Wald-type tests when rank conditions fail: a smooth regularization approach *

Jean-Marie Dufour † Pascale Valéry ‡
McGill University HEC Montréal

February 4, 2011

*We are grateful to Jean-Claude Cosset, Director of Research at HEC Montreal, and to Marine Carrasco, Russell Davidson, John Galbraith and Raymond Kan for helpful comments as well as seminar participants in Marcel Dagenais econometric seminars at Université de Montréal. We are grateful to Sebastian Jaimungal as well as seminar participants in Actuarial Sciences and Mathematical Finance seminar series at the Fields Institute, and to the Department of Statistics, University of Toronto. This work has been presented at the NBER-NSF conference in Aarhus, Denmark in September 2008 and at the 26th Canadian Econometric Study Group in Ottawa in September 2009. This work has been presented to the 2010 Econometric Society World Congress in Shanghai, China. This work is supported by 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 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; 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

ABSTRACT

This paper examines Wald-type tests in presence of (possibly) singular covariance matrices. Two different types of singularity are addressed: *first*, the sample matrix has *full rank* but converges to a *singular* covariance matrix; in this case, the Wald statistic is still computable, but usual regularity conditions do not hold anymore, which modifies its asymptotic distribution. This asymptotic singularity causes the rank condition of Andrews (1987) to be violated at the limit due to isolated values of the parameter. *Second*, the sample matrix does not have full rank, but converges to a possibly nonsingular population matrix. This finite sample singularity may be due to redundant restrictions. To address such difficulties, we introduce a novel mathematical object: the *regularized inverse* that can be contrasted with the well-known *generalized* inverse, with its specific properties. A new class of *regularized* inverses can be defined that exploits *total eigenprojection* techniques, [Kato (1966), Tyler (1981)], together with a *variance regularizing function* (VRF) that modifies the small eigenvalues that fall below a certain threshold c so that their inverse is well defined. Under specific regularity conditions, the new regularized inverse converges to its regularized counterpart. This class of regularized inverses nests the spectral cut-off type inverse used by Lutkepohl and Burda (1997), and the Tikhonov-type inverse. We define *three* regularized Wald statistics: the first statistic admits a nonstandard asymptotic distribution, which corresponds to a linear combination of χ^2 variables if the restrictions are Gaussian. An *upper bound* is derived that corresponds to a χ^2 variable with *full rank*. The second regularized statistic relies on a *superconsistent* estimator of the eigenvalues at the threshold c whose distribution can be simulated. The third statistic lets the threshold vary with the sample size leading to the spectral cut-off modified Wald statistic of Lutkepohl and Burda (1997). The regularized statistics are consistent against global alternatives, with a loss of power for the spectral cut-off Wald statistic relative to the other statistics, as illustrated in a simulation exercise.

Key words: Regularized Wald test; Moore-Penrose inverse; spectral cut-off and Tikhonov regularizations; superconsistent estimator.
JEL classification: C1, C13, C12, C15, C32

Contents

List of Assumptions, Propositions and Theorems

List of Tables

1. Introduction

This paper examines Wald-type tests in presence of (possibly) singular covariance matrices. More specifically, we address two different types of singularity: *first*, the sample matrix has *full rank* but converges to a *singular* covariance matrix; in this case, the Wald statistic is still computable, but usual regularity conditions do not hold anymore, which modifies its asymptotic distribution. The claim made by Andrews (1987), and used by Lutkepohl and Burda (1997), is that if the sample matrix is consistent for a singular covariance matrix, then the use of a generalized inverse of the sample matrix instead of the g-inverse based on the population matrix will not affect the asymptotic distribution of the quadratic form, provided the sample matrix has the same rank as the population matrix with probability converging to one. Otherwise, the asymptotic distribution of the quadratic form is modified. Andrews's rank condition may be violated at the limit due to isolated values of the parameter. For instance, in the case of (highly) nonlinear restrictions, the rank of the derivative matrix of the restrictions may be lower for certain values of the parameter than for others. If this isolated value is true, the rank of the derivative matrix based on the consistent estimator will generally exceed that of the derivative matrix evaluated at the true value with probability bounded away from zero. Therefore, the weight matrix of the Wald statistic based on the estimator will not satisfy the rank condition when this isolated value is true, thus modifying the asymptotic distribution of the test statistic; see Dufour and Valéry (2009) for the stochastic volatility model with a Jacobian matrix that is degenerated (*i.e.* reduced rank) at an isolated value of the parameter.

Asymptotic singularity can arise due to highly nonlinear restrictions, as encountered in impulse response functions in VAR models, or when testing multi-step noncausality in VAR models, or testing Granger noncausality in VARMA models. Peñaranda and Sentana (2008) also faced an asymptotic singularity problem in the context of spanning tests in the return-mean-variance-frontier; in a GMM framework, the asymptotic covariance matrix of the sample moment conditions is singular under the null of spanning. Consequently, the Wald-type test does not have its standard asymptotic distribution anymore. One can also face asymptotic singularity with asymptotically redundant restrictions when testing candidate stochastic discount factors in the Hansen-Jaganathan distance, see Kan and Robotti (2009, p. 3461). Asymptotic singularity can be caused also by *superconsistent* estimators or any estimators that do not exhibit the conventional parametric speed of convergence. In this case, the Jacobian matrix can display a lower rank when the estimator is not appropriately scaled. While some authors investigate the possibility of multiple convergence rates, we adopt a systematic approach by regularizing the matrix. For multiple convergence speed estimators, see Antoine and Renault (2010*a*). Also, the presence of a cointegrating relationship implies the singularity of the VAR coefficient matrix $\Phi(1)$, and this can be related to tests of rank. More generally, any situations in a linear regression, where the matrix of the cross product of the covariates $\big(X'X\big)/T$ does converge to a singular population matrix are potential applications.

The *second* type of singularity our methodology can deal with corresponds to the case where the sample matrix does not have full rank, but converges to a possibly nonsingular population matrix. This finite sample singularity may be due to redundant restrictions. When dealing with highly nonlinear conditional moment restrictions as in Gallant and Tauchen (1989) in the I-CAPM framework, many of the parametric restrictions turn out to be redundant, creating thereby collinearity problems for the Jacobian matrix. Redundant moment restrictions also arise with the dynamic panel GMM estimator, when linear moment conditions imply nonlinear moment conditions under additional initial conditions on the dependent variable, see Arellano and Bond (1991), Ahn and Schmidt (1995), and Blundell, Bond and Windmeijer (2000).

To overcome the problem of asymptotic singularity, Lutkepohl and Burda (1997) proposes to reduce the rank of the matrix estimator in order to satisfy Andrews's rank condition. In so doing, they set to zero the small problematic eigenvalues to produce a consistent estimator for the rank of the population matrix. In the same vein, Gill and Lewbel (1992), Cragg and Donald (1996, 1997), Robin and Smith (2000) focus on tests for the rank of a matrix that is unobserved, but for which a $\sqrt{n}$ consistent estimator is available. In contrast, we tackle this problem differently by regularizing the matrix estimator, *i.e.* perturbing its small problematic eigenvalues. Whereas all those various procedures focus on the detection of zero eigenvalues to consistently estimate the rank of the asymptotic covariance matrix, we provide a smooth approach based on regularization tools that can benefit from not canceling the small problematic eigenvalues. In other words, by not dropping some restrictions, additional information can be exploited to increase power. While Gill and Lewbel (1992), Cragg and Donald (1996, 1997) require to *know* the rank of the asymptotic covariance matrix, Robin and Smith (2000) relaxes this assumption. However, unlike Cragg and Donald (1996, 1997) and Robin and Smith (2000) who assume Gaussianity for the limiting distribution of the covariance matrix estimator, our methodology based on Eaton and Tyler (1994) condition is more general, as Gaussianity is not required, nor the conventional $\sqrt{n}$ convergence speed; in this respect, compare our Assumption **2.3**, page 5, to Robin and Smith (2000), Assumption 2.2, page 154. In addition, a rank condition, (Assumption 2.2, page 155) relating the asymptotic covariance matrix to the characteristic vector matrices has to be satisfied for Robin and Smith (2000) tests of rank to hold, which requires more information on the characteristic vector structure of the matrix of interest; in practice, however, there is no guarantee for this assumption to be satisfied. Moreover, our methodology is simple and transparent compared to that of Robin and Smith (2000) that is more difficult to implement. Thus, their methodology can be viewed as an alternative to that of Lutkepohl and Burda (1997) to provide a consistent estimator for the rank of the population covariance matrix, or to consistently estimate the small problematic eigenvalues. Although our methodology can be applied to any procedure providing a consistent estimate for the rank of the population matrix, the availability of such a procedure is not necessary for the validity of our approach.

Knight and Fu (2000) have tackled the asymptotic singularity problem differently by working on the null space of the singular matrix on which there exists a positive definite matrix. More specifically, they study the asymptotic behavior of Bridge estimators in nearly singular designs and find that the resulting estimators have a slower rate of convergence than the usual root-n convergence rate. Therefore, the regularization helps preserve the usual root-n convergence rate of the estimators; see for instance Carrasco and Florens (2000),Carrasco, Chernov, Florens and Ghysels (2007).

Further advantages of our *regularization* approach that reinforces its generality are the following: the regularization technique we propose does not require a re-parametrization of the initial parameters, it is systematic for econometricians who want to apply it without efforts. Finding suitable transformations of the parameters that surmount the singularity problems can reveal tricky, almost infeasible, in highly nonlinear models for econometricians, as pointed out by Gallant and Tauchen (1989) in the I-CAPM framework. This is the approach proposed by Peñaranda and Sentana (2008), where the authors exploit some implicit restrictions on the initial parameters to reduce the number of parameters to identify. Simultaneously, they use a Moore-Penrose inverse, as Lutkepohl and Burda (1997), in the GMM criterion to reduce the number of moment conditions. In this way, the reduced set of moment conditions will locally identify a subset of the initial parameter vector. In so doing, the authors assume that the strongest collinearity in the design is restricted to the conditions that have no influence on the response. They also derive a reduced rank Wald test statistic in the GMM framework similar to that of Lutkepohl and Burda (1997).

While our main concern is testing, some authors make use of the related spectral decomposition based-tools, [Engl, Hanke and Neubauer (2000), Kress (1999)], to regularize estimators when a continuum of moments is used in a GMM or IV framework; see Carrasco and Florens (2000), Carrasco, Chernov, Florens and Ghysels (2007), Carrasco, Florens and Renault (2007), Carrasco (2007). In particular, Carrasco (2007) proposes some *modified IV estimators* based on different ways of inverting the covariance matrix of instruments. Indeed, when the number of instruments is very large with respect to the sample size or even infinite, the covariance matrix of the moment conditions becomes singular and some non-standard inverses are required. Also, when they are more moment conditions than observations, the covariance matrix of the moment conditions involved in the GMM criterion is singular. As noticed by Satchachai and Schmidt (2008), using a generalized inverse to overcome the singularity is not a good idea, as the value of the two step GMM criterion function is always less or equal to one. The problem is even worse for the continuous updating GMM, as its criterion function equals one for all parameter values.

Similarly, the GMM estimator of the parameters of the Consumption-based CAPM model based on the Euler equations involves the inverse of the covariance matrix of the moment conditions as a weighting matrix. There exist situations where this weighting matrix turns out to be singular. Indeed, the more asset returns are used in cross-section, the more information is available to identify the model parameters whose identification can be tricky when dealing with important nonlinearities. Indeed, an internal nonlinear habit function based on current and lagged consumption as the one specified in Chen and Ludvigson (2009) requires a lot of cross-sectional information to empirically identify the unknown habit function. However, the more assets used in cross-section, the more chance of collinearity, the higher the probability to end up with a singular weighting matrix which completely invalidates the usual tests. The same problem arises when assessing the pricing errors of different candidate Stochastic Discount Factors models through the *Hansen-Jagannathan distance*. Indeed, this measure of asset pricing model misspecification involves a sample second moment matrix of the N assets as a weighting matrix. A large number of assets can create collinearity and hence singularity difficulties arise that break down standard inference. From an asset pricing perspective, the availability of an inverse that does not amplify *excessively* the pricing errors is crucial for portfolio allocation. In contrast, risk management focuses instead on a precise estimator of the covariance matrix. See Fan, Fan and Lv (2006) for an examination of the properties of high dimensional covariance matrix estimators in the context of observable factor models.

By contrast, we provide valid asymptotic or simulation-based *regularized Wald* test procedures that can deal with such problems. Moreover, our *regularization* approach goes beyond the GMM framework and can accommodate any consistent estimator as the Sieve Minimum Distance estimator used in Chen and Ludvigson (2009) for the habit-based asset pricing model.

It is important to stress another situation where the Jacobian matrix of the moment conditions in a GMM framework can have a deficient rank due to (first-order) underidentification. This is the problem studied by Dovonon and Renault (2009), where the authors overcome the weak identification problems by going a step further and examine a second-order identification condition. Deficient rank problems due to identification issues go beyond the scope of the present paper. Even though we allow the underlying parameter θ to be unidentified, unlike Lutkepohl and Burda (1997), we assume that a transformation of it, that is $\psi(\theta)$, is identified. So the kind of rank deficiency we consider in this paper does not come from (weak) identification problems; see also Antoine and Renault (2009), and Antoine and Renault (2010*b*) for such issues.

When dealing with singular covariance matrices, usual inverses are discarded and replaced with *generalized* inverses, or *g-inverses* [see Moore (1977), Andrews (1987) for the generalized Wald tests] or modified inverses proposed by Lutkepohl and Burda (1997). However, when using non-standard inverses, econometricians are not always aware of two difficulties. *First*, the well-known continuous mapping theorem so widely used by econometricians to derive asymptotic distributional results for test statistics does not apply anymore because g-inverses are not (necessarily) continuous. This fact has been observed by Andrews (1987). In addition, eigenvectors are not continuous functions in the elements of the matrix unlike the eigenvalues. *Second*, when performing the singular value decomposition of a matrix, the eigenvectors corresponding to eigenvalues with multiplicity larger than one, are not uniquely defined, which may rule out the convergence of the estimates towards population quantities. Ignoring such concerns may lead to distributional results that are strictly speaking *wrong*.

