)$ denotes the $q \times q$ matrix of the cofactors of A. As long as $\text{vec}[\text{cof}(A)] \neq 0$, *i.e.*, $A \neq 0$, the Wald statistic:

$$W_n = n\det(\hat{A}_n)\Big(\text{vec}[\text{cof}(\hat{A}_n)]'\hat{\Omega}_n\text{vec}[\text{cof}(\hat{A}_n)]\Big)^{-1}\det(\hat{A}_n)\,,$$

where $\hat{\Omega}_n$ is a consistent estimator of Ω, has a standard chi-square distribution. However, when $A = 0$,

$$n^{q/2}\det(\hat{A}_n) \xrightarrow[n\to\infty]{\mathscr{L}} \det(A_\infty)\,.$$

When $q \geq 2$, the convergence rate is $1/(n^{q/2})$ that is greater than $1/\sqrt{n}$ so that the limiting distribution is not Gaussian anymore; it is a determinant transformation (nonlinear) of a multivariate Gaussian distribution. The asymptotic distribution of the Wald test statistic is clearly modified.

4. Regularized inverses

The methodology proposed in this paper is based on replacing the inverse of the parameter covariance matrix (or the generalized inverse) by a regularized inverse. We will now define these. Let $\boldsymbol{\Sigma}$ be a $q \times q$ positive-semidefinite matrix with eigenvalues $\lambda_1(\boldsymbol{\Sigma}) \geq \lambda_2(\boldsymbol{\Sigma}) \geq \cdots \geq \lambda_q(\boldsymbol{\Sigma}) \geq 0$, so $\boldsymbol{\Sigma}$ has the spectral decomposition $\boldsymbol{\Sigma} = V\Lambda V'$, where $\Lambda = diag[\lambda_1, \ldots, \lambda_q]$ where $\lambda_i \equiv \lambda_i(\boldsymbol{\Sigma})$, $i = 1, \ldots, q$, and V is an orthogonal matrix whose columns are eigenvectors of $\boldsymbol{\Sigma}$. Clearly, $\boldsymbol{\Sigma}V = V\Lambda$. Let $m(\lambda_i)$ denote the multiplicity of the eigenvalue λ_i, *i.e.* exactly $m(\lambda_i)$ eigenvalues are equal to λ_i (including itself). Although the matrix Λ is uniquely defined, the eigenvector matrix V is not uniquely defined when there is an eigenvalue with multiplicity larger than one $[m(\lambda_i) > 1]$, for the eigenvectors associated with an eigenvalue with multiplicity $m(\lambda_i) > 1$ are uniquely defined only up to post-multiplication by an $m(\lambda_i) \times m(\lambda_i)$ orthogonal matrix. Similarly, let $\boldsymbol{\Sigma}_n$ be a consistent estimator of $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1(\boldsymbol{\Sigma}_n) \geq \lambda_2(\boldsymbol{\Sigma}_n) \geq \cdots \geq \lambda_q(\boldsymbol{\Sigma}_n)$ and $\Lambda_n = diag\big[\lambda_1(\boldsymbol{\Sigma}_n), \ldots, \lambda_q(\boldsymbol{\Sigma}_n)\big]$, so that $\boldsymbol{\Sigma}_n = V_n\Lambda_n V_n'$ where V_n is an orthogonal matrix of eigenvectors.

If $rank(\boldsymbol{\Sigma}_n) = rank(\boldsymbol{\Sigma}) = q$ a.s., *i.e.* if both matrices are a.s. nonsingular, the inverses $\boldsymbol{\Sigma}^{-1} = V\Lambda^{-1}V'$ and $\boldsymbol{\Sigma}_n^{-1} = V_n\Lambda_n^{-1}V_n'$ are a.s. well defined. However, if $\text{rank}(\boldsymbol{\Sigma}) < q$ and/or $\text{rank}(\boldsymbol{\Sigma}_n) \leq q$, we need to make adjustments. For this, we consider a *regularized* inverse as follows. where $\mathbb{R}_0^+$ represents the nonnegative real numbers.

Definition 4.1 REGULARIZED INVERSE. *Let $\boldsymbol{\Sigma}$ be a $q \times q$ real symmetric positive semidefinite matrix admitting the spectral decomposition $\boldsymbol{\Sigma} = V\Lambda V'$ where $\Lambda = diag[\lambda_1(\boldsymbol{\Sigma}), \ldots, \lambda_q(\boldsymbol{\Sigma})]$, $c \in \mathbb{R}_0^+$, $g(\lambda; c)$ a real-valued nonnegative function of λ and c such that $g(\lambda; c)$ is bounded and non-increasing in λ, and $\Lambda^\dagger(c) \equiv \Lambda^\dagger(\boldsymbol{\Sigma}; c) \equiv diag\big[g(\lambda_1(\boldsymbol{\Sigma}); c), \ldots, g(\lambda_q(\boldsymbol{\Sigma}); c)\big]$.Then a matrix of the form $\boldsymbol{\Sigma}^R(c) = V\Lambda^\dagger(c)V'$ is called a* regularized inverse *of $\boldsymbol{\Sigma}$.*

We will call $g(\lambda; c)$ the *variance regularization function* (VRF). VRF perturbs the small eigenvalues in order to stabilize their inverse, preventing them from exploding. For $c > 0$, we denote $q(\boldsymbol{\Sigma}, c)$ the number of eigenvalues $\lambda_i(\boldsymbol{\Sigma})$ such that $\lambda_i(\boldsymbol{\Sigma}) > c$, and $q(\boldsymbol{\Sigma}_n, c)$ the number of eigenvalues $\lambda_i(\boldsymbol{\Sigma}_n)$ such that $\lambda_i(\boldsymbol{\Sigma}_n) > c$.

****

We now introduce a partition of the matrix $\Lambda^\dagger(c)$ into three submatrices where c represents a threshold which may depend on the sample, *i.e.* $c = c(n, Y_n)$:

$$\Lambda^\dagger(c) = \begin{pmatrix} \Lambda_1^\dagger[\bar{\lambda}; c] & 0 & 0 \\ 0 & \Lambda_2^\dagger[\bar{\lambda}; c] & 0 \\ 0 & 0 & \Lambda_3^\dagger[\bar{\lambda}; c] \end{pmatrix}. \tag{4.1}$$

Let $q_i = \dim(\Lambda_i^\dagger[\bar{\lambda}; c])$, for $i = 1, 2, 3$, with $q_1 = q(\boldsymbol{\Sigma}, c)$, $q_2 = m(c)$ and $q_3 = q - q_1 - q_2$. $m(c)$ denotes the multiplicity of the eigenvalue $\lambda = c$ (if any). The three components correspond to $\Lambda_1^\dagger[\bar{\lambda}; c] = diag\big[g(\lambda_1; c), \ldots, g(\lambda_{q_1}; c)\big]$ for $\lambda > c$, $\Lambda_2^\dagger[\bar{\lambda}; c] = g(c; c)I_{q_2}$ for $\lambda = c$,

$\Lambda_3^{\dagger}[\bar{\lambda};c] = diag\big[g(\lambda_{q_1+q_2+1};c),\ldots,g(\lambda_q;c)\big]$ for $\lambda < c$. More specifically, the large eigenvalues that fall above the threshold c may remain unchanged whereas those equal to or smaller than the threshold are (possibly) modified to stabilize their inverse. Thus, the first component is "regular" and remains unmodified (except for a ridge-type inverse), while the others may not be "regular". In particular, the third component requires a regularization. Though our variance regularization function (VRF) does nest several regularization schemes (*i.e.*, variants around the Tikhonov, spectral cut-off, ridge regularization schemes), the one we advocate in the paper and defined later produces a novel full-rank regularized inverse; we shall call it the Dufour-Valery (DV) scheme. Indeed, because of the invertibility difficulties raised from small values of λ, we shall replace the latter with eigenvalues bounded away from zero, *i.e.,* $\frac{1}{c}$; this yields the Dufour-Valery *Dufour-Valery full-rank* regularized matrix. By contrast, the modified Moore Penrose inverse (or spectral cut-off type) sets to zero all small problematic eigenvalues, *i.e.* $\Lambda_2^{\dagger}[\bar{\lambda};c] = \Lambda_3^{\dagger}[\bar{\lambda};c] = 0$, yielding a *reduced-rank* matrix.

Let V_1 be a $q \times q_1$ matrix whose columns are the eigenvectors associated with the eigenvalues $\lambda > c$ arranged in the same order as the eigenvalues. The eigenvectors associated with $\lambda > c$ form a basis for the eigenspace corresponding to λ. If $m(\lambda) = 1$, these eigenvectors are uniquely defined, otherwise not. The same holds for the $q \times q_2$ matrix V_2 whose columns are the eigenvectors associated with the eigenvalues $\lambda = c$ and for the $q \times q_3$ matrix V_3 whose columns are the eigenvectors associated with the eigenvalues $\lambda < c$. $\Lambda_1^{\dagger}[\lambda(\boldsymbol{\Sigma}_n);c]$, $\Lambda_2^{\dagger}[\lambda(\boldsymbol{\Sigma}_n);c]$,$\Lambda_3^{\dagger}[\lambda(\boldsymbol{\Sigma}_n);c]$, V_{1n}, V_{2n} and V_{3n} denote the corresponding quantities based on the sample analog $\boldsymbol{\Sigma}_n$, with $\dim(\Lambda_1[\lambda(\boldsymbol{\Sigma}_n);c]) = \hat{q}_1 = \text{card}\{i \in I : \lambda_i(\boldsymbol{\Sigma}_n) > c\}$, $\dim(\Lambda_2[\lambda(\boldsymbol{\Sigma}_n);c]) = \hat{q}_2 = \text{card}\{i \in I : \lambda_i(\boldsymbol{\Sigma}_n) = c\}$, $\dim(\Lambda_3[\lambda(\boldsymbol{\Sigma}_n);c]) = \hat{q}_3 = \text{card}\{i \in I : \lambda_i(\boldsymbol{\Sigma}_n) < c\}$, respectively. Using (4.1), the *regularized* inverse can be decomposed as follows:

$$\boldsymbol{\Sigma}^R(c) = V\Lambda^{\dagger}(c)V' = [V_1\ V_2\ V_3]\begin{pmatrix}\Lambda_1^{\dagger}[\bar{\lambda};c] & 0 & 0\\ 0 & \Lambda_2^{\dagger}[\bar{\lambda};c] & 0\\ 0 & 0 & \Lambda_3^{\dagger}[\bar{\lambda};c]\end{pmatrix}\begin{bmatrix}V_1'\\ V_2'\\ V_3'\end{bmatrix} = \sum_{i=1}^{3}\boldsymbol{\Sigma}_{ii}^R(c) \tag{4.2}$$

where $\boldsymbol{\Sigma}_{ii}^R(c) = V_i\Lambda_i^{\dagger}(c)V_i'$ $i = 1,2,3$ and $\Lambda_i^{\dagger}(c) = \Lambda_i^{\dagger}[\bar{\lambda};c]$ for the sake of notational simplicity. Likewise, $\boldsymbol{\Sigma}$ can be decomposed as:

$$\boldsymbol{\Sigma} = V\Lambda V' = \sum_{i=1}^{3}\boldsymbol{\Sigma}_{ii}(c) = \sum_{i=1}^{3}V_i\Lambda_i(c)V_i' \tag{4.3}$$

where $\boldsymbol{\Sigma}_{ii}(c) = V_i\Lambda_i(c)V_i'$; $\Lambda_1(c) = diag(\lambda)_{\lambda>c}$, $\Lambda_2(c) = diag(\lambda)_{\lambda=c}$ and $\Lambda_3(c) = diag(\lambda)_{\lambda<c}$. In the absence of zero eigenvalues, the usual inverse can be computed as $\boldsymbol{\Sigma}^{-1} = V\Lambda^{-1}V' = \sum_{i=1}^{3}\boldsymbol{\Sigma}_{ii}^{-1}(c) = \sum_{i=1}^{3}V_i\Lambda_i^{-1}(c)V_i'$. Let us establish some useful properties for the regularized inverses, with I_q denoting a conformable identity matrix.

Proposition 4.1 PROPERTY OF THE REGULARIZED INVERSES. *Let $\boldsymbol{\Sigma} = V\Lambda V'$ be a positive semidefinite matrix, such that $\lambda_1 \geq \cdots \geq \lambda_q \geq 0$. Let $\lambda g(\lambda;c) \leq 1\ \forall\ \lambda$. Then, the regularized inverse $\boldsymbol{\Sigma}^R(c)$ of $\boldsymbol{\Sigma}$, defined in 4.1, satisfies the following relations.*

i) $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^R(c) = \boldsymbol{\Sigma}^R(c)\boldsymbol{\Sigma} \leq I_q$;

ii) $T\boldsymbol{\Sigma}^R(c)T' \leq I_q$, *where $T = V\Lambda^{1/2}V'$ is the square root of $\boldsymbol{\Sigma}$;*

iii) $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^R(c)\boldsymbol{\Sigma} \leq \boldsymbol{\Sigma}$;

iv) if $g(\lambda;c) > 0$, then $\big(\boldsymbol{\Sigma}^R(c)\big)^{-1} \geq \boldsymbol{\Sigma}$;

v) if $\lambda > 0$ then $g(\lambda;c) > 0$ and rank$\big(\boldsymbol{\Sigma}^R(c)\big) \geq$rank$(\boldsymbol{\Sigma})$.

It is important to notice that any transformation of the original matrix $\boldsymbol{\Sigma}$ that diminishes the inverse $\boldsymbol{\Sigma}^R(c)$ satisfies relation *iv*). Note that the generalized inverses usually denoted by $\boldsymbol{\Sigma}^-$ share properties *i*) and *iii*) with the *regularized* inverses. By contrast, property *iii*) appears as a dominance relation for the *regularized* inverse as opposed to *g*-inverses for which $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^-\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$. Result *v*) is well known for *g*-inverses and is related to the generalized inverse with maximal rank. See Rao and Mitra (1971, Lemmas 2.2.1 and 2.2.3 page 20-21)] for results *iii*) and *v*) regarding *g*-inverses. Finally, note that *ii*) is another way of formulating *i*), and can be useful for sandwich estimators.

5. Regularized Wald statistic

In this section, we introduce the concept of regularized test statistic which embeds three possible cases. *Case 1* corresponds to the regular setup where the estimator of the covariance matrix converges to a full-rank fixed matrix. In this case, regularizing is useless and (relatively) innocuous, although the asymptotic distribution is modified in the fixed threshold case. *Case 2* corresponds to a sample covariance matrix that converges to a singular limiting matrix but satisfies Andrews's rank condition. In such a case, the limiting distribution is modified only through an adjustment of the degree of freedom when the threshold decays to zero with the sample size. Finally *case 3* makes use of a sample covariance matrix which violates the typical rank condition. Also, the regularized weighting matrix converges to an object that differs from the original population matrix. This yields a valid test but at the cost of a *modified* asymptotic distribution with a fixed threshold.

Based on decomposition (4.3), the original Wald statistic W_n defined in equation (2.1) enjoys the following decomposition:

$$W_n = W_{1n}(c) + W_{2n}(c) + W_{3n}(c) \,, \tag{5.1}$$

$W_{in}(c) = \big[H_n(\hat{\psi}_n - \psi_0)\big]' \boldsymbol{\Sigma}_{ii,n}^{-1}(c) \big[H_n(\hat{\psi}_n - \psi_0)\big]$, with $\boldsymbol{\Sigma}_{ii,n}^{-1}(c) = V_{in}\Lambda_{in}^{-1}(c)V_{in}'$ for $i = 1,2,3$, and $\Lambda_{in}^{-1}(c) = \Lambda_i^{-1}[\lambda(\boldsymbol{\Sigma}_n); c]$. For $i = 2,3$, $W_{in}(c) = 0$, eventually. Note that decomposition (4.3) results in three independent random variables. When Andrews's rank condition does not hold, the Wald test statistic has to be *regularized* to account for such irregularities as introduced next.

Definition 5.1 DEFINITION OF THE REGULARIZED WALD STATISTIC. *The regularized Wald statistic is* $W_n^R(c) = X_n' \boldsymbol{\Sigma}_n^R(c) X_n = \big[H_n(\hat{\psi}_n - \psi_0)\big]' \boldsymbol{\Sigma}_n^R(c) \big[H_n(\hat{\psi}_n - \psi_0)\big]$, *where* $\boldsymbol{\Sigma}_n^R(c)$ *satisfies decomposition* (4.2).

Using decomposition (4.2), the *regularized* Wald statistic can be decomposed as follows:

$$\begin{aligned} W_n^R(c) = X_n' \boldsymbol{\Sigma}_n^R(c) X_n &= \big[H_n(\hat{\psi}_n - \psi_0)\big]' \boldsymbol{\Sigma}_n^R(c) \big[H_n(\hat{\psi}_n - \psi_0)\big] = \big[H_n(\hat{\psi}_n - \psi_0)\big]' \sum_{i=1}^{3} \boldsymbol{\Sigma}_{ii,n}^R(c) \big[H_n(\hat{\psi}_n - \psi_0)\big] \\ &= W_{1n}^R(c) + W_{2n}^R(c) + W_{3n}^R(c) \,, \end{aligned} \tag{5.2}$$

where $W_{in}^R(c) = \big[H_n(\hat{\psi}_n - \psi_0)\big]' \boldsymbol{\Sigma}_{ii,n}^R(c) \big[H_n(\hat{\psi}_n - \psi_0)\big]$; $\boldsymbol{\Sigma}_{ii,n}^R(c) = V_{in}\Lambda_{in}^{\dagger}(c)V_{in}'$ for $i = 1,2,3$, denotes the sample analog of the elements in decomposition (4.2).

By partitioning the inverse of the eigenvalue matrix $\Lambda^{\dagger}(c)$ into three blocks, $\Lambda_1^{\dagger}(c)$ for $\lambda > c$, $\Lambda_2^{\dagger}(c)$ for $\lambda = c$ and $\Lambda_3^{\dagger}(c)$ for $\lambda < c$, we have identified a convenient decomposition for the statistic into three components: the first component builds on the "large" eigenvalues that remain unchanged; the second component gathers the eigenvalues exactly equal to the threshold c (if any), while the third incorporates the small modified eigenvalues. This decomposition sheds light on the structure of the distribution of the *regularized* test statistic. By contrast, Lütkepohl and Burda (1997) only keep the eigenvalues greater than the threshold c, which cancels out the last two components, *i.e.* $W_{2n}^R(c) = W_{3n}^R(c) = 0$. Thus discarding the small eigenvalues might reduce information. However, as Lütkepohl and Burda (1997) use a χ^2 distribution with fewer degrees of freedom, a deeper investigation is required to gauge power. More importantly, in finite samples it will be

difficult to disentangle the estimates that really correspond to $\lambda = c$ from those close to but distinct from c. This complicates the asymptotic distribution and the estimation procedure. Note that $W_{1n}(c) = W_{1n}^R(c)$ for this is the regular component common to both statistics. Moreover, when there is no eigenvalues exactly equal to c, $m(c) = 0$, and the second component vanishes.

6. Eigenprojections

6.1. Discontinuities of eigenvectors: an illustration

We now discuss some non-uniqueness and discontinuity issues regarding the eigenvectors of a given matrix. It is well-known in spectral theory that eigenvectors corresponding to multiple eigenvalues are not uniquely defined (only up to the post multiplication by an $m(\lambda) \times m(\lambda)$ orthogonal matrix with $m(\lambda)$ indicating the multiplicity of the eigenvalue). However, econometricians are not always aware of such technical details that could jeopardize asymptotic results. Further, whereas eigenvalues are generally known to be continuous functions in the elements of the matrix, eigenvectors not. The main pitfall consists of deriving convergence results for the estimates of the eigenvectors based on the consistency of the sample matrix; this critically hinges on the continuity assumption of eigenvectors (w.r.t. the elements of the matrix). Even in the deterministic case, eigenvectors are not necessarily continuous functions of the elements of the matrix. We illustrate such a discontinuity in a simple counter-example.[7]

Example 6.1 Let $A(x)$ be the matrix function defined as:

$$A(x) = 1_{\{x \geq 0\}} \begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix} + 1_{\{x<0\}} \begin{bmatrix} 1+x & 0 \\ 0 & 1-x \end{bmatrix} \tag{6.1}$$

where $1_A = 1$ if $x \in A$ and $1_A = 0$ otherwise. This matrix function is clearly continuous at $x = 0$, with $A(0) = I_2$. However, for $x < 0$, the spectral decomposition of $A(x)$ is:

$$A(x) = (1+x) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + (1-x) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{6.2}$$

with $(1+x)$ and $(1-x)$ being the eigenvalues and $(1,0)'$ and $(0,1)'$ the eigenvectors, $P(1+x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $P(1-x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ the eigenprojections, while for $x > 0$, it is

$$A(x) = \frac{1}{\sqrt{2}}(1+x) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{1}{\sqrt{2}}(1-x) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{6.3}$$

with $(1+x)$ and $(1-x)$ being the eigenvalues and $\frac{1}{\sqrt{2}}(1,1)'$ and $\frac{1}{\sqrt{2}}(1,-1)'$ the eigenvectors, $P(1+x) = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $P(1-x) = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ the eigenprojections. Clearly, the eigenvalues $(1+x)$ and $(1-x)$ are continuous at $x = 0$ whereas the eigenvectors and the eigenprojections are not the same whether $x \to 0^+$ or $x \to 0^-$.

Being unaware of this caveat may lead to *wrong* distributional results by mistakenly applying the continuous mapping theorem to objects that are *not* continuous. Nevertheless, there exist functions of eigenvectors

[7] We are grateful to Russell Davidson for this example.

that are continuous w.r.t. the elements of the matrix. Specifically, for an eigenvalue λ, the projection matrix $P(\lambda)$ that projects onto the space spanned by the eigenvectors associated with λ - the *eigenspace* $V(\lambda)$ - may still be not continuous in the elements of the matrix, but the *total* eigenprojection that corresponds to the sum of the individual eigenprojections over a subset of the spectral set, will share the nice continuity property with the eigenvalues. For further discussion of this important property, see Rellich (1953), Kato (1966) and Tyler (1981).

6.2. Continuity properties of eigenvalues and total eigenprojections

In order to derive the asymptotic distribution of the regularized test statistics, it will be useful to review and adapt some results on spectral theory used in Tyler (1981). Let $\mathscr{S}(\boldsymbol{\Sigma})$ denote the spectral set of $\boldsymbol{\Sigma}$, *i.e.* the set of all eigenvalues of $\boldsymbol{\Sigma}$. The *eigenspace* of $\boldsymbol{\Sigma}$ associated with λ is defined as all the linear combinations from a basis of eigenvectors $\mathbf{x}_i$, $i = 1, \ldots, m(\lambda)$, *i.e.*

$$V(\lambda) = \{\mathbf{x}_i \in \mathbb{R}^q | \boldsymbol{\Sigma}\mathbf{x}_i = \lambda \mathbf{x}_i\} \,. \tag{6.4}$$

Clearly, $\dim V(\lambda) = m(\lambda)$. Since $\boldsymbol{\Sigma}$ is a $q \times q$ matrix symmetric in the metric of a real positive definite symmetric matrix $\mathbf{T}$, *i.e.* $\mathbf{T}\boldsymbol{\Sigma}$ is symmetric [see Tyler (1981, p.725)], we have:

$$\mathbb{R}^q = \sum_{\lambda \in \mathscr{S}(\boldsymbol{\Sigma})} V(\lambda) \,. \tag{6.5}$$

The *eigenprojection* of $\boldsymbol{\Sigma}$ associated with λ, denoted $P(\lambda)$, is the projection operator onto $V(\lambda)$ w.r.t. decomposition (6.5) of $\mathbb{R}^q$. For any set of vectors $\mathbf{x}_i$ in $V(\lambda)$ such that $\mathbf{x}_i'\mathbf{T}\mathbf{x}_j = \delta_{ij}$, where δ_{ij} denotes the Kronecker' s delta, $P(\lambda)$ has the representation

$$P(\lambda) = \sum_{j=1}^{m(\lambda)} \mathbf{x}_j \mathbf{x}_j' \mathbf{T} \,. \tag{6.6}$$

$P(\lambda)$ is symmetric in the metric of $\mathbf{T}$. This yields

$$\boldsymbol{\Sigma} = \sum_{\lambda \in \mathscr{S}(\boldsymbol{\Sigma})} \lambda P(\lambda) \,, \qquad \boldsymbol{\Sigma}_n = \sum_{\lambda(\boldsymbol{\Sigma}_n) \in \mathscr{S}(\boldsymbol{\Sigma}_n)} \lambda(\boldsymbol{\Sigma}_n) P[\lambda(\boldsymbol{\Sigma}_n)] \,. \tag{6.7}$$

If v is any subset of the spectral set $\mathscr{S}(\boldsymbol{\Sigma})$, then the *total* eigenprojection for $\boldsymbol{\Sigma}$ associated with the eigenvalues in v is defined to be $\sum_{\lambda \in v} P(\lambda)$. Below we report a lemma given by Tyler (1981, Lemma 2.1, p. 726) that states an important continuity property for eigenvalues and *total* eigenprojections (*i.e.*, the sum of the eigenprojections over a subset of the spectral set) on eigenspaces for non-random symmetric matrices from which consistency of sample regularized inverses will follow.