To address such difficulties, we introduce a class of *regularized* inverses that exploits *total eigenprojection* techniques, *i.e.* an eigenprojection operator taken over a subset of the spectral set. Following Kato (1966) and Tyler (1981), we work with the *eigenprojections* in order to overcome the discontinuity and non-uniqueness features of eigenvectors. The eigenprojection projects onto the *invariant* (to the choice of the basis) eigenspace, *i.e.* the subspace generated by the eigenvectors. A lemma given by Tyler (1981) states the continuity property for the *total eigenprojection*. In this way, the important continuity property is preserved for eigenvalues and eigenprojections even though eigenvectors are *not* continuous. In addition to this total eigenprojection technique, we define a perturbation function of the inverse of the eigenvalues called *variance regularizing function* (VRF) that modifies the small eigenvalues that fall below a certain threshold so that their inverse is well defined whereas the large eigenvalues remain unchanged. The class of admissible VRF has to satisfy certain continuity and boundedness properties with additional regularity conditions so that the regularized inverse does converge to its regularized counterpart. Otherwise the convergence result (stated with a fixed value of the threshold) may break down. Our regularized inverse does nest the spectral cut-off type inverse used by Lutkepohl and Burda (1997), and other modified inverses as in Valéry (2005). The distributional theory of the test statistics then expressed as a transformation of the regularized inverse, hence of the total eigenprojections, will be greatly simplified and valid.

Our contributions can be summarized as follows. *First*, we introduce a novel mathematical object: the *regularized inverse* that can be contrasted with the well-known *generalized* inverse, with its specific properties. This new class of inverses can have *full rank* or reduced rank, and satisfies a decomposition result: a *regular* component built on large eigenvalues while the others involving the small eigenvalues may not be *regular*. This block decomposition of the inverse, coming from spectral decomposition tools, is important insofern as it is carried over to the test statistic itself, and is useful to get an insight on the structure of the distribution. *Second*, under specific regularity conditions on the VRF, the regularized inverse is shown to converge to its regularized full rank counterpart, with the convergence holding component by component. Besides, our regularized inverse class is general and *does nest* the spectral cut-off type inverse, or the modified Moore-Penrose inverse proposed by Lutkepohl and Burda (1997), or the Tikhonov regularized inverse. *Third*, we define *three* regularized Wald statistics: the first two statistics rely on a fixed value for the threshold in the VRF $g(\lambda; c)$ while the third one lets the threshold vary with the sample size, but requires more information about the sample behavior of the eigenvalues, see Eaton and Tyler (1994) for the distributional theory of the sample eigenvalues of a matrix. *Fourth*, 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* is then derived for this first regularized statistic under general laws for the restrictions; such a bound corresponds to a χ^2 variable with *full rank* under Gaussianity. Hence, the test is *asymptotically valid*, meaning that the usual critical point (given by the χ^2 variable with *full rank*) can be used, but is conservative. *Fifth*, the second regularized statistic relies on a *superconsistent* estimator of the eigenvalues at the threshold c whose distribution can be simulated. Interestingly, we observe that simulating the distribution of the superconsistent estimator-based regularized statistic makes it unsensitive to the choice of the threshold. In other words, simulating the distribution makes the regularized statistic less sensitive to the tuning parameters. *Sixth*, when the threshold goes to zero with the sample size, we obtain the spectral cut-off modified Wald statistic of Lutkepohl and Burda (1997) as a special case. Under normality, the test has the asymptotic χ^2 distribution with a reduced rank, *i.e.* the number of eigenvalues greater than zero. Note that Lutkepohl and Burda (1997) result only holds for distinct eigenvalues whereas our result accounts for eigenvalues with multiplicity larger than one. *Seventh*, we also show that the regularized statistics are consistent against global alternatives, but the spectral cut-off Wald test used by Lutkepohl and Burda (1997) has reduced power in some directions of the alternative, as illustrated in a Monte Carlo simulation.

Finally, we investigate, in a Monte Carlo experiment, the finite sample properties of the (regularized) test statistics under two different designs: *first*, under Gaussianity, the full-rank regularized statistic using the conservative bound tends to underreject the null hypothesis in singular designs, while the full-rank regularized statistic based on the superconsistent estimator of the eigenvalues displays the right level *asymptotically* (for a sufficient large value of the threshold). In contrast, the spectral cut-off modified Wald statistic proposed by Lutkepohl and Burda (1997) tends to overreject the null hypothesis in small samples, with severe size distortions when the process approaches the nonstationary region. Using a reduced critical point in a singular design, their statistic reaches the right level asymptotically. As for the standard Wald statistic, its behavior is clearly modified in singular designs, either suffering from severe overrejections in small samples (especially for parameter values set to -0.99), or underrejections in large samples. Regarding power properties, although the bound is conservative, it does *not* entail a loss of power under the alternative, which makes it attractive. Further, our regularization approach is systematic and robust to both designs, regular and irregular, whereas the modified Moore-Penrose statistic has reduced power in regular designs. Indeed, by setting to zero the small eigenvalues, the modified Moore-Penrose statistic does not exploit the additional information unlike the full-rank regularized statistics. *Second*, when deviating from normality, the standard Wald statistic along with the spectral cut-off statistic strongly overreject the null hypothesis, with empirical size frequencies varying between 0.17 and 0.50 compared to a 0.05 level test. In contrast, the full-rank regularized statistics that allow different probability distributions achieve to control for the size without losing power. Overall, the full-rank regularized statistic that uses the bound is very appealing, as it always controls for size, does not imply reduced power, is robust to both designs, regular and irregular, and is easier to implement compared to the simulation-based superconsistent estimator full-rank competitor. Moreover, the standard Wald statistic and the modified Moore-Penrose Wald statistic are *infeasible* tests in practice, as they overreject the null when the process is close to the nonstationary region. Besides, the modified Moore-Penrose Wald statistic requires to know whether we are in a singular or nonsingular design to choose the reduced critical point; this makes it less attractive in practice.

The paper is organized as follows. In Section 2 we describe a general framework with minimal assumptions. In Section 3, we provide specific examples found in the literature, where the researcher can face (asymptotic) singularity covariance matrices that modify the asymptotic distribution of the standard Wald test statistic. Then, we introduce the class of *regularized* inverses as opposed to *generalized* inverse in Section 4 followed by the *regularized* test statistic in Section 5. More specifically, a decomposition of the test statistic is identified through the corresponding decomposition of the covariance matrix. In Section 6 we review and adapt some results on total eigenprojections to derive the convergence results for the regularized inverses. In particular, we emphasize some (non)uniqueness and (dis)continuity properties related to eigenvectors of a given matrix and resort to total eigenprojection techniques to surmount such difficulties. In Section 7, we establish the asymptotic properties of the new regularized inverse based on fixed threshold. In Section 8, we state new asymptotic distributional results for the regularized Wald test statistic using a fixed threshold and exploit the decomposition of the regularized statistic to derive an upper bound. In Section 9, we propose a new statistic based on a superconsistent estimator at c of the eigenvalues. In Section 10, we find as a special case the Lutkepohl and Burda (1997) result in the Gaussian case. Finally an application to causality testing is provided in Section 3 followed by simulation results in Section 11. Concluding remarks follow while the proofs are gathered in the appendix.

2. Framework

We want to test a null hypothesis of the form

$$H_0(\psi_0) : \psi(\theta) = \psi_0 \tag{2.1}$$

where $\psi(\theta) \in \Omega \in \mathbb{R}^q$ is the parameter of interest with the parameter θ identifying the true underlying data generating process. A usual test statistic for testing the null hypothesis is the Wald statistic as soon as we can find a consistent estimator $\hat{\psi}_n$ of the restrictions no matter where it comes from, *i.e.*,

$$W_n(\psi_0) = a_n^2[\hat{\psi}_n - \psi_0]'\Sigma_n^{-1}[\hat{\psi}_n - \psi_0] \tag{2.2}$$

provided the inverse of the weighting matrix exists. a_n represents a convergence rate that may be *different* from the conventional $\sqrt{n}$ to precisely allow situations where some components of $\hat{\psi}_n$, or linear combinations of them, may converge faster or slower than $\sqrt{n}$. It is well-known in the faster case that *superconsistent* estimators can raise asymptotic singularity problems, when not suitably scaled. Usually, Σ_n is a consistent estimator of the restriction covariance matrix Σ in order to get a chi square distribution, as specified in Assumption **2.4** below. For another choice of Σ_n, the Wald test will not have the standard chi square distribution, but can still be conducted. In this paper, we shall place ourselves under weak assumptions contrary to the ones usually made in the econometric literature to conduct such a test. First, we will *not* assume the restrictions $\psi(\theta)$ to be differentiable with respect to (w.r.t.) the underlying parameter θ. Such a differentiability assumption unnecessarily restricts the set of admissible restrictions and can be avoided. To do so, we assume that a consistent estimator $\hat{\psi}_n$ is available satisfying the following assumption, where the notation $\underset{n\to\infty}{\overset{\mathcal{L}}{\to}}$ denotes the usual convergence in law, and $\mathcal{L}(X)$ the law of X.

Assumption 2.1 CONVERGENCE IN LAW OF THE RESTRICTIONS. *a_n is a sequence of real constants such that $a_n \to \infty$, and*

$$X_n = a_n(\hat{\psi}_n - \psi) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} X \tag{2.3}$$

where $\mathcal{L}(X)$ is known.

This assumption significantly enlarges the family of admissible laws for $\hat{\psi}_n$, or $\psi(\hat{\theta}_n)$, $\hat{\theta}_n$ being a consistent estimator of θ. For instance, the typical Gaussian distribution for X can easily be replaced by a chi-square distribution. Generally speaking, any distribution that can be consistently estimated by simulations is admissible. Therefore, if $\mathcal{L}(X)$ 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 that ψ is identified; in other words, θ can be unidentified, but there exist transformations of θ, *i.e.* $\psi(\theta)$, that can be identified. Whereas Lutkepohl and Burda (1997) assume the availability of an asymptotic gaussian estimator of θ, as in equation (2.10), that restricts unnecessarily to situations where θ is identified, we relax this assumption here. Note that ψ will alternately equal ψ_0 under the null hypothesis, or ψ_1 under the alternative. Further assumptions are required on the limiting weighting matrix Σ to obtain a componentwise characterization of the *modified* Wald statistic.

Assumption 2.2 EIGENSPACE AND EIGENPROJECTION. *The $q \times q$ matrix Σ is such that: $\forall\ j = 1, \ldots, k$, with $1 \leq k \leq q$,*

$$B(d_j) = \Big(v(d_j)_l\Big)_{l=1,\ldots,m(d_j)} \tag{2.4}$$

forms an orthonormate basis for the eigenspace

$$\mathcal{V}(d_j) = \{v \in \mathbb{R}^q, |\ \ \Sigma v = d_j v\} \tag{2.5}$$

such as

$$\Sigma = \sum_{d_j} d_j P_j(\Sigma) \tag{2.6}$$

where

$$P_j(\Sigma) = P(d_j)(\Sigma) = B(d_j)B(d_j)' \tag{2.7}$$

where the d_j's denote the k distinct eigenvalues of Σ with multiplicity $m(d_j)$ such that $q = \sum_{j=1}^{k} m(d_j)$.

Most of the time, the weighting matrix Σ, as well as its sample analog Σ_n, is interpreted as a covariance matrix. Nevertheless, such an interpretation is very restrictive and discards distributions whose moments do not exist, e.g., the Cauchy distribution. Therefore, Assumptions **2.1** and **2.3** are purposedly formulated to allow such degenerate distributions. A general condition, given by Eaton and Tyler (1994), states the convergence result for this set of parameters.

Assumption 2.3 EATON-TYLER CONDITION. *Σ_n is a sequence of $p \times q$ real random matrices and Σ is a $p \times q$ real nonstochastic matrix such that*

$$Q_n = b_n(\Sigma_n - \Sigma) \overset{\mathcal{L}}{\to} Q \tag{2.8}$$

where b_n is a sequence of real constants such that $b_n \to +\infty$ and Q a random matrix.

Again, this assumption is general and allows situations, unlike Robin and Smith (2000), where the matrix estimator is not asymptotically Gaussian. Eaton-Tyler condition is stated for rectangular matrices, but most of the time we will consider square matrices that are symmetric matrices with real eigenvalues. Assumptions **2.1** and **2.3**, together with relaxing the assumption of convergence of ranks, will define the cornerstone for the validity of the distributional results developed further. In addition, it is important to note that the generality of Assumption **2.3** allows for 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, specifically for non-continuous distributions of matrices and their eigenvalues, to provide a thorough distributional theory.

A special case of Assumptions **2.1** and **2.3** that is usually encountered in the econometric literature consists in specifying a Gaussian distribution for X whose parameterization hinges on Σ with $a_n = \sqrt{n}$ as in Lutkepohl and Burda (1997).

Assumption 2.4 ROOT-n ASYMPTOTIC NORMALITY.

$$X_n = \sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta)) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} X = N(0, \Sigma) \tag{2.9}$$

where Σ is a $q \times q$ matrix.

Note that the most degenerate case corresponding to $\Sigma = 0$ is allowed by Assumption **2.4**. In this case, $d_j = 0$, with $m(0) = q$. Usually, the asymptotic normality of the restrictions is deduced from the root-n asymptotic normality of the estimator $\hat{\theta}_n$ of the underlying parameter θ through the delta method, *i.e.*,

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \Sigma_\theta) . \tag{2.10}$$

This requires the differentiability of the restrictions unlike Assumption **2.1**. In so doing, 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 (2.10), the covariance matrix has the typical form

$$\Sigma = P(\theta)\Sigma_\theta P(\theta)' \tag{2.11}$$

which critically hinges on the differentiability of the restrictions, *i.e.*

$$P(\theta) = \partial\psi/\partial\theta'$$

as in Lutkepohl 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 statistic but only the *covariance* matrix of the restrictions yields the standard chi-squared distribution. If a different weighting matrix is used instead, the distribution may be modified as seen further.

Further, among usual regularity conditions made, when conducting tests based on quadratic forms such as Wald-type tests, is the well-known rank condition for the covariance matrix. When Σ and Σ_n have full ranks, we are in the regular case with the $q \times q$-weighting matrix Σ being nonsingular, and therefore $W_n(\psi_0)$ has an asymptotic $\chi^2(q)$ distribution. This is not necessarily true, however, if Σ is singular. In this case, Σ does not admit a usual inverse, but can still be inverted by means of a generalized inverse, or a *regularized* inverse as shown later on. However, when the population matrix Σ has a reduced rank, additional conditions are required. This is the case covered by Andrews (1987).

Assumption 2.5 CONVERGENCE OF THE RANKS. *Σ and Σ_n are matrices such that*

$$P\big[\mathrm{rank}(\Sigma_n) = \mathrm{rank}(\Sigma)\big] \to 1, \ \textit{with } |\Sigma| \geq 0$$

and n growing to infinity, where $|\cdot|$ stands for the determinant.

In other words, the rank of the sample matrix has to converge almost surely (a.s.) towards the *reduced rank* of the population matrix in order for the quadratic form to have a limiting χ^2 distribution, with fewer degrees of freedom, under Gaussianity. We shall relax this assumption in the paper.

To tackle the problem of ranks that do not converge, unlike Moore (1977), Andrews (1987) and Lutkepohl and Burda (1997) who use a reduced rank estimator for the covariance matrix, such as the spectral cut-off Moore-Penrose inverse, we shall eventually increase the rank, by regularizing the smallest eigenvalues instead. In so doing, the modified matrix will converge to a different object, affecting thereby the limiting distribution. It is important to note that the regularization approach exposed next embed all rank possibilities, including the spectral cut-off reduced rank. Also, the regularization techniques proposed to deal with incomplete ranks, when (possibly) combined with simulated testing procedure, holds under weak assumptions as Assumptions **2.1** and **2.3**. In Section 5, we introduce the *regularized* Wald test statistic based on *regularized* inverses of the covariance matrix as a way to handle such difficulties. Let us introduce before the class of *regularized* inverses, as opposed to the class of *generalized* inverses.

3. Examples

In this section, we provide examples where the econometrician can face (asymptotic) singularity of the covariance matrix that will affect the asymptotic distribution of the Wald test statistic.

3.1. Multistep noncausality under Gaussianity

As already observed by Lutkepohl and Burda (1997), testing noncausality restrictions may raise some singularity problems for the Wald test. We shall reconsider the example provided by Lutkepohl and Burda (1997) in our specific regularization design. A VAR(1) process is considered for the (3×1)vector $\mathrm{y}_t = [x_t \ y_t \ z_t]'$ as follows:

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

Consider

$$Y \equiv (\mathrm{y}_1, \ldots, \mathrm{y}_n) \quad (3 \times n)$$
$$B \equiv (A_1) \quad (3 \times 3)$$
$$Z_t \equiv [\mathrm{y}_t] \ \ (3 \times 1) \quad Z \equiv (Z_0, \ldots, Z_{n-1}) \quad (3 \times n)$$
$$U = (u_1, \ldots, u_n) \quad (3 \times n)$$

where $u_t = [u_{x,t} \ u_{y,t} \ u_{z,t}]'$ is a white noise with (3×3) nonsingular covariance matrix Σ_u. Let $\alpha = \mathrm{vec}(A_1) = \mathrm{vec}(B)$. Testing $H_0: \ y_t \overset{(\infty)}{\not\to} x_t$ requires to test $h = pK_z + 1 = 2$ restrictions on α [see Dufour and Renault (1998)] of the form:

$$r(\alpha) = \begin{bmatrix} \alpha_{xy} \\ \alpha_{xx}\alpha_{xy} + \alpha_{xy}\alpha_{yy} + \alpha_{xz}\alpha_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

These restrictions are fulfilled in the following three different parameter settings

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

But we can observe that the first-order partial derivative of the restrictions leads to a singular matrix

$$\frac{\partial r}{\partial \alpha'} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{xy} & 0 & 0 & \alpha_{xx} + \alpha_{yy} & \alpha_{xy} & \alpha_{xz} & \alpha_{zy} & 0 & 0 \end{bmatrix} \tag{3.2}$$

if (3.1) holds. Under such circumstances, the Wald test does not have the standard χ^2-distribution under the null. To perform the Wald test, let us consider the multivariate LS estimator of $\alpha = \mathrm{vec}(A_1) = \mathrm{vec}(B)$. Using the column stacking operator vec we have:

$$Y = BZ + U \tag{3.3}$$

or

$$\mathrm{vec}(Y) = \mathrm{vec}(BZ) + \mathrm{vec}(U) \tag{3.4}$$
$$y = \big(Z' \otimes I_3\big)\mathrm{vec}(B) + \mathrm{vec}(U) \tag{3.5}$$
$$y = \big(Z' \otimes I_3\big)\alpha + u \tag{3.6}$$

where $\mathrm{E}(uu') = I_3 \otimes \Sigma_u$. The multivariate LS estimator $\hat{\alpha}$ is given by:

$$\hat{\alpha} = \Big((ZZ')^{-1}Z \otimes I_3\Big)y \ . \tag{3.7}$$

The asymptotic distribution of the multivariate LS estimator:

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N\big(0, \Gamma^{-1} \otimes \Sigma_u)\big) \tag{3.8}$$

implies the asymptotic distribution for the restrictions:

$$\sqrt{n}(r(\hat{\alpha}) - r(\alpha)) \xrightarrow{\mathcal{L}} N\big(0, \Sigma_{r(\alpha)})\big) \tag{3.9}$$

where

$$\hat{\Sigma}_{r(\alpha)} = \frac{\partial r}{\partial \alpha'}(\hat{\alpha})\hat{\Sigma}_\alpha \frac{\partial r'}{\partial \alpha}(\hat{\alpha}) \tag{3.10}$$

is a consistent estimator for $\Sigma_{r(\alpha)}$ and

$$\hat{\Sigma}_\alpha = \hat{\Gamma}^{-1} \otimes \hat{\Sigma}_u \tag{3.11}$$

is a consistent estimator for Σ_α with

$$\hat{\Gamma} = \frac{1}{n}ZZ' \tag{3.12}$$

and

$$\hat{\Sigma}_u = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t \hat{u}_t' = \frac{1}{n}Y\big[I_n - Z'(ZZ')^{-1}Z\big]Y' \ . \tag{3.13}$$

From the asymptotic distribution (3.9), a Wald-type test is easily obtained to test the null $H_0 : r(\alpha) = 0$, *i.e.*

$$W_\psi = nr(\hat{\alpha})' \hat{\Sigma}^R_{r(\alpha)} r(\hat{\alpha}) \tag{3.14}$$

where a regularization is required under parameter setting (3.1).

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:

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

based on the following moment conditions:

$$\mu_2(\theta_1) = \mathsf{E}(u_t^2) = r_y^2 \exp[(1/2) r_w^2/(1 - a_w^2)] \,,$$

$$\mu_4(\theta_1) = \mathsf{E}(u_t^4) = 3 r_y^4 \exp[2 r_w^2/(1 - a_w^2)] \,,$$

$$\mu_{2,\,2}(1|\theta_1) = \mathsf{E}[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1 - a_w)] \,.$$

When testing for homoskedasticity ($a_w = r_w = 0$), in this model, which 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 appears to be 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 shown that

$$\frac{\partial \mu}{\partial \theta'} = \begin{bmatrix} 0 & 0 & 2r_y \\ 0 & 0 & 12 r_y^3 \\ 0 & 0 & 4 r_y^3 \end{bmatrix} \tag{3.15}$$

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). An important regularity condition is violated. This raises estimation difficulties and was handled by redefining the estimator in this case: we set $a_w = r_w = 0$ and $r_y = \sqrt{\mu_2(\theta_1)}$ when $\kappa \leq 3$. Further, $\partial\mu/\partial\theta'$ 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 θ, 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 is again a violation of a standard regularity condition, and the Wald statistic has a non-regular asymptotic distribution.

3.3. Asymptotic singularity for $(X'X)/T$ in (linear) regressions

More generally, each time the matrix of the cross product of the covariates , $(X'X)/T$, does converge to a singular population matrix, the standard Wald test will fail to have its conventional distribution, as illustrated in the two examples below.

3.3.1. Degenerate factors

Suppose we want to test whether macroeconomic fundamentals, like real and inflation factors in Ludvigson and Ng (2009), have no forecasting power for future excess returns on U.S. government bonds beyond the predictive power contained in forward rates and yield spreads as in Cochrane and Piazzesi (2005). The forecasting regression of excess bond returns on estimated common factors, and possibly nonlinear functions of those factors has the following form:

$$y_{t+1} = \gamma_0 + \gamma_1 F_{1t} + \gamma_2 F_{2t} + \gamma_3 F_{3t} + \gamma_4 F_{4t} + \gamma_5 F_{5t} + \gamma_6 F_{6t} + \gamma_7 CP_t + \epsilon_{t+1} \tag{3.16}$$

where $F_{2t} = F_{1t}^{\beta}$, and $\epsilon_{t+1} \overset{i.i.d.}{\sim} N(0,1)$; the corresponding forecasting regression of excess bond returns averaged across maturity, that is $\frac{1}{4}\sum_{N=2}^{5} rx_{t+1}^{(N)}$, on a linear combination of factors, proposed by Ludvigson and Ng (2009), and Cochrane and Piazzesi (2005), is given by:

$$\frac{1}{4}\sum_{N=2}^{5} rx_{t+1}^{(N)} = \gamma_0 + \gamma_1 F_{1t} + \gamma_2 F_{1t}^3 + \gamma_3 F_{2t} + \gamma_4 F_{3t} + \gamma_5 F_{4t} + \gamma_6 F_{8t} + \gamma_7 CP_t + \epsilon_{t+1} \ , \tag{3.17}$$

where CP_t is the Cochrane and Piazzesi (2005) factor that is a linear combination of five forward spreads. Suppose we want to test the null hypothesis that the sixth macroeconomic factor has no predictive power, *i.e.* $H_0: \ \gamma_6 = 0$ against $H_1: \ \gamma_6 \neq 0$ in equation (3.16), when the true data generating process (DGP) corresponds to $\beta = 0$. Thus, for some isolated value of the parameter space, the factor loading γ_2 on the second factor is not identified, as the constant term turns out to be $\tilde{\gamma}_0 = \gamma_0 + \gamma_2$. Therefore, the sample covariance matrix $\varSigma_n$ based on the finite sample estimates will converge to a singular population matrix $\varSigma_F$ under the true DGP, *i.e.*,

$$\varSigma_n = \frac{1}{n}\sum_{t=0}^{n} F_t F_t' \ \xrightarrow{p} \ \varSigma_F \geq 0 \ ,$$

where $F_t = (\iota, F_{1t}, F_{1t}^{\hat{\beta}}, F_{3t}, F_{4t}, F_{5t}, F_{6t})'$, with $\hat{\beta}$ a consistent estimate of β. Provided we can find a consistent estimate $\hat{\varGamma}_n$ of $\varGamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6)'$, the asymptotic distribution of the Wald statistic

$$n\hat{\gamma}_6[\hat{\varSigma}_{\hat{\gamma}_6}]^{-1}\hat{\gamma}_6$$

will be modified due to the reduced rank of $\varSigma_F$, when $\beta = 0$.

3.3.2. Asymptotic singularity in event studies

Finally, we can also face asymptotic singularity problems when conducting event studies in the following specification:

$$r_t = \alpha + \beta r_{mt} + \gamma d_t + \epsilon_t, \quad t = 1, \ldots, T \ ,$$

where r_t denotes the stock return at time t, r_{mt} the return on the market portfolio at time t, and d_t is the event dummy such that: $d_t = 1$ for the event day, and zero otherwise. If the event occurs only once in the sample, then the matrix of the cross product of the covariates , $(X'X)/T$, will converge to a singular population matrix. Hence, the standard Wald statistic for testing some hypotheses on the parameters will not have its conventional distribution.

3.4. Singularity issues when testing SDF candidates with the Hansen-Jagannathan distance

Let $y(\gamma)$ be a stochastic discount factor (SDF) candidate involving some unknown parameters γ, and R be a vector of gross returns on N test portfolios; see Hansen and Jagannathan (1991) for SDF. $y(\gamma)$ is said to be misspecified if for all values of γ, the pricing errors $e(\gamma)$ is nonzero, *i.e.*,

$$e(\gamma) = E[Ry(\gamma)] - 1_N \neq 0_N \; . \tag{3.18}$$

The famous Hansen and Jagannathan (1997) distance, henceforth HJ-distance, for assessing specification errors in stochastic discount factor models, is defined as the square root of a quadratic form of the pricing errors:

$$\delta = \left[e(\gamma)' U^{-1} e(\gamma) \right]^{1/2}, \tag{3.19}$$

where $U = E[RR']$. When the model is misspecified, the HJ-distance is defined as

$$\delta = \left[\underset{\gamma}{min}\; e(\gamma)' U^{-1} e(\gamma) \right]^{1/2} . \tag{3.20}$$

in the empirical asset pricing literature. Kan and Zhou (2006) show that for linear factor models it is equivalent to use the inverse of the covariance matrix of gross returns instead of the second sample moments, *i.e.* V_{22}^{-1} instead of U^{-1} in the HJ-distance. Kan and Robotti (2009) also focus on linear factor asset pricing models:

$$y(\gamma) = \gamma' x$$

where $x = [1, F']'$, and the pricing errors of the N assets are given by:

$$e(\gamma) = E[Ry(\gamma)] - 1_N = E[Rx'\gamma] - 1_N = D\gamma - 1_N \; ,$$

where $D = E[Rx'] = [\mu_2, V_{21} + \mu_2 \mu_1']$. Kan and Robotti (2009) use V_{22}^{-1}as a weighting matrix in the squared HJ-distance:

$$\begin{aligned} \delta^2 &= \underset{\gamma}{min} \big[D\gamma - 1_N \big]' V_{22}^{-1} \big[D\gamma - 1_N \big] \\ &= 1_N' V_{22}^{-1} 1_N - 1_N' V_{22}^{-1} D \Big(D' V_{22}^{-1} D \Big)^{-1} \Big(D' V_{22}^{-1} 1_N \Big) . \end{aligned} \tag{3.21}$$

The unique value of γ that minimizes $e(\gamma)' V_{22}^{-1} e(\gamma)$ is given by:

$$\gamma_{HJ} = \Big(D' V_{22}^{-1} D \Big)^{-1} \Big(D' V_{22}^{-1} 1_N \Big) , \tag{3.22}$$

provided that V_{21} is of *full column rank*, which implies that D is also of *full column rank*. However, if some factors in F are nonpervasive and do not contribute to the variance of the gross returns R, V_{21} may be singular along with D. As a result, γ_{HJ} and the minimized value of δ^2 are not defined anymore because the matrix $\Big(D' V_{22}^{-1} D \Big)^{-1}$ is singular.