Lemma 6.1 CONTINUITY OF EIGENVALUES AND EIGENPROJECTIONS. *Let $\boldsymbol{\Sigma}_n$ be a $q \times q$ real matrix symmetric in the metric of a real positive definite symmetric matrix T_n with eigenvalues $\lambda_1(\boldsymbol{\Sigma}_n) \geq \lambda_2(\boldsymbol{\Sigma}_n) \geq \cdots \geq \lambda_q(\boldsymbol{\Sigma}_n)$. Let $P_{k,t}(\boldsymbol{\Sigma}_n)$ represent the total eigenprojection for $\boldsymbol{\Sigma}_n$ associated with $\lambda_k(\boldsymbol{\Sigma}_n) \ldots \lambda_t(\boldsymbol{\Sigma}_n)$ for $t \geq k$. If $\boldsymbol{\Sigma}_n \to \boldsymbol{\Sigma}$ as $n \to \infty$, then:*

i) $\lambda_k(\boldsymbol{\Sigma}_n) \to \lambda_k(\boldsymbol{\Sigma})$*;*

ii) $P_{k,t}(\boldsymbol{\Sigma}_n) \to P_{k,t}(\boldsymbol{\Sigma})$ *provided* $\lambda_{k-1}(\boldsymbol{\Sigma}) \neq \lambda_k(\boldsymbol{\Sigma})$ *and* $\lambda_t(\boldsymbol{\Sigma}) \neq \lambda_{t+1}(\boldsymbol{\Sigma})$.

This lemma tells us that the eigenvalues are continuous functions in the elements of the matrix. The same continuity property holds for the *total* eigenprojection. Thus, no matter what the multiplicity of the

eigenvalues, this continuity property holds for the *total* eigenprojection $P_{k,t}(\boldsymbol{\Sigma})$ provided that one can find one eigenvalue before and one after that are distinct. It will be useful to extend Lemma 6.1 to random symmetric matrices. To the best of our knowledge, these results are not explicitly stated elsewhere.

Lemma 6.2 CONTINUITY OF EIGENVALUES AND EIGENPROJECTIONS: ALMOST SURE CONVERGENCE. *Let $\boldsymbol{\Sigma}_n$ be a $q \times q$ real random matrix symmetric in the metric of a real positive definite symmetric random matrix T_n and with eigenvalues $\lambda_1(\boldsymbol{\Sigma}_n) \geq \lambda_2(\boldsymbol{\Sigma}_n) \geq \cdots \geq \lambda_q(\boldsymbol{\Sigma}_n)$. Let $P_{k,t}(\boldsymbol{\Sigma}_n)$ represent the total eigenprojection for $\boldsymbol{\Sigma}_n$ associated with $\lambda_k(\boldsymbol{\Sigma}_n) \ldots \lambda_t(\boldsymbol{\Sigma}_n)$ for $t \geq k$. If $\boldsymbol{\Sigma}_n \xrightarrow{a.s.} \boldsymbol{\Sigma}$ as $n \to \infty$, then:*

i) $\lambda_k(\boldsymbol{\Sigma}_n) \xrightarrow{a.s.} \lambda_k(\boldsymbol{\Sigma})$;

ii) $P_{k,t}(\boldsymbol{\Sigma}_n) \xrightarrow{a.s.} P_{k,t}(\boldsymbol{\Sigma})$ *provided* $\lambda_{k-1}(\boldsymbol{\Sigma}) \neq \lambda_k(\boldsymbol{\Sigma})$ *and* $\lambda_t(\boldsymbol{\Sigma}) \neq \lambda_{t+1}(\boldsymbol{\Sigma})$.

We can now show that the continuity property of the eigenvalues and eigenprojections established in the a.s. case, remain valid in the case of convergence in probability .

Lemma 6.3 CONTINUITY OF EIGENVALUES AND EIGENPROJECTIONS: CONVERGENCE IN PROBABILITY. *Let $\boldsymbol{\Sigma}_n$ be a $q \times q$ real random matrix symmetric in the metric of a real positive definite symmetric random matrix T_n with eigenvalues $\lambda_1(\boldsymbol{\Sigma}_n) \geq \lambda_2(\boldsymbol{\Sigma}_n) \geq \cdots \geq \lambda_q(\boldsymbol{\Sigma}_n)$. Let $P_{k,t}(\boldsymbol{\Sigma}_n)$ represent the total eigenprojection for $\boldsymbol{\Sigma}_n$ associated with $\lambda_k(\boldsymbol{\Sigma}_n), \ldots, \lambda_t(\boldsymbol{\Sigma}_n)$ for $t \geq k$. If $\boldsymbol{\Sigma}_n \xrightarrow{p} \boldsymbol{\Sigma}$ as $n \to \infty$, then:*

i) $\lambda_k(\boldsymbol{\Sigma}_n) \xrightarrow{p} \lambda_k(\boldsymbol{\Sigma})$;

ii) $P_{k,t}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_{k,t}(\boldsymbol{\Sigma})$ *provided* $\lambda_{k-1}(\boldsymbol{\Sigma}) \neq \lambda_k(\boldsymbol{\Sigma})$ *and* $\lambda_t(\boldsymbol{\Sigma}) \neq \lambda_{t+1}(\boldsymbol{\Sigma})$.

7. Asymptotic properties of the regularized inverse

In this section, we derive asymptotic results for the *regularized* inverse that hold for a general variance regularization function (VRF) family. More specifically, in Subsection 7.1, we introduce a family of general variance regularization functions that exploits a threshold. This VRF family is general as it embeds both cases, continuous VRFs (see equation (7.3)), or discontinuous VRFs (see equation (7.2) with $c > 0$). Such a regularization approach based on a cut-off point to disentangle large eigenvalues from small eigenvalues enables us to recover an important strand of the statistical literature that estimates the rank of a matrix; see Gill and Lewbel (1992), Cragg and Donald (1996, 1997) , Robin and Smith (2000) and others. In the same vein, the approach introduced by Lütkepohl and Burda (1997) yields a modified reduced-rank estimator for the covariance matrix; we generalize it to non-Gaussian estimators in the presence of possible multiple eigenvalues.Lütkepohl and Burda (1997) propose to reduce the rank of the matrix estimator to satisfy Andrews's rank condition. The asymptotic rank is meaningful, especially if one wants to recover the asymptotic chi-square distribution for the test statistic. Basically, we wanted to be ecumenical by allowing all rank possibilities, from reduced ranks to full ranks. Besides, the threshold method is attractive because it leads to a genuine bound for the nonstandard distribution. Finally, Subsection 7.2 reviews well-known continuous regularization schemes extensively used in ill-posed inverse problems. Such continuous VRFs do not make use of a threshold, hence the resulting distributional theory is easier. Those regularization tools can be cast into the $\mathscr{G}_c$ VRF family for a specific choice of the threshold. See Carrasco, Florens and Renault (2007) for a comprehensive review on regularization tools in ill-posed inverse problems in structural econometrics.

7.1. The family of admissible Variance Regularization Function (VRF)

We now define the VRF family, and provide a few examples.

Definition 7.1 THE FAMILY OF ADMISSIBLE VRF. *$\mathcal{G}_c$ is the class of admissible scalar VRFs, such as for a real scalar $c \geq 0$:*

$$\begin{array}{rccc} g(.,c): & \mathbb{R}_+ & \rightarrow & \mathbb{R}_+ \\ & \lambda & \rightarrow & g(\lambda;c) \end{array}$$

$g(\lambda;c)$ is continuous almost everywhere (a.e.) w.r.t. λ, except possibly at $\lambda = c$, (w.r.t. the Lebesgue measure); g is a function that takes bounded values everywhere; g is non-increasing in λ; $\lim_{c \to 0^+} g(\lambda;c) = g(\lambda;0)$

Note that we allow a discontinuity at $\lambda = c$ to precisely embed a spectral cut-off type regularization such as a modified Moore-Penrose inverse that is clearly *not* continuous around $\lambda = c$ for $c > 0$, see (7.2). Some possible choices for the VRF could be:

$$g(\lambda;c) = \begin{cases} \frac{1}{\lambda+\varepsilon_1} & \text{if } \lambda > c \\ \frac{1}{\varepsilon_2+\gamma(c-\lambda)} & \text{if } \lambda \leq c \end{cases} \tag{7.1}$$

with $\gamma \geq 0$. This VRF can be viewed as a *modified* Hodges' estimator applied to the eigenvalues. See Hodges and Lehmann (1950), LeCam (1953). Interesting special cases include:

1. $\varepsilon_1 = 0$, $\gamma = \infty$, $c \geq 0$, hence
$$g(\lambda;c) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > c \\ 0 & \text{if } \lambda \leq c \end{cases} \tag{7.2}$$
and therefore $\Lambda^{\dagger}(c) = \Lambda^{+}(c)$, where
$$\Lambda^{+}(c) = diag[1/\lambda_1 I(\lambda_1 > c), \ldots, 1/\lambda_{q_1} I(\lambda_{q_1} > c), 0, \ldots, 0\,]$$
corresponds to a spectral cut-off regularization scheme [see Carrasco (2012), Carrasco, Florens and Renault (2007) and the references therein]; $I(s)$ is equal to 1 if the relation s is satisfied. In particular, $\Lambda^{+}(c)$ is a *modified version* of the Moore-Penrose inverse of
$$\Lambda = diag[\lambda_1 I(\lambda_1 > 0), ..., \lambda_{q_1} I(\lambda_{q_1} > 0), \lambda_{q_1+1} I(\lambda_{q_1+1} > 0), \ldots, \lambda_q I(\lambda_q > 0)]$$
used by Lütkepohl and Burda (1997). We also consider the case where some eigenvalues may be smaller than the threshold c, with $c \neq 0$.

2. $\varepsilon_1 = 0$, $\gamma = 0$ and $\varepsilon_2 = c$, with $c \neq 0$, hence
$$g(\lambda;c) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > c \\ \frac{1}{c} & \text{if } \lambda \leq c\,. \end{cases} \tag{7.3}$$
We shall call this regularization scheme, the Dufour-Valery (DV) regularization scheme, which will define the full-rank regularized inverse. We advocate this method in the paper and use it in the simulations. This method is more refined than the following ridge-type one.

3. $\gamma = 0$ and $\varepsilon_2 = \lambda + \varepsilon_1$, with $\varepsilon_1 \geq 0$ and $c \geq 0$, hence

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda+\varepsilon_1} & \text{if } \lambda > c \\ \frac{1}{\lambda+\varepsilon_1} & \text{if } \lambda \leq c\,. \end{cases} \tag{7.4}$$

This regularization corresponds to a uniform (ridge-type) regularization (see *e.g.,* Hoerl and Kennard (1970)); uniformity means that all eigenvalues are being modified regardlessly. Thus, this regularization scheme is somewhat crude because it does not distinguish the large eigenvalues from the small ones. Usually this type of regularization is applied to the estimator of the regression coefficients in ridge regressions, not to covariance matrices, *i.e.,* $\boldsymbol{\Sigma}^R(c) = (\boldsymbol{\Sigma} + \varepsilon_1 Id)^{-1}$. Note that this case embeds the standard inverse when $\varepsilon_1 = 0$. This ridge-type regularization of the covariance matrix corresponds to a non-optimal linear shrinkage estimator of the covariance matrix; see Ledoit and Wolf (2004b) for optimal linear shrinkage estimators of high-dimensional covariance matrices.[8]

4. $\varepsilon_1 = 0$, $\gamma > 0$ with $\gamma = \frac{\alpha}{\lambda(c-\lambda)}$, $\alpha > 0$, and $\varepsilon_2 = \lambda$, with $c \neq 0$, hence

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > c \\ \frac{\lambda}{\lambda^2+\alpha} & \text{if } \lambda \leq c\,, \end{cases} \tag{7.5}$$

which corresponds to a variation around the Tikhonov regularization since $\frac{1}{\lambda+\gamma(c-\lambda)} = \frac{1}{\lambda+\alpha/\lambda} = \frac{\lambda}{\lambda^2+\alpha}$.

Based on the spectral decomposition defined in equation (6.7), we immediately deduce a spectral decomposition for the regularized inverses:

$$\boldsymbol{\Sigma}^R(c) = V\Lambda^{\dagger}(c)V' = \sum_{\lambda \in \mathscr{S}(\boldsymbol{\Sigma})} g(\lambda; c) P(\lambda)\,, \quad \boldsymbol{\Sigma}_n^R(c) = V_n \Lambda_n^{\dagger}(c) V_n' = \sum_{\lambda(\boldsymbol{\Sigma}_n) \in \mathscr{S}(\boldsymbol{\Sigma}_n)} g\big[\lambda(\boldsymbol{\Sigma}_n); c\big] P\big[\lambda(\boldsymbol{\Sigma}_n)\big]\,. \tag{7.6}$$

Thus, the dependence on c of the regularized inverses comes from the VRF $g(\lambda; c)$. The threshold c may be size-dependent, *i.e.*, $g(\lambda, c_n)$. This is a special case of c fixed and will be studied in Section 9.

7.2. The Variance Regularization Functions: the continuous case without threshold

Well-known continuous regularization schemes that do not use any threshold are the Tikhonov regularization and the Landweber Fridman iterative regularization. For readers interested in regularization tools in ill-posed inverse problems in structural econometrics, see Carrasco, Florens and Renault (2007), Carrasco (2012).The Tikhonov regularization scheme is closely related to the ridge regression. In this case, $\bar{g}(\lambda) = \frac{\lambda}{\lambda^2+\alpha}$, $\alpha > 0$.

For the Landweber Fridman iterative regularization scheme, $\bar{g}(\lambda) = \frac{1-\left(1-\gamma\lambda^2\right)^{1/\alpha}\lambda}{\lambda}$, $\gamma > 0,\ \alpha > 0$. This class of VRF that does not make use of a threshold can be recast into the $\mathscr{G}_c$ family by selecting the threshold c such that $c > \lambda_{max}$, where λ_{max} denotes the largest eigenvalue of $\boldsymbol{\Sigma}$, *i.e.* $\bar{g}(\lambda) = g(\lambda; \bar{c})$ with $\bar{c} > \lambda_{max}$.

Without a threshold, the convergence of the regularized inverse is straightforward; it follows from the continuity property of $\bar{g}(\cdot)$ and of the total eigenprojection technique. However, there is a trade-off between

[8] For readers interested in ridge regressions, and more generally in Bridge and Lasso-type regressions see Tibshirani (1996), Knight and Fu (2000), Knight (2008); see also Luo (2010) and Luo and Zuo (2011) for tests in the context of large dimensional regression coefficients with ridge estimators. More specifically, Knight and Fu (2000) derive the asymptotic behavior of Lasso-type estimators in regression settings, including asymptotics for "nearly singular" designs, an alternative terminology for asymptotic singularity. Thus, Knight and Fu (2000), Knight (2008) have tackled the asymptotic singularity problem – or nearly singular design – by working on the null space of the singular matrix on which there exists a positive definite matrix. Under the nearly-singular design, the authors show that collinearity slows down the rate of convergence of Bridge estimators relatively to $\sqrt{n}$. See also Caner (2008) for the sample behavior of (2)LS estimators, GMM estimators under nearly-singular design.

the simplicity of the continuous regularization schemes above - that simplifies the asymptotic theory - and the maintained hypothesis of a chi-square distribution with reduced rank. Indeed, the threshold allows us to disentangle the large eigenvalues from the small problematic ones. This observation enables us to still exploit the chi-square distribution, especially when the rank of the limiting matrix is reduced. Estimating the reduced rank of a matrix is a tricky problem that has drawn much attention in the statistical and econometric literature. Our approach is general as it encompasses the two limiting cases: the reduced rank statistic that still follows a chi-square distribution and the modified full-rank statistic that has a nonstandard distribution (a linear combination of chi-squares). A simple alternative in between is to use the standard chi-square distribution (with full rank) as an upper bound: thus one can use the standard critical point instead. Although the chi-square upper bound is conservative, it enjoys good power properties as shown later on in the simulations.

7.3. Asymptotic properties of the regularized inverse when c is fixed

Because the random objects considered here are matrices, we must choose a norm suitable to matrices. For this reason, we consider the finite dimensional inner product space $(\mathscr{S}_q, <\cdot,\cdot>)$, where $\mathscr{S}_q$ is the vector space of $q\times q$ symmetric matrices. $\mathscr{S}_q$ is equipped with the inner product $<\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2>= tr[\boldsymbol{\Sigma}_1'\boldsymbol{\Sigma}_2]$, where tr denotes the trace operator. Let $\|\cdot\|_F$ denote the Frobenius norm induced by this inner product, *i.e.* $\|\boldsymbol{\Sigma}\|_F^2 = tr[\boldsymbol{\Sigma}'\boldsymbol{\Sigma}]$. The notion of consistency for matrix estimators will be defined relative to the Frobenius norm. Recall that $A^R(c)$ denote the regularized inverse of a $q\times q$ symmetric matrix A. In the sequel, let $I=\{1,2,\ldots,q\}$ denote the set of indices such that $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_q$, and $J=\{1,2,\ldots,k\}$ the subset of I corresponding to the indices associated with the distinct eigenvalues of $\boldsymbol{\Sigma}$, *i.e.* $d_1>d_2>\cdots>d_j>\cdots>d_k$, so that $\sum_{j=1}^{k} m(d_j)=q\geq 1$ and $1\leq k\leq q$, with $m(d_j)$ denoting the multiplicity of d_j. Let us define a partition of I, denoted $\mathscr{P}(I)$ such that:

$$\mathscr{P}(I)=\{I_j\subset I,\, j\in J : I_j\bigcap_{j\neq l} I_l=\emptyset,\, \bigcup_{j=1}^{k} I_j=I\}\ ,\ \ I=\{1,\ldots,q\}, \tag{7.7}$$

with

$$I_j=\{i\in I:\ \lambda_i=d_j\}\ ,\quad \text{card}\, I_j=m(d_j) \tag{7.8}$$

and

$$I(c)=\{i\in I:\ \lambda_i=d_j=c\}\ ,\ ,\quad \text{card}\, I(c)=m(c) \tag{7.9}$$

We adopt the convention that $I(c)=\emptyset$, if there is no eigenvalues equal to c. Recall that if $\boldsymbol{\Sigma}v=d_jv$ for some $v\neq 0$, then d_j is an eigenvalue of $\boldsymbol{\Sigma}$, and v an eigenvector of $\boldsymbol{\Sigma}$ associated with d_j, $v\in\mathbb{R}^q$. The eigenspace of $\boldsymbol{\Sigma}$ associated with d_j is $\mathscr{V}(d_j)=\{v\in\mathbb{R}^q,|\ \boldsymbol{\Sigma}v=d_jv\}$. The dimension of $\mathscr{V}(d_j)$ is given by the multiplicity of d_j, *i.e.* $m(d_j)$. See Tyler (1981, Section 2, Preliminaries, page 726) for a review of spectral theory results. The vector space $\mathbb{R}^q$ can be decomposed as $\mathbb{R}^q=\mathscr{V}(d_1)\oplus\cdots\oplus\mathscr{V}(d_j)\oplus\cdots\oplus\mathscr{V}(d_k)$. Each $u\in\mathbb{R}^q$ can be expressed in the form $u=u_1+\cdots+u_j+\cdots+u_k$, with $u_j\in\mathscr{V}(d_j)$, $j\in J$ in a unique way. The operator $P_j=P(d_j)$ is such that: $P_ju=u_j$ is the eigenprojection operator that projects onto the eigenspace $\mathscr{V}(d_j)$ along $N_j=\mathscr{V}(d_1)\oplus\cdots\oplus\mathscr{V}(d_{j-1})\oplus\mathscr{V}(d_{j+1})\oplus\cdots\oplus\mathscr{V}(d_k)$. Thus, $P_j(\boldsymbol{\Sigma})=P(d_j)(\boldsymbol{\Sigma})$, projects $\boldsymbol{\Sigma}$ onto the eigenspace $\mathscr{V}(d_j)$ along N_j. For all $j=1,\ldots,k,$ with $1\leq k\leq q,$ the $B(d_j)$'s, such that $B(d_j)=[v(d_j)_l]_{l=1,\ldots,m(d_j)}$ form an orthonormal basis for the eigenspace $\mathscr{V}(d_j)=\{v\in\mathbb{R}^q,|\ \boldsymbol{\Sigma}v=d_jv\}$. Let

$$P_j(\boldsymbol{\Sigma})=P(d_j)(\boldsymbol{\Sigma})=B(d_j)B(d_j)'\,, \tag{7.10}$$

when it is expressed in the Euclidean metric. The Euclidean metric specified here implies that the metric T in equation (6.6) is equal to the identity matrix, that is $P(\lambda)=\sum_{j=1}^{m(\lambda)}\mathbf{x}_j\mathbf{x}_j'\mathbf{T}$, with $T=Id$. Furthermore,

$\sum_{j=1}^{k} P_j = I_q, \quad P_k P_j = \delta_{jk} P_j$, with $\delta_{jk} = 0$ for $j \neq k$ and $\delta_{jk} = 1$ for $j = k$. There is a one-to-one mapping from J to $\mathscr{P}(I)$ such that:

$$j \longmapsto I_j \,, \tag{7.11}$$

where the total eigenprojection operator $P_{I_j}(\bullet)$ applied to $\boldsymbol{\Sigma}_n$, with $\boldsymbol{\Sigma}_n \xrightarrow{p} \boldsymbol{\Sigma}$, yields by Lemma 6.3 ii)

$$P_{I_j}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_j(\boldsymbol{\Sigma}) = P(d_j)(\boldsymbol{\Sigma}) \tag{7.12}$$

and

$$\dim P_{I_j} = \dim P_j = m(d_j) = \dim \mathscr{V}(d_j) \;\; with \;\; 1 = \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} P_{I_j} \,. \tag{7.13}$$

Proposition 7.1 UNIQUE REPRESENTATION OF THE REGULARIZED INVERSE. *For a given VRF $g(.,c)$ in the $\mathscr{G}_c$ family, the regularized inverse $\boldsymbol{\Sigma}^R(c) = V\Lambda^{\dagger}(c)V'$ of a symmetric matrix $\boldsymbol{\Sigma}$ and its sample analog $\boldsymbol{\Sigma}_n^R(c) = V_n\Lambda_n^{\dagger}(c)V_n'$ admit an unique representation of the form:*

$$\boldsymbol{\Sigma}^R(c) = \sum_{j=1}^{k} g(d_j; c) P_j(\boldsymbol{\Sigma}) \tag{7.14}$$

and

$$\boldsymbol{\Sigma}_n^R(c) = \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i \in I_j} g(\hat{\lambda}_i; c) \tag{7.15}$$

where the d_j's denote the distinct eigenvalues of $\boldsymbol{\Sigma}$ with multiplicity $m(d_j)$, $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$; $P_{I_j}(\boldsymbol{\Sigma}_n)$ and $P_j(\boldsymbol{\Sigma})$ are defined at equations (7.10)-(7.13) *with I_j defined at equation* (7.8). *If $\boldsymbol{\Sigma} = 0$, $P(0)(\boldsymbol{\Sigma}) = I_q$, and $\boldsymbol{\Sigma}^R(c) = g(0;c)P(0)(\boldsymbol{\Sigma}) = g(0;c)I_q$.*

The uniqueness of the representation of the regularized inverse immediately follows from the uniqueness of the decomposition involving only distinct eigenvalues. In particular, this representation exploits the Spectral Theorem; see Eaton (2007, Theorem 1.2a, p.53), and the references therein. Thus, there is a one-to-one relation between the regularized inverse and the VRF $g(.,c)$ in the $\mathscr{G}_c$ family. An interesting case producing a nonstandard asymptotic distribution corresponds to a fixed threshold c; an upper bound can be derived in the Gaussian case (see Corollary 8.3).

Let us first define a superconsistent estimator of the eigenvalues at c. The estimator $\hat{\boldsymbol{\lambda}}(c) = (\hat{\lambda}_i(c))_{i=1,\ldots,q}$ of the eigenvalues of a $q \times q$ positive semidefinite matrix $\boldsymbol{\Sigma}$ satisfies:

$$\hat{\lambda}_i(c) = \begin{cases} \hat{\lambda}_i & \text{if } |\hat{\lambda}_i - c| > \nu \frac{e_n}{b_n} \\ c & \text{if } |\hat{\lambda}_i - c| \leq \nu \frac{e_n}{b_n} \,, \end{cases} \tag{7.16}$$

for each $i = 1, \ldots, q$ where b_n is the speed of convergence of the sample eigenvalues as defined in Theorem A.2; e_n is chosen such that $e_n \to \infty$ with $\frac{e_n}{b_n} \to 0$ as n grows to infinity, and ν is an arbitrary strictly positive constant. $\hat{\lambda}_i(c)$ corresponds to a Hodges estimator; see Hodges and Lehmann (1950), LeCam (1953), Lehmann and Casella (1998), Leeb and Pötscher (2008).

Assumption 7.1 REGULARITY CONDITIONS FOR THE CONVERGENCE OF THE REGULARIZED INVERSE. *The VRF $g \in \mathscr{G}_c$, and for $i = 1, \ldots, q$, $\lambda_i = \lambda_i(\boldsymbol{\Sigma})$ are the eigenvalues of a $q \times q$ positive semidefinite matrix $\boldsymbol{\Sigma}$. At least, one of the following conditions holds:*

i) the VRF g is continuous at $\lambda_i = c$;

ii) $\nexists\, \lambda_i :\ \lambda_i = c$*;*

iii) the estimator $\hat{\lambda}_i(c)$ *of* λ_i *defined in equation* (7.16) *is superconsistent at c,* i.e. $\mathbb{P}\left[\hat{\lambda}_i(c) = c\right] \underset{n\to\infty}{\to} 1$.

As long as one of the above conditions holds, both convergence results of the regularized inverse (Propositions 7.2 and 7.3) will hold, otherwise they may break down. Let us now state the a.s. convergence for the regularized inverse when c is fixed.

Proposition 7.2 ALMOST SURE CONVERGENCE OF THE REGULARIZED INVERSE. *Let* $g \in \mathscr{G}_c$ *with c fixed. Suppose* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_n$ *are* $q \times q$ *symmetric matrices with* $\operatorname{rank}(\boldsymbol{\Sigma}) = r \leq q$. *Let the regularized inverses satisfy equations* (7.14) *and* (7.15). *Let Assumption 7.1 hold. If* $\boldsymbol{\Sigma}_n \overset{a.s.}{\to} \boldsymbol{\Sigma}$*, then*

$$\boldsymbol{\Sigma}_n^R(c) \overset{a.s.}{\to} \boldsymbol{\Sigma}^R(c)\,. \tag{7.17}$$

Proposition 7.3 CONVERGENCE IN PROBABILITY OF THE REGULARIZED INVERSE. *Suppose* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_n$ *are* $q \times q$ *symmetric matrices such that* $\operatorname{rank}(\boldsymbol{\Sigma}) = r \leq q$. *Suppose Assumption 2.2 holds with* $p = q$*, and Assumption 7.1 holds and c is fixed. Let the regularized inverses satisfy equations* (7.14) *and* (7.15)*, and decomposition* (4.2). *Then*

$$\boldsymbol{\Sigma}_n^R(c) = \boldsymbol{\Sigma}_{11,n}^R(c) + \boldsymbol{\Sigma}_{22,n}^R(c) + \boldsymbol{\Sigma}_{33,n}^R(c) \overset{p}{\to} \boldsymbol{\Sigma}^R(c) \tag{7.18}$$

where

$$\boldsymbol{\Sigma}_{11,n}^R(c) = \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c) \overset{p}{\to} \sum_{j=1}^{k_1} g(d_j; c) P_j(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}_{11}^R(c)\,, \tag{7.19}$$

$$\boldsymbol{\Sigma}_{22,n}^R(c) = P_{I(c)}(\boldsymbol{\Sigma}_n)\frac{1}{m(c)}\sum_{i\in I(c)} g(\hat{\lambda}_i; c) \overset{p}{\to} g(c; c) 1_{\{d_j = c\}} P_{j(c)}(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}_{22}^R(c)\,, \tag{7.20}$$

$$\boldsymbol{\Sigma}_{33,n}^R(c) = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c) \overset{p}{\to} \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j; c) P_j(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}_{33}^R(c)\,, \tag{7.21}$$

$k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$*,* k *is the number of distinct eigenvalues of* $\boldsymbol{\Sigma}$*, and* $P_{j(c)}(\boldsymbol{\Sigma}) = P(d_j)(\boldsymbol{\Sigma})$ *for* $d_j = c$*, where* $P_j(\boldsymbol{\Sigma}) = P(d_j)(\boldsymbol{\Sigma})$ *is defined at equation* (7.10). I_j *and* $I(c)$ *are defined in* (7.8) *and* (7.9). $m(d_j)$ *and* $m(c)$ *denote the multiplicity of* d_j *and c respectively.*

The problematic component for the convergence of the regularized inverse is the second one involving the eigenvalue $\lambda_i = d_j = c$. If the VRF g is continuous at $\lambda_i = d_j = c$, equation (7.20) holds; if there are no eigenvalues $\lambda_i = d_j = c$, $I(c) = \emptyset$, $1_{\{d_j = c\}} = 0$, and the convention adopted is to set $\boldsymbol{\Sigma}_{22,n}^R(c) = \boldsymbol{\Sigma}_{22}^R(c) = 0$; if there exists a superconsistent estimator of the eigenvalue at c, (7.20) holds. Otherwise, $\boldsymbol{\Sigma}_n^R(c)$ may not converge to $\boldsymbol{\Sigma}^R(c)$. In other words, the conditions stated in Assumption 7.1 are necessary conditions for (7.17) and (7.18) to hold.

8. Asymptotic distribution of the regularized Wald tests with a fixed threshold

In this section, we characterize the asymptotic distribution of the regularized Wald statistic for general distributions, before presenting the Gaussian case. More specifically, we characterize the asymptotic distribution

of the regularized Wald statistic for the class of VRFs defined in 7.1. In particular, this sheds light on the (relatively unknown) inferential properties of ridge-type regularization techniques, which are instead applied to regression coefficients' estimation. Further, the decomposition of the regularized statistic into three independent components provides an insight on the structure of the distribution; an upper bound can easily be derived in the Gaussian case. Power and consistency properties of the test are next established.

Proposition 8.1 CHARACTERIZATION OF THE REGULARIZED WALD STATISTIC WHEN THE THRESHOLD IS FIXED. *Suppose $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ are $q \times q$ symmetric matrices such that* $\text{rank}(\boldsymbol{\Sigma}) = r \leq q$ *. Suppose Assumptions* **2.1** *with $\psi = \psi_0$,* **2.2** *with $p = q$, and* **7.1** *hold. Let the regularized inverses satisfy equations* (7.14) *and* (7.15)*, decomposition* (4.2)*, and the eigenprojection is expressed as in equation* (7.10)*. Let $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ be the number of distinct eigenvalues of $\boldsymbol{\Sigma}$ larger than c, and $W_n^R(c)$ is defined in* **5.1***. Then $W_n^R(c) \xrightarrow{\mathscr{L}} W^R(c)$, where*

$$W^R(c) = X' \boldsymbol{\Sigma}^R(c) X = \sum_{j=1}^{k} g(d_j; c) X' B(d_j) B(d_j)' X = W_1^R(c) + W_2^R(c) + W_3^R(c)\,, \tag{8.1}$$

$$W_1^R(c) = X' \boldsymbol{\Sigma}_{11}^R(c) X = \sum_{j=1}^{k_1} g(d_j; c) X' B(d_j) B(d_j)' X\,, \tag{8.2}$$

$$W_2^R(c) = X' \boldsymbol{\Sigma}_{22}^R(c) X = g(c; c) 1_{\{d_j = c\}} X' B(c) B(c)' X\,, \tag{8.3}$$

$$W_3^R(c) = X' \boldsymbol{\Sigma}_{33}^R(c) X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j; c) X' B(d_j) B(d_j)' X\,. \tag{8.4}$$

Interestingly when $\boldsymbol{\Sigma} = 0$ the distribution of $W^R(c)$ can still be characterized; the regularized weighting matrix is given by $\boldsymbol{\Sigma}^R(c) = g(0; c) I_q$, so the regularized Wald statistic simplifies to $g(0; c) X'X$ in the general case. In the Gaussian case, when $\boldsymbol{\Sigma} = 0$, $d_j = 0$ with multiplicity q, the limiting statistic is equal to zero (see equation (8.5), where $W^R(c) = 0$). Note also that the components are independent due to the specific decomposition of the regularized weighting matrix. We can now easily consider the special case where X is Gaussian, with the Lütkepohl and Burda (1997)'s result obtained as a special case of Corollary 8.2. Besides, if there is no eigenvalues such that $\lambda_i = d_j = c$, $W_2^R(c) = 0$ due to the indicator function, and $W^R(c) = W_1^R(c) + W_3^R(c)$ for all the subsequent results stated in this section.

Corollary 8.2 THE REGULARIZED WALD STATISTIC WITH A FIXED THRESHOLD: THE GAUSSIAN CASE. *Suppose $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ are $q \times q$ symmetric matrices such that* $\text{rank}(\boldsymbol{\Sigma}) = r \leq q$ *. Under Assumptions 2.2 with $p = q$, 2.4 with $\psi(\theta) = \psi_0$, and 7.1, let the regularized inverses satisfy equations* (7.14) *and* (7.15)*, decomposition* (4.2)*, and the eigenprojection is expressed as in equation* (7.10)*. Let $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ be the number of distinct eigenvalues of $\boldsymbol{\Sigma}$ larger than c, and $W_n^R(c)$ is defined in (5.1). Let $B(d_j)'X = x_j$, where $x_j \sim N[0, d_j I_{m(d_j)}]$, for $j = 1, \ldots, k$, or equivalently $x_j = \sqrt{d_j} u_j$, with $u_j \sim N[0, I_{m(d_j)}]$.*

(i) If $\boldsymbol{\Sigma} = 0$, $d_j = 0$ with $m(0) = q$, then

$$W_n^R(c) \xrightarrow{\mathscr{L}} W^R(c) = X' \boldsymbol{\Sigma}^R(c) X = d_j u_j' g(0; c) I_q u_j = 0\,. \tag{8.5}$$

(ii) If $\boldsymbol{\Sigma} \neq 0$*, then* $W_n^R(c) \xrightarrow{\mathscr{L}} W^R(c)$*, where*

$$W^R(c) = X'\boldsymbol{\Sigma}^R(c)X = \sum_{j=1}^{k} g(d_j;c)d_j v_j = W_1^R(c) + W_2^R(c) + W_3^R(c)\,, \tag{8.6}$$

$$W_1^R(c) = X'\boldsymbol{\Sigma}_{11}^R(c)X = \sum_{j=1}^{k_1} g(d_j;c)d_j v_j\,, \quad W_2^R(c) = X'\boldsymbol{\Sigma}_{22}^R(c)X = g(c;c)1_{\{d_j=c\}}cv_{j(c)}\,, \tag{8.7}$$

$$W_3^R(c) = X'\boldsymbol{\Sigma}_{33}^R(c)X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)d_j v_j\,, \tag{8.8}$$

and $v_j \sim \chi^2\big(m(d_j)\big), v_{j(c)} \sim \chi^2\big(m(c)\big)$.

We can see from this corollary that the three components can be interpreted as a linear combination of chi-square variables with the degree of freedom given by the multiplicity of the distinct eigenvalues. Note that when $\boldsymbol{\Sigma}$ has rank $r < q$, the last component $W_3^R(c)$ contains a zero eigenvalue, *i.e.* $d_k = 0$, when $c \neq 0$. When $c = 0$, in this case $W_2^R(0) = W_3^R(0) = 0$ and $W_1^R(0) = W^+(0)$; we obtain the Lütkepohl and Burda (1997) result as a special case. Note that their result only holds for distinct eigenvalues.

Corollary 8.3 CHARACTERIZATION OF THE BOUND: THE GAUSSIAN CASE. *Suppose* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_n$ *are* $q \times q$ *symmetric matrices such that* $\operatorname{rank}(\boldsymbol{\Sigma}) = r \leq q$ *. Under Assumptions 2.2 with* $p = q$*, 2.4 with* $\psi(\theta) = \psi_0$*, and 7.1, let the regularized inverses satisfy equations* (7.14) *and* (7.15)*, decomposition* (4.2)*, and the eigenprojection is expressed as in equation* (7.10)*. Let* $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ *be the number of distinct eigenvalues of* $\boldsymbol{\Sigma}$ *larger than c, and* $W_n^R(c)$ *is defined in (5.1). Let* $B(d_j)'X = x_j$*, where* $x_j \sim N[0, d_j I_{m(d_j)}]$*, for* $j = 1, \ldots, k$*, . Let* $g(.;c) \in \mathscr{G}_c$*, with a fixed threshold c such that*

$$g(d_j;c)d_j \leq 1 \;\; \forall\; j = 1, \ldots, k$$

then

$$W_1^R(c) \leq \chi^2(q_1)\,, \quad W_2^R(c) \leq \chi^2(m(c)),\; W_3^R(c) \leq \chi^2(q_3)$$

and

$$W^R(c) \leq \sum_{j=1}^{k} v_j \;\sim\; \chi^2(q)$$

where $v_j \sim \chi^2\big(m(d_j)\big)$*,* $q_1 = \sum_{j=1}^{k_1} m(d_j)$*,* $q_3 = q - q_1 - m(c)$*, and* $q = \sum_{j=1}^{k} m(d_j)$*.*

In the Gaussian case we obtain a chi-square as an upper bound for the *regularized* statistic, when c is fixed. Each component is distributed as a chi-square variable with the degree of freedom given by the sum of the multiplicities of the distinct eigenvalues involved in the sum. As the decomposition involves three independent chi-square variables, the resulting distribution for the overall statistic is also chi-square due to its stability; the degree of freedom is then given by the sum of the degrees of freedom of each component. As a result, the critical point given by the standard chi-square distribution (if X is Gaussian) can be used to provide an *asymptotically valid* test. However, improved power over this conservative bound could be achieved by simulations. For comparison, the bound provided by Dufour et al. (2017) is not applicable in general (*e.g.*, the case of several polynomial restrictions is excluded), except for the restrictive case of a *single* polynomial restriction. In the specific latter case, they provide a tight bound (*i.e.*, one fourth of the standard chi-square

distribution) for the original Wald statistic. In contrast, we derive an upper bound for the modified Wald statistic that works through multiple nonlinear restrictions. In this respect, the bound we derive is universal.

We shall now show that the regularized statistic is consistent against a global alternative when X_n follows a general distribution.

Proposition 8.4 CONSISTENCY PROPERTY OF THE TEST. *Suppose $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ are $q \times q$ symmetric matrices such that* $\text{rank}(\boldsymbol{\Sigma}) = r \leq q$ *. Suppose Assumptions 2.2 with $p = q$ and 7.1 hold. Let the regularized inverses satisfy Property 7.1, decomposition* (4.2)*, and the eigenprojection is expressed as in equation* (7.10)*. Let $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ be the number of distinct eigenvalues of $\boldsymbol{\Sigma}$ larger than c, and $W_n^R(c)$ is defined in 5.1. Suppose also that there exist some eigenvalues of the limiting matrix $\boldsymbol{\Sigma}$ such that $d_j \neq 0$ under the alternative. Suppose further $X_n = H_n(\hat{\psi}_n - \psi_1)$ satisfies Assumption 2.1, with $\psi = \psi_1$. If $\psi_1 - \psi_0 = \Delta \neq 0$, and $\Delta' \boldsymbol{\Sigma}^R(c)\Delta > 0$, then*

$$W_n^R(c) \underset{n\to\infty}{\to} \infty \,, \tag{8.9}$$

which means that for any positive values of ε as large as possible, we have $\mathbb{P}\left[W_n^R(c) > \varepsilon\right] \underset{n\to\infty}{\to} 1$.

We also characterize the behavior the regularized Wald statistic under local alternatives as in the next proposition.

Proposition 8.5 LOCAL POWER CHARACTERIZATION. *Suppose $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ are $q \times q$ symmetric matrices such that* $\text{rank}(\boldsymbol{\Sigma}) = r \leq q$ *. Under Assumption 2.2 with $p = q$, and under Assumption 7.1, let the regularized inverses satisfy Property 7.1. Let $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ be the number of distinct eigenvalues of $\boldsymbol{\Sigma}$ larger than c, and $W_n^R(c)$ is defined in 5.1. Suppose there exist some eigenvalues of the limiting matrix $\boldsymbol{\Sigma}$ such that $d_j \neq 0$ under the alternative. Suppose further $X_n = H_n(\hat{\psi}_n - \psi_{1n})$ satisfies Assumption 2.1. If $H_n(\psi_{1n} - \psi_0) \to \Delta \neq 0$, and $\Delta' \boldsymbol{\Sigma}^R(c)\Delta > 0$, then*

$$W_n^R(c) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X' \boldsymbol{\Sigma}^R(c) X + 2X' \boldsymbol{\Sigma}^R(c)\Delta + \Delta' \boldsymbol{\Sigma}^R(c)\Delta \,. \tag{8.10}$$

We can observe from this result that the limiting quantity involves three components: the first component is still a quadratic form in X in accordance with the null hypothesis; the second component is a linear form in X; the third one represents a noncentrality parameter. Only the last two components will contribute to power. Note that in the Lütkepohl and Burda (1997) case, the noncentrality parameter based on the modified Moore-Penrose inverse $\Delta' \boldsymbol{\Sigma}_c^+ \Delta$ is expected to be smaller than the noncentrality parameter $\Delta' \boldsymbol{\Sigma}^R(c)\Delta$, which may entail a loss of power even though the chi-square distribution with reduced degrees of freedom yields a smaller critical point. Indeed, there may exist some directions for the alternative, where a spectral cut-off type Moore-Penrose inverse that sets to zero the small eigenvalues, may destroy power as shown in the next corollary.

Corollary 8.6 LOCAL POWER CHARACTERIZATION: DELTA IN THE NULL EIGENSPACE. *Suppose the assumptions of Proposition 8.5 are satisfied. Suppose further that $\Delta \in \mathscr{V}(0)$, then*

$$W_n^R(c) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X' \boldsymbol{\Sigma}^R(c) X + 2g(0;c) X'\Delta + g(0;c)\Delta'\Delta \,. \tag{8.11}$$

We do not expect the test to be consistent against all types of alternatives. There may exist some directions where power is reduced or eventually destroyed, whether Δ lies in the eigenspace $\mathscr{V}(0)$ associated with the null eigenvalue or not. In such a case, the choice of $g(0;c)$ is critical for power considerations. By setting

$g(0;c) = 0$, the spectral cut-off Moore Penrose inverse used by Lütkepohl and Burda (1997) will destroy power.

9. The case with a varying threshold c_n

We shall now present the convergence results for the regularized inverse based on a varying threshold as well as the regularized test statistic. In this section, we concentrate on the modified Moore-Penrose inverse (*i.e.*, spectral cut-off) and exclude the DV scheme. Let $\lambda_i = \lambda_i(\boldsymbol{\Sigma})$ and $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$ for notational simplicity. First when designing the VRF $g(\lambda; c_n)$, the varying threshold c_n must be selected so that

$$\mathbb{P}\left[|\hat{\lambda}_i - \lambda_i| > c_n\right] = \mathbb{P}\left[|b_n(\hat{\lambda}_i - \lambda_i)| > b_n c_n\right] \underset{n\to\infty}{\to} 0 \tag{9.1}$$

with $c_n \to 0$ and $b_n c_n \to \infty$ as n grows to infinity. Thus, c_n declines to 0 slower than $1/b_n$, and $b_n c_n \to \infty$ slower than b_n. Indeed, the threshold must not decline to zero either too fast, or too slow. Selecting c_n in this way ensures that the nonzero eigenvalues of the covariance matrix will eventually be greater than the threshold, while the true zero eigenvalues will fall below the threshold and are set to zero at least in large samples. In most cases, a natural choice for $b_n = \sqrt{n}$ and a suitable choice for c_n is $c_n = n^{-1/3}$. This convergence rate plays a crucial role in Proposition 9.1 below. For a deeper insight on the limiting distribution of the sample eigenvalues, please refer to Appendix A where we summarize some general results from Eaton and Tyler (1994). Note, however, the knowledge of the entire limiting distribution of the sample eigenvalues is unnecessary for our purpose; only the convergence rate b_n plays an important role in the asymptotic properties of the regularized covariance matrices based on a varying threshold c_n. Thus, a varying threshold requires more information.

Proposition 9.1 CONVERGENCE OF THE REGULARIZED INVERSE WHEN THE THRESHOLD VARIES WITH THE SAMPLE SIZE. *Let $\boldsymbol{\Sigma}$ be a $q \times q$ real symmetric positive semidefinite nonstochastic matrix and $\boldsymbol{\Sigma}_n$ a sequence of $q \times q$ real symmetric random matrices. Let $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ satisfy Assumption 2.2 with $p = q$ and let $g \in \mathscr{G}_c$, with $g(0;0) = 0$. Let $\lambda_i = \lambda_i(\boldsymbol{\Sigma})$ and $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$, with $\lambda_{i+1} \geq \lambda_i \geq 0$, $i = 1, \ldots, q$ and d_j's denote the distinct eigenvalues of $\boldsymbol{\Sigma}$. Suppose further that $c_n \underset{n\to\infty}{\to} 0$ and $b_n c_n \underset{n\to\infty}{\to} \infty$. If $\boldsymbol{\Sigma}^R(0)$ and $\boldsymbol{\Sigma}_n^R(c_n)$ have the representations* (7.14) *and* (7.15) *respectively, then*

$$\boldsymbol{\Sigma}_n^R(c_n) \xrightarrow{\text{p}} \boldsymbol{\Sigma}^R(0)\,. \tag{9.2}$$

In other words, if $\boldsymbol{\Sigma}_n \to \boldsymbol{\Sigma}$ in probability, then the regularized inverse of $\boldsymbol{\Sigma}_n$ will converge towards the regularized inverse of $\boldsymbol{\Sigma}$. In the following, we establish a *characterization* of the asymptotic distribution of the *regularized* test statistic in the general case. This characterization makes use of the decomposition of the *regularized* statistic into a regular component and a regularized one.

Proposition 9.2 ASYMPTOTIC CHARACTERIZATION OF THE REGULARIZED WALD STATISTIC WITH VARYING THRESHOLD. *Let $\boldsymbol{\Sigma}$ be a $q \times q$ real symmetric positive semidefinite nonstochastic matrix and $\boldsymbol{\Sigma}_n$ a sequence of $q \times q$ real symmetric random matrices. Let $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ satisfy Assumption 2.2 with $p = q$ and $g \in \mathscr{G}_c$, with $g(0;0) = 0$. Let $g(\lambda; c)$ be defined in equation* 7.2. *Suppose $c_n \underset{n\to\infty}{\to} 0$ and $b_n c_n \underset{n\to\infty}{\to} \infty$. Let $\boldsymbol{\Sigma}^R(0)$ and $\boldsymbol{\Sigma}_n^R(c_n)$ have the representations* (7.14) *and* (7.15) *respectively. Suppose also Assumption 2.1 holds, and rank$(\boldsymbol{\Sigma}) = q_1$. Let k_1 be the number of non-zero distinct eigenvalues d_j of $\boldsymbol{\Sigma}$,* i.e.,

$\sum_{j=1}^{k_1} m(d_j) = q_1 \geq 1$, $g(d_j;0) = 0$, $\forall\ j \geq k_1+1$, *and* $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$. *Then, under* $H_0(\boldsymbol{\psi}_0) : \boldsymbol{\psi}(\boldsymbol{\theta}_0) = \boldsymbol{\psi}_0$,

$$W_n^R(c_n) = X_n' \boldsymbol{\Sigma}_n^R(c_n) X_n \xrightarrow{\mathscr{L}} X' \boldsymbol{\Sigma}^R(0) X = W^R(0) \tag{9.3}$$

$$W_n^R(c_n) = W_{1n}^R(c_n) + W_{2n}^R(c_n) \tag{9.4}$$

$$W_{1n}^R(c_n) = X_n' \boldsymbol{\Sigma}_{11,n}^R(c_n) X_n \xrightarrow{\mathscr{L}} X' \boldsymbol{\Sigma}_{11}^R(0) X \equiv W_1^R(0) \tag{9.5}$$

$$W_{2n}^R(c_n) = X_n' \boldsymbol{\Sigma}_{22,n}^R(c_n) X_n \ \textit{ such that } \ \mathbb{P}\big[W_{2n}^R(c_n) = 0\big] \to 1\ . \tag{9.6}$$

Thus, when the threshold c_n converges to zero at an appropriate rate, based on the sample eigenvalues' convergence rate, the limiting *regularized* inverse boils down to the modified Moore-Penrose inverse, which cancels the nonregular component $W_2^R(0)$. Moreover, if we restrict the convergence in law above to the sole standard Gaussian distribution, *i.e.*, $\big[X_n = a_n(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) = \sqrt{n}[\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0] \to N[0, \boldsymbol{\Sigma}]\big]$, we obtain the result given by Lütkepohl and Burda (1997, Proposition 2, page 318) as a special case (see Corollary 9.3). In this case, the regularized Wald test is asymptotically distributed as a $\chi^2(q_1)$ variable with $q_1 < q$. Further, note that Lütkepohl and Burda (1997, Proposition 2, page 318)'s result only holds for distinct eigenvalues, unlike Proposition 9.2 that is valid for multiple eigenvalues.

Corollary 9.3 ASYMPTOTIC DISTRIBUTION OF THE REGULARIZED WALD STATISTIC IN THE GAUSSIAN CASE WITH VARYING THRESHOLD. *Let* $\boldsymbol{\Sigma}$ *be a* $q \times q$ *real symmetric positive semidefinite nonstochastic matrix and* $\boldsymbol{\Sigma}_n$ *a sequence of* $q \times q$ *real symmetric random matrices. Suppose Assumption 2.1 holds, and* $rank(\boldsymbol{\Sigma}) = q_1$. *Suppose also that Assumptions 2.2 with* $p = q$, *and 2.4 hold. Let* $g \in \mathscr{G}_c$, *with* $g(0;0) = 0$, *and* $g(\lambda; c)$ *be defined in equation* 7.2. *Suppose* $c_n \underset{n\to\infty}{\to} 0$ *and* $b_n c_n \underset{n\to\infty}{\to} \infty$. *Let* $\boldsymbol{\Sigma}^R(0)$ *and* $\boldsymbol{\Sigma}_n^R(c_n)$ *have the representations* (7.14) *and* (7.15) *respectively. Let the eigenprojection be expressed as in equation* (7.10). *Let* k_1 *be the number of non-zero distinct eigenvalues* d_j *of* $\boldsymbol{\Sigma}$, i.e., $\sum_{j=1}^{k_1} m(d_j) = q_1 \geq 1$, $g(d_j;0) = 0$, $\forall\ j \geq k_1+1$, *and* $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$. *Let* $B(d_j)'X = x_j$, *with* $x_j \sim N\big[0, d_j I_{m(d_j)}\big]$ *for all* j, *or equivalently* $x_j = \sqrt{d_j} u_j$, $u_j \sim N\big[0, I_{m(d_j)}\big]$. *Let* $g(d_j;0) = \frac{1}{d_j}$, $\forall\ j \leq k_1$ *and 0 otherwise. Then, under* $H_0(\boldsymbol{\psi}_0) : \boldsymbol{\psi}(\boldsymbol{\theta}_0) = \boldsymbol{\psi}_0$

$$W_n^R(c_n) = n[\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0]' \boldsymbol{\Sigma}_n^R(c_n) [\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0] = W_{1n}^R(c_n) + W_{2n}^R(c_n)\ ,$$

with

$$W_{1n}^R(c_n) = n[\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0]' \boldsymbol{\Sigma}_{11,n}^R(c_n) [\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0]\ , \tag{9.7}$$

$$W_{2n}^R(c_n) = n[\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0]' \boldsymbol{\Sigma}_{22,n}^R(c_n) [\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\psi}_0]\ , \tag{9.8}$$

and

$$W_{1n}^R(c_n) \xrightarrow{\mathscr{L}} W_1^R(0) \ \sim\ \chi^2(q_1) \ \textit{ and } \ \mathbb{P}\big[W_{2n}^R(c_n) = 0\big] \to 1\ . \tag{9.9}$$

When the threshold goes to zero at the appropriate speed, the limiting regularized statistic has a standard chi square distribution with the degree of freedom given by the multiplicity of the nonzero eigenvalues. Meanwhile, the nonregular component collapses to zero due to the spectral cut-off Moore-Penrose inverse.