Kan and Robotti (2009) consider two competing SDF models: SDF of model 1 is given by $y_1 = \eta' x_1$, with $x_1 = [1, f_1', f_2']$, while SDF of model 2 is given by $y_2 = \beta' x_2$, with $x_2 = [1, f_1', f_3']$. When the dimension K_2 of the second factor is equal to zero, model 2 nests model 1. For non-nested models, Kan and Robotti (2009) shows that testing equality of two SDF, $y_1 = y_2$, imposes restrictions on η and β: $y_1 = y_2$ holds if and only if $\eta_1 = \beta_1$,

$\eta_2 = 0_{K_2}$, and $\beta_2 = 0_{K_3}$. However, the restriction $\eta_1 = \beta_1$ is redundant because it is implied by $\eta_2 = 0_{K_2}$ and $\beta_2 = 0_{K_3}$. Let $\psi = [\eta_2', \beta_2']'$. Hence, Kan and Robotti (2009) notes in a footnote 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 $\sqrt{n}\big[\hat{\eta}_1' - \hat{\beta}_1', \hat{\psi}'\big]'$ 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."

3.5. Spanning tests in the Return Mean Variance Frontier with asymptotic singularity

Peñaranda and Sentana (2008) examine spanning tests in the Return Mean Variance Frontier (RMVF). They test if there is simultaneous tangency at two points. They denote c_i^{-1} and c_{ii}^{-1} two arbitrary expected returns. The null of spanning can be written as:

$$H_0 \ : \ a(c_i) = 0, \ \ a(c_{ii}) = 0$$

where the regression intercepts $a(c_i)$ and $a(c_{ii})$ are defined by the moment conditions:

$$E\big\{H_L\big[R; a(c_i), b(c_i), a(c_{ii}, b(c_{ii})\big]\big\} = 0 \ .$$

But it has been pointed out by Marin (1996), Peñaranda and Sentana (2008) and that the asymptotic covariance matrix of the sample analog of the moment conditions is singular under the null. Hence, the conventional distributional theory of the Wald-type test does not hold anymore. To deal with this issue, Penaranda and Sentana (2008) propose for spanning tests in RMVF, and in Stochastic Discount Frontier Mean-Variance frontiers introduced by Hansen and Jagannathan (1991)), a modified GMM estimator under singularity of the covariance matrix (GMMS). Their methodology consists in replacing the ordinary inverse of Σ by a generalized inverse, the Moore-Penrose, while imposing parametric restrictions in order to work with a smaller number of parameters. By decreasing both the number of parameters and the number of moment conditions, they avoid singularity. Hence, they propose, like Lutkepohl and Burda (1997), a Wald test statistic with *reduced rank* based on a modified GMM estimator.

3.6. Deviation from Normality: the Delta method breaks down

Our approach is not restricted to Gaussian distributions; we allows situations where the restrictions to be tested are not Gaussian anymore. Indeed, we examine a situation where the restrictions are asymptotically distributed as a chi-squared variable. Suppose the underlying parameter θ is a $p \times 1$ vector such as

$$\sqrt{n}(\hat{\theta}_n - \theta) \sim N[0, I_p] \ , \tag{3.23}$$

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

$$H_0(\psi_0) : \psi(\theta) = \theta'\theta = 0 \ . \tag{3.24}$$

The data generating process corresponding to (3.23) is:

$$Y = \theta\iota + u \ , \quad u \sim N[0, I_p] \ ,$$

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

$$\hat{\theta}_n = [(\iota\iota')^{-1}\iota \otimes I_p]y = \frac{1}{n}(\iota \otimes I_p)y \tag{3.25}$$

where $y = vec(Y)$ is $pn \times 1$. Under the null, it is easily seen that the restrictions do not have the conventional $\sqrt{n}$ convergence speed as usual. Thus, its distribution under the null, where $\theta = 0$, is

$$n\psi(\hat{\theta}_n) = (\sqrt{n}\hat{\theta}_n)'(\sqrt{n}\hat{\theta}_n) \overset{a}{\sim} \chi^2(p) \, . \tag{3.26}$$

The weighting matrix used in the quadratic form is:

$$\Sigma = P(\theta)\Sigma_\theta P(\theta)', \quad \Sigma_\theta = I_p \tag{3.27}$$

with

$$P(\theta) = \frac{\partial \psi}{\partial \theta'} = 2\theta' \, .$$

One difficulty introduced by such a restriction is a deficiency of the rank of the weighting matrix when shifting from Σ_θ with full rank p to Σ with rank 1. More importantly, although the restriction is differentiable w.r.t. θ, the delta method completely 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 Σ_θ and Σ are given by:

$$\hat{\Sigma}_\theta = \frac{1}{n}\hat{u}\hat{u}', \quad \text{with } \hat{u} = Y - \hat{\theta}_n \iota$$

and

$$\hat{\Sigma} = P(\hat{\theta}_n)\hat{\Sigma}_\theta P(\hat{\theta}_n)' \, . \tag{3.28}$$

We will apply the regularization techniques introduced in sections 7 and 10 to $\hat{\Sigma}$ to get $\hat{\Sigma}^R(c)$. Hence, the appropriate statistic to test this null hypothesis should be:

$$W_\psi^R(c) = n\psi(\hat{\theta}_n)'\hat{\Sigma}^R(c)n\psi(\hat{\theta}_n) = n^2\psi(\hat{\theta}_n)'\hat{\Sigma}^R(c)\psi(\hat{\theta}_n) \tag{3.29}$$

instead of the standard Wald statistic

$$W = \sqrt{n}\psi(\hat{\theta}_n)'\hat{\Sigma}^{-1}\sqrt{n}\psi(\hat{\theta}_n) = n\psi(\hat{\theta}_n)'\hat{\Sigma}^{-1}\psi(\hat{\theta}_n) \, ; \tag{3.30}$$

neither the Moore-Penrose modified Wald statistic proposed by Lutkepohl and Burda (1997) is suitable:

$$W^+(c_n) = \sqrt{n}\psi(\hat{\theta}_n)'\hat{\Sigma}^+(c_n)\sqrt{n}\psi(\hat{\theta}_n) = n\psi(\hat{\theta}_n)'\hat{\Sigma}^+(c_n)\psi(\hat{\theta}_n) \, ; \tag{3.31}$$

both of them do not use the *right* convergence speed neither the *right* distribution since $\psi(\hat{\theta}_n)$ is not Gaussian anymore.

4. Regularized inverses

The methodology introduced in this section applies to any symmetric matrices and more specifically to covariance matrices. We first introduce some notations. Let $\bar{\lambda} = (\lambda_1, \ldots, \lambda_q)'$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q$ are the eigenvalues of a $q \times q$ (covariance) matrix Σ, and V an orthogonal matrix such that $\Sigma = V\Lambda V'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_q)$. Specifically, V consists of eigenvectors of the matrix Σ ordered so that $\Sigma V = V\Lambda$. Let $m(\lambda)$ be the multiplicity of the eigenvalue λ. Although the matrix Λ is uniquely defined, the matrix V consisted of the eigenvectors is not uniquely defined when there is an eigenvalue with multiplicity $m(\lambda) > 1$. The eigenvectors which correspond to eigenvalues with $m(\lambda) > 1$ are uniquely defined only up to post-multiplication by an $m(\lambda) \times m(\lambda)$ orthogonal matrix. Moreover, let Σ_n be a consistent estimator of Σ with eigenvalues $\lambda_1(\Sigma_n) \geq \lambda_2(\Sigma_n) \geq \ldots \geq \lambda_q(\Sigma_n)$ and V_n an orthogonal matrix such that $\Sigma_n = V_n\Lambda_n V_n'$ where $\Lambda_n = \text{diag}\big[\lambda_1(\Sigma_n), \ldots, \lambda_q(\Sigma_n)\big]$. For $c > 0$, we denote $q(\Sigma, c)$ the number of eigenvalues λ such that $\lambda > c$ and $q(\Sigma_n, c)$ the number of eigenvalues $\lambda(\Sigma_n)$ such that $\lambda(\Sigma_n) > c$.

If $\text{rank}(\Sigma_n) = \text{rank}(\Sigma) = q$ with probability 1, *i.e.* both matrices are almost surely (a.s.) nonsingular, so the inverses $\Sigma^{-1} = V\Lambda^{-1}V'$ and $\Sigma_n^{-1} = V_n\Lambda_n^{-1}V_n'$ are a.s. well defined. However, if $\text{rank}(\Sigma) < q$ and $\text{rank}(\Sigma_n) \leq q$, we need to make adjustments. For this, we define a *regularized* inverse of a (covariance) matrix Σ as below.

Definition 4.1 DEFINITION OF THE REGULARIZED INVERSE . *Σ is a $q \times q$ real symmetric semi-definite positive matrix with rank$(\Sigma) \leq q$. Its regularized inverse is:*

$$\Sigma^R(c) = V\Lambda^{\dagger}(c)V' \tag{4.1}$$

where

$$\Lambda^{\dagger}(c) = \Lambda^{\dagger}[\bar{\lambda}; c] = \begin{pmatrix} g(\lambda_1; c) & & 0 \\ & \ddots & \\ 0 & & g(\lambda_q; c) \end{pmatrix} \tag{4.2}$$

$g(\lambda; c) \geq 0$, with $c \geq 0$, and $g(\lambda; c)$ bounded.

The scalar function $g(\lambda; c)$ modifies the inverse of the eigenvalues in order to make the inverse well-behaved in a neighborhood of the true eigenvalues. We shall call it the *(variance) regularization function* (VRF). The VRF perturbs the small eigenvalues in order to stabilize their inverse, preventing them from exploding.

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 size and possibly on the sample itself, *i.e.* $c = c[n, Y_n]$:

$$\Lambda^{\dagger}(c) = \begin{pmatrix} \Lambda_1^{\dagger}[\bar{\lambda}; c] & 0 & 0 \\ 0 & \Lambda_2^{\dagger}[\bar{\lambda}; c] & 0 \\ 0 & 0 & \Lambda_3^{\dagger}[\bar{\lambda}; c] \end{pmatrix} . \tag{4.3}$$

Let $q_i = \dim \Lambda_i^{\dagger}[\bar{\lambda}; c]$, for $i = 1, 2, 3$, with $q_1 = q(\Sigma, c)$, $q_2 = m(c)$ 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:

$$\Lambda_1^{\dagger}[\bar{\lambda}; c] = \text{diag}\big[g(\lambda_1; c), \ldots, g(\lambda_{q_1}; c)\big] \;\text{ for }\; \lambda > c \;\;, \tag{4.4}$$

$$\Lambda_2^{\dagger}[\bar{\lambda}; c] = g(c; c)I_{q_2} \;\text{ for }\; \lambda = c \;\;, \tag{4.5}$$

$$\Lambda_3^{\dagger}[\bar{\lambda}; c] = \text{diag}\big[g(\lambda_{q_1+q_2+1}; c), \ldots, g(\lambda_q; c)\big] \;\text{ for }\; \lambda < c \;. \tag{4.6}$$

More specifically, the large eigenvalues that fall above the threshold c remain unchanged whereas those equal to or smaller than the threshold are inflated to make their inverse well-behaved. Thus, the first component is "regular" and remains unmodified, while the others may not be "regular". In particular, the third component requires a regularization. Indeed, because of the invertibility difficulties raised from small values of λ, we shall replace the latter with eigenvalues bounded away from zero. Instead of using a spectral cut-off Moore Penrose inverse, we propose alternatively a *full-rank* regularized matrix. This regularization contains the spectral cut-off type regularization as a special case. Indeed, the spectral cut-off Moore Penrose inverse sets to zero all the 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 with 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 $\lambda < c$. $\Lambda_1^{\dagger}[\lambda(\Sigma_n); c]$,

$\Lambda_2^{\dagger}[\lambda(\Sigma_n);c]$,$\Lambda_3^{\dagger}[\lambda(\Sigma_n);c]$, V_{1n}, V_{2n} and V_{3n} denote the corresponding quantities based on the sample analog Σ_n, with $\dim \Lambda_1[\lambda(\Sigma_n);c] = \hat{q}_1 = \text{card}\{i \in I : \lambda_i(\Sigma_n) > c\}$, $\dim \Lambda_2[\lambda(\Sigma_n);c] = \hat{q}_2 = \text{card}\{i \in I : \lambda_i(\Sigma_n) = c\}$, $\dim \Lambda_3[\lambda(\Sigma_n);c] = \hat{q}_3 = \text{card}\{i \in I : \lambda_i(\Sigma_n) < c\}$, respectively.

Using (4.3), the *regularized* inverse can be decomposed as follows:

$$\Sigma^R(c) \quad = \quad V\Lambda^{\dagger}(c)V' = [V_1 \ V_2 \ V_3] \begin{pmatrix} \Lambda_1^{\dagger}[\bar{\lambda};c] & 0 & 0 \\ 0 & \Lambda_2^{\dagger}[\bar{\lambda};c] & 0 \\ 0 & 0 & \Lambda_3^{\dagger}[\bar{\lambda};c] \end{pmatrix} \begin{bmatrix} V_1' \\ V_2' \\ V_3' \end{bmatrix} = \sum_{i=1}^{3} \Sigma_i^R(c) \tag{4.7}$$

where

$$\Sigma_i^R(c) = V_i \Lambda_i^{\dagger}(c) V_i' \quad i = 1,2,3 \tag{4.8}$$

$\Lambda_i^{\dagger}(c) = \Lambda_i^{\dagger}[\bar{\lambda};c]$ for the sake of notational simplicity. Note that the original matrix Σ can be decomposed similarly as:

$$\Sigma = V\Lambda V' = \sum_{i=1}^{3} \Sigma_i = \sum_{i=1}^{3} V_i \Lambda_i V_i' \ . \tag{4.9}$$

with $\Lambda_1(c) = \{\lambda : \lambda > c\}$, $\Lambda_2(c) = \{\lambda : \lambda = c\}$ and $\Lambda_3(c) = \{\lambda : \lambda < c\}$. In the absence of problematic zero eigenvalues, the usual inverse can be computed as:

$$\Sigma^{-1} = V\Lambda^{-1}V' = \sum_{i=1}^{3} \Sigma_i^{-1} = \sum_{i=1}^{3} V_i \Lambda_i^{-1} V_i' \ . \tag{4.10}$$

Let I_q and I_{q_i} denote conformable identity matrices. Let us establish some useful properties for the regularized inverses.

Property 1 PROPERTY OF THE REGULARIZED INVERSES. *Let $\Sigma = V\Lambda V'$ be a positive semi definite 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 $\Sigma^R(c)$ of Σ, defined in* **4.1**, *satisfies the following relations.*

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

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

iii) $\Sigma\Sigma^R(c)\Sigma \ \leq \ \Sigma$;

iv) If $g(\lambda;c) > 0$, then $\left(\Sigma^R(c)\right)^{-1} \ \geq \ \Sigma$;

v) If $\left(\lambda > 0 \ \Rightarrow \ g(\lambda;c) > 0\right)$, then $\left(\text{rank}\left(\Sigma^R(c)\right) \geq \text{rank}(\Sigma)\right)$.

It is important to notice that any transformation of the original matrix Σ that diminishes the inverse $\Sigma^R(c)$ satisfies relation $iv)$. Note that the generalized inverses usually denoted by Σ^- 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 $\Sigma\Sigma^-\Sigma = \Sigma$. Result $v)$ is well known for g-inverses and is related to 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 tests which embed 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 with decomposition (4.9) and (4.10) boiling down to single block when $c = 0$. *Case 2* corresponds to a sample covariance matrix which converges to a singular limiting matrix but satisfying the rank condition **2.5**. In such a case, the limiting distribution is modified only through an adjustment of the degree of freedom; this is the case covered by Andrews (1987) and Lutkepohl and Burda (1997). 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 is different from the original population matrix. Regularizing then yields a valid test but at the cost of a *fully modified* asymptotic distribution. This is the route investigated here. We consider situations where the finite sample rank generally exceeds the asymptotic one. This type of singularities can be encountered when the derivative matrix of the restrictions has a lower rank only at the true value of the parameter, or in the presence of superconsistent estimators or estimators that do not converge at the expected parametric speed. In this respect, Antoine and Renault (2009), Antoine and Renault (2010*a*) pointed out that although all parameters are identified, but some rates of convergence are as slow as $n^{1/4}$, the standard GMM estimator asymptotics are modified. The regularized Wald statistic can also handle cases where the finite sample matrix is singular possibly due to redundant restrictions that are difficult to detect analytically.

Based on decomposition (4.9), the original Wald statistic $W_n(\psi_0)$ defined in equation (2.2) enjoys the following decomposition

$$W_n(\psi_0) = W_{1n} + W_{2n} + W_{3n} \ , \tag{5.1}$$

where $W_{in} = a_n^2 \ [\hat{\psi}_n - \psi_0]' \Sigma_{in}^{-1} [\hat{\psi}_n - \psi_0]$, with $\Sigma_{in}^{-1} = V_{in} \Lambda_{in}^{-1} V_{in}'$ for $i = 1, 2, 3$, and $\Lambda_{in}^{-1} = \Lambda_i^{-1}[\lambda(\Sigma_n); c]$. For $i = 2, 3$, $W_{in} = 0$, eventually.

The specific irregular setup here consists in allowing singular covariance matrices that violates Assumption **2.5** of Andrews (1987). As a consequence, the Wald test statistic has to be modified or *regularized* to account for such irregularities. Let us introduce the *regularized* Wald statistic in the next definition.

Definition 5.1 DEFINITION OF THE REGULARIZED WALD STATISTIC. *The regularized Wald statistic is*

$$\begin{aligned} W_n^R(c) &= X_n' \Sigma_n^R(c) X_n \\ &= a_n \ [\hat{\psi}_n - \psi_0]' \Sigma_n^R(c) a_n [\hat{\psi}_n - \psi_0] \ . \end{aligned} \tag{5.2}$$

Built on the *regularized* inverse of Section 4 and its decomposition (4.7)-(4.8), the *regularized* Wald statistic can be decomposed as follows.

$$\begin{aligned} W_n^R(c) &= X_n' \Sigma_n^R(c) X_n = a_n^2 \ [\hat{\psi}_n - \psi_0]' \Sigma_n^R(c) [\hat{\psi}_n - \psi_0] \\ &= a_n^2 \ [\hat{\psi}_n - \psi_0]' \sum_{i=1}^{3} \Sigma_{in}^R(c) [\hat{\psi}_n - \psi_0] \\ &= W_{1n}^R(c) + W_{2n}^R(c) + W_{3n}^R(c) \ , \end{aligned} \tag{5.3}$$

where

$$W_{in}^R(c) = a_n^2 \ [\hat{\psi}_n - \psi_0]' \Sigma_{in}^R(c) [\hat{\psi}_n - \psi_0]$$

with $\Sigma_{in}^R(c) = V_{in} \Lambda_{in}^{\dagger}(c) V_{in}'$ for $i = 1, 2, 3$.

By partitioning the inverse of the eigenvalue matrix $\Lambda^{\dagger}$ 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 of the statistic into three components: a first component involving the "large" eigenvalues remains unchanged; a second component gathers the eigenvalues exactly equal to the threshold c, while a third one incorporates the small eigenvalues. As we shall see in Section **8.1**, this decomposition helps one to better understand the structure of the distribution of the *regularized* test statistic. By contrast, Lutkepohl and Burda (1997) only keep the eigenvalues greater than the threshold c, which cancels out the last two components, *i.e.* $W_{2n}^R(c) = 0$ and $W_{3n}^R(c) = 0$. Thus discarding the small eigenvalues may result in a loss of information. However, as Lutkepohl and Burda (1997) use a χ^2 distribution with fewer degrees of freedom, a deeper investigation must be conducted for power assessment. More importantly, in finite samples it will be difficult to disentangle between the estimates which really correspond to $\lambda = c$ from those close to c, but distinct from c. This makes the estimation procedure trickier and the asymptotic distribution more complicated. Note that $W_{1n} = W_{1n}^R(c)$ for this is the regular component common to both statistics, the usual Wald and the regularized Wald statistics. Moreover, when there is no eigenvalues exactly equal to c, $m(c) = 0$, and the second component collapses to zero.

6. Results on eigenprojections

6.1. Discontinuities of eigenvectors: an illustration

We discuss now some non-uniqueness and discontinuity issues regarding the eigenvectors of a given matrix. While it is well-known in spectral theory that eigenvectors corresponding to multiple eigenvalues are not uniquely defined (but only up to the post multiplication by an $m(\lambda) \times m(\lambda)$ orthogonal matrix with $m(\lambda)$ indicating the multiplicity of the eigenvalue), econometricians are not cautious about such considerations that could entail convergence problems. Second, whereas the eigenvalues are generally known to be continuous functions of the elements of the matrix, this statement does not necessarily hold for the eigenvectors. The main pitfall consists of drawing convergence results for the estimates of the eigenvectors based on the consistency of the sample matrix which critically hinges on the continuity assumption of eigenvectors (w.r.t. the elements of the matrix). Even in the deterministic case, the eigenvectors are not necessarily continuous functions of the elements of the matrix. To see their discontinuity, we consider a simple counter-example[1].

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

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

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], 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]. Clearly, the

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

eigenvalues $(1+x)$ and $(1-x)$ are continuous at $x = 0$ whereas the eigenvectors are not the same whether $x \to 0^+$ or $x \to 0^-$.

Being unaware of this caveat may lead to *wrong* distributional results through mistakenly applying the continuous mapping theorem to objects that are *not* continuous. Nevertheless, there exists functions of the eigenvectors 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* $V(\lambda)$ - is continuous in the elements of the matrix. This follows from the fact that $V(\lambda)$ is invariant to the choice of a basis. 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 statistics, it will be useful to review and adapt some results on spectral theory used by Tyler (1981). Let $\mathcal{S}(\Sigma)$ denote the spectral set of Σ, *i.e.* the set of all eigenvalues of Σ. The *eigenspace* of Σ associated with λ is defined as all the linear combinations from a basis of eigenvectors x_i, $i = 1, \ldots, m(\lambda)$, *i.e.*

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

Clearly, $\dim V(\lambda) = m(\lambda)$. Since Σ is a $q \times q$ matrix symmetric in the metric of a real positive definite symmetric matrix T (*i.e.* $\mathrm{T}\Sigma$ is symmetric), we have:

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

The *eigenprojection* of Σ 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 x_i in $V(\lambda)$ such that $\mathrm{x}_i' \mathrm{T} \mathrm{x}_j = \delta_{ij}$, where δ_{ij} denotes the Kronecker' s delta, $P(\lambda)$ has the representation

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

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

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

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

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

i) $\lambda_k(\Sigma_n) \to \lambda_k(\Sigma)$, *and*

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

This lemma tells us that the eigenvalues are continuous functions in the elements of the matrix. The same continuity property holds for the projection operators [or equivalently for the projection matrices for there exists a one-to-one mapping relating the operator to the matrix w.r.t. the bases] associated with the eigenvalues and transmitted to their sum. No matter what the multiplicity of the eigenvalues involved in the total eigenprojection $P_{k,t}(\Sigma)$, this continuity property holds provided that we can find one eigenvalue before and one after that are distinct.

It will be useful to extend Lemma **6.2** to random symmetric matrices. To do so, we first consider a.s. convergence (in symbol $\overset{a.s.}{\to}$) and then convergence in probability (in symbol $\overset{p}{\to}$). To the best of our knowledge, these results are not explicitly stated elsewhere. In the following we will tacitly assume that a probability space $(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P)$ is given and that all random variables are defined on this space.

Lemma 6.3 CONTINUITY OF EIGENVALUES AND EIGENPROJECTIONS: ALMOST SURE CONVERGENCE. *Let Σ_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 $\lambda_1(\Sigma_n) \geq \lambda_2(\Sigma_n) \geq \ldots \geq \lambda_q(\Sigma_n)$. Let $P_{k,t}(\Sigma_n)$ represent the total eigenprojection for Σ_n associated with $\lambda_k(\Sigma_n) \ldots \lambda_t(\Sigma_n)$ for $t \geq k$. If $\Sigma_n \overset{a.s.}{\to} \Sigma$ as $n \to \infty$, then:*

i) $\lambda_k(\Sigma_n) \overset{a.s.}{\to} \lambda_k(\Sigma)$, *and*

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

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 6.4 CONTINUITY OF EIGENVALUES AND EIGENPROJECTIONS: CONVERGENCE IN PROBABILITY.
Let Σ_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 $\lambda_1(\Sigma_n) \geq \lambda_2(\Sigma_n) \geq \ldots \geq \lambda_q(\Sigma_n)$. Let $P_{k,t}(\Sigma_n)$ represent the total eigenprojection for Σ_n associated with $\lambda_k(\Sigma_n), \ldots, \lambda_t(\Sigma_n)$ for $t \geq k$. If $\Sigma_n \overset{p}{\to} \Sigma$ as $n \to \infty$, then:

i) $\lambda_k(\Sigma_n) \overset{p}{\to} \lambda_k(\Sigma)$, *and*

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

6.3. Asymptotic distribution of eigenvalues

In this subsection, we summarize general results on 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 Σ_n satisfying the condition $b_n(\Sigma_n - \Sigma) \overset{\mathcal{L}}{\to} Q$ where Σ is a nonstochastic matrix, then the sample eigenvalues will have the same convergence rate, with $b_n[\Psi(\Sigma_n) - \Psi(\Sigma)] \overset{\mathcal{L}}{\to} \left[H_D\left(\frac{1}{2}[Q'_{11} + Q_{11}]\right), \quad \Psi(Q_{22})\right]'$. $H_D(.)$ 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.

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, $i = 1, \ldots, k$. 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

$$H(C) = \begin{pmatrix} \rho(C_{11}) \\ \rho(C_{22}) \\ \vdots \\ \rho(C_{kk}) \end{pmatrix} \tag{6.8}$$

$H(C)$ 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}, $i = 1, \dots, k$. Let Γ be an orthogonal matrix such that

$$\Gamma A \Gamma' = D, \tag{6.9}$$

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

Lemma 6.5 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*

$$W_n = b_n(S_n - A) \xrightarrow{\mathcal{L}} W \; . \tag{6.10}$$

Then

$$b_n\big(\rho(S_n) - \rho(A)\big) \xrightarrow{\mathcal{L}} H\big(\Gamma W \Gamma'\big) \; . \tag{6.11}$$

For any $p \times q$ real matrix Σ, the $\Psi(.)$ function is a vector-valued function that stacks the eigenvalues of the corresponding object as defined below:

$$\Psi(\Sigma) = f(\rho(\Sigma'\Sigma)) = \begin{pmatrix} \sqrt{\xi_1} \\ \vdots \\ \sqrt{\xi_q} \end{pmatrix} \quad \text{with } f(x) = \begin{pmatrix} \sqrt{x_1} \\ \vdots \\ \sqrt{x_q} \end{pmatrix} \tag{6.12}$$

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

Let

$$T = \big(df(\xi)\big) = \frac{1}{2}\text{diag}(\xi_1^{-1/2}, \dots, \xi_q^{-1/2}) \; . \tag{6.13}$$

In the first part of the theorem below, we gather the special cases where the matrix Σ 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 Σ in the form

$$\Sigma = \Gamma_1' \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Gamma_2' \tag{6.14}$$

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 Σ. Partition the matrix Σ_n as

$$\Sigma_n = \begin{pmatrix} \Sigma_{n11} & \Sigma_{n12} \\ \Sigma_{n21} & \Sigma_{n22} \end{pmatrix} \tag{6.15}$$

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

Theorem 6.6 DISTRIBUTION OF THE EIGENVALUES OF RECTANGULAR MATRICES IN THE GENERAL CASE. *Let $\Psi(\cdot)$ be defined as in (6.12), and suppose Assumption* **2.3** *holds.*

i) If $\Sigma = 0$, then

$$b_n\big(\Psi(\Sigma_n) - \Psi(\Sigma)\big) \xrightarrow{\mathcal{L}} \Psi(Q) \,. \tag{6.16}$$

ii) If Σ has full rank q, then

$$b_n\big(\Psi(\Sigma_n) - \Psi(\Sigma)\big) \xrightarrow{\mathcal{L}} TH\big(\Gamma\big[\Sigma' Q + Q'\Sigma\big]\Gamma'\big) \tag{6.17}$$

where H, Γ and T are defined in (6.8),(6.9) and (6.13).

iii) If $rank(\Sigma) = r$, $1 \le r < q$, then

$$b_n\big[\Psi(\Sigma_n) - \Psi(\Sigma)\big] \xrightarrow{\mathcal{L}} \begin{bmatrix} H_D\big(\frac{1}{2}[Q_{11}' + Q_{11}]\big) \\ \Psi(Q_{22}) \end{bmatrix} \tag{6.18}$$

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 $D = \mathrm{diag}(\xi_1^{1/2}, \ldots, \xi_r^{1/2})$ consisted of the strictly positive singular values of Σ defines a function H_D on $r \times r$ symmetric matrices as H is defined in (6.8) on $q \times q$ symmetric matrices.