10. Empirical application to Multistep noncausality: saving-to-growth causation

In this section, we conduct noncausality tests to assess any causal relation between investment, saving and growth. Indeed, there is no consensus in the literature whether higher saving results in higher growth or the other way around in cross-country data. Especially, East Asian economies had experienced high growth rates long before they had high saving rates. Levine and Renelt (1992) argue that the investment rate is the key variable that is correlated with growth. They claim that the saving-to-growth causation reflects the same causal channel, but with the additional linkage that high saving leads to high investment. We shall investigate this relation in a single-country data set, focusing on U.S. data. The data come from the World Development Indicator's database (WDI), and are yearly observations spanning from 1972 to 2012. The data have been differenced once to account for the presence of unit roots since the Augmented Dickey-Fuller tests detected the presence of unit roots at a 0.05 significance level. We use Saving that represents the gross domestic saving (in % of GDP), Investment that corresponds to gross capital formation (in % of GDP) and GDP growth (in annual %). The gross capital formation consists of additions to the fixed assets of the economy plus net changes in the level of inventories. We also use Foreign direct Investment (FDI) (in % of GDP); FDI are the net inflows of investment to acquire 10% or more of voting stocks in an enterprise operating in an economy other than that of the investor.

In this section, we conduct noncausality tests of the form:

$$H_0 : \psi(\theta) = 0 \,, \tag{10.1}$$

for several horizons, *i.e.*, at horizons $H = 1, 2, 3, 4$ and 5. Dufour and Renault (1998, Proposition 4.5) state that in a VAR(1) model it is sufficient to have noncausality up to horizon 2 for noncausality to hold at all horizons; therefore testing for noncausality at horizons 3, 4 and 5 is superfluous and adds redundancy uselessly.

The Monte Carlo tests are simulated under the null of noncausality using $N = 99$ simulated statistics. The estimate of the parameters are based on the real data; we then construct an ad-hoc restricted estimate by zeroing the corresponding parameters such that $\psi(\hat{\theta}) = 0$. Using an unrestricted estimator $\text{vec}(\hat{\theta})$, we built the restricted version of the estimator, *i.e.*, $\text{vec}(\tilde{\theta}) = (\hat{\theta}_1,\ \hat{\theta}_2,\ \hat{\theta}_3,\ 0,\ \hat{\theta}_5,\ 0,\ 0,\ \hat{\theta}_8,\ \hat{\theta}_9)'$. We use this ad-hoc restricted estimate to simulate the distribution of the test statistic under the null hypothesis. Recall that the Wald test employs an unrestricted estimator, although its distribution is simulated under the null in the Monte Carlo procedure. The nominal level used in the test has been fixed at $\alpha = 0.05$.

In addition to Panels A and B of Table 1, in which no redundant restrictions are added to the genuine restrictions, we purposely add redundant restrictions to assess their effect on the testing procedures; see panels C, D and E. More specifically, Panel A only tests $\psi_1(\theta) = \theta_{xy} = 0$ while Panel B focuses on testing two restrictions:

$$\psi_2(\theta) = \begin{bmatrix} \theta_{xy} \\ \theta_{xx}\theta_{xy} + \theta_{xy}\theta_{yy} + \theta_{xz}\theta_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{10.2}$$

which corresponds to the case of no redundant restrictions with the following Jacobian

$$\frac{\partial \psi_2}{\partial \theta'} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \theta_{xy} & 0 & 0 & \theta_{xx}+\theta_{yy} & \theta_{xy} & \theta_{xz} & \theta_{zy} & 0 & 0 \end{bmatrix} .$$

In the trivariate VAR(1) model, in which

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} Growth \\ Saving \\ Investment \end{bmatrix},$$

the corresponding unrestricted estimates of the parameters and their estimated standard deviation are the following:

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_{xx} & \hat{\theta}_{yx} & \hat{\theta}_{zx} & \hat{\theta}_{xy} & \hat{\theta}_{yy} & \hat{\theta}_{zy} & \hat{\theta}_{xz} & \hat{\theta}_{yz} & \hat{\theta}_{zz} \\ -0.1466 & -0.8969 & -0.4203 & 0.3928 & 0.3176 & 0.5392 & -0.4411 & -0.4741 & -0.3438 \end{bmatrix}$$

$$\hat{\sigma} = \begin{bmatrix} \hat{\sigma}_1 & \hat{\sigma}_2 & \hat{\sigma}_3 & \hat{\sigma}_4 & \hat{\sigma}_5 & \hat{\sigma}_6 & \hat{\sigma}_7 & \hat{\sigma}_8 & \hat{\sigma}_9 \\ 2.149 & 3.2311 & 2.5876 & 0.6313 & 0.9505 & 0.7612 & 1.8284 & 2.7531 & 2.2048 \end{bmatrix}.$$

In Table 1, we test for noncausality between Saving, Investment and Growth. In panel A, the results for W, $W_{DV}(bound)$ and W_{LB} coincide regardless of the procedure used, asymptotic or simulated, since regularization is unnecessary in this case. We next observe that when redundant restrictions are added, the reported determinant of the estimated covariance matrix diminishes. The behavior of the standard Wald test statistic W seriously deteriorates. This poor behavior is striking in Panel E about the Investment-Growth causation; the value of the statistic jumps from 3.2388 (Panel D: Investment-Growth causation) to 11.7251 (Panel E: Investment-Growth causation) forcing the standard statistic to erroneously reject the null of noncausation. Similarly, the standard Wald test statistic W steadily misbehaves as the determinant approaches zero in the Growth-Saving causality analysis. While the asymptotic standard test still rejects the null of noncausation from Growth to Saving with a value of 40.5742 (Panel E: Growth-to-Saving), its simulated counterpart fails to reject the null with a p-value of 0.12 (Panel E: Growth-to-Saving). Thus, simulating from a misbehaved statistic does not produce reliable inference; a severe contradiction arises between the decision based on the asymptotic critical value and the simulated procedure. Further, the discrepancy between the standard Wald statistic W and the full-rank regularized Wald statistic $W_{DV}(bound)$ widens with the number of redundant restrictions added (Panel E: Investment-to-Growth, Growth-to-Saving). Note also the puzzling conclusion produced by the simulated test based on the spectral cut-off statistic W_{LB}. When redundant restrictions are added, the simulated procedure inverts the decision of the test when one moves from Panel B to panel C and so on in the Saving-to-Growth causation.

While most of the procedures are not able to reject the null hypothesis that Saving does not cause Growth at all horizons, we unambiguously find that Growth causes Saving for U.S. data. Our findings support the original literature by Houthakker (1961, 1965), and Modigliani (1970) at the cross-country level. However, our single-country results on U.S. data do not support Levine and Renelt (1992)'s cross-country findings that high investment causes high growth. Importantly, in the presence of redundant restrictions the simulated version of the full-rank regularized test statistic $W_{DV}(bound)$ steadily produces results consistent with those obtained without redundant restrictions. These results confirm those predicted from the theory: as stated in Dufour and Renault (1998, Proposition 4.5), in a VAR(1) model it is sufficient to have noncausality up to horizon 2 for noncausality to hold at all horizons. In other words, our findings at horizons 3,4 and 5 corroborate the results obtained at horizon 2.

Next, when replacing Saving by FDI in Table 2, all tests fail to reject the null that FDI does not cause Growth, nor that Growth does not cause FDI. Nevertheless, all tests regardless of the approach used, asymptotic or simulated, unambiguously reject the null that Investment does not cause Growth at all horizons; this finding does support that Investment is crucial for fostering Growth. As predicted by the theory in a VAR(1) model, decisions obtained at horizon 2 are not reversed at higher horizons. Again, singularity critically im-

pacts the behavior of the standard Wald statistic, triggering an erroneous rejection of the null that FDI does not cause Growth in panel E.

11. Conclusion

In this paper, we examine and propose Wald-type tests statistics that deal with asymptotic singular covariance matrices. To do so, we introduce a new class of *regularized* inverses, as opposed to generalized inverses, that embeds the spectral cut-off and Tikhonov regularized inverses known in the statistical literature. We propose a regularized Wald statistic that produces valid inference under fairly weak assumptions: the full-rank statistic relies on a fixed value for the threshold in the VRF $g(\lambda; c)$ and does not require the knowledge of the asymptotic rank nor the Gaussianity distribution. In contrast, the reduced rank Wald statistic that lets the threshold vary with the sample size requires more information about the sample behavior of the eigenvalues. By exploiting eigenprojection techniques, we show that the first regularized Wald statistic admits a nonstandard asymptotic distribution in the general case, which corresponds to a linear combination of χ^2 variables if the restrictions are Gaussian. An *upper bound*, which is invariant to the degree of rank deficiency, is then derived for the full-rank regularized statistic that corresponds to a χ^2 variable with *full rank* under Gaussianity. Hence, the test is *asymptotically valid*, meaning that the usual critical point can be used, but is conservative. Instead of using the asymptotic bound, we propose three ways to conduct the regularized Wald test by simulations through the technique of Monte Carlo tests: one may simulate under the DGP if available, or from the distribution of the estimator of the parameters (or of the restrictions) to correct for size distortions. One can also simulate from the linear combination of chi-square variables to produce an asymptotically valid test for the full-rank regularized statistic. Finally, when the threshold goes to zero with the sample size, we obtain the spectral cut-off modified Wald statistic of Lütkepohl and Burda (1997) as a special case. Under normality, the test has the usual asymptotic distribution whose reduced rank is given by the number of eigenvalues greater than zero. Note that Lütkepohl and Burda (1997)'s result only holds for distinct eigenvalues whereas our result accounts for multiple eigenvalues. We also show that the regularized statistics are consistent against global alternatives, but the spectral cut-off Wald statistic has reduced power in some directions of the alternative. Besides, our approach is easy to implement: it only requires to compute eigenvalues and eigenvectors. It is therefore simple, systematic, and robust to all kinds of setups. More generally, the regularization techniques developed in this paper to deal with asymptotic singularity and deficient rank problems are not restricted to the sole Wald statistic, but can easily be applied to other statistics such as the Lagrange multiplier statistic, or score-type test statistics.

Table 1. Testing for noncausality between Saving, Investment and Growth

$H_0: \ y_t \overset{(H)}{\not\to} x_t$, nominal size $= 0.05$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$

Panel A: Testing for noncausality at Horizon $H=1$												
	$H_0: \ Saving \overset{(1)}{\not\to} Growth$				$H_0: \ Investment \overset{(1)}{\not\to} Growth$				$H_0: \ Growth \overset{(1)}{\not\to} Saving$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 10.4400$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 7.5800$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.3985$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	3.0050	Not rej.	0.07	Not rej.	1.1712	Not rej.	0.28	Not rej.	15.1027	Rej.	0.01	Rej.
W_{LB}	3.0050	Not rej.	0.07	Not rej.	1.1712	Not rej.	0.28	Not rej.	15.1027	Rej.	0.01	Rej.
$W_{DV}\,(bound)$	3.0050	Not rej.	0.07	Not rej.	1.1712	Not rej.	0.28	Not rej.	15.1027	Rej.	0.01	Rej.
Panel B: Testing for noncausality at Horizon $H=2$												
	$H_0: \ Saving \overset{(2)}{\not\to} Growth$				$H_0: \ Investment \overset{(2)}{\not\to} Growth$				$H_0: \ Growth \overset{(2)}{\not\to} Saving$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 52.9062$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 27.9013$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0725$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	3.115	Not rej.	0.07	Not rej.	2.1668	Not rej.	0.15	Not rej.	15.1053	Rej.	0.01	Rej.
W_{LB}	3.1150	Not rej.	0.06	Not rej.	2.1668	Not rej.	0.13	Not rej.	13.298	Rej.	0.01	Rej.
$W_{DV}\,(bound)$	3.115	Not rej.	0.07	Not rej.	2.1668	Not rej.	0.15	Not rej.	15.1053	Rej.	0.01	Rej.
Panel C: Testing for noncausality at Horizon $H=3$												
	$H_0: \ Saving \overset{(3)}{\not\to} Growth$				$H_0: \ Investment \overset{(3)}{\not\to} Growth$				$H_0: \ Growth \overset{(3)}{\not\to} Saving$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 22.2189$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 8.3861$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.00198$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	3.2886	Not rej.	0.19	Not rej.	2.1802	Not rej.	0.23	Not rej.	38.1558	Rej.	0.01	Rej.
W_{LB}	3.1947	Not rej.	0.03	Rej.	2.1673	Not rej.	0.10	Not rej.	12.0253	Rej.	0.01	Rej.
$W_{DV}\,(bound)$	3.2886	Not rej.	0.15	Not rej.	2.1802	Not rej.	0.21	Not rej.	15.4005	Rej.	0.01	Rej.
Panel D: Testing for noncausality at Horizon $H=4$												
	$H_0: \ Saving \overset{(4)}{\not\to} Growth$				$H_0: \ Investment \overset{(4)}{\not\to} Growth$				$H_0: \ Growth \overset{(4)}{\not\to} Saving$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 2.6445$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.3841$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.000124$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	3.3415	Not rej.	0.52	Not rej.	3.2388	Not rej.	0.46	Not Rej.	40.4650	Rej.	0.01	Rej.
W_{LB}	3.191	Not rej.	0.02	Rej.	2.0726	Not rej.	0.10	Not rej.	8.9015	Rej.	0.01	Rej.
$W_{DV}\,(bound)$	3.3415	Not rej.	0.16	Not rej.	2.5492	Not rej.	0.23	Not rej.	18.9391	Rej.	0.01	Rej.
Panel E: Testing for noncausality at Horizon $H=5$												
	$H_0: \ Saving \overset{(5)}{\not\to} Growth$				$H_0: \ Investment \overset{(5)}{\not\to} Growth$				$H_0: \ Growth \overset{(5)}{\not\to} Saving$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0204$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.00331$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.000001019$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	4.8147	Not rej.	0.65	Not rej.	11.7251	Rej.	0.39	Not rej.	40.5742	Rej.	0.12	Not rej.
W_{LB}	3.1948	Not rej.	0.01	Rej.	1.9985	Not rej.	0.09	Not Rej.	8.6573	Rej.	0.01	Rej.
$W_{DV}\,(bound)$	3.331	Not rej.	0.21	Not rej.	3.4582	Not Rej.	0.19	Not rej.	18.3798	Rej.	0.01	Rej.

Note: For panels A, B, C, D and E, the full-rank statistics, *i.e.* W and W_{DV} use the $\chi^2_{1-\alpha}(1) = 3.84$, $\chi^2_{1-\alpha}(2) = 5.99$, $\chi^2_{1-\alpha}(3) = 7.81$, $\chi^2_{1-\alpha}(4) = 9.49$ and $\chi^2_{1-\alpha}(5) = 11.07$ respectively.

W_{LB} is based on the $\chi^2_{1-\alpha}(1) = 3.84$, and $\chi^2_{1-\alpha}(2) = 5.99$ because the other restrictions are redundant and are dropped for the reduced-rank statistic.

Table 2. Testing for noncausality between FDI, Investment and Growth

$H_0: y_t \overset{(H)}{\not\to} x_t$, nominal size = 0.05, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$												
Panel A: Testing for noncausality at Horizon $H = 1$												
	$H_0: FDI \overset{(1)}{\not\to} Growth$				$H_0: Investment \overset{(1)}{\not\to} Growth$				$H_0: Growth \overset{(1)}{\not\to} FDI$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 16.7032$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 2.1792$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0673$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	0.1076	Not rej.	0.75	Not rej.	22.1267	Rej.	0.01	Rej.	0.7103	Not rej.	0.40	Not rej.
W_{LB}	0.1076	Not rej.	0.75	Not rej.	22.1267	Rej.	0.01	Rej.	0.7103	Not rej.	0.40	Not rej.
W_{DV} (*bound*)	0.1076	Not rej.	0.75	Not rej.	22.1267	Rej.	0.01	Rej.	0.4784	Not rej.	0.40	Not rej.
Panel B: Testing for noncausality at Horizon $H = 2$												
	$H_0: FDI \overset{(2)}{\not\to} Growth$				$H_0: Investment \overset{(2)}{\not\to} Growth$				$H_0: Growth \overset{(2)}{\not\to} FDI$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 91.6564$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 2.8886$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.00153$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	0.1179	Not rej.	0.85	Not rej.	23.801	Rej.	0.01	Rej.	0.7592	Not rej.	0.50	Not rej.
W_{LB}	0.1161	Not rej.	0.74	Not rej.	13.596	Rej.	0.01	Rej.	0.7592	Not rej.	0.40	Not rej.
W_{DV} (*bound*)	0.1179	Not rej.	0.85	Not rej.	23.801	Rej.	0.01	Rej.	0.4979	Not rej.	0.47	Not rej.
Panel C: Testing for noncausality at Horizon $H = 3$												
	$H_0: FDI \overset{(3)}{\not\to} Growth$				$H_0: Investment \overset{(3)}{\not\to} Growth$				$H_0: Growth \overset{(3)}{\not\to} FDI$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 2.5523$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.7248$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.00858$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	0.1179	Not rej.	0.86	Not rej.	49.3885	Rej.	0.01	Rej.	0.8615	Not rej.	0.44	Not rej.
W_{LB}	0.1145	Not rej.	0.74	Not rej.	15.4611	Rej.	0.01	Rej.	0.7479	Not rej.	0.39	Not rej.
W_{DV} (*bound*)	0.1179	Not rej.	0.85	Not rej.	49.3885	Rej.	0.01	Rej.	0.5435	Not rej.	0.44	Not rej.
Panel D: Testing for noncausality at Horizon $H = 4$												
	$H_0: FDI \overset{(4)}{\not\to} Growth$				$H_0: Investment \overset{(4)}{\not\to} Growth$				$H_0: Growth \overset{(4)}{\not\to} FDI$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0086$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0044$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.000000000138$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	0.1221	Not rej.	0.87	Not rej.	51.5662	Rej.	0.01	Rej.	0.974	Not rej.	0.48	Not rej.
W_{LB}	0.1182	Not rej.	0.74	Not rej.	12.7066	Rej.	0.01	Rej.	0.7352	Not rej.	0.40	Not rej.
W_{DV} (*bound*)	0.1185	Not rej.	0.85	Not rej.	50.7662	Rej.	0.01	Rej.	0.5612	Not rej.	0.43	Not rej.
Panel E: Testing for noncausality at Horizon $H = 5$												
	$H_0: FDI \overset{(5)}{\not\to} Growth$				$H_0: Investment \overset{(5)}{\not\to} Growth$				$H_0: Growth \overset{(5)}{\not\to} FDI$			
	$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0000024$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.00063$				$\det(\hat{\Sigma}_{\hat{\psi}}) = 0.0000000000000157$			
Statistics	Asy		Simul-E		Asy		Simul-E		Asy		Simul-E	
$\hat{V}(\hat{\theta})$	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision	numerical value	Decision
W	50.5629	Rej.	0.09	Not rej.	84.2611	Rej.	0.02	Rej.	0.9898	Not Rej.	0.61	Not rej.
W_{LB}	0.1147	Not rej.	0.74	Not rej.	12.953	Rej.	0.01	Rej.	0.7343	Not rej.	0.40	Not rej.
W_{DV} (*bound*)	0.1696	Not rej.	0.84	Not rej.	60.035	Rej.	0.01	Rej.	0.5629	Not rej.	0.43	Not rej.

Note: For panels A, B, C, D and E, the full-rank statistics, *i.e.* W and W_{DV} use the $\chi^2_{1-\alpha}(1) = 3.84$, $\chi^2_{1-\alpha}(2) = 5.99$, $\chi^2_{1-\alpha}(3) = 7.81$, $\chi^2_{1-\alpha}(4) = 9.49$ and $\chi^2_{1-\alpha}(5) = 11.07$ respectively.

W_{LB} is based on the $\chi^2_{1-\alpha}(1) = 3.84$, and $\chi^2_{1-\alpha}(2) = 5.99$ because the other restrictions are redundant and are dropped for the reduced-rank statistic.

References

Ahn, S. and Schmidt, P. (1995), 'Efficient estimation of models for dynamic panel data', *Journal of Econometrics* **68**, 5–28.

Al-Sadoon, M. M. (2017), 'A unifying theory of tests of rank', *Journal of Econometrics* **199**, 49–62.

Anderson, T. W. (1963), 'Asymptotic theory for principal component analysis', *Annals of Mathematical Statistics* **34**, 122–148.

Anderson, T. W. (1987), The asymptotic distribution of characteristic roots and vectors in multivariate components of variance, Technical report 23, Econometric workshop, Stanford University.

Andrews, D. W. K. (1987), 'Asymptotic results for generalized Wald tests', *Econometric Theory* **3**, 348–358.

Antoine, B. and Renault, E. (2009), 'Efficient GMM with nearly-weak identification', *The Econometric Journal* **12**, 135–171.

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to emplyment equations', *Review of Economic Studies* **58**, 277–297.

Bickel, P. J. and Levina, E. (2004), 'Some theory for Fisher's linear discriminat function, "naive Bayes", and some more alternatives when there are many more variables than observations', *Bernoulli* **10**, 989–1010.

Bickel, P. J. and Levina, E. (2008*a*), 'Covariance regularization by thresholding', *Annals of Statistics* **36**, 2577–2604.

Bickel, P. J. and Levina, E. (2008*b*), 'Regularized estimation of large covariance matrices', *Annals of Statistics* **36**, 199–227.

Blundell, R., Bond, S. and Windmeijer, F. (2000), 'Estimation in dynamic panel data models: improving on the performance of the standard GMM estimators', *Advances in Econometrics* **15**.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate ARMA models', *Journal of the American Statistical Association* **87**(420), 1082–1090.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994), 'Simplified conditions for non-causality between two vectors in multivariate ARMA models', *Journal of Econometrics* **63**, 271–287.

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

Bühlmann, P. (2013), 'Statistical significance in high-dimensional linear models', *Bernoulli* **19**(4), 1212–1242.

Caner, M. (2008), 'Nearly-singular design in GMM and generalized empirical likelihood estimators', *Journal of Econometrics* **144**, 511–523.

Carrasco, M. (2012), 'A regularization approach to the many instrument problem', *Journal of Econometrics* **170**, 383–398.

Carrasco, M., Chernov, M., Florens, J.-P. and Ghysels, E. (2007), 'Efficient estimation of general dynamic models with a continuum of moment conditions', *Journal of Econometrics* **140**, 529–573.

Carrasco, M. and Florens, J.-P. (2000), 'Generalization of GMM to a continuum of moment conditions', *Econometric Theory* **16**, 797–834.

Carrasco, M., Florens, J.-P. and Renault, E. (2007), *Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization*, Vol. 6 of *Handbook of Econometrics*, North Holland, Amsterdam.

Carrasco, M. and Noumon, N. (2011), Optimal portfolio selection using regularization, CIREQ working paper, University of Montreal, University of Montreal.

Cragg, J. and Donald, S. (1996), 'On the asymptotic properties of LDU-based tests of the rank of a matrix', *Journal of the American Statistical Association* **91**, 1301–1309.

Cragg, J. and Donald, S. (1997), 'Inferring the rank of a matrix', *Journal of Econometrics* **76**, 223–250.

Doran, H. E. and Schmidt, P. (2006), 'GMM estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model', *Journal of Econometrics* **133**, 387–409.

Dovonon, P. and Renault, E. (2009), GMM overidentification test with first order underidentification. CIREQ working paper, University of Montreal.

Drton, M., Massam, H. and Olkin, I. (2008), 'Moments of minor of Wishart matrices', *The Annals of Statistics* **36**, 2261–2283.

Drton, M., Sturmfels, B. and Sullivant, S. (2007), 'Algebraic factor analysis: Tetrads, pentads, and beyond', *Probability Theory and Related Fields* **138**(3), 463–493.

Duchesne, P. and Francq, C. (2015), 'Multivariate hypothesis testing using generalized and 2-inverses – with applications', *Statistics* **49**(3), 475–496.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2006), 'Monte carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**, 443–477.

Dufour, J.-M., Pelletier, D. and Renault, É. (2006), 'Short run and long run causality in time series: Inference', *Journal of Econometrics* **132**(2), 337–362.