For our purposes, we do not need the knowledge of the whole distribution but only the convergence rate b_n of the sample eigenvalues for the convergence property of regularized inverse when c varies with the sample size. See Eaton and Tyler (1994, Propositions 3.1 and 3.4 and Theorem 4.2) for a proof.

7. Asymptotic properties of the regularized inverse

In this section, we derive asymptotic results for the *regularized* inverse that hold for a relatively general variance regularization function (VRF) family.

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 VRF, such as for a real scalar, $c \ge 0$:*

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

$g(\lambda; c)$ is continuous almost everywhere w.r.t. λ, except possibly at $\lambda = c$, (w.r.t. the Lebesgue measure). g is a function that takes bounded values everywhere, and g is non-increasing in λ.

Note importantly that we allow a discontinuity at $\lambda = c$ to precisely embed a spectral-cutoff type regularization such as a modified Moore-Penrose inverse that is clearly *not* continuous around $\lambda = c$, see (7.2).

Some possible choices for the VRF could be:

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

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:

i) $\gamma = \infty$, $c \geq 0$, 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

$$\Lambda^{+}(c) = \text{diag}[1/\lambda_1 I(\lambda_1 > c), \ldots, 1/\lambda_{q_1} I(\lambda_{q_1} > c), 0, \ldots, 0\,]$$

corresponds to a spectral cut-off regularization scheme [see Carrasco (2007), Carrasco, Florens and Renault (2007) and the references therein]; $I(s)$ is equal to 1 if the relation s is satisfied. In particular, Λ^{+}_{c} is a *modified version* of the Moore-Penrose inverse of

$$\Lambda = \text{diag}[\lambda_1 I(\lambda_1 > 0), ..., \lambda_{q_1} I(\lambda_{q_1} > 0), \lambda_{q_1+1} I(\lambda_{q_1+1} > 0) \ldots, \lambda_q I(\lambda_q > 0)]$$

used by Lutkepohl and Burda (1997, henceforth LB). We also consider the case where some eigenvalues may be smaller than the threshold c, with $c \neq 0$.

ii) $\gamma = 0$ and $\epsilon = c$, with $c \neq 0$, hence

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > c \\ \frac{1}{c} & \text{if } \lambda \leq c\,. \end{cases} \tag{7.3}$$

iii) $\gamma > 0$ with $\gamma = \frac{\alpha}{\lambda \times (c-\lambda)}$, $\alpha > 0$, and $\epsilon = \lambda$, with $c \neq 0$, hence

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > c \\ \frac{\lambda}{\lambda^2+\alpha} & \text{if } \lambda \leq c\,, \end{cases} \tag{7.4}$$

which corresponds to a variation around the Tikhonov regularization (related to the ridge regression) 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:

$$\Sigma^{R}(c) = V\Lambda^{\dagger}(c)V' = \sum_{\lambda \in \mathcal{S}(\Sigma)} g(\lambda; c)P(\lambda)\,, \quad \Sigma_n^{R}(c) = V_n\Lambda_n^{\dagger}(c)V_n' = \sum_{\lambda(\Sigma_n) \in \mathcal{S}(\Sigma_n)} g\big[\lambda(\Sigma_n); c\big] P\big[\lambda(\Sigma_n)\big]\,. \tag{7.5}$$

Thus, the dependence on c of the regularized inverses comes from the VRF $g(\lambda; c)$. Besides, the threshold c may be size-dependent, *i.e.*, $g(\lambda, c_n)$. This case is a special case of c fixed and will be studied in Section 10.

7.2. 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 $(\mathcal{S}_q, <\cdot,\cdot>)$, where $\mathcal{S}_q$ is the vector space of $q \times q$ symmetric matrices. $\mathcal{S}_q$ is equipped with the inner product $<\Sigma_1, \Sigma_2> = tr[\Sigma_1'\Sigma_2]$, where tr denotes the trace. Let $\|\cdot\|$ denote the Frobenius norm induced by this inner product, *i.e.* $\|\Sigma\|_F^2 = tr[\Sigma'\Sigma]$. Let A^R denote the regularized inverse of a $q \times q$ real symmetric matrix A. In the subsequent propositions, let $I = \{1, 2, \ldots, q\}$, denote the set of indices such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q$, and $J = \{1, 2, \ldots, k\}$ the subset of I corresponding to the indices associated with the distinct eigenvalues of Σ, *i.e.* $d_1 > d_2 > \ldots > d_j > \ldots > d_k$, 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 $\mathcal{P}(I)$ such that:

$$\mathcal{P}(I) = \{I_j \subset I, j \in J : I_j \bigcap_{j \neq l} I_l = \emptyset, \bigcup_{j=1}^{k} I_j = I\} \ , \quad I = \{1, \ldots, q\}, \tag{7.6}$$

with

$$I_j = \{i \in I : \ \lambda_i = d_j\} \ , \quad \text{card } I_j = m(d_j) \tag{7.7}$$

and

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

We adopt the convention that $I(c) = \emptyset$, if there is no eigenvalues equal to c. The vector space $\mathbb{R}^q$ can be decomposed as $\mathbb{R}^q = \mathcal{V}(d_1) \oplus \cdots \oplus \mathcal{V}(d_j) \oplus \cdots \oplus \mathcal{V}(d_k)$. Each $u \in \mathbb{R}^q$ can be expressed in the form $u = u_1 + \cdots + u_j + \cdots + u_k$, with $u_j \in \mathcal{V}(d_j)$, $j \in J$ in a unique way. The operator $P_j = P(d_j)$ is such that: $P_j u = u_j$ is the eigenprojection operator on the eigenspace $\mathcal{V}(d_j)$ along $N_j = \mathcal{V}(d_1) \oplus \cdots \oplus \mathcal{V}(d_{j-1}) \oplus \mathcal{V}(d_{j+1}) \oplus \cdots \oplus \mathcal{V}(d_k)$. Thus,

$$P_j(\Sigma) = P(d_j)(\Sigma) \tag{7.9}$$

projects Σ onto the eigenspace $\mathcal{V}(d_j)$ along N_j. Furthermore, $\sum_{j=1}^{k} P_j = 1$, $\quad P_k P_j = \delta_{jk} P_j$. There is a one-to-one mapping from J to $\mathcal{P}(I)$ such that:

$$\forall j \in J : \ j \ \longmapsto \ I_j \tag{7.10}$$

where the total eigenprojection operator $P_{I_j}(\bullet)$ applied to Σ_n, with $\Sigma_n \xrightarrow{p} \Sigma$, yields by Lemma **6.4** ii)

$$P_{I_j}(\Sigma_n) \xrightarrow{p} \ P_j(\Sigma) = P(d_j)(\Sigma) \tag{7.11}$$

and

$$\dim \, P_{I_j} = \dim \, P_j = m(d_j) = \dim \, \mathcal{V}(d_j) \ \ with \ \ 1 = \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} P_{I_j} \ . \tag{7.12}$$

Property 2 UNIQUE REPRESENTATION OF THE REGULARIZED INVERSE. *For a given $g(.,c)$ VRF in the $\mathcal{G}_c$ family, the regularized inverse $\Sigma^R(c) = V \Lambda^\dagger(c) V'$ of a symmetric matrix Σ and its sample analog $\Sigma_n^R(c) = V_n \Lambda_n^\dagger(c) V_n'$ admit an unique representation of the form:*

$$\Sigma^R(c) = \sum_{j=1}^{k} g(d_j; c) P_j(\Sigma) \tag{7.13}$$

and

$$\Sigma_n^R(c) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i \in I_j} g(\hat{\lambda}_i; c) \tag{7.14}$$

where the d_j's denote the distinct eigenvalues of Σ with multiplicity $m(d_j)$, $\hat{\lambda}_i = \lambda_i(\Sigma_n)$; $P_{I_j}(\Sigma_n)$ and $P_j(\Sigma)$ are defined at equations (7.9)-(7.12) with I_j defined at equation (7.7). If $\Sigma = 0$, $P(0)(\Sigma) = I_q$, and $\Sigma^R(c) = g(0; c) P(0)(\Sigma) = g(0; c) I_q$.

The uniqueness of the representation of the regularized inverse immediately follows from the uniqueness of the decomposition involving only distinct eigenvalues. Thus, there is a one-to-one relation between the regularized inverse and the VRF $g(.,c)$ in the $\mathcal{G}_c$ family. An interesting case producing a nonstandard asymptotic distribution corresponds to using a fixed threshold. In this case, the asymptotic distribution of the regularized test statistic involves a nonstandard component that can be bounded above as shown in Corollary **8.3**.

Assumption 7.2 REGULARITY CONDITIONS FOR THE CONVERGENCE OF THE REGULARIZED INVERSE. *The VRF $g \in \mathcal{G}_c$, and for $i = 1, \ldots, q$, $\lambda_i = \lambda_i(\Sigma)$ are the eigenvalues of a $q \times q$ semi definite matrix Σ. 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 is superconsistent at c, i.e. $P\big[\hat{\lambda}_i(c) = c\big] \underset{n\to\infty}{\to} 1$.

As soon as one of the three above conditions hold, both convergence results of the regularized inverse (Propositions **7.3** and **7.4**) will hold, otherwise it may break down. Let us now state the a.s. convergence for the regularized inverse when c is fixed.

Proposition 7.3 ALMOST SURE CONVERGENCE OF THE REGULARIZED INVERSES. *Let $g \in \mathcal{G}_c$. Suppose Σ and Σ_n are $q \times q$ symmetric matrices with $rank(\Sigma) = r \leq q$. Let the regularized inverses be defined at equations* (7.13) *and* (7.14). *Let Assumption* **7.2** *hold. If $\Sigma_n \overset{a.s.}{\to} \Sigma$, then*

$$\Sigma_n^R(c) \overset{a.s.}{\to} \Sigma^R(c) \ . \tag{7.15}$$

Proposition 7.4 CONVERGENCE IN PROBABILITY OF THE REGULARIZED INVERSES. *Suppose Σ and Σ_n are $q \times q$ symmetric matrices such that $rank(\Sigma) = r \leq q$. Assumption* **2.3** *holds with $p = q$, and Assumption* **7.2** *holds. Let the regularized inverses satisfy Property 2, and decomposition* (4.7)-(4.8). *Then*

$$\Sigma_n^R(c) = \Sigma_{11,n}^R(c) + \Sigma_{22,n}^R(c) + \Sigma_{33,n}^R(c) \tag{7.16}$$

where

$$\Sigma_{11,n}^R(c) = \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i, c) \ \overset{p}{\to} \ \sum_{j=1}^{k_1} g(d_j; c) P_j(\Sigma) \equiv \ \Sigma_{11}^R(c) \tag{7.17}$$

$$\Sigma_{22,n}^R(c) = P_{I(c)}(\Sigma_n) \frac{1}{m(c)} \sum_{i\in I(c)} g(\hat{\lambda}_i, c) \ \overset{p}{\to} \ g(c; c) 1_{\{d_j = c\}} P_{j(c)}(\Sigma) \equiv \ \Sigma_{22}^R(c) \tag{7.18}$$

$$\Sigma_{33,n}^R(c) = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i\in I_j} g(\hat{\lambda}_i, c) \ \overset{p}{\to} \ \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j; c) P_j(\Sigma) \equiv \ \Sigma_{33}^R(c) \ . \tag{7.19}$$

$$\Sigma_n^R(c) \overset{p}{\to} \Sigma^R(c) \ . \tag{7.20}$$

$k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$, *$k$ is the number of distinct eigenvalues of Σ, and $P_{j(c)}(\Sigma) = P(d_j)(\Sigma)$ for $d_j = c$, where $P_j(\Sigma) = P(d_j)(\Sigma)$ is defined at equation* (7.9). *I_j and $I(c)$ are defined in* (7.7) *and* (7.8).

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 (7.18) 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 $\Sigma_{22,n}^R(c) = \Sigma_{22}^R(c) = 0$; if there exists a superconsistent estimator of the eigenvalue at c, (7.18) holds. Otherwise, $\Sigma_n^R(c)$ may not converge to $\Sigma^R(c)$. In other words, the conditions stated in Assumption **7.2** are necessary conditions for (7.15) and (7.20) to hold.

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

In this section, we characterize (Proposition **8.1**) the asymptotic distribution of the regularized Wald statistic for general distributions for X_n, before giving its specific expression for the Gaussian case (Corollary **8.2**). The decomposition of the regularized statistic into three independent components, one regular and two nonregular ones, provides an insight on the structure of the distribution which yields an upper bound for the test statistic in the general case (Corollary **8.3**); its expression in the special Gaussian case follows (Corollary **8.4**).