Dufour, J.-M. and Renault, E. (1998), 'Short run and long run causality in time series: theory', *Econometrica* **66**, 1099–1125.

Dufour, J.-M., Renault, E. and Zinde-Walsh, V. (2017), Wald tests when restrictions are locally singular. CIREQ working paper, McGill University.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**, 1351–1365.

Dufour, J.-M. and Taamouti, M. (2007), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* **139**, 133–153.

Dufour, J.-M. and Valéry, P. (2009), 'Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models', *Journal of Econometrics* **150**, 193–206.

Eaton, M. L. (2007), *Multivariate Statistics: A Vector Space Approach*, Institute of Mathematical Statistics, Lecture Notes - Monograph Series, volume 53, printed in Lithuania.

Eaton, M. L. and Tyler, D. E. (1991), 'On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix', *Annals of Statistics* **19**(1), 260–271.

Eaton, M. L. and Tyler, D. E. (1994), 'The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis', *Journal of Multivariate Analysis* **50**, 238–264.

Fan, J., Fan, Y. and Lv, J. (2008), 'High Dimensional Covariance Matrix Estimation Using a Factor Model', *Journal of Econometrics* **147**, 186–197.

Fan, J., Liao, Y. and Mincheva, M. (2011), 'High Dimensional Covariance Matrix Estimation in Approximate Factor Models', *Annals of Statistics* **39**, 3320–3356.

Galbraith, J. W. and Zinde-Walsh, V. (1997), 'On some simple autoregression-based estimation and identification techniques for ARMA models', *Biometrika* **84**, 685–696.

Gallant, A. R. and Tauchen, G. (1989), 'Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications', *Econometrica* **57**, 1091–1120.

Gill, L. and Lewbel, A. (1992), 'Testing the rank and definiteness of estimated matrices with applications to factor, state-space, and ARMA models', *Journal of the American Statistical Association* **87**, 766–776.

Glonek, G. F. V. (1993), 'On the behaviour of Wald statistics for the disjunction of two regular hypotheses', *Journal of The Royal Statistical Society Series B* **55**(3), 749–755.

Gonzalo, J. and Lee, T.-H. (1998), 'Pitfalls in testing for long run relationships', *Journal of Econometrics* **86**, 129–154.

Gouriéroux, C. and Jasiak, J. (2009), Size distortion in the analysis of volatility and covolatility effects, Technical report, CREST and University of Toronto.

Gouriéroux, C., Monfort, A. and Renault, E. (1989), 'Testing for common roots', *Econometrica* **57**, 171–185.

Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press Inc., New Jersey.

Hodges, J. L. and Lehmann, E. L. (1950), 'Some Problems in Minimax Point Estimation', *The Annals of Mathematical Statistics* **21**, 182–197.

Hoerl, A. and Kennard, R. (1970), 'Ridge regression biased estimation for nonorthogonal problems', *Technometrics* **12**, 55–67.

Houthakker, H. (1961), 'An international comparison of personal saving', *Bulletin of the International Statistical Institute* **38**, 55–60.

Houthakker, H. S. (1965), *On some determinants of Saving in developed and underdeveloped countries*, Problems in Economic Development, ed. by E. Robinson, MacMillan, London.

Kan, R. and Robotti, C. (2009), 'Model comparison using the Hansen-Jagannathan distance', *Review of Financial Studies* **22**, 3449–3490.

Kasahara, H. and Shimotsu, K. (2014), 'Non-parametric identification and estimation of the number of components in multivariate mixtures', *Journal of the Royal Statistical Society: Series B* **76**, 97–111.

Kato, T. (1966), *Perturbation theory for linear operators*, Springer-Verlag, New York.

Kleibergen, F. and Paap, R. (2006), 'Generalized reduced rank tests using the singular value decomposition', *Journal of Econometrics* **133**, 97–126.

Knight, K. (2008), 'Shrinkage estimation for nearly-singular designs', *Econometric Theory* **24**, 323–337.

Knight, K. and Fu, W. (2000), 'Asymptotics for LASSO-type estimators', *Annals of Statistics* **28**, 1356–1378.

LeCam, L. (1953), 'On some asymptotic properties of maximum likelihood estimates and related Bayes's estimates', *Univ. of Calif. Publ. in Statist.* **1**, 277–330.

Ledoit, O. and Wolf, M. (2004a), 'Honey, I shrunk the sample covariance matrix', *Journal of Portfolio Management* **30**, 110–119.

Ledoit, O. and Wolf, M. (2004b), 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis* **88**, 365–411.

Leeb, H. and Pötscher, B. M. (2003), 'The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations', *Econometric Theory* **19**, 100–142.

Leeb, H. and Pötscher, B. M. (2005), 'Model selection and inference: Facts and fiction', *Econometric Theory* **21**, 21–59.

Leeb, H. and Pötscher, B. M. (2008), 'Sparse estimators and the oracle property, or the return of Hodges' estimator', *Journal of Econometrics* **142**, 201–211.

Lehmann, E. L. and Casella, G. (1998), *Theory of Point Estimation*, 2nd edn, Springer-Verlag, New York.

Levine, R. E. and Renelt, D. (1992), 'A sensitivity analysis of cross-country growth regressions', *American Economic Review* **82**, 942–963.

Li, W. K. and McLeod, A. (1981), 'Distribution of the residual autocorrelations in multivariate ARMA time series models', *Journal of the Royal Statistical Society, Series B* **43**, 231–239.

Ljung, G. M. (1986), 'Diagnostic testing of univariate time series models', *Biometrika* **73**, 725–730.

Lukacs, E. (1975), *Stochastic convergence*, second edn, Academic Press, inc., New York.

Luo, J. (2010), 'The discovery of mean square error consistency of a ridge estimator', *Statistics and Prpbability letters* **80**, 343–347.

Luo, J. and Zuo, Y.-J. (2011), 'A new test for large dimensional regression coefficients', *Open Journal of Statistics* **1**, 212–216.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.

Lütkepohl, H. and Burda, M. M. (1997), 'Modified wald tests under nonregular conditions', *Journal of Econometrics* **78**, 315–332.

MacKinnon, D. P. (2008), *Introduction to Statistical Mediation Analysis*, Lawrence Erlbaum Associates, New York and London.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. and Sheets, V. (2002), 'A comparison of methods to test mediation and other intervening variable effects', *Journal of Personality and social Psychology* **7**(1), 83–104.

MacKinnon, D. P., Lockwood, C. M. and Williams, J. (2006), 'Confidence limits for the indirect effect: Distribution of the product and resampling methods', *Multivariate Behavioral Research* **39**(1), 99–128.

Marin, L. (1996), Efectos de la diversificacion internacional sobre los limites de volatilidad de las tasas de descuento estocasticas, Master's thesis, CEMFI.

Modigliani, F. (1970), *The life cycle hypothesis of saving and inter-country differences in the saving ratio*, Induction, Growth, and Trade: Essays in Honor of Sir Roy Harrod, ed. by W. A. Eltis., Clarendon Press, London.

Moore, D. S. (1977), 'Generalized Inverses, Wald's Method, and the Construction of Chi-Square Tests of Fit', *Journal of the American Statistical Association* **72**(357), 131–137.

Newey, W. K. and McFadden, D. (1994), *Large Sample Estimation and Hypothesis Testing*, Vol. 4 of *Handbook of Econometrics*, North Holland, Amsterdam.

Peñaranda, F. and Sentana, E. (2012), 'Spanning tests in return and stochastic discount factor mean-variance frontiers: a unifying approach', *Journal of Econometrics* **170**, 303–324.

Pötscher, B. (1985), 'The behaviour of the Lagrange Multiplier test in testing the orders of an ARMA-model', *Metrika* **32**, 129–150.

Rao, R. C. and Mitra, S. K. (1971), *Generalized Inverses of Matrices and its Applications*, John Wiley & Sons, inc., Toronto.

Rellich, F. (1953), *Perturbation theory of eigenvalues problems*, Lecture Notes, New York University.

Robin, J.-M. and Smith, R., J. (2000), 'Tests of rank', *Econometric Theory* **16**(2), 151–175.

Sargan, J. D. (1980), 'Some tests of dynamic specification for a single equation', *Econometrica* **48**(4), 879–897.

Satchachai, P. and Schmidt, P. (2008), 'GMM with more moment conditions than observations', *Economics Letters* **99**, 252–255.

Sims, C. A., Stock, J. H. and Watson, M. W. (1990), 'Inference in linear time series model with some unit roots', *Econometrica* **58**, 113–144.

Sobel, M. E. (1982), 'Asymptotic confidence intervals for indirect effects in structural equation models', *Sociological Methodology* **13**, 290–312.

Sobel, M. E. (1986), 'Some new results on indirect effects and their standard errors in covariance structure models', *Sociological Methodology* **16**, 159–186.

Stock, J. H. and Wright, J. (2000), 'GMM with weak identification', *Econometrica* **68**, 1055–1096.

Stock, J. H., Wright, J. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic statistics* **20**, 518–529.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the Lasso', *Journal of the Royal Statistical Society Series B* **58**, 267–288.

Tofighi, D. and MacKinnon, D. P. (2016), 'Monte Carlo confidence intervals for complex functions of indirect effects', *Structural Equation modeling: A Multidisciplinary Journal* **23**(2), 194–205.

Tyler, D. E. (1981), 'Asymptotic inference for eigenvectors', *Annals of Statistics* **9**(4), 725–736.

Xu, K.-L. (2016), Global estimation and inference of regression discontinuity design with discrete ordered or duration outcomes, Technical report, Texas AM University.

Zwiernik, P. and Smith, J. Q. (2012), 'Tree cumulants and the geometry of binary tree models', *Bernoulli* **18**, 290–321.

A. Appendix: Asymptotic distribution of eigenvalues

In this section, we summarize general results on the sample eigenvalue behavior established by Eaton and Tyler (1991, 1994) . Before establishing convergence results for the regularized covariance matrices and the regularized tests statistics, we shall first study the convergence rate of the eigenvalues in the general case where the covariance matrix may be singular with (possibly) multiple eigenvalues. To do so, we shall apply a general result given by Eaton and Tyler (1994) where they generalize classical results due to Anderson (1963, 1987) on the behavior of the sample roots (of a determinantal equation). Specifically under relatively weak conditions, Eaton and Tyler (1994) show the following: if a sequence of random $(p \times q)-$matrices $\boldsymbol{\Sigma}_n$ satisfies the condition $b_n(\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} Q$ where $\boldsymbol{\Sigma}$ is a nonstochastic matrix, then the sample eigenvalues will have the same convergence rate, with $b_n[\Psi(\boldsymbol{\Sigma}_n) - \Psi(\boldsymbol{\Sigma})] \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \left[H_D\left(\frac{1}{2}[Q'_{11} + Q_{11}]\right), \quad \Psi(Q_{22})\right]'$. $H_D(.)$ and $\Psi(.)$ are vector-valued functions stacking the eigenvalues of the corresponding objects. A more detailed definition of those vectors will follow. For our purpose, the convergence rate b_n of the sample eigenvalues is the only thing we need in deriving the convergence property of the regularized covariance matrices that employ a varying threshold c_n.

Let $d_1 > d_2 > \cdots > d_k$ denote the distinct eigenvalues of a $q \times q$ symmetric matrix C and let m_i be the multiplicity of d_i, $i = 1, \ldots, k$, $1 \leq k \leq q$. Given the eigenvalue multiplicities of C, it is possible to partition the matrix C into blocks such as C_{ii} is the $m_i \times m_i$ diagonal block of C and C_{ij} the $m_i \times m_j$ off-diagonal blocks, $i, j = 1, \ldots, k$. Thus, a function H on $q \times q$ symmetric matrices can be defined by

$$H(C) = \begin{pmatrix} \rho(C_{11}) \\ \rho(C_{22}) \\ \vdots \\ \rho(C_{kk}) \end{pmatrix} \tag{A.1}$$

$H(C)$ takes values in $\mathbb{R}^q$ and $\rho(C_{ii})$ consists of the m_i-vector of ordered eigenvalues of the diagonal block C_{ii}, $i = 1, \ldots, k$. Let Γ be an orthogonal matrix such that

$$\Gamma A \Gamma' = D, \tag{A.2}$$

where the diagonal matrix D consists of the ordered eigenvalues of a nonrandom symmetric matrix A. Eaton and Tyler (1991) first establish the distributional theory for symmetric matrices before extending it to general $p \times q$ matrices.

Lemma A.1 DISTRIBUTION OF THE EIGENVALUES OF A SYMMETRIC SQUARE MATRIX. *Let S_n be a sequence of $q \times q$ random symmetric matrices. Suppose there exists a nonrandom symmetric matrix A and a sequence of constants $b_n \to +\infty$ such that*

$$W_n = b_n(S_n - A) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} W \, . \tag{A.3}$$

Then

$$b_n\big(\rho(S_n) - \rho(A)\big) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} H\big(\Gamma W \Gamma'\big) \, . \tag{A.4}$$

For any $p \times q$ real matrix $\boldsymbol{\Sigma}$, the $\Psi(.)$ function is a vector-valued function that stacks the eigenvalues of

the corresponding object as defined below:

$$\Psi(\boldsymbol{\Sigma}) = f(\rho(\boldsymbol{\Sigma}'\boldsymbol{\Sigma})) = \begin{pmatrix} \sqrt{\xi_1} \\ \vdots \\ \sqrt{\xi_q} \end{pmatrix} \quad \text{with } f(x) = \begin{pmatrix} \sqrt{x_1} \\ \vdots \\ \sqrt{x_q} \end{pmatrix} \tag{A.5}$$

where $\xi_1 \geq \cdots \geq \xi_q > 0$ are the eigenvalues of $\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$. Let

$$T = \big(df(\xi)\big) = \frac{1}{2}\text{diag}(\xi_1^{-1/2}, \ldots, \xi_q^{-1/2}) \,. \tag{A.6}$$

In the first part of the theorem below, we gather the special cases where the matrix $\boldsymbol{\Sigma}$ may have rank $r = 0$ or $r = q$ before giving the general result in the second part. In the second part of the theorem, write the $p \times q$ matrix $\boldsymbol{\Sigma}$ in the form

$$\boldsymbol{\Sigma} = \Gamma_1' \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Gamma_2' \tag{A.7}$$

where Γ_1 (Γ_2) is a $p \times p$ (resp. $q \times q$) orthogonal matrix, and D is a $r \times r$ diagonal matrix. D consists of the strictly positive singular values of $\boldsymbol{\Sigma}$. Partition the matrix $\boldsymbol{\Sigma}_n$ as

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \boldsymbol{\Sigma}_{n11} & \boldsymbol{\Sigma}_{n12} \\ \boldsymbol{\Sigma}_{n21} & \boldsymbol{\Sigma}_{n22} \end{pmatrix} \tag{A.8}$$

where $\boldsymbol{\Sigma}_{n11}$ is $r \times r$, $\boldsymbol{\Sigma}_{n12}$ is $r \times (q-r)$, $\boldsymbol{\Sigma}_{n21}$ is $(p-r) \times r$ and $\boldsymbol{\Sigma}_{n22}$ is $(p-r) \times (q-r)$. Partition the random limit matrix Q accordingly. The $r \times r$ diagonal matrix $D = diag(\xi_1^{1/2}, \ldots, \xi_r^{1/2})$ defines a function H_D on $r \times r$ symmetric matrices. Let $T_D = \frac{1}{2} diag(\xi_1^{-1/2}, \ldots, \xi_r^{-1/2})$. The general case $1 \leq r < q$ can be thought as gluing together the two special cases $r = 0$ and $r = q$.

Theorem A.2 DISTRIBUTION OF THE EIGENVALUES OF RECTANGULAR MATRICES IN THE GENERAL CASE. *Let $\Psi(\cdot)$ be defined as in* (A.5)*, and suppose Assumption 2.2 holds with $A_n = \boldsymbol{\Sigma}_n$ and $A = \boldsymbol{\Sigma}$.*

i) If $\boldsymbol{\Sigma} = 0$, then

$$b_n\big(\Psi(\boldsymbol{\Sigma}_n) - \Psi(\boldsymbol{\Sigma})\big) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \Psi(Q) \,. \tag{A.9}$$

ii) If $\boldsymbol{\Sigma}$ has full rank q, then

$$b_n\big(\Psi(\boldsymbol{\Sigma}_n) - \Psi(\boldsymbol{\Sigma})\big) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} TH\big(\Gamma\big[\boldsymbol{\Sigma}'Q + Q'\boldsymbol{\Sigma}\big]\Gamma'\big) \tag{A.10}$$

where H, Γ and T are defined in (A.1),(A.2) *and* (A.6).

iii) If $\text{rank}(\boldsymbol{\Sigma}) = r$, $1 \leq r < q$, *then*

$$b_n\big[\Psi(\boldsymbol{\Sigma}_n) - \Psi(\boldsymbol{\Sigma})\big] \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \begin{bmatrix} H_D\big(\frac{1}{2}[Q'_{11} + Q_{11}]\big) \\ \Psi(Q_{22}) \end{bmatrix} \tag{A.11}$$

where $Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$ is a well-defined random element, with Q_{11} being an $r \times r$ matrix and Q_{22} a $(p-r) \times (q-r)$ matrix. The $r \times r$ diagonal matrix $D = diag(\xi_1^{1/2}, \ldots, \xi_r^{1/2})$ consisted of the strictly

positive singular values of $\boldsymbol{\Sigma}$ defines a function H_D on $r \times r$ symmetric matrices as H is defined in (A.1) *on $q \times q$ symmetric matrices.*

Note the generality of the theorem that allows for general convergence rates rather than the standard root-n one; it does not critically hinge upon the normal limiting distribution for the sample eigenvalues, a desirable feature for positive eigenvalues. For our purposes, we do not need to know the whole distribution but only the convergence rate b_n of the sample eigenvalues to establish the convergence property of the regularized inverse when c varies with the sample size. Again, the knowledge of the sample convergence rate is unnecessary for the regularized inverse based upon the fixed threshold case. See Eaton and Tyler (1994, Propositions 3.1 and 3.4 and Theorem 4.2) for a proof of the theorem.

The rare cases where the asymptotic distribution of the empirical eigenvalues could be uniform would correspond to situations where all the population eigenvalues are greater than zero (Theorem A.2, case ii), or all are equal to zero (Theorem A.2, case i). Otherwise, the distribution cannot be uniform: the inspection of Theorem A.2 case iii that examines a strictly positive but incomplete rank shows that the structure of the distribution is different on the first r singular values than on the last $q-r$ ones. Similarly, the finite-sample distribution of the sample eigenvalues will depend on the rank of the sample matrix; if the sample matrix has full rank, the probability to have a zero sample eigenvalue is zero. Yet, the number of the sample eigenvalues greater than the threshold (c or c_n) will vary with the sample size.

B. Appendix: Proofs

PROOF OF LEMMA 2.1 By Assumption 2.2, $\left(b_n(A_n - A) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} Q\right) \Rightarrow \left(A_n \underset{n\to\infty}{\overset{p}{\to}} A\right)$ and by Assumption 2.1 we have:

$$\begin{aligned} W_n(\psi_0) &= \left[H_n(\hat{\psi}_n - \psi_0)\right]'(A_n - A)\left[H_n(\hat{\psi}_n - \psi_0)\right] + \left[H_n(\hat{\psi}_n - \psi_0)\right]' A \left[H_n(\hat{\psi}_n - \psi_0)\right] \\ &\underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X'0X + X'AX \,. \end{aligned}$$

□

PROOF OF PROPOSITION 4.1 Using Definition 4.1 and (4.3), we have $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^R(c) = V\Lambda V'V\Lambda^\dagger(c)V' = V\Lambda\Lambda^\dagger(c)V'$ since the V_i's are orthogonal. For all λ, $0 \le \lambda g(\lambda;c) \le 1$, so that $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^R(c) = V diag\left[\lambda_j g(\lambda_j;c)\right]_{j=1,\cdots,q} V' \le I_q$. Regarding *ii*), we have:

$$T\boldsymbol{\Sigma}^R(c)T' = V\Lambda^{1/2}V'V\Lambda^\dagger(c)V'V\Lambda^{1/2}V' = V\Lambda^{1/2}\Lambda^\dagger(c)\Lambda^{1/2}V' = V\,\mathrm{diag}\left[\lambda_j g(\lambda_j;c)\right]_{j=1,\ldots,q} V' \le I_q$$

since $0 \le \lambda g(\lambda;c) \le 1$ for all λ. Regarding *iii*), we have:

$$\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^R(c)\boldsymbol{\Sigma} \ge 0 \Leftrightarrow \boldsymbol{\Sigma}\left(I_q - \boldsymbol{\Sigma}^R(c)\boldsymbol{\Sigma}\right) \ge 0 \Rightarrow I_q - \boldsymbol{\Sigma}^R(c)\boldsymbol{\Sigma} \ge 0$$

since $\boldsymbol{\Sigma}$ is semidefinite positive. The last implication holds by *i*). As for *iv*), for all $\lambda \ge 0$, $g(\lambda;c)$ is bounded; if $g(\lambda;c) > 0$, we have $\left(\lambda g(\lambda;c) \le 1\right) \Rightarrow \left(0 < g(\lambda;c) \le \frac{1}{\lambda} \le \infty\right) \Rightarrow \left([g(\lambda;c)]^{-1} - \lambda \ge 0\right)$. Hence, $\left(\boldsymbol{\Sigma}^R(c)\right)^{-1} - \boldsymbol{\Sigma} = V diag\left[\left(g(\lambda_j;c)\right)^{-1} - \lambda_j\right]_{j=1,\cdots,q} V' \ge 0$. Finally for *v*), the rank is given by the number of eigenvalues greater than zero (which accounts for multiplicity by repeating the eigenvalue as many times as its multiplicity). As $\boldsymbol{\Sigma}^R(c) = V diag\left[g(\lambda_j;c)\right]_{j=1,\cdots,q} V'$, hence $\left(\lambda > 0 \Rightarrow g(\lambda;c) > 0\right) \Rightarrow \left(rank\left(\boldsymbol{\Sigma}^R(c)\right) \ge rank(\boldsymbol{\Sigma})\right)$. □

PROOF OF LEMMA 6.2 If $\boldsymbol{\Sigma}_n \overset{a.s.}{\to} \boldsymbol{\Sigma}$, then the event $A = \{\omega : \boldsymbol{\Sigma}_n(\omega) \underset{n\to\infty}{\to} \boldsymbol{\Sigma}\}$ has probability one, *i.e.* $\mathbb{P}(A) = 1$. For any $\omega \in A$, we have by Lemma 6.1:

$$[\boldsymbol{\Sigma}_n(\omega) \underset{n\to\infty}{\to} \boldsymbol{\Sigma}] \Rightarrow [\lambda_j(\boldsymbol{\Sigma}_n(\omega)) \to \lambda_j(\boldsymbol{\Sigma}), \quad j = 1, \ldots, J] .$$

Denoting $B = \{\omega : \lambda_j(\boldsymbol{\Sigma}_n(\omega)) \underset{n\to\infty}{\to} \lambda_j(\boldsymbol{\Sigma})\}$, we have $A \subseteq B$, hence we have with probability one result *i*). By the same argument, we have result *ii*) for the eigenprojections. □

PROOF OF LEMMA 6.3 If $\boldsymbol{\Sigma}_n \overset{p}{\to} \boldsymbol{\Sigma}$ with eigenvalues $\{\lambda_j(\boldsymbol{\Sigma}_n)\}$, then every subsequence $\{\boldsymbol{\Sigma}_{n_k}\}$ with eigenvalues $\{\lambda(\boldsymbol{\Sigma}_{n_k})\}$, also satisfies $\boldsymbol{\Sigma}_{n_k} \overset{p}{\to} \boldsymbol{\Sigma}$. By Lukacs (1975, theorem 2.4.3, page 48), there exists $\{\boldsymbol{\Sigma}_{m_l}\} \subseteq \{\boldsymbol{\Sigma}_{n_k}\}$ such that $\boldsymbol{\Sigma}_{m_l} \overset{a.s.}{\to} \boldsymbol{\Sigma}$. Hence by Lemma 6.2, we have:

1. $\lambda_j(\boldsymbol{\Sigma}_{m_l}) \overset{a.s.}{\to} \lambda_j(\boldsymbol{\Sigma})$;
2. $P_{j,t}(\boldsymbol{\Sigma}_{m_l}) \overset{a.s.}{\to} P_{j,t}(\boldsymbol{\Sigma})$ provided $\lambda_{j-1}(\boldsymbol{\Sigma}) \neq \lambda_j(\boldsymbol{\Sigma})$ and $\lambda_t(\boldsymbol{\Sigma}) \neq \lambda_{t+1}(\boldsymbol{\Sigma})$.

As $\{\boldsymbol{\Sigma}_{m_l}\} \subseteq \{\boldsymbol{\Sigma}_{n_k}\} \subseteq \{\boldsymbol{\Sigma}_n\}$ with the corresponding eigenvalues $\{\lambda_j(\boldsymbol{\Sigma}_{m_l})\} \subseteq \{\lambda_j(\boldsymbol{\Sigma}_{n_k})\} \subseteq \{\lambda_j(\boldsymbol{\Sigma}_n)\}$, by Lukacs (1975, theorem 2.4.4 page 49) it suffices that every subsequence $\{\lambda_j(\boldsymbol{\Sigma}_{n_k})\}$ of $\{\lambda_j(\boldsymbol{\Sigma}_n)\}$ contains a subsequence $\{\lambda_j(\boldsymbol{\Sigma}_{m_l})\}$ which converges a.s. to get $\lambda_j(\boldsymbol{\Sigma}_n) \overset{p}{\to} \lambda_j(\boldsymbol{\Sigma})$. By the same argument, we have $P_{j,t}(\boldsymbol{\Sigma}_n) \overset{p}{\to} P_{j,t}(\boldsymbol{\Sigma})$. □

PROOF OF PROPOSITION 7.2 If $\boldsymbol{\Sigma}_n \overset{a.s.}{\to} \boldsymbol{\Sigma}$, then by Lemma 6.2 i), we have $\hat{\lambda}_i \overset{a.s.}{\to} d_j, \ \forall i \in I_j$, where $I_j = \{i \in I : \lambda_i = d_j\}$. Under the additional Assumption 7.1, and the a.e. continuity of $g(.,c)$, we have $g(\hat{\lambda}_i; c) \overset{a.s.}{\to} g(d_j; c) \ \forall i \in I_j$. Moreover, by Lemma 6.2 ii), we have $P_{I_j}(\boldsymbol{\Sigma}_n) \overset{a.s.}{\to} P_j(\boldsymbol{\Sigma})$. Hence,

$$\begin{aligned}
\boldsymbol{\Sigma}_n^R(c) &= \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c) = \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \left[g(d_j; c) - g(d_j; c) + \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c) \right] \\
&= \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) g(d_j; c) + \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} \left[g(\hat{\lambda}_i; c) - g(d_j; c) \right] \overset{a.s.}{\to} \sum_{j=1}^{k} P_j(\boldsymbol{\Sigma}) g(d_j; c)
\end{aligned}$$

since $g(d_j; c) = \frac{1}{m(d_j)} \times m(d_j) g(d_j; c) = \frac{1}{m(d_j)} \sum_{i\in I_j} g(d_j; c)$ and $g(\hat{\lambda}_i; c) \overset{a.s.}{\to} g(d_j; c) \ \forall i \in I_j$, and $P_{I_j}(\boldsymbol{\Sigma}_n) = O_p(1)$. □

PROOF OF PROPOSITION 7.3 Using decomposition (4.2) and equation (7.15), we have:

$$\begin{aligned}
\boldsymbol{\Sigma}_n^R(c) &= \sum_{i=1}^{3} \boldsymbol{\Sigma}_{ii,n}^R(c) = \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i, c) \text{ where} \\
\boldsymbol{\Sigma}_{11,n}^R(c) &= \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i, c) \text{ for } d_j > c, \ k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}} \\
\boldsymbol{\Sigma}_{22,n}^R(c) &= P_{I(c)}(\boldsymbol{\Sigma}_n) \frac{1}{m(c)} \sum_{i\in I(c)} g(\hat{\lambda}_i, c), \text{ for } d_j = c \\
\boldsymbol{\Sigma}_{33,n}^R(c) &= \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i, c) \text{ for } d_j < c .
\end{aligned}$$

Note that Assumption 2.2 implies $\boldsymbol{\Sigma}_n \xrightarrow{p} \boldsymbol{\Sigma}$, hence by Lemma 6.3 i) and ii), eigenvalues and total eigenprojections are continuous; together with Assumption 7.1, we have: $\forall\ i \in I_j\ , g(\hat{\lambda}_i, c) \xrightarrow{p} g(d_j; c)$, and $P_{I_j}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_j(\boldsymbol{\Sigma})$. Also,

$$\begin{aligned}\boldsymbol{\Sigma}^R_{11,n}(c) &= \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c) = \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\left[g(d_j; c) - g(d_j; c) + \frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c)\right] \\ &= \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n) g(d_j; c) + \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i; c) - g(d_j; c)\right] \xrightarrow{p} \sum_{j=1}^{k_1} g(d_j; c) P_j(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}^R_{11}(c)\end{aligned}$$

since $g(d_j; c) = \frac{1}{m(d_j)} \times m(d_j) g(d_j; c) = \frac{1}{m(d_j)} \sum_{i\in I_j} g(d_j; c)$, and $P_{I_j}(\boldsymbol{\Sigma}_n) = O_p(1)$. Again by Lemma 6.3 i) and ii, and under Assumption 7.1, if $\lambda_i = d_j = c$, with multiplicity $m(c)$, $g(\hat{\lambda}_i, c) \xrightarrow{p} g(c; c)$, hence $\boldsymbol{\Sigma}^R_{22,n}(c) = P_{I(c)}(\boldsymbol{\Sigma}_n)\frac{1}{m(c)}\sum_{i\in I(c)} g(\hat{\lambda}_i, c) \xrightarrow{p} g(c; c) 1_{\{d_j = c\}} P_{j(c)}(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}^R_{22}(c)$. The proof for $\boldsymbol{\Sigma}^R_{33,n}(c)$ is similar to that of $\boldsymbol{\Sigma}^R_{11,n}(c)$. Hence, $\boldsymbol{\Sigma}^R_n(c) \xrightarrow{p} \boldsymbol{\Sigma}^R(c) = \boldsymbol{\Sigma}^R_{11}(c) + \boldsymbol{\Sigma}^R_{22}(c) + \boldsymbol{\Sigma}^R_{33}(c)$. □

PROOF OF PROPOSITION 8.1 By Proposition 7.3, we have $\boldsymbol{\Sigma}^R_n(c) \xrightarrow{p} \boldsymbol{\Sigma}^R(c)$ and under Assumption 2.1, $X_n \xrightarrow[n\to\infty]{\mathscr{L}} X$, hence $W^R_n(c) = X_n' \boldsymbol{\Sigma}^R_n(c) X_n \xrightarrow[n\to\infty]{\mathscr{L}} X' \boldsymbol{\Sigma}^R(c) X = W^R(c)$. Using representation (7.14) for $\boldsymbol{\Sigma}^R(c)$, and the form $P_j(\boldsymbol{\Sigma}) = B(d_j)B(d_j)'$, we can write:

$$W^R(c) = X'\boldsymbol{\Sigma}^R(c)X = X'\left(\sum_{j=1}^{k} g(d_j; c) P_j(\boldsymbol{\Sigma})\right)X = \sum_{j=1}^{k} g(d_j; c) X' P_j(\boldsymbol{\Sigma}) X = \sum_{j=1}^{k} g(d_j; c) X' B(d_j) B(d_j)' X\ .$$

We can further decompose the overall statistic into three blocks depending on whether the eigenvalues are larger (or smaller) than c, with $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$, *i.e.*,

$$\begin{aligned} W^R_1(c) &= X'\boldsymbol{\Sigma}^R_{11}(c)X = \sum_{j=1}^{k} g(d_j; c) 1_{\{d_j > c\}} X' P_j(\boldsymbol{\Sigma}) X \\ &= \sum_{j=1}^{k_1} g(d_j; c) X' P_j(\boldsymbol{\Sigma}) X = \sum_{j=1}^{k_1} g(d_j; c) X' B(d_j) B(d_j)' X\ . \end{aligned}$$

Similarly, $W^R_2(c) = X'\boldsymbol{\Sigma}^R_{22}(c)X = g(c; c) 1_{\{d_j = c\}} X' P_{j(c)}(\boldsymbol{\Sigma}) X = g(c; c) 1_{\{d_j = c\}} X' B(c) B(c)' X$. And

$$W^R_3(c) = X'\boldsymbol{\Sigma}^R_{33}(c)X = \sum_{j=1}^{k} g(d_j; c) 1_{\{d_j < c\}} X' P_j(\boldsymbol{\Sigma}) X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j; c) X' B(d_j) B(d_j)' X\ .$$

□

PROOF OF COROLLARY 8.2 In the Gaussian case, we have: $B(d_j)'X = x_j$, where $x_j \sim N\left[0, d_j I_{m(d_j)}\right]$, or

equivalently $x_j = \sqrt{d_j}u_j$ with $u_j \sim N\big[0, I_{m(d_j)}\big]$, hence

$$W^R(c) = X'\boldsymbol{\Sigma}^R(c)X = X'\Big(\sum_{j=1}^{k} g(d_j;c)P_j(\boldsymbol{\Sigma})\Big)X = \sum_{j=1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X = \sum_{j=1}^{k} g(d_j;c)d_j u_j'u_j$$

with the three blocks corresponding to

$$W_1^R(c) = X'\boldsymbol{\Sigma}_{11}^R(c)X = \sum_{j=1}^{k_1} g(d_j;c)X'B(d_j)B(d_j)'X = \sum_{j=1}^{k_1} g(d_j;c)d_j u_j'u_j\,,$$

$$W_2^R(c) = X'\boldsymbol{\Sigma}_{22}^R(c)X = g(c;c)1_{\{d_j=c\}}X'B(c)B(c)'X = g(c;c)1_{\{d_j=c\}}cu_j'u_j\,,$$

and $W_3^R(c) = X'\boldsymbol{\Sigma}_{33}^R(c)X = \sum\limits_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X = \sum\limits_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)d_j u_j'u_j\,.$ □

PROOF OF PROPOSITION 8.4 The quantity $H_n\big[\hat{\psi}_n - \psi_0\big]$ can be written as:

$$H_n\big[\hat{\psi}_n - \psi_0\big] = H_n\big[\hat{\psi}_n - \psi_1 + \psi_1 - \psi_0\big] = H_n[\hat{\psi}_n - \psi_1\big] + H_n\big[\psi_1 - \psi_0\big]\,. \tag{B.1}$$

As $X_n = H_n[\hat{\psi}_n - \psi_1\big]$ satisfies Assumption 2.1, we have

$$\begin{aligned}
W_n^R(c) &= \big[H_n(\hat{\psi}_n - \psi_1) + H_n(\psi_1 - \psi_0)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\hat{\psi}_n - \psi_1) + H_n(\psi_1 - \psi_0)\big] \\
&= \big[H_n(\hat{\psi}_n - \psi_1)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\hat{\psi}_n - \psi_1)\big] + \big[H_n(\hat{\psi}_n - \psi_1)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_1 - \psi_0)\big] \\
&\quad + \big[H_n(\psi_1 - \psi_0)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\hat{\psi}_n - \psi_1)\big] + \big[H_n(\psi_1 - \psi_0)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_1 - \psi_0)\big] \\
&= X_n'\boldsymbol{\Sigma}_n^R(c)X_n + X_n'\boldsymbol{\Sigma}_n^R(c)(H_n\Delta) + (H_n\Delta)'\boldsymbol{\Sigma}_n^R(c)X_n + (H_n\Delta)'\boldsymbol{\Sigma}_n^R(c)(H_n\Delta) \\
&\xrightarrow[n\to\infty]{\mathscr{L}} X'\boldsymbol{\Sigma}^R(c)X + X'\boldsymbol{\Sigma}^R(c)(H_n\Delta) + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)X + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)(H_n\Delta) \;\to\; \infty
\end{aligned} \tag{B.2}$$

since $X_n \xrightarrow[n\to\infty]{\mathscr{L}} X$, $\boldsymbol{\Sigma}_n^R(c) \xrightarrow{p} \boldsymbol{\Sigma}^R(c)$, and the typical (i,j)-element of $(H_n\Delta)'\boldsymbol{\Sigma}^R(c)(H_n\Delta)$ is given by: $\Delta_j h_{n,ij}\boldsymbol{\Sigma}_{ij}^R(c)h_{n,ji}\Delta_i = h_{n,ij}\underbrace{\Delta_j\boldsymbol{\Sigma}_{ij}^R(c)\Delta_i}_{O(1)}h_{n,ji} \underset{n\to\infty}{\to} \infty$ as $h_{n,ij}$ grows to infinity for $i,j = 1,...,q$. Since this quantity $h_{n,ij}\Delta_j\boldsymbol{\Sigma}_{ij}^R(c)\Delta_i h_{n,ji} \underset{n\to\infty}{\to} \infty$ with probability 1, hence $W_n^R(c)$ converges to infinity with probability 1. The quantity

$$X'\boldsymbol{\Sigma}^R(c)X + 2X'\boldsymbol{\Sigma}^R(c)(H_n\Delta) + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)(H_n\Delta)$$

is asymptotically equivalent to $X'\boldsymbol{\Sigma}^R(c)X + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)(H_n\Delta)$ due to the dominance principle of $(H_n\Delta)'\boldsymbol{\Sigma}^R(c)$ over $2X'\boldsymbol{\Sigma}^R(c)$, *i.e.*, $X'\boldsymbol{\Sigma}^R(c)X + 2X'\boldsymbol{\Sigma}^R(c)(H_n\Delta) + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)(H_n\Delta) = X'\boldsymbol{\Sigma}^R(c)X + \big[2X'\boldsymbol{\Sigma}^R(c) + (H_n\Delta)'\boldsymbol{\Sigma}^R(c)\big](H_n\Delta)$. □

PROOF OF PROPOSITION 8.5 As $X_n = H_n(\hat{\psi}_n - \psi_{1n})$, under the local alternative $H_n(\psi_{1n} - \psi_0) \to \Delta \neq 0$, then

$$\begin{aligned}
W_n^R(c) &= [H_n(\hat{\psi}_n - \psi_{1n})\big]'\boldsymbol{\Sigma}_n^R(c)[H_n(\hat{\psi}_n - \psi_{1n})\big] + 2[H_n(\hat{\psi}_n - \psi_{1n})\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_{1n} - \psi_0)\big] \\
&\quad + \big[H_n(\psi_{1n} - \psi_0)\big]'\boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_{1n} - \psi_0)\big]
\end{aligned}$$

$$
\begin{aligned}
&= X_n' \boldsymbol{\Sigma}_n^R(c) X_n + 2X_n' \boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_{1n}-\psi_0)\big] + \big[H_n(\psi_{1n}-\psi_0)\big]' \boldsymbol{\Sigma}_n^R(c)\big[H_n(\psi_{1n}-\psi_0)\big] \\
&\underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X' \boldsymbol{\Sigma}^R(c) X + 2X' \boldsymbol{\Sigma}^R(c)\Delta + \Delta' \boldsymbol{\Sigma}^R(c)\Delta \qquad \text{(B.3)}
\end{aligned}
$$

since $X_n \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X$, $\boldsymbol{\Sigma}_n^R(c) \overset{p}{\to} \boldsymbol{\Sigma}^R(c)$. □

PROOF OF COROLLARY 8.6 From Proposition 8.5, we have:

$$
W_n^R(c) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X' \boldsymbol{\Sigma}^R(c) X + 2X' \boldsymbol{\Sigma}^R(c)\Delta + \Delta' \boldsymbol{\Sigma}^R(c)\Delta \ .
$$

As $\Delta \in \mathscr{V}(0)$, $P(0)(\boldsymbol{\Sigma})\Delta = \Delta$, and we have: $\boldsymbol{\Sigma}^R(c)\Delta = \sum_{d_j} g(d_j;c) P_j(\boldsymbol{\Sigma})\Delta = g(0;c)P(0)(\boldsymbol{\Sigma})\Delta = g(0;c)\Delta$ since $P_j(\boldsymbol{\Sigma})\Delta = 0$ for all eigenprojections on the eigenspaces different from $\mathscr{V}(0)$. Hence,

$$
W_n^R(c) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X' \boldsymbol{\Sigma}^R(c) X + 2g(0;c) X'\Delta + g(0;c)\Delta'\Delta \ .
$$

□

PROOF OF PROPOSITION 9.1 We need to show that $lim_{n\to\infty} \mathbb{P}\big[\|\boldsymbol{\Sigma}_n^R(c_n) - \boldsymbol{\Sigma}^R(0)\| \geq \varepsilon\big] = 0$ for every $\varepsilon > 0$. Let r denote the rank of the matrix of interest $\boldsymbol{\Sigma}$. Three possible cases will be considered in the proof: $r = q$, $r = 0$ and $1 \leq r < q$. Let $I = \{1, 2, \ldots, q\}$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \geq \cdots \geq \lambda_q \geq 0$, and $J = \{1, 2, \ldots, k\}$ the subset of I corresponding to the indices of the distinct eigenvalues of $\boldsymbol{\Sigma}$: $d_1 > d_2 > \cdots > d_j > \cdots > d_k \geq 0$ where the multiplicity of the distinct eigenvalue d_j is denoted $m(d_j)$, so that $\sum_{j=1}^{k} m(d_j) = q \geq 1$ and $1 \leq k \leq q$. For $j \in J$, let I_j denote the subset of I such that $I_j = \{i \in I : \ \lambda_i = d_j\}$, hence the I_j's are disjoint sets such as $\bigcup_{j=1}^{k} I_j = \{1, \ldots, q\}$. If zero is an eigenvalue, then $d_k = 0$. Let $P_j(\boldsymbol{\Sigma}) = P(d_j)(\boldsymbol{\Sigma})$ represent the eigenprojection operator projecting onto the eigenspace $\mathscr{V}(d_j)$ associated with d_j. First we show that

$$
\lim_{n\to\infty} \mathbb{P}[|g(\hat{\lambda}_i; c_n) - g(d_j;0)| \geq \varepsilon] = 0 \ \ \forall\, i \in I_j \ , \ \forall\, \varepsilon > 0 \qquad \text{(B.4)}
$$

as it is used later on in the proof. By Lemma 6.3 i), we have for all $i \in I_j$, $\hat{\lambda}_i \overset{p}{\to} d_j$. Besides, as $c_n \underset{n\to\infty}{\to} 0$, we have

$$
\mathbb{P}\big[|\hat{\lambda}_i - d_j| > c_n\big] = \mathbb{P}\big[|b_n(\hat{\lambda}_i - d_j)| > b_n c_n\big] \underset{n\to\infty}{\to} 0 \qquad \text{(B.5)}
$$

since $b_n c_n \to \infty$ and $b_n\big(\hat{\lambda}_i - d_j\big)$ converges in distribution by Theorem A.2. Note that for $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n)$, we can write

$$
\lim_{n\to\infty} \mathbb{P}[|g[\lambda_i(\boldsymbol{\Sigma}_n); c_n] - g(d_j;0)| > \varepsilon] = \lim_{n\to\infty,\, m\to\infty} \mathbb{P}[|g[\lambda_i(\boldsymbol{\Sigma}_n); c_m] - g(d_j;0)| > \varepsilon] \ . \qquad \text{(B.6)}
$$

It is equivalent to write

$$
\begin{aligned}
|g[\lambda_i(\boldsymbol{\Sigma}_n); c_m] - g(d_j;0)| &= |g[\lambda_i(\boldsymbol{\Sigma}_n); c_m] - g[\lambda_i(\boldsymbol{\Sigma}_n);0] + g[\lambda_i(\boldsymbol{\Sigma}_n);0] - g(d_j;0)| \\
&\leq |g[\lambda_i(\boldsymbol{\Sigma}_n); c_m] - g[\lambda_i(\boldsymbol{\Sigma}_n);0]| + |g[\lambda_i(\boldsymbol{\Sigma}_n);0] - g(d_j;0)| \ .
\end{aligned} \qquad \text{(B.7)}
$$

Hence, $\lim_{n\to\infty,\, m\to\infty} \mathbb{P}\{|g[\lambda_i(\boldsymbol{\Sigma}_n);c_m] - g[\lambda_i(\boldsymbol{\Sigma}_n);0]| > \varepsilon\} = 0$ since $\lim_{c\to 0^+} g(\lambda;c) = g(\lambda;0)$. Further, $\lim_{n\to\infty} \mathbb{P}\{|g[\lambda_i(\boldsymbol{\Sigma}_n);0] - g[d_j;0]| > \varepsilon\} = 0$, since $\hat{\lambda}_i = \lambda_i(\boldsymbol{\Sigma}_n) \xrightarrow{p} d_j$, $\forall\, i \in I_j$ and $g \in \mathscr{G}_c$ is continuous a.e. w.r.t. λ, hence (B.4) follows.

Consider first the case where the limiting matrix $\boldsymbol{\Sigma}$ has full rank, *i.e.* $rank(\boldsymbol{\Sigma}) = r = q$. For all $j \in J:\ d_j > 0$ since $r = q$, then by (B.4) and by Lemma 6.3 i) and ii), we have:

$$g(\hat{\lambda}_i; c_n) \xrightarrow{p} g(d_j;0)\ , \text{and}\ P_{I_j}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_j(\boldsymbol{\Sigma})\ ,$$

provided $\lambda_{i-1} \neq \lambda_i$ and $\lambda_j \neq \lambda_{j+1}$. Since $g(d_j;0) = \frac{1}{m(d_j)} \times m(d_j) g(d_j;0) = \frac{1}{m(d_j)} \sum_{i\in I_j} g(d_j;0)$, we have after adding and substracting the quantity $\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) g(d_j;0)$ simultaneously:

$$\begin{aligned}
\boldsymbol{\Sigma}_n^R(c_n) &= \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c_n) \\
&= \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \Big[g(d_j;0) - g(d_j;0) + \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c_n) \Big] \\
&= \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \Big[g(d_j;0) + \frac{1}{m(d_j)} \sum_{i\in I_j} \big[g(\hat{\lambda}_i; c_n) - g(d_j;0) \big] \Big] \xrightarrow{p} \sum_{j=1}^{k} P_j(\boldsymbol{\Sigma}) g(d_j;0) = \boldsymbol{\Sigma}^R(0)\ ,
\end{aligned}$$

since $P_{I_j}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_j(\boldsymbol{\Sigma})$, $P_{I_j}(\boldsymbol{\Sigma}_n) = O_p(1)$ and $|g(\hat{\lambda}_i; c_n) - g(d_j;0)| \xrightarrow{p} 0$ by (B.4).

Second, consider the case where $d_1 = 0$ with multiplicity $m(0) = q$. In this case, $\boldsymbol{\Sigma}_n \xrightarrow{p} \boldsymbol{\Sigma} = 0$, *i.e.* $\boldsymbol{\Sigma}_n$ converges to a zero matrix so that the range of $\boldsymbol{\Sigma}$ is $\mathscr{R}(\boldsymbol{\Sigma}) = \{0\}$ and its null-space is $\mathrm{N}(\boldsymbol{\Sigma}) = \mathbb{R}^q$. Let $P_1(\boldsymbol{\Sigma})$ denote the eigenprojection operator of $\boldsymbol{\Sigma}$ associated with its zero eigenvalue ($d_1 = 0$) which projects onto the corresponding eigenspace $\mathscr{V}(0)$, with $\dim\big[\mathscr{V}(0)\big] = q$. After adding and substracting the quantity $P_{I_1}(\boldsymbol{\Sigma}_n) g(0;0)$ simultaneously, we have:

$$\begin{aligned}
\boldsymbol{\Sigma}_n^R(c_n) &= P_{I_1}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_1)} \sum_{i\in I_1} g(\hat{\lambda}_i; c_n) = P_{I_1}(\boldsymbol{\Sigma}_n) \big[g(0;0) - g(0;0) + \frac{1}{m(0)} \sum_{i\in I_1} g(\hat{\lambda}_i; c_n) \big] \\
&= P_{I_1}(\boldsymbol{\Sigma}_n) g(0;0) + P_{I_1}(\boldsymbol{\Sigma}_n) \frac{1}{m(0)} \sum_{i\in I_1} \big[g(\hat{\lambda}_i; c_n) - g(0;0) \big] \\
&\xrightarrow{p} g(0;0) P_1(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^R(0)\ , \qquad \text{(B.8)}
\end{aligned}$$

since by Lemma 6.3 ii), we have $P_{I_1}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_1(\boldsymbol{\Sigma})$, $P_{I_1}(\boldsymbol{\Sigma}_n) = O_p(1)$ and by (B.4), we have with $d_1 = 0$: $|g(\hat{\lambda}_i; c_n) - g(0;0)| \xrightarrow{p} 0\ , \forall\, i \in I_1$.