Proposition 8.1 CHARACTERIZATION OF THE REGULARIZED WALD STATISTIC WHEN THE THRESHOLD IS FIXED. *Suppose the assumptions of Proposition* **7.4** *hold together with Assumption* **2.1***, with* $\psi = \psi_0$*. Suppose the* $q \times q$ *limiting weighting matrix* Σ *satisfies Assumption* **2.2***, let* $k_1 = \sum_{j=1}^{k} 1_{\{d_j > c\}}$ *be the number of distinct eigenvalues of* Σ *larger than* c*, and* $W_n^R(c)$ *is defined in (*5.2*). Then*

$$W_n^R(c) \xrightarrow{\mathcal{L}} W^R(c) \tag{8.1}$$

where

$$\begin{aligned} W^R(c) &= X'\Sigma^R(c)X = \sum_{j=1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X \\ &= W_1^R(c) + W_2^R(c) + W_3^R(c)\ , \end{aligned} \tag{8.2}$$

$$W_1^R(c) = X'\Sigma_{11}^R(c)X = \sum_{j=1}^{k_1} g(d_j;c)X'B(d_j)B(d_j)'X\ , \tag{8.3}$$

$$W_2^R(c) = X'\Sigma_{22}^R(c)X = g(c;c)1_{\{d_j=c\}}X'B(c)B(c)'X\ , \tag{8.4}$$

$$W_3^R(c) = X'\Sigma_{33}^R(c)X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X\ . \tag{8.5}$$

Interestingly, when $\Sigma = 0$, the distribution of $W^R(c)$ is not degenerate: the regularized weighting matrix is given by $\Sigma^R(c) = g(0;c)I_q$, so the regularized Wald statistic simplifies to $g(0;c)X'X$ in the general case; in the Gaussian case, when $\Sigma = 0$, $d_j = 0$ with multiplicity q, and the limiting statistic is equal to zero (see equation (8.6), 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 Lutkepohl and Burda (1997) result obtained as a special case of the Corollary **8.2**. Besides, if there is no eigenvalues such that $\lambda = d_j = c$, $W_2^R(c) = 0$ 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 the assumptions of Proposition* **8.1** *hold, but replace Assumption* **2.1** *with* **2.4** *, with* $\psi = \psi_0$*, and* $B(d_j)'X = \sqrt{d_j}x_j$*, where* $x_j = N[0, I_{m(d_j)}]$*, for* $j = 1, \ldots, k$*.*

i) If $\Sigma = 0$*, then*

$$W_n^R(c) \xrightarrow{\mathcal{L}} W^R(c) = X'\Sigma^R(c)X = \sum_{j=1}^{k} g(d_j;c)d_jx_j'x_j = 0\ . \tag{8.6}$$

ii) If $\Sigma \neq 0$, then

$$W_n^R(c) \xrightarrow{\mathcal{L}} W^R(c) \tag{8.7}$$

where $W^R(c) = X'\Sigma^R(c)X = \sum_{j=1}^{k} g(d_j;c)d_j v_j = W_1^R(c) + W_2^R(c) + W_3^R(c)$ *with*

$$W_1^R(c) = X'\Sigma_{11}^R(c)X = \sum_{j=1}^{k_1} g(d_j;c)d_j v_j \ , \tag{8.8}$$

$$W_2^R(c) = X'\Sigma_{22}^R(c)X = g(c;c)1_{\{d_j=c\}} c v_{j(c)} \ , \tag{8.9}$$

$$W_3^R(c) = X'\Sigma_{33}^R(c)X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)d_j v_j \ , \tag{8.10}$$

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

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 Σ has rank $r < q$, the last component $W_3^R(c)$ that corresponds to the eigenvalues less than c, will contain a zero eigenvalue, *i.e.* $d_k = 0$, when $c \neq 0$. When $c = 0$, in this case $W_2^R(0) = W_3^R(0) = 0$, $W_1^R(0) = W^+(0)$, and we obtain the Lutkepohl and Burda (1997) result as a special case. Note that their result only holds for distinct eigenvalues.

Corollary 8.3 CHARACTERIZATION OF AN UPPER BOUND IN THE GENERAL CASE. *Suppose the assumptions of Proposition* **8.1** *hold. Let $Y_j = B(d_j)'X$. Let $g \in \mathcal{G}_c$, with a fixed threshold c such that*

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

then,

$$W^R(c) \leq \sum_{j=1}^{k} Y_j' Y_j \ .$$

The proof is straightforward. We obtain a characterization of an *upper bound* for general distributions for the *regularized* Wald statistic, when c is fixed. Such a *valid* bound will yield a *consistent* test under the alternative. However, using the standard chi square critical point corresponding to the Gaussian case will also produce consistent test under the alternative, yet at the cost of size distortions under the null.

Corollary 8.4 CHARACTERIZATION OF THE BOUND: THE GAUSSIAN CASE. *Suppose the assumptions of Corollary* **8.2** *hold. Let $g \in \mathcal{G}_c$, 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

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

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 multiplicity 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. A convenient choice for g could be $g(d_j; c)d_j = 1$, for all $j = 1, \ldots, k_1 + m(c)$, which yields $W_1^R(c) \sim \chi^2(q_1)$, and $W_2^R(c) \sim \chi^2(m(c))$. Although a fixed threshold leads to a nonstandard asymptotic distribution for the regularized statistic, the decomposition of the statistic into three components naturally provides an upper bound for the nonregular components. In consequence, the critical points given by the standard chi square distribution (if X is Gaussian) can be used to provide an *asymptotically valid* test. However, improved power over those conservative bounds can be achieved by simulations.

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

Proposition 8.5 CONSISTENCY PROPERTY OF THE TEST. *Suppose the assumptions of Proposition* **8.1** *hold, and* $W_n^R(c)$ *as defined in (*5.2*). Suppose there exist some eigenvalues of the limiting matrix* Σ *such that* $d_j \neq 0$ *under the alternative. Suppose further* $X_n = a_n(\hat{\psi}_n - \psi_1)$ *satisfies Assumption* **2.1***, with* $\psi = \psi_1$*. If* $\psi_1 - \psi_0 = \Delta \neq 0$*, and* $\Delta' \Sigma^R(c) \Delta > 0$*, then*

$$W_n^R(c) \underset{n\to\infty}{\to} \infty \,. \tag{8.11}$$

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

Proposition 8.6 LOCAL POWER CHARACTERIZATION. *Suppose the assumptions of Proposition* **8.1** *hold, and* $W_n^R(c)$ *as defined in (*5.2*). Suppose there exist some eigenvalues of the limiting matrix* Σ *such that* $d_j \neq 0$ *under the alternative, with* $\psi = \psi_1$*. If* $a_n(\psi_{1n} - \psi_0) \to \Delta \neq 0$*, and* $\Delta' \Sigma^R(c) \Delta > 0$*, then*

$$W_n^R(c) \overset{\mathcal{L}}{\underset{n\to\infty}{\to}} X' \Sigma^R(c) X + 2X' \Sigma^R(c) \Delta + \Delta' \Sigma^R(c) \Delta \,. \tag{8.12}$$

We can observe from this result that the limiting quantity involve three components: the first component, a quadratic form in X, still satisfies the null hypothesis; the second component is a linear form in X; the third one represents a noncentrality parameter. Only the last two component will contribute to power. Note that in the Lutkepohl and Burda (1997) case, their noncentrality parameter based on the modified Moore-Penrose inverse $\Delta' \Sigma_c^+ \Delta$ is expected to be smaller than the noncentrality parameter $\Delta' \Sigma^R(c) \Delta$, which may entail a loss of power despite a smaller critical point, due to a chi square distribution with reduced degree of freedom. 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 stated in the next corollary.

Corollary 8.7 LOCAL POWER CHARACTERIZATION: DELTA IN THE NULL EIGENSPACE. *Suppose the assumptions of Proposition* **8.6** *are satisfied. Suppose further that* $\Delta \in \mathcal{V}(0)$*, then*

$$W_n^R(c) \overset{\mathcal{L}}{\underset{n\to\infty}{\to}} X' \Sigma^R(c) X + 2g(0; c) X' \Delta + g(0; c) \Delta' \Delta \,. \tag{8.13}$$

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 $\mathcal{V}(0)$ 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 Lutkepohl and Burda (1997) will destroy power.

9. Regularized Wald statistic based on a super consistent eigenvalue estimator

We now introduce a new regularized Wald statistic $\tilde{W}_n^R(\hat{\lambda}(c); c)$ built on a *superconsistent* estimator of the eigenvalues at c, denoted $\hat{\lambda}(c)$. In so doing, a second layer of regularization is introduced through a *modified* Hodges-Lehnman estimator applied to the eigenvalues. More importantly, the superconsistency property of the eigenvalue estimator can accommodate a discontinuity of the VRF g at $\lambda = c$. The regularization is henceforth twofold: first, we modify the estimator of the eigenvalue as in (9.1), second the Hodges-Lehnman estimator of the eigenvalues, $\hat{\lambda}(c)$, is plugged into the weighting matrix and then regularized. Such a *superconsistent* estimator at c can be designed as follows.

The modified estimator $\hat{\lambda}(c) = (\hat{\lambda}_i(c))_{i=1,\ldots,q}$ of the eigenvalues of a $q \times q$ semi definite positive matrix Σ such that for each $i = 1, \ldots, q$, $\hat{\lambda}_i(c)$ satisfies:

$$\hat{\lambda}_i(c) = \begin{cases} \hat{\lambda}_i & \text{if } |\hat{\lambda}_i - c| > \nu \frac{e_n}{b_n} \\ c & \text{if } |\hat{\lambda}_i - c| \le \nu \frac{e_n}{b_n} \end{cases}, \tag{9.1}$$

where b_n is the speed of convergence of the sample eigenvalues as defined in Theorem **6.6**, 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 positive constant. $\hat{\lambda}_i(c)$ corresponds to a *modified* Hodges's estimator; see Hodges and Lehmann (1950), LeCam (1953), Lehmann and Casella (1998), Leeb and Potscher (2008). This estimator enjoys the *superconsistency* property 3 $ii)$. The sign function is defined as:

$$s[x] = \begin{cases} 1, & \text{if } \quad x > 0 \\ 0 & \text{if } \quad x = 0 \\ -1, & \text{if } \quad x < 0\,. \end{cases} \tag{9.2}$$

Property 3 SUPERCONSISTENT ESTIMATOR. *Under the assumptions given in Theorem* **6.6**, *the estimator* $\hat{\lambda}_i(c)$, *defined in* (9.1), *of* λ_i *of the* $q \times q$ *semi definite positive matrix* Σ *has the following properties for each* $i,\ 1 \le i \le q$,

i) $\hat{\lambda}_i(c) \xrightarrow{p} \lambda_i$

ii) $P\big[\hat{\lambda}_i(c) = c\big] \underset{n \to \infty}{\to} 1$, *if* $\lambda_i = c$

iii) $P\big\{s[\hat{\lambda}_i(c) - c] = s[\lambda_i - c]\big\} \underset{n \to \infty}{\to} 1$, *where* $s[.]$ *denotes the sign function defined in* (9.2).

Property 3$i)$ states the usual convergence in probability while 3$ii)$ states the *superconsistency* property of the modified estimator at c. Finally, Property 3$iii)$ states that the modified estimator falls in the appropriate class depending on whether $\lambda_i > c$, $\lambda_i = c$, or $\lambda_i < c$. As emphasized in Assumption **7.2**, $\lambda_i = c$ deserves a careful treatment, specifically if a mixture of a continuous distribution and of a Delta-Dirac distribution at c is considered. Although it is rather unlikely to encounter situations where $\lambda_i = c$ in finite samples, we wanted to provide a comprehensive thorough study of all possibilities. Thus, to circumvent the complications aroused by such a case, we rely on a *superconsistent* estimator. Recall that $I(c)$ is defined in (7.8), and its estimator is given by

$$\hat{I}(c) = \{i \in I :\ \hat{\lambda}_i(c) = c\}\ , \tag{9.3}$$

with $I(c) = \hat{I}(c) = \emptyset$, if there exist no eigenvalues $\lambda_i = c$. Then,

$$P\big[\hat{I}(c) = I(c)\big] = P\big[\hat{\lambda}_i(c) = c, \forall i \in I(c)\big] = P\Big[\bigcap_{i \in I(c)} \{\hat{\lambda}_i(c) = c\}\Big] \to\ 1\ , \tag{9.4}$$

since $P\big[\hat{\lambda}_i(c) = c\big] \to\ 1$ for all i. Note that only the modified estimator $\hat{\lambda}_i(c)$ satisfies (9.4) unlike estimators with continuous distributions for which $P\big[\hat{\lambda}_i = c\big] = 0$.

In the special case where $b_n = n^{1/2}$, we can take $e_n = n^{1/2-\delta}$ with $0 < \delta < 1/2$, so that:

$$\hat{\lambda}_i(c) = \begin{cases} \hat{\lambda}_i & \text{if } |\hat{\lambda}_i - c| > \frac{\nu}{n^{\delta}} \\ c & \text{if } |\hat{\lambda}_i - c| \leq \frac{\nu}{n^{\delta}} \end{cases} . \tag{9.5}$$

Thus, if $\lambda_i = c$, we have:

$$P\big[\hat{\lambda}_i(c) = c\big] = P\big[\sqrt{n}|\hat{\lambda}_i - c| \leq \nu n^{1/2-\delta}\big] \geq \ P\big[\sqrt{n}|\hat{\lambda}_i - \lambda_i| \leq \nu n^{1/2-\delta}\big] \underset{n\to\infty}{\to} 1 \tag{9.6}$$

since $\sqrt{n}(\hat{\lambda}_i - \lambda_i) = O_p(1)$. If $\lambda_i \neq c$, $\forall\, \epsilon > 0$, , then

$$P\big[|\hat{\lambda}_i(c) - \lambda_i| \leq \epsilon\big] = P\big[|\hat{\lambda}_i - c| > \nu \frac{n^{1/2-\delta}}{n^{1/2}}\big] = P\big[\sqrt{n}|\hat{\lambda}_i - c| > \nu n^{1/2-\delta}\big] \underset{n\to\infty}{\to} 1 \ . \tag{9.7}$$

Finally, if a consistent estimator of the number of eigenvalues greater than c is available, we will be able to simulate the distribution of the superconsistent estimator-based regularized statistic $\tilde{W}_n^R(\hat{\lambda}(c); c)$. Therefore, the simulated distribution will converge to the right distribution (that of $W^R(c)$), so that the level of the simulation-based test is controlled asymptotically. Let us define now an estimator of the possible multiplicity of c:

$$\hat{m}(c) = \sum_{i=1}^{q} 1_{\{i \in \hat{I}(c)\}} \ . \tag{9.8}$$

The number of eigenvalues greater than c is given by $k_1 = \sum_{j=1}^{k} 1_{\{d_j - c > 0\}} = \sum_{i=1}^{q} 1_{\{\lambda_i - c > 0\}}$, and its estimator corresponding to $\hat{k}_1 = \sum_{i=1}^{q} 1_{\{\hat{\lambda}_i(c) - c > 0\}}$ satisfies the following relation: $\forall\, \epsilon > 0 : \ P\big[|\hat{k}_1 - k_1| \leq \epsilon\big] \underset{n\to\infty}{\to} 1$.