Finally, suppose $d_k = 0$ and $d_1 \neq 0$. Then

$$\begin{aligned}
\|\boldsymbol{\Sigma}_n^R(c_n) - \boldsymbol{\Sigma}^R(0)\| &= \|\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c_n) - \sum_{j=1}^{k} P_j(\boldsymbol{\Sigma}) g(d_j;0)\| \\
= \|\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n) \Big[g(d_j;0) - g(d_j;0) + \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i; c_n) \Big] &- \sum_{j=1}^{k} P_j(\boldsymbol{\Sigma}) g(d_j;0)\|
\end{aligned}$$

$$
\begin{aligned}
&= \|\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i;c_n)-g(d_j;0)\big] + \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)g(d_j;0) - \sum_{j=1}^{k} P_j(\boldsymbol{\Sigma})g(d_j;0)\| \\
&\leq \|\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i;c_n)-g(d_j;0)\big]\| + \|\sum_{j=1}^{k} g(d_j;0)\big[P_{I_j}(\boldsymbol{\Sigma}_n)-P_j(\boldsymbol{\Sigma})\big]\| \\
&\leq \|\sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i;c_n)-g(d_j;0)\big]\| + \sum_{j=1}^{k}|g(d_j;0)|\|P_{I_j}(\boldsymbol{\Sigma}_n)-P_j(\boldsymbol{\Sigma})\| \\
&\leq \sum_{j=1}^{k}\|P_{I_j}(\boldsymbol{\Sigma}_n)\|\frac{1}{m(d_j)}\sum_{i\in I_j}|g(\hat{\lambda}_i;c_n)-g(d_j;0)|\| + \sum_{j=1}^{k}|g(d_j;0)|\|P_{I_j}(\boldsymbol{\Sigma}_n)-P_j(\boldsymbol{\Sigma})\|
\end{aligned}
\tag{B.9}
$$

since $P_{I_j}(\boldsymbol{\Sigma}_n) = O_p(1)$, $|g(\hat{\lambda}_i;c_n)-g(0;0)| \xrightarrow{p} 0\,, \forall i \in I_j$ by (B.4), $g(d_j;0) = O(1)$ and $\|P_{I_j}(\boldsymbol{\Sigma}_n)-P_j(\boldsymbol{\Sigma})\| \xrightarrow{p} 0$, by Lemma 6.3 ii). We can finally conclude that:

$$
\lim_{n\to\infty}\mathbb{P}\big[\|\boldsymbol{\Sigma}_n^R(c_n)-\boldsymbol{\Sigma}^R(0)\|\geq\varepsilon\big]=0\,.
$$

□

PROOF OF PROPOSITION 9.2 By Proposition 9.1, we have $\boldsymbol{\Sigma}_n^R(c_n) \xrightarrow{p} \boldsymbol{\Sigma}^R(0)$ and by Assumption 2.1, $X_n \xrightarrow[n\to\infty]{\mathscr{L}} X$, hence

$$
W_n^R(c_n) = X_n'\boldsymbol{\Sigma}_n^R(c_n)X_n \xrightarrow[n\to\infty]{\mathscr{L}} X'\boldsymbol{\Sigma}^R(0)X\ . \tag{B.10}
$$

The statistic can be decomposed as:

$$
W_n^R(c_n) = W_{1n}^R(c_n) + W_{2n}^R(c_n)
$$

where $W_{in}^R(c_n) = X_n'\boldsymbol{\Sigma}_{ii,n}^R(c_n)X_n$, for $i = 1,2$ and

$$
\boldsymbol{\Sigma}_n^R(c_n) = \sum_{j=1}^{k} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i,c_n) = \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i,c_n) + \sum_{j\geq k_1+1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i,c_n)\,.
$$

Let's focus on the first component:

$$
\boldsymbol{\Sigma}_{11,n}^R(c_n) = \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i,c_n) = \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\left[g(d_j;0)-g(d_j;0)+\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i,c_n)\right] \tag{B.11}
$$

$$
= \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)g(d_j;0) + \sum_{j=1}^{k_1} P_{I_j}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i;c_n)-g(d_j;0)\big] \tag{B.12}
$$

since $g(d_j;0) = \frac{1}{m(d_j)}\sum_{i\in I_j} g(d_j;0)$. Using the continuity property of the eigenvalues and total eigenprojections given in Lemma 6.3 i) and ii) provided we can find distinct eigenvalue before and after, we have $P_{I_j}(\boldsymbol{\Sigma}_n) \xrightarrow{p} P_j(\boldsymbol{\Sigma})$, $P_{I_j}(\boldsymbol{\Sigma}_n) = O_p(1)$, and by (B.4) $\forall\varepsilon>0$, $\lim_{n\to\infty}\mathbb{P}\big[|g(\hat{\lambda}_i;c_n)-g(d_j;0)|\geq\varepsilon\big] = 0\ \forall i\in I_j$.

Hence, we have:

$$\boldsymbol{\Sigma}^R_{11,n}(c_n) \xrightarrow{p} \sum_{j=1}^{k_1} g(d_j;0)P_j(\boldsymbol{\Sigma}) \equiv \boldsymbol{\Sigma}^R_{11}(0)\ , \text{with } \sum_{j=1}^{k_1} m(d_j) = q_1 = \text{rank}\big[\boldsymbol{\Sigma}^R_{11}(0)\big] = \dim\mathscr{V}(q_1) \tag{B.13}$$

and,

$$W^R_{1n}(c_n) = X'_n\boldsymbol{\Sigma}^R_{11,n}(c_n)X_n \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} X'\boldsymbol{\Sigma}^R_{11}(0)X \equiv W^R_1(0)\ .$$

For the second part of the statistic, the $q\times q$ matrix $\boldsymbol{\Sigma}$ is such that $rank(\boldsymbol{\Sigma}) = q_1$, so $d_{k_1+1} = 0$ with multiplicity $m(d_{k_1+1}) = q - q_1$. The regularization operates such that:

$$g(\hat\lambda_i; c_n) = \begin{cases} \frac{1}{\hat\lambda_i} & \text{if } \hat\lambda_i > c_n \\ 0 & \text{if } \hat\lambda_i \le c_n \end{cases} \tag{B.14}$$

If $\lambda_i = d_{k_1+1} = 0$, then

$$\mathbb{P}\big[g(\hat\lambda_i; c_n) = 0\big] = \mathbb{P}\big[b_n|\hat\lambda_i| \le b_n c_n\big] \underset{n\to\infty}{\to} 1\ \forall\, i \in I_{k_1+1}\ ,\ \text{card}(I_{k_1+1}) = q - q_1 < \infty\ ,$$

since $b_n(\hat\lambda_i - \lambda_i) = O_p(1)\ \forall i,$ and $b_n c_n \underset{n\to\infty}{\to} \infty$. A fortiori, it still holds for $\mathbb{P}\big[\sum_{i\in I_{k_1+1}} g(\hat\lambda_i, c_n) = 0\big] \underset{n\to\infty}{\to} 1$.

$$W^R_{2n}(c_n) = X'_n\boldsymbol{\Sigma}^R_{22,n}(c_n)X_n \text{ with } \boldsymbol{\Sigma}^R_{22,n}(c_n) = P_{I_{k_1+1}}(\boldsymbol{\Sigma}_n)\frac{1}{m(d_{k_1+1})}\sum_{i\in I_{k_1+1}} g(\hat\lambda_i, c_n)$$

Since $P_{I_{k_1+1}}(\boldsymbol{\Sigma}_n) = O_p(1)$, then $\mathbb{P}\big[P_{I_{k_1+1}}(\boldsymbol{\Sigma}_n)\sum_{i\in I_{k_1+1}} g(\hat\lambda_i; c_n) = 0\big] \underset{n\to\infty}{\to} 1$; this implies that $\mathbb{P}\big[\boldsymbol{\Sigma}^R_{22,n}(c_n) = 0\big] \underset{n\to\infty}{\to} 1$, hence, we have: $\mathbb{P}\big[W^R_{2n}(c_n) = 0\big] \underset{n\to\infty}{\to} 1$. □

PROOF OF PROPOSITION 9.3 Apply the results of Proposition 9.2 with $X_n = \sqrt{n}\big[\psi(\hat\theta_n) - \psi_0\big] \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N\big[0, \boldsymbol{\Sigma}\big] = X$. Following equation (7.10), $P_j(\boldsymbol{\Sigma}) = B(d_j)B(d_j)'$ and $B(d_j)'X = x_j$, where $x_j \sim N\big[0, d_j I_{m(d_j)}\big]$, or equivalently $x_j = \sqrt{d_j}u_j$, with $u_j \sim N(0, I_{m(d_j)})$. Recall that $\lim_{n\to\infty} g(\hat\lambda_i; c_n) = g(d_j;0)$ by definition of g in the $\mathscr{G}_c$ family (see 7.1) and $\lim_{n\to\infty} c_n = 0$. Hence, we can write:

$$\begin{aligned} W^R_1(0) &= X'\boldsymbol{\Sigma}^R_{11}(0)X = X'\big(\sum_{j=1}^{k_1} g(d_j;0)P_j(\boldsymbol{\Sigma})\big)X = \sum_{j=1}^{k_1} g(d_j;0)X'P_j(\boldsymbol{\Sigma})X \\ &= \sum_{j=1}^{k_1} g(d_j;0)X'B(d_j)B(d_j)'X = \sum_{j=1}^{k_1} g(d_j;0)x'_jx_j = \sum_{j=1}^{k_1}\frac{1}{d_j}d_ju'_ju_j = \sum_{j=1}^{k_1} u'_ju_j\ , \end{aligned}$$

where $u_j \sim N(0, I_{m(d_j)})$. Hence, $u'_ju_j \sim \chi(m(d_j))$. As $\sum_{j=1}^{k_1} m(d_j) = q_1$, hence $W^R_1(0) \sim \chi(q_1)$. □

C. Appendix: Alternative simulation-based approaches

In this section, we propose three alternative simulation-based approaches that rely on the technique of Monte Carlo tests to enhance the performance of the (regularized) Wald test; see Dufour (2006) and the references therein for a detailed presentation of the technique of Monte Carlo tests. To test the null hypothesis $H_0 : \psi(\theta) = \psi_0$, we consider different ways of simulating the asymptotic distribution of the (regularized) Wald statistic. The approaches differ through the strength of the assumptions made on the asymptotic distribution. They can be described as follows.

1. Simul-R approach: This approach requires the minimal assumption, and relies on the asymptotic distribution of the restrictions without the need to specify that of the parameter θ. By focusing on the restrictions, this approach can accommodate situations where some components of θ are not identified but whose transformations are. Thus, we simulate from the distribution of the restrictions, *i.e.*, $\sqrt{n}(\hat{\psi}_n - \psi_0) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N(0, \boldsymbol{\Sigma})$, with $\hat{\psi}_n = \psi_0 + \frac{1}{\sqrt{n}}\hat{U}'_\psi \times \tilde{v}$, where $\tilde{v} \sim N[0, I]$. $\hat{U}_\psi$ refers to the Cholesky decomposition of the estimate of $\boldsymbol{\Sigma}$, namely $\boldsymbol{\Sigma}_n = \hat{U}'_\psi \times \hat{U}_\psi$. We can then easily build the regularized Wald statistic as:
$$S_n(\hat{\psi}_n) = \sqrt{n}[\hat{\psi}_n - \psi_0]' \boldsymbol{\Sigma}_n^R(c) \sqrt{n}[\hat{\psi}_n - \psi_0] ,$$
where $\boldsymbol{\Sigma}_n^R(c)$ denotes the regularized inverse of $\boldsymbol{\Sigma}_n$.

2. Simul-E approach: This approach is more restrictive than Simul-R to the extent that it requires the identification of the whole parameter vector θ and situations for which the delta method applies. Nevertheless, it can accommodate some discontinuities in the restrictions (*e.g.*, ratios of parameters with null values in the denominator). Thus, we simulate from the distribution of the estimator of θ: $\sqrt{n}(\hat{\theta}_n - \theta_0) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} \mathrm{N}(0, \boldsymbol{\Sigma}_\theta)$, using $\hat{\theta}_n = \theta_0 + \frac{1}{\sqrt{n}}\hat{U}' \times \tilde{v}$ and the Cholesky decomposition $\hat{U}$: $\hat{U}' \times \hat{U} = \boldsymbol{\Sigma}_{\theta,n}$, where $\boldsymbol{\Sigma}_{\theta,n}$ is an estimator of $\boldsymbol{\Sigma}_\theta$, and $\tilde{v} \sim N[0, I]$. Applying the delta method, we can deduce the distribution of the restrictions, *i.e.* $\sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta_0)) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N(0, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \Gamma \boldsymbol{\Sigma}_\theta \Gamma'$, and Γ corresponds to the derivative of the restrictions w.r.t. θ. Using consistent estimators of the above quantities, we can then easily build the regularized Wald statistic statistic as:
$$S_n(\hat{\theta}_n) = \sqrt{n}[\psi(\hat{\theta}_n) - \psi(\theta_0)]' \boldsymbol{\Sigma}_n^R(c) \sqrt{n}[\psi(\hat{\theta}_n) - \psi(\theta_0)] .$$

3. Simul-DGP approach: This approach is the most restrictive since it requires the highest level of information. Thus, when the full DGP is specified, one can simulate from it; y can be expressed as a function of θ, *i.e.* $y_j = f(\theta, \tilde{v}_j)$, $j = 1, \ldots, n$ where $\tilde{v}_j$ is a random variable and $y_1^n = (y_1, \ldots, y_n)$. For instance, one can simulate from a parametric Gaussian model under the null (as we do in the next section, see equation (D.1)) and build the statistic such as:
$$S_n(y_1^n, \hat{\theta}_n) = n[\psi(\hat{\theta}_n(y_1^n)) - \psi(\theta_0)]' \boldsymbol{\Sigma}_n^R(c)[\psi(\hat{\theta}_n(y_1^n)) - \psi(\theta_0)]$$

In the following, we shall denote $S^{(i)}$ the i-th replication of the simulated statistic associated with the i-th random vector $\tilde{v}^{(i)}$, for $i = 1, \ldots, N$. Please note that n refers to the sample size while N to the number of replications of the Monte Carlo test. For $i = 0$, let $S^{(0)} = S^{(0)}(\psi_0)$ refer to the test statistic computed from observed data when the true parameter vector is $\psi(\theta_0) = \psi_0$. Note that the technique of Monte Carlo tests does not require the number of replications N to be large, and the validity of the procedure holds for N fixed; for example $N = 19$ is sufficient to control the level of the test irrespective of the sample size. In other words, if one simulates from the exact distribution of the test statistic instead of the asymptotic approximation, the Monte Carlo test would yield an exact test.

In the Simul-R approach, we draw for $i = 1, \ldots, N$, $\tilde{v}^{(i)} \sim N(0, I)$ such that:

$$\hat{\psi}_n^{(i)} = \psi_0 + \frac{1}{\sqrt{n}} \hat{U}_\psi' \tilde{v}^{(i)} \text{ or equivalently } \sqrt{n}[\hat{\psi}_n^{(i)} - \psi_0] = \hat{U}_\psi' \tilde{v}^{(i)}$$

with $\Sigma_n = \hat{U}_\psi' \hat{U}_\psi = V_n' \Lambda[\lambda(\Sigma_n)] V_n$ based on its spectral decomposition. Using its regularized counterpart, *i.e.* $\Sigma_n^R(c) = V_n' \Lambda^\dagger[\lambda(\Sigma_n); c] V_n$, we get N replications of the regularized Wald statistic under the null,

$$S_n^{(i)}(\hat{\psi}_n) = \sqrt{n}[\hat{\psi}_n^{(i)} - \psi_0]' \Sigma_n^R(c) \sqrt{n}[\hat{\psi}_n^{(i)} - \psi_0] \, , i = 1, \ldots, N \, .$$

The *i.i.d.* assumption usually made for $(S_n^{(1)}(\psi), \ldots, S_n^{(N)}(\psi))$ can be relaxed to the exchangeability assumption. Let $S_n(N, \psi) = (S_n^{(1)}(\psi), \ldots, S_n^{(N)}(\psi))$, and the sample p-value functions be defined as:

$$\hat{G}_{nN}(x|\psi) \equiv \hat{G}_{nN}[x; S_n(N, \psi)] = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(S_n^{(i)}(\psi) \geq x) \tag{C.1}$$

$$\hat{p}_{nN}(x|\psi) = \frac{N \hat{G}_{nN}(x|\psi) + 1}{N + 1} \, . \tag{C.2}$$

Recall that $S_n^{(0)}$ refers to the test statistic that is computed from observed data. The test rejects the null hypothesis at level α if $\hat{p}_{nN}(S_n^{(0)}|\hat{\psi}_n) \leq \alpha$.

The asymptotic validity of bootstrap p-values based on a consistent point estimate (of the parameter ψ) is established in Dufour (2006, Proposition 6.1, p.464). The proposition also states the validity of bootstrap p-values for general sequences of random variables with (possibly discrete) distributions (when ties may have nonzero probability). Under appropriate conditions that can be found in Dufour (2006), let the random variables $S_n^{(0)}$ and $\hat{\psi}_n$ be independent of $S_n(N, \psi)$. If $\hat{\psi}_n \xrightarrow{p} \psi_0$ then for $0 \leq \alpha \leq 1$,

$$\lim_{n \to \infty} \{ \mathbb{P}\left[\hat{p}_{nN}(S_n^{(0)}|\hat{\psi}_n) \leq \alpha\right] - \mathbb{P}\left[\hat{p}_{nN}(S_n^{(0)}|\psi_0) \leq \alpha\right] \} = 0 \, . \tag{C.3}$$

D. Appendix: Simulation results: Multi-step noncausality

In this section, we perform Monte Carlo experiments to assess the empirical behavior of the (regularized) Wald statistics in the presence of asymptotic singularity. We consider the following VAR(1) process

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = A_1 \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + u_t = \begin{bmatrix} \theta_{xx} & \theta_{xy} & \theta_{xz} \\ \theta_{yx} & \theta_{yy} & \theta_{yz} \\ \theta_{zx} & \theta_{zy} & \theta_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + u_t \, , \tag{D.1}$$

for $t = 1, \ldots, n$, where $u_t = [u_{x,t} \; u_{y,t} \; u_{z,t}]'$ is a Gaussian noise with a (3×3) nonsingular covariance matrix Σ_u. We are interested in testing for multi-step noncausality *i.e.,*

$$H_0 : \psi(\theta) = \begin{bmatrix} \theta_{xy} \\ \theta_{xx}\theta_{xy} + \theta_{xy}\theta_{yy} + \theta_{xz}\theta_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \, . \tag{D.2}$$

using three different versions of the Wald statistic, *i.e.*, $W_n^R(c) = n\psi(\hat{\theta}_n)' \Sigma_n^R(c) \psi(\hat{\theta}_n)$. As pointed out in Section 3.1, singularity problems arise under parameter setting (3.1). Let $\mathbf{y}_t = [x_t \; y_t \; z_t]'$, $Y \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, $B \equiv (A_1) \; Z_t \equiv [\mathbf{y}_t], \quad Z \equiv (Z_0, \ldots, Z_{n-1})$, $U \equiv [u_t]_{t=1,\ldots,n} = (u_1, \ldots, u_n)$ Using the standard column stacking operator *vec*, let $\theta = vec(A_1) = vec(B)$, where B is (3×3) and Y, Z and U are $(3 \times n)$. We use the multivariate

LS estimator of θ. Applying the column stacking operator *vec* on:

$$Y = BZ + U \tag{D.3}$$

we have:

$$\text{vec}(Y) = \text{vec}(BZ) + \text{vec}(U) \tag{D.4}$$
$$y = \big(Z' \otimes I_3\big)\text{vec}(B) + \text{vec}(U) \tag{D.5}$$
$$y = \big(Z' \otimes I_3\big)\theta + u \tag{D.6}$$

where $\text{E}(uu') = I_n \otimes \boldsymbol{\Sigma}_u$. The multivariate LS estimator $\hat{\theta}_n$ is given by:

$$\hat{\theta}_n = \left((ZZ')^{-1}Z \otimes I_3\right)y\,, \tag{D.7}$$

such that:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} \text{N}\big(0,\, \boldsymbol{\Sigma}_\theta\big) \tag{D.8}$$

where $\boldsymbol{\Sigma}_\theta = \Omega^{-1} \otimes \boldsymbol{\Sigma}_u$; see Lütkepohl (2005, Proposition 3.1 p. 74, eq. (3.2.15)). Provided the delta method applies, the restrictions are also asymptotically Gaussian:

$$\sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta_0)) \xrightarrow{\mathscr{L}} \text{N}\big(0,\, \boldsymbol{\Sigma}\big) \tag{D.9}$$

where

$$\boldsymbol{\Sigma} = \frac{\partial \psi}{\partial \theta'}(\theta)\boldsymbol{\Sigma}_\theta \frac{\partial \psi'}{\partial \theta}(\theta)\,. \tag{D.10}$$

A consistent estimator of $\boldsymbol{\Sigma}$ is easily obtained as:

$$\boldsymbol{\Sigma}_n = \frac{\partial \psi}{\partial \theta'}(\hat{\theta}_n)\boldsymbol{\Sigma}_{\theta,n} \frac{\partial \psi'}{\partial \theta}(\hat{\theta}_n) \tag{D.11}$$

by plugging in a consistent estimator of $\boldsymbol{\Sigma}_\theta$, *i.e.*, $\boldsymbol{\Sigma}_{\theta,n} = \hat{\Omega}^{-1} \otimes \hat{\boldsymbol{\Sigma}}_u$ with $\hat{\Omega} = \frac{1}{n}ZZ'$ and $\hat{\boldsymbol{\Sigma}}_u = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t\hat{u}_t' = \frac{1}{n}Y\big[I_n - Z'(ZZ')^{-1}Z\big]Y'$. We examine three different parameter settings for the VAR(1) coefficients $A_1 = \begin{bmatrix} \theta_{xx} & \theta_{xy} & \theta_{xz} \\ \theta_{yx} & \theta_{yy} & \theta_{yz} \\ \theta_{zx} & \theta_{zy} & \theta_{zz} \end{bmatrix}$. The first two parameter setups correspond to:

$$A_1 = A_{10} = \begin{bmatrix} -0.99 & \theta_{xy} & \theta_{xz} \\ 0 & -0.99 & 0.5 \\ 0 & 0 & -0.99 \end{bmatrix}, \quad A_1 = A_{20} = \begin{bmatrix} -0.9 & \theta_{xy} & \theta_{xz} \\ 0 & -0.9 & 0.5 \\ 0 & 0 & -0.9 \end{bmatrix},$$

where the problem of singularity is obtained for $\theta_{xy} = \theta_{xz} = \theta_{zy} = 0$. The key parameter to disentangle between the regularity point and singularity point under this setup is θ_{xz}, with $\theta_{xz} = 0$ corresponding to a singularity point, and $\theta_{xz} \neq 0$ to a regularity point. A third parameter setup is examined, *i.e.*, $A_1 = A_{11} = \begin{bmatrix} 0.3 & \theta_{xy} & \theta_{xz} \\ 0.7 & 0.3 & 0.25 \\ 0.5 & 0.4 & 0.3 \end{bmatrix}$ where $\theta_{xy} = \theta_{xz} = 0$, and $\theta_{zy} = 0.4 \neq 0$ yields a regular setup. The first two parameter settings involve parameters close to the nonstationary region, whereas the third one falls inside the stationary region. $u_t = [u_{x,t}\ u_{y,t}\ u_{z,t}]'$ is a Gaussian noise with nonsingular covariance matrix $\boldsymbol{\Sigma}_u$, whose values have

Table 3. Notations of the statistics

Notations of the statistics	
Notations	Definition
W	Standard Wald statistic using the standard critical point
$W_{DV}(bound)$	Full-rank regularized Wald statistic using the asymptotic bound and a fixed threshold
W_{LB}	LB Reduced-rank Wald statistic based on the modified Moore-Penrose inverse and a threshold that varies with the sample size
W_{Noise}	Modified Wald statistic resulting from adding a noise to the restrictions; using the the standard critical point
W_{Ridge}	Wald statistic whose covariance matrix is regularized by adding $0.1\times$Identity matrix.
Simul-R	Monte Carlo tests - simulated version of the corresponding statistic using the distribution of the restrictions
Simul-E	Monte Carlo tests - simulated version of the corresponding statistic using the distribution of the estimator of the parameter
Simul-DGP	Monte Carlo tests - simulated version of the corresponding statistic using a specified DGP
Simul-Mixt	Simulated version of the linear combination of modified chi-square variables as in eq. (ii): $\sum_{j=1}^{2} g(\hat{\lambda}_j;c)\hat{\lambda}_j v_j$, where the v_j's are independent and random draws from a $\chi^2(1)$.

been set to

$$\boldsymbol{\Sigma}_u = \begin{pmatrix} 1.5 & -0.7 & 0.3 \\ -0.7 & 0.5 & -0.4 \\ 0.3 & -0.4 & 1 \end{pmatrix}$$

in the simulation design. Its determinant is different from zero, *i.e.,* $\det(\boldsymbol{\Sigma}_u) = 0.143$. The threshold values have been set to $c_n = \hat{\lambda}_1 n^{-1/3}$ in the case of a varying threshold and to $c = 0.1$ for the fixed threshold. We also use $c_n = \hat{\lambda}_1 n^{-1/2}$ sporadically; it performs better in the regular setup in terms of power because it regularizes less often. Note that the choice of $c_n = \hat{\lambda}_1 n^{-1/3}$, (or $c_n = \hat{\lambda}_1 n^{-1/2}$) only applies to the spectral cut-off regularized Wald statistic recommended by Lütkepohl and Burda (1997), whereas we propose the fixed value of $c = 0.1$ for the full-rank regularized statistic. Concerning c_n, it has been normalized by the largest eigenvalues to account for scaling issues of the data. We use 5000 replications in all simulation experiments. The nominal size to perform the tests has been fixed to 0.05, with critical points for the chi-square distribution with full rank given by $\chi^2_{95\%}(2) = 5.99$, or with reduced rank given by $\chi^2_{95\%}(1) = 3.84$ for the spectral cut-off regularized Wald statistic. In the tables below, W denotes the standard Wald statistic, $W_{DV}(bound)$ the full-rank regularized Wald statistic that uses the bound and the fixed threshold c; W_{LB} denotes the spectral cut-off Wald statistic that uses the varying threshold c_n. For comparison purposes, we also report the modified Wald statistic that results from adding noise to the restrictions to make them less efficient; it is denoted W_{noise}. See Lütkepohl and Burda (1997, Proposition 1, page 317) for its form. Note that W_{LB} and W_{Noise} are the two modified Wald statistics proposed by Lütkepohl and Burda (1997). We propose to implement the LB reduced-rank statistic through Monte Carlo tests (Simul-R, simul-E, Simul-DGP) that help to reduce size distortions in finite samples.

D.1. Level assessment

We study the empirical behavior of the test statistics under the null hypothesis:

$$H_0: \ \psi(\theta) = \begin{bmatrix} \theta_{xy} \\ \theta_{xx}\theta_{xy} + \theta_{xy}\theta_{yy} + \theta_{xz}\theta_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

first in irregular setups (see Table 4, panels $A: \ A_1 = A_{10}$ and $B: \ A_1 = A_{20}$), then in a regular setup (see Table 4, panel $C: \ A_1 = A_{11}$). It is clear from Table 4, panels A and B that the standard Wald statistic, W, does not have its usual asymptotic distribution in non-regular setups, either suffering from severe over-rejections in small samples, or from under-rejections in large samples. Its behavior gets worse when parameter values

approach the nonstationary region (Table 4, Panel A). Similarly, the reduced rank Wald statistic, W_{LB}, displays the same finite sample behavior as W in non-regular setups, with severe size distortions when parameter values get close to the nonstationary region, but exhibits good size properties asymptotically. In contrast, the full-rank regularized statistic that uses the bound, $W_{DV}(bound)$, does not suffer from over-rejection under the null hypothesis, but under-rejects instead. Nevertheless, if one simulates directly from the DGP provided it is specified, one can mitigate the underrejection of the bound by using the Simul-DGP approach. The Simul-DGP approach for W_{DV} remarkably dominates its competitors W and W_{LB} particularly in small samples (see Table 4, panel $A: A_1 = A_{10}$, $n = 50$). Thus, it is crucial to simulate from a well-behaved statistic to produce a reliable test. To the extent that all testing procedures are asymptotically justified, including the version of the Monte Carlo tests used here, it is not surprising that all tests approach the nominal level of 0.05 for sufficiently large sample sizes. In particular, the level is controlled for all three simulation-based approaches in large samples. Further, for the sake of comparison, we have also implemented the uniform (ridge-type) regularization scheme defined in equation (7.4). Unlike the DV scheme defined in equation (7.3) that only modifies the small problematic eigenvalues, the uniform (ridge-type) regularization scheme modifies all eigenvalues regardless of their magnitude. By modifying the large eigenvalues uselessly, this regularization scheme tends to exacerbate the conservative feature of the asymptotic bound relative to the DV one, especially in small samples. This pattern is even more striking in the regular setup; this sheds light on the lack of robustness (to regular settings unlike the DV scheme) of this crude regularization scheme. Regarding the regular setup shown in panel C of Table 4, all statistics display the correct expected level of 0.05. Note also that we have tried different values for the fixed threshold c, and we recommend $c = 0.1$. Its impact on power will be examined next. Thus, the less one regularizes, *i.e.* one chooses $c = 0.01$ instead of $c = 0.1$, the more the full-rank regularized statistic behaves like the standard Wald statistic. Selecting a fixed value for the threshold in an optimal way – that might enhance power – could be considered through data-driven procedures, *e.g.,* cross-validation methods. As for the reduced rank statistic, the asymptotic W_{LB} test statistic behaves slightly differently depending on the choice of the varying threshold c_n in regular setups; in nonregular setups, regardless of c_n, *i.e.*, $c_n = \hat{\lambda}_1 n^{-1/3}$ or $c_n = \hat{\lambda}_1 n^{-1/2}$, the results are identical. However, the threshold $c_n = \hat{\lambda}_1 n^{-1/3}$ that exploits the convergence rate of the sample eigenvalues towards their population analogs permits to control the level of the test (that relies on the W_{LB} statistic) in the simulation-based approach. More specifically, in the simulation-based approaches of W_{LB}, using $c_n = \hat{\lambda}_1 n^{-1/2}$ leads to over-rejections in small-to-moderately large sample sizes relative to $c_n = \hat{\lambda}_1 n^{-1/3}$; see Table 4, panel C. Also, power will differ markedly w.r.t. c_n in the regular setup as shown in the power exercise. Note also the correct asymptotic level of the simulated version of the linear combination of chi-square variables as in Corollary 8.2(ii): $\sum_{j=1}^{2} g(\hat{\lambda}_j; c)\hat{\lambda}_j v_j$, where the v_j's are independent and random draws from a $\chi^2(1)$. In the regular setup, the level of the corresponding procedure is controlled for all sample sizes. Finally, although W_{Noise} enables us to control size under the null, this procedure is not recommendable from the viewpoint of power as shown next.

D.2. Power assessment

We also study the empirical power for alternatives close to a singularity point $\theta_{xz} = 0$:

$$H_1: \ \psi(\theta) = \begin{bmatrix} \delta \\ (\theta_{xx} + \theta_{yy})\delta \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

with $\theta_{xy} = \delta$, ($\delta = 0.1264$ or $\delta = 0.04$) whose empirical power is reported in panels A and B of Table 5. We also consider a second type of alternative for a violation of the second restriction only, while maintaining

Table 4. Empirical levels of Multistep noncausality tests $H_0 : \psi(\theta) = 0$

$H_0 : \psi(\theta) = 0$; nominal size= 0.05, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
Panel A: irregular setup															
$H_0 : \psi(\theta) = 0$ with with $\theta_{xy} = \theta_{xz} = \theta_{zy} = 0$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = -0.99$, $A_1 = A_{10}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.3234	0.2996	0.2326	0.1870	-	0.2496	0.2274	0.1956	0.1448	-	0.0858	0.0736	0.0946	0.0588	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.2820	0.2376	0.2450	0.1074	-	0.2308	0.1894	0.1966	0.0860	-	0.1116	0.0914	0.0902	0.0536	-
W_{Noise}	0.0534	-	-	-	-	0.0566	-	-	-	-	0.0534	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	0.0030/0.0306	0.2380/0.2384	0.2154/0.2264	0.0700/0.0832	-	0.0004/0.007	0.1892/0.1892	0.1798/0.1828	0.0492/0.0576	-	0.000/0.000	0.0914/0.0914	0.0870/0.0870	0.0340/0.0356	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.2380/0.2382	0.2136/0.2230	0.0676/0.0878	0.2942	-	0.1892/0.1892	0.1786/0.1810	0.0476/0.0564	0.2416	-	0.0914/0.0914	0.0870/0.0870	0.0340/0.0342	0.1222
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.0078/0.065	-	-	-	-	0.0006/0.0224	-	-	-	-	0.000/0.000	-	-	-	-
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.0600	0.0506	0.0758	0.0550	-	0.0370	0.0300	0.0554	0.0428	-	0.0178	0.0142	0.0384	0.0284	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.0862	0.0678	0.0688	0.0470	-	0.0670	0.0530	0.0534	0.0438	-	0.0528	0.0368	0.0380	0.0400	-
W_{Noise}	0.0532	-	-	-	-	0.0536	-	-	-	-	0.0530	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	0.0000/0.0000	0.0678/0.0678	0.0680/0.0680	0.0318/0.0328	-	0.0000/0.0000	0.0530/0.0530	0.0530/0.0530	0.0324/0.0332	-	0.000/0.000	0.0368/0.0368	0.0378/0.0378	0.0366/0.0370	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.0678/0.0678	0.0680/0.0680	0.0318/0.0318	0.0960	-	0.0530/0.0530	0.0530/0.0530	0.0324/0.0324	0.0732	-	0.0368/0.0368	0.0378/0.0378	0.0366/0.0366	0.0594
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.000/0.000	-	-	-	-	0.000/0.000	-	-	-	-	0.000/0.000	-	-	-	-
Panel B: irregular setup															
$H_0 : \psi(\theta) = 0$ with $\theta_{xy} = \theta_{xz} = \theta_{zy} = 0$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = -0.9$, $A_1 = A_{20}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.1044	0.0904	0.1146	0.0690	-	0.0576	0.0488	0.0760	0.0528	-	0.0260	0.0200	0.0470	0.0392	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.1404	0.1156	0.1178	0.0608	-	0.0948	0.0742	0.0732	0.0498	-	0.0614	0.0442	0.0470	0.0406	-
W_{Noise}	0.0614	-	-	-	-	0.0544	-	-	-	-	0.0566	-	-	-	-
$W_{ridge-type}$; $c = 0.1/\ c = 0.01$	0.0318/0.0638	0.1154/0.1152	0.1142/0.1190	0.0672/0.0626	-	0.0126/0.0346	0.0742/0.0738	0.0730/0.0728	0.0524/0.0504	-	0.0032/0.0152	0.0442/0.0442	0.0470/0.0468	0.0414/0.0412	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.1154/0.1142	0.1180/0.1194	0.0694/0.0596	0.1494	-	0.0742/0.0738	0.0734/0.0728	0.0540/0.0494	0.1032	-	0.0442/0.0442	0.0470/0.470	0.0410/0.0406	0.0660
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.0662/0.0728	-	-	-	-	0.0378/0.0388	-	-	-	-	0.0194/0.0194	-	-	-	-
Panel B: irregular setup															
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.0224	0.0176	0.0452	0.0426	-	0.0202	0.0168	0.0446	0.0460	-	0.0162	0.0130	0.0352	0.0348	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.0582	0.0428	0.0464	0.0440	-	0.0574	0.0416	0.0438	0.0474	-	0.0468	0.0338	0.0364	0.0374	-
W_{Noise}	0.0516	-	-	-	-	0.0534	-	-	-	-	0.0518	-	-	-	-
$W_{ridge-type}$; $c = 0.1/\ c = 0.01$	0.0048/0.0162	0.0428/0.0428	0.0464/0.0468	0.0444/0.0442	-	0.0024/0.0136	0.0416/0.0416	0.0436/0.0436	0.0472/0.0474	-	0.0026/0.0118	0.0338/0.0338	0.0364/0.0364	0.0368/0.0374	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.0428/0.0428	0.0464/0.0466	0.0440/0.0440	0.0656	-	0.0416/0.0416	0.0438/0.0438	0.0474/0.0474	0.0648	-	0.0338/0.0338	0.0364/0.0364	0.0374/0.0374	0.0520
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.0184/0.0184	-	-	-	-	0.0166/0.0166	-	-	-	-	0.0138/0.0138	-	-	-	-
Panel C: regular setup															
$H_0 : \psi(\theta) = 0$ with $\theta_{xz} = \theta_{xy} = 0$, $\theta_{zy} = 0.4$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = 0.3$, $A_1 = A_{11}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.0438	0.0350	0.0570	0.0406	-	0.0422	0.0330	0.0508	0.0434	-	0.0508	0.0394	0.0422	0.0420	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}/c_n = \hat{\lambda}_1 n^{-1/2}$	0.0576/0.0516	0.0426/0.0634	0.0458/0.0656	0.0398/0.0542	-	0.0590/0.050	0.0446/0.0810	0.0432/0.0772	0.0398/0.0670	-	0.0538/0.0508	0.0400/0.1060	0.0358/0.0986	0.0410/0.1074	-
$W_{ridge-type}$; $c = 0.1/\ c = 0.01$	0.0264/0.0414	0.0404/0.0364	0.0566/0.0568	0.0410/0.0406	-	0.0232/0.0376	0.0396/0.0340	0.0488/0.0514	0.0428/0.0418	-	0.0212/0.0446	0.0440/0.0398	0.0398/0.0436	0.0442/0.0434	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.0352/0.0350	0.0564/0.0570	0.0404/0.0406	0.0488	-	0.0338/0.0330	0.0524/0.0508	0.0426/0.0434	0.0508	-	0.0394/0.0394	0.0434/0.0422	0.0434/0.0420	0.0560
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.0424/0.0438	-	-	-	-	0.0398/0.0422	-	-	-	-	0.0484/0.0508	-	-	-	-
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.0476	0.0354	0.0374	0.0376	-	0.0450	0.0348	0.0360	0.0346	-	0.0452	0.0336	0.0336	0.0386	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}/c_n = \hat{\lambda}_1 n^{-1/2}$	0.0534/0.0476	0.0366/0.1060	0.0332/0.0926	0.0326/0.0936	-	0.0478/0.0450	0.0368/0.0750	0.0356/0.0622	0.0324/0.0562	-	0.0452/0.0452	0.0336/0.0336	0.0338/0.0338	0.0386/0.0386	-
$W_{ridge-type}$; $c = 0.1/\ c = 0.01$	0.0166/0.0412	0.0392/0.0348	0.0382/0.0374	0.0372/0.0392	-	0.0178/0.0396	0.0370/0.0352	0.0370/0.0360	0.0340/0.0332	-	0.0158/0.0372	0.0336/0.0336	0.0340/0.0326	0.0422/0.0384	-
W_{DV}; $c = 0.1/\ c = 0.01$	-	0.0354/0.0354	0.0382/0.0374	0.0384/0.0376	0.0556	-	0.0346/0.0348	0.0358/0.0360	0.0348/0.0346	0.0524	-	0.0338/0.0336	0.0346/0.0336	0.0398/0.0386	0.0528
W_{DV} (*bound*); $c = 0.1/\ c = 0.01$	0.0464/0.0476	-	-	-	-	0.0444/0.0450	-	-	-	-	0.0444/0.0452	-	-	-	-

See Table 3 for the definition of the acronyms.

Table 5. Locally-level corrected empirical power of tests for multistep noncausality $H_1 : r(\theta) \neq 0$

$H_1 : \psi(\theta) \neq 0$; nominal size= 0.05, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
Panel A: irregular setup															
$H_1 : \psi(\theta) \neq 0$ with $\theta_{xy} = \delta = 0.1264$, $\theta_{xz} = 0$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = -0.99$, $A_1 = A_{10}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.9006	0.24130	0.5210	0.4411	-	0.9994	0.3472	0.6647	0.7131	-	1.00	1.00	1.00	1.00	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.9604	0.4621	0.4618	0.9518	-	0.9996	0.7488	0.6709	0.9996	-	1.00	1.00	1.00	1.00	-
W_{Noise}	0.3798	-	-	-	-	0.6976	-	-	-	-	1.00	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	0.9722	0.4702	0.5707	0.9684	-	0.9996	0.7463	0.7812	0.9998	-	1.00	1.00	1.00	1.00	-
W_{DV}; $c = 0.1$	-	0.4696	0.5935	0.9708	-	-	0.7463	0.8032	0.9998	-	-	1.00	1.00	1.00	-
$W_{DV}(bound)$; $c = 0.1$	0.9812	-	-	-	-	0.9996	-	-	-	-	1.00	-	-	-	-
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{Noise}	1.00	-	-	-	-	1.00	-	-	-	-	1.00	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{DV}; $c = 0.1$	-	1.00	1.00	1.00	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00
$W_{DV}(bound)$; $c = 0.1$	1.00	-	-	-	-	1.00	-	-	-	-	1.00	-	-	-	-
Panel B: irregular setup															
$H_1 : \psi(\theta) \neq 0$ with $\theta_{xy} = \delta = 0.04$, $\theta_{xz} = 0$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = -0.99$, $A_1 = A_{10}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.4726	0.2314	0.3723	0.1900	-	0.9256	0.3460	0.6363	0.5983	-	1.00	1.00	1.00	1.00	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	0.6454	0.3406	0.3484	0.5604	-	0.9744	0.7276	0.6557	0.9694	-	1.00	1.00	1.00	1.00	-
W_{Noise}	0.0828	-	-	-	-	0.1130	-	-	-	-	0.4110	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	0.0926	0.4230	0.5127	0.3496	-	0.7684	0.7418	0.7769	0.9672	-	1.00	1.00	1.00	1.00	-
W_{DV}; $c = 0.1$	-	0.4268	0.5380	0.3020	-	-	0.7424	0.7989	0.9622	-	-	1.00	1.00	1.00	-
$W_{DV}(bound)$; $c = 0.1$	0.2118	-	-	-	-	0.8710	-	-	-	-	1.00	-	-	-	-
Panel B: irregular setup															
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}$	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{Noise}	0.7124	-	-	-	-	0.9514	-	-	-	-	1.00	-	-	-	-
$W_{ridge-type}$; $c = 0.1$	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{DV}; $c = 0.1$	-	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-
$W_{DV}(bound)$; $c = 0.1$	1.00	-	-	-	-	1.00	-	-	-	-	1.00	-	-	-	-
Panel C: regular setup															
$H_1 : \psi(\theta) \neq 0$ with $\theta_{xz} = \delta = 0.1264$, $\theta_{xy} = 0$, $\theta_{zy} = 0.4$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = 0.3$, $A_1 = A_{11}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;															
	$n = 50$					$n = 100$					$n = 500$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.0922	0.0726	0.1152	0.0888	-	0.1882	0.1524	0.2050	0.1808	-	0.8280	0.7874	0.7936	0.7920	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}/c_n = \hat{\lambda}_1 n^{-1/2}$	0.0612/0.0716	0.0466	0.0486	0.0412	-	0.0650/0.1494	0.0516	0.0490	0.0456	-	0.1206/0.8270	0.1028	0.0972	0.0984	-
$W_{ridge-type}$; $c = 0.1$	0.0378	0.0592	0.0810	0.0600	-	0.0672	0.1062	0.1268	0.1050	-	0.5114	0.6296	0.6348	0.6396	-
W_{DV}; $c = 0.1$	-	0.0704	0.1096	0.0850	0.0972	-	0.1458	0.1896	0.1670	0.1970	-	0.7750	0.7794	0.7796	0.8310
$W_{DV}(bound)$; $c = 0.1$	0.0854	-	-	-	-	0.1700	-	-	-	-	0.8064	-	-	-	-
	$n = 1000$					$n = 2000$					$n = 5000$				
Statistics	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt	Asy	Simul-R	Simul-E	Simul-DGP	Simul-Mixt
W	0.9912	0.9846	0.9854	0.9866	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-
W_{LB}; $c_n = \hat{\lambda}_1 n^{-1/3}/c_n = \hat{\lambda}_1 n^{-1/2}$	0.3116/0.9912	0.2938	0.2892	0.2866	-	0.8838/1.00	0.8784	0.8736	0.8730	-	1.00/1.00	1.00	1.00	1.00	-
$W_{ridge-type}$; $c = 0.1$	0.9106	0.9508	0.9492	0.9508	-	0.9988	0.9994	0.9992	0.9996	-	1.00	1.00	1.00	1.00	-
W_{DV}; $c = 0.1$	-	0.9830	0.9840	0.9846	0.9920	-	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00
$W_{DV}(bound)$; $c = 0.1$	0.9898	-	-	-	-	1.00	-	-	-	-	1.00	-	-	-	-

See Table 3 for the definition of the acronyms.

fulfilled the first restriction as in Lütkepohl and Burda (1997), *i.e.*

$$H_1: \ \psi(\theta) = \begin{bmatrix} 0 \\ (\theta_{xz} \times \theta_{zy}) \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

with $\theta_{xz} = \delta = 0.1264$, $\theta_{zy} = 0.4$ and $\theta_{xy} = 0$, under a regular design:

$$A_1 = A_{11} = \begin{bmatrix} 0.3 & 0 & \theta_{xz} \\ 0.7 & 0.3 & 0.25 \\ 0.5 & 0.4 & 0.3 \end{bmatrix};$$

see panel C of Table 5. First of all, all power frequencies reported in Table 5 have been locally corrected for level distortions (only for over-rejections). See Table 6 in appendix E for level correction.

First note that panels A and B of Table 5 correspond to parameter values on the nonstationary region, *i.e.* $A_1 = A_{10}$. In Table 5, though conservative, the full-rank regularized test statistic that uses the bound, *i.e.,* $W_{DV}(bound)$ exhibits higher power than its oversized competitors W and W_{LB} for alternatives sufficiently far from the null, *i.e.* for values of δ sufficiently different from zero (see Table 5, panel A, $n = 50$ that corresponds to $\delta = 0.1264$). However, when δ is close to zero, power is reduced for $W_{DV}(bound)$ (see Table 5, panel B, $n = 50$ with $\delta = 0.04$). Indeed for alternatives close to the null, W_{LB} benefits from a reduced critical point. Nevertheless, the simulated versions of the full rank statistic, especially the Simul-DGP version of W_{DV} has as much power as W_{LB} as soon as the sample size reaches $n = 100$ (see Table 5, panel B, $n = 100$ with $\delta = 0.04$). In particular for W_{DV}, we can observe as of $n = 100$ that power tends to increase when moving from Simul-R to Simul-E to Simul-DGP, with the highest power achieved for Simul-DGP which is the most demanding procedure in terms of information. More importantly, the locally-level corrected performance of the statistics W and W_{LB} corresponds to *infeasible* tests in practice, because this level correction requires the knowledge of the true, unknown parameter values unlike $W_{DV}(bound)$ whose level is controlled in all scenarios. The superiority of the simulated version of W_{DV} over the simulated version of the standard Wald statistic in small samples (*i.e.,* $n = 50, 100$ in panels A and B) is remarkable. Furthermore, the asymptotic test based upon $W_{DV}(bound)$, which uses a more refined regularization scheme than the crude uniform (ridge-type) regularization has more power than $W_{ridge-type}$ in small samples (see panel B, $n = 50, 100$ for alternatives close to the null). In other words, the uniform (ridge-type) regularization scheme tends to be overly conservative over the regularization we recommend (the DV-type one) and amplifies power loss in small samples. In the same vein, Bühlmann (2013) has emphasized that the ridge-type regularization – usually encountered in regression settings for coefficients' estimation – is not theoretically rate optimal in terms of power. Nonetheless, the conservativeness of the asymptotic tests – which is more acute for $W_{ridge-type}$ than for $W_{DV}(bound)$ – can be mitigated using simulations to increase power. Further, the behavior of the modified Wald statistic that results from adding noise to the restrictions to make them less efficient, as suggested by Lütkepohl and Burda (1997, Proposition 1, page 317), displays correct level under the null. However, such a noise tends to destroy power under the alternative and is not the approach we would recommend; compare W_{noise}'s performance in panel B, for n=50,..., 1000 relative to its competitors. Finally, the most striking result is the severe *under-performance* of the reduced rank statistic W_{LB} in a regular setup (panel C) when $c_n = \hat{\lambda}_1 n^{-1/3}$. As already mentioned by Lütkepohl and Burda (1997), by underestimating the true rank of the covariance matrix, this reduced rank statistic puts more weight on the first restriction that remains fulfilled in this case. A violation of the null hypothesis coming from the second restriction will not be detected by a statistic that underestimates the rank; a full-rank regularized statistic dominates in this respect. Thus, these results on power reinforce the better properties of the full-rank regularized statistics over the spectral cut-off type. However, when $c_n = \hat{\lambda}_1 n^{-1/2}$, power is restored for W_{LB} in regular setups. Indeed, in regular setups where regularization is unnecessary, dropping some restrictions might damage power significantly. Thus, the choice of c_n is critical in regular setups because it can diminish power substantially. The contrasting results

displayed for W_{LB} in panel C highlights the superiority of full-rank statistics over reduced-rank ones. Overall, we recommend $W_{DV}(bound)$ along with the Simul-DGP version of W_{DV}, as both procedures control level while achieving reasonably good power in small samples under both setups (regular and irregular).

E. Appendix: Rejection rules to correct size distortions

Insert Table 6 that is displayed below here.

Table 6. Empirical levels of Multistep noncausality tests $H_0 : \psi(\theta) = 0$ and modified rejection rules.

$H_0 : \psi(\theta) = 0$; nominal size= 0.05, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;					
Panel A: irregular setup					
$H_0 : \psi(\theta) = 0$ with with $\theta_{xy} = \theta_{xz} = \theta_{zy} = 0$ and $\theta_{xx} = \theta_{yy} = \theta_{zz} = -0.99$, $A_1 = A_{10}$, $c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;					
	$n = 50$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/4.13 when pv. ≤ 0.01	0.0499	1/1.81 when pv. ≤ 0.01	0.0499	0.0515
W_{DV}	1/2.11 when pv. ≤ 0.01	0.0499	1/1.67 when pv. ≤ 0.01	0.0500	0.0430
W_{LB}	1/ 2.10 when pv. ≤ 0.01	0.0500	1/ 2.108 when pv. ≤ 0.01	0.0500	0.0358
	$n = 100$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/2.88 when pv. ≤ 0.01	0.0500	1/1.503 when pv. ≤ 0.01	0.0499	0.0527
W_{DV}	1/1.34 when pv. ≤ 0.01	0.0500	1/1.245 when pv. ≤ 0.01	0.0499	0.0476
W_{LB}	1/ 1.335 when pv. ≤ 0.01	0.0500	1/ 1.49 when pv. ≤ 0.01	0.0500	0.0486
	$n = 500$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/1 when pv. ≤ 0.03	0.0502	1/1 when pv. ≤ 0.02	0.0342	0.0486
W_{DV}	1/1 when pv. ≤ 0.02	0.0238	1/1 when pv. ≤ 0.02	0.0290	0.0340
W_{LB}	1/ 1 when pv. ≤ 0.02	0.0238	1/ 1 when pv. ≤ 0.02	0.0302	0.0436
	$n = 1000$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/1 when pv. ≤ 0.05	0.0506	1/1 when pv. ≤ 0.03	0.0418	0.0436
W_{DV}	1/1 when pv. ≤ 0.04	0.0496	1/1 when pv. ≤ 0.03	0.0370	0.0318
W_{LB}	1/1 when pv. ≤ 0.04	0.0496	1/1 when pv. ≤ 0.03	0.0372	0.0470
	$n = 2000$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/1 when pv. ≤ 0.05	0.0300	1/1 when pv. ≤ 0.04	0.0440	-
W_{DV}	1/1 when pv. ≤ 0.04	0.0414	1/1 when pv. ≤ 0.04	0.0414	-
W_{LB}	1/1 when pv. ≤ 0.04	0.0414	1/1 when pv. ≤ 0.04	0.0418	-
	$n = 5000$				
Statistics	Rejection Rule	Simul-R	Rejection Rule	Simul-E	Simul-DGP
W	1/1 when pv. ≤ 0.05	0.0142	1/1 when pv. ≤ 0.05	0.0384	-
W_{DV}	1/1 when pv. ≤ 0.05	0.0368	1/1 when pv. ≤ 0.05	0.0378	-
W_{LB}	1/1 when pv. ≤ 0.05	0.0368	1/1 when pv. ≤ 0.05	0.0380	-

See Table 3 for the definition of the acronyms.