The *regularized* inverse based on the *superconsistent* estimator $\hat{\lambda}(c) = (\hat{\lambda}_i(c))_{i=1,\ldots,q}$, with $V_n = [\hat{x}_i]_{i=1,\ldots,q}$ the matrix of eigenvectors, corresponds to:

$$\begin{aligned} \tilde{\Sigma}_n^R\big(\hat{\lambda}(c); c\big) &= V_n\big[\mathrm{diag}(g(\hat{\lambda}_i(c); c))\big]_{i=1,\ldots,q} V_n' \\ &= \tilde{\Sigma}_{11,n}^R\big(\hat{\lambda}(c); c\big) + \tilde{\Sigma}_{22,n}^R\big(\hat{\lambda}(c); c\big) + \tilde{\Sigma}_{33,n}^R\big(\hat{\lambda}(c); c\big) \ , \end{aligned} \tag{9.9}$$

where

$$\tilde{\Sigma}_{11,n}^R\big(\hat{\lambda}(c); c\big) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i \in I_j} g(\hat{\lambda}_i(c); c) 1_{\{(\hat{\lambda}_i(c) - c) > 0\}} \ , \tag{9.10}$$

$$\tilde{\Sigma}_{22,n}^R\big(\hat{\lambda}(c); c\big) = g(c; c) \frac{1}{\hat{m}(c)} \sum_{i \in \hat{I}(c)} 1_{\{\hat{\lambda}_i(c) = c\}} \times \ P_{\hat{I}(c)}(\Sigma_n) \ , \tag{9.11}$$

$$\tilde{\Sigma}_{33,n}^R\big(\hat{\lambda}(c); c\big) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i \in I_j} g(\hat{\lambda}_i(c); c) 1_{\{(\hat{\lambda}_i(c) - c) < 0\}} \ , \tag{9.12}$$

since $\bigcup_{j=1}^{k} I_j = \{1, \ldots, q\}$, and where $P_{I_j}(\Sigma_n) = \sum_{i \in I_j} \hat{x}_i \big(\hat{x}_i' \hat{x}_i\big)^{-1} \hat{x}_i'$, $P_{\hat{I}(c)}(\Sigma_n) = \sum_{i \in \hat{I}(c)} \hat{x}_i \big(\hat{x}_i' \hat{x}_i\big)^{-1} \hat{x}_i'$. The $\hat{x}_i$'s do not have norm equal to 1, and $\hat{m}(c)$ is defined in (9.8).

Proposition 9.1 DISTRIBUTION OF THE SIMULATION BASED TEST. *Let* $\hat{\lambda}(c) = (\hat{\lambda}_i(c))_{i=1,\ldots,q}$, *the vector of* *the superconsistent estimators of the eigenvalues at* c*, satisfy Property 3, and* $W_n^R(c)$ *satisfy* (8.1)*. Let the VRF* $g \in \mathcal{G}_c$ *, and the superconsistent estimator-based regularized statistic be:*

$$\tilde{W}_n^R(\hat{\lambda}(c); c) = X_n' \tilde{\Sigma}_n^R(\hat{\lambda}(c); c) X_n \tag{9.13}$$

where $\tilde{\Sigma}_n^R(\hat{\lambda}(c); c)$ *is defined in equations* (9.9)-(9.12)*. Then*

$$\underset{n \to \infty}{\text{plim}} \{\tilde{W}_n^R(\hat{\lambda}(c); c) - W_n^R(c)\} = 0 \ . \tag{9.14}$$

Thus, the simulated distribution of $\tilde{W}_n^R(\hat{\lambda}(c); c)$ will converge to the right distribution asymptotically so that the level of the simulation-based test will be controlled asymptotically.

10. The case with a varying threshold c_n

We shall now present the convergence results for the regularized inverses which are fundamental to obtain well-behaved regularized test statistics when the threshold varies with the sample size. Let $\hat{\lambda}_i = \lambda_i(\Sigma_n)$ and $\lambda_i = \lambda_i(\Sigma)$ for notational simplicity. First when designing the VRF $g(\lambda; c_n)$, the varying threshold c_n must be selected so that

$$Pr\big[|\hat{\lambda}_i - \lambda_i| > c_n\big] = Pr\big[|b_n(\hat{\lambda}_i - \lambda_i)| > b_n c_n\big] \underset{n\to\infty}{\to} 0 \tag{10.1}$$

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 does. 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 **10.1** below.

Proposition 10.1 CONVERGENCE OF THE REGULARIZED INVERSE WHEN THE THRESHOLD VARIES WITH THE SAMPLE SIZE. *Let* Σ *be a* $q \times q$ *real symmetric positive semidefinite nonstochastic matrix and* Σ_n *a sequence of* $q \times q$ *real symmetric random matrices. Suppose the assumptions of Theorem* **6.6** *hold with* $p = q$ *and let* $g \in \mathcal{G}_c$*. Let* $\hat{\lambda}_i = \lambda_i(\Sigma_n)$*. Suppose further that* $c_n \underset{n\to\infty}{\to} 0$ *and* $b_n c_n \underset{n\to\infty}{\to} \infty$*. If* $\Sigma^R(0)$ *and* $\Sigma_n^R(c_n)$ *have the representation (*7.13*) and (*7.14*) respectively, then*

$$\Sigma_n^R(c_n) \overset{\text{p}}{\to} \Sigma^R(0) \ . \tag{10.2}$$

Thus, an important continuity property for the regularized inverse (unlike g-inverses) is established in this proposition that contributes to the econometric literature.

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 a decomposition of the *regularized* statistic into a regular component and a regularized one. Recall that we want to test the null hypothesis given in equation (2.1), *i.e.* $H_0(\psi_0) : \psi(\theta_0) = \psi_0$.

Proposition 10.2 ASYMPTOTIC CHARACTERIZATION OF THE REGULARIZED WALD STATISTIC WITH VARYING THRESHOLD. *Suppose the assumptions of Proposition* **10.1** *are satisfied. Suppose, also, Assumption* **2.1** *holds, and* $\text{rank}(\Sigma) = q_1$*. Let* k_1 *be the number of non-zero distinct eigenvalues,* i.e., $\sum_{j=1}^{k_1} m(d_j) = q_1 \geq 1$*, and* $g(d_j; 0) = 0, \forall\, j \geq k_1 + 1$*. Let* $\mathcal{V}_1(\Sigma)$*, and* $\mathcal{V}_1(\Sigma_n)$*, be the eigenspace associated with the total eigenprojections*

$\sum_{j=1}^{k_1} P_j(\Sigma)$, $\sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)$ *respectively, and* $\mathcal{V}_2(\Sigma)$, $\mathcal{V}_2(\Sigma_n)$, *their complements in* $\mathbb{R}^q$. *Then, under* $H_0(\psi_0)$,

$$W_n^R(c_n) = X_n' \Sigma_n^R(c_n) X_n \xrightarrow{\mathcal{L}} X' \Sigma^R(0) X = W^R(0) \tag{10.3}$$

$$W_n^R(c_n) = W_{1n}^R(c_n) + W_{2n}^R(c_n) \tag{10.4}$$

$$W_{1n}^R(c_n) = X_n' \Sigma_{11,n}^R(c_n) X_n \xrightarrow{\mathcal{L}} X' \Sigma_{11}^R(0) X \equiv W_1^R(0) \tag{10.5}$$

$$W_{2n}^R(c_n) = X_n' \Sigma_{22,n}^R(c_n) X_n \xrightarrow{\mathcal{L}} 0 \,. \tag{10.6}$$

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 spectral cut-off Moore-Penrose inverse, which annihilates the nonregular component $W_2^R(0)$. Moreover, if we restrict the convergence in law above to the sole standard Gaussian distribution, *i.e.*, $\big[X_n = a_n(\hat{\psi}_n - \psi_0) = \sqrt{n}[\psi(\hat{\theta}) - \psi_0] \to N[0, \Sigma]\big]$, we obtain the result given by Lutkepohl and Burda (1997, Proposition 2, page 318) as a special case (Corollary **10.3**). In this case, the regularized Wald test is asymptotically distributed as a $\chi^2(q_1)$ variable, with q_1 denoting the number of nonzero eigenvalues greater than the threshold. It is important to note, also, that Lutkepohl and Burda (1997, Proposition 2, page 318) result holds only for distinct eigenvalues, unlike Proposition **10.2** that is valid under multiple eigenvalues.

Corollary 10.3 ASYMPTOTIC DISTRIBUTION OF THE REGULARIZED WALD STATISTIC IN THE GAUSSIAN CASE WITH VARYING THRESHOLD. *Suppose the assumptions of Proposition* **10.2** *hold. Replace Assumption* **2.1** *with* **2.4**. *Suppose further Assumption* **2.2** *holds with* $X_j = N\big[0, I_{m(d_j)}\big]$ *for all* j. $g(d_j; 0) = \frac{1}{d_j}, \forall\, j \leq k_1$ *and 0 otherwise. Then, under* $H_0(\psi_0)$,

$$W_n^R(c_n) = n[\psi(\hat{\theta}) - \psi_0]' \Sigma_n^R(c_n) [\psi(\hat{\theta}) - \psi_0] = W_{1n}^R(c_n) + W_{2n}^R(c_n) \,,$$

with

$$W_{1n}^R(c_n) = n[\psi(\hat{\theta}) - \psi_0]' \Sigma_{11,n}^R(c_n) [\psi(\hat{\theta}) - \psi_0] \,, \tag{10.7}$$

$$W_{2n}^R(c_n) = n[\psi(\hat{\theta}) - \psi_0]' \Sigma_{22,n}^R(c_n) [\psi(\hat{\theta}) - \psi_0] \,, \tag{10.8}$$

and

$$W_{1n}^R(c_n) \xrightarrow{\mathcal{L}} W_1^R(0) \sim \chi^2(q_1) \text{ and } W_{2n}(c_n) \xrightarrow{\mathcal{L}} 0. \tag{10.9}$$

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.

The *regularized* test has power against local alternatives:

$$H_1 : \sqrt{n}\big(\psi_{1n}(\theta) - \psi_0\big) \to \Delta \,, \Delta \neq 0 \,. \tag{10.10}$$

Under this alternative, the *regularized* test has an asymptotic noncentral χ^2 distribution, *i.e.*

$$W_n^R(c_n) \xrightarrow{\mathcal{L}} \chi^2\big(q_1, \Delta' \Sigma_{11}^R \Delta\big) \,. \tag{10.11}$$

For example, in the LB case the modified statistic corresponds to:

$$W_n^+(c_n) = n\psi(\hat{\theta})' \hat{V} \hat{\Lambda}^+(c_n) \hat{V}' \psi(\hat{\theta}), \tag{10.12}$$

with

$$W_n^R(c_n) = W_{1n}^R(c_n) + W_{2n}^R(c_n) \geq \ W_n^+(c_n) \quad \text{in finite samples} \tag{10.13}$$

where $\Lambda^+(c_n) = diag(\lambda_1^{-1}, \ldots, \lambda_{q_1}^{-1}, 0, \ldots, 0)$ represents a modified version of the Moore-Penrose inverse of $\Lambda = diag(\lambda_1, \ldots, \lambda_{q_1}, \lambda_{q_1+1}, \ldots, \lambda_q)$. $\Lambda^+(c_n)$ corresponds to a spectral cut-off regularization scheme.

11. Simulation results

In this section, we perform Monte Carlo experiments to assess the empirical behavior of the (regularized) Wald statistics in two different situations: first, we conduct a multi-step noncausality test under the normality assumption, then we test nonlinear restrictions on parameters in a non-Gaussian case, where the delta method breaks down.

11.1. Multi-step noncausality under Gaussianity

To test the null of multi-step noncausality $H_0 : r(\alpha) = 0$, we use four different versions of the Wald statistic *i.e.*

$$W = nr(\hat{\alpha})'\hat{\Sigma}_{r(\alpha)}^R r(\hat{\alpha}) \tag{11.1}$$

where singularity problems arise under parameter setting (3.1).

11.1.1. Simulation design

We examine three different kinds of parameter settings for the VAR(1) coefficients

$$A_1 = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} .$$

The first two parameter setups correspond to:

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

where the problem of singularity is obtained for $\alpha_{xy} = \alpha_{xz} = \alpha_{zy} = 0$. The key parameter here to disentangle between the regularity point and singularity point under this configuration is α_{xz}, with $\alpha_{xz} = 0$ corresponding to a singularity point, and $\alpha_{xz} \neq 0$ to a regularity point.

A third parameter setup is examined:

$$A_1 = A_{11} = \begin{bmatrix} 0.3 & \alpha_{xy} & \alpha_{xz} \\ 0.7 & 0.3 & 0.25 \\ 0.5 & 0.4 & 0.3 \end{bmatrix},$$

where $\alpha_{xy} = \alpha_{xz} = 0$, and $\alpha_{zy} = 0.4 \neq 0$ 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.

Let $u_t = [u_{x,t} \ u_{y,t} \ u_{z,t}]'$ be a Gaussian noise with nonsingular covariance matrix Σ_u. The threshold values have been set to

$$c_n = \hat{\lambda}_1 n^{-1/3}, \quad c = 0.1, 0.001 .$$

Concerning c_n, it has been normalized by the largest eigenvalues to account for scaling issues of the data. For the fixed threshold c, we study a weak and a stronger regularization to investigate its impact on the results. We use

5000 replications in the simulation experiment. The nominal size to perform the tests has been fixed to 5%, 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 $\chi^2_{95\%}(1) = 3.84$. In the tables below, let W_1 denote the standard Wald statistic, W_2 the spectral cut-off regularized Wald statistic, W_3 the full-rank regularized Wald statistic using the conservative bound, and W_4 the regularized Wald statistic based on a super-consistent estimator of the eigenvalues at c whose distribution is simulated.

11.1.2. Level assessment

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

$$H_0: \quad r(\alpha) = \begin{bmatrix} \alpha_{xy} \\ \alpha_{xx}\alpha_{xy} + \alpha_{xy}\alpha_{yy} + \alpha_{xz}\alpha_{zy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} ,$$

first in irregular setups (see Table 1, panels $A: \ A_1 = A_{10}$ and $B: \ A_1 = A_{20}$), then in a regular setup (se Table 1, panel $C: \ A_1 = A_{11}$).

It is clear from Table 1, panels A and B that the standard Wald statistic, W_1, does not have its conventional asymptotic distribution in non-regular setups, either suffering from over-rejections in small samples, or from under-rejections in large samples; its behavior is more critical when parameter values approach the nonstationary region (Table 1, Panel A). The reduced rank Wald statistic, W_2, displays the same finite sample behavior as W_1, in the non-regular setups, with more and more size distortions when parameters values get close to the nonstationary region, but reaches the right asymptotic size when the sample size increases. In contrast, the full-rank regularized statistic that uses the bound, W_3, is conservative, as it under-rejects the null hypothesis, whereas the full-rank regularized statistic based on the superconsistent estimator of the eigenvalues, W_4, reaches the right nominal level of 0.05 for large sample sizes, providing evidence that the level is controlled *at least* asymptotically. We also report in the last column the empirical frequency [denoted by $freq(\hat{\lambda}(c))$] of the superconsistent estimator for the smallest eigenvalue. Regarding the regular setup shown in Table 1, panel C, all statistics display the correct expected size of 0.05 at least asymptotically. However, for the regular setup, the modified Moore-Penrose Wald statistic, W_2, proposed by Lutkepohl and Burda (1997) should use the critical point given by the full-rank chi-squared distribution, *i.e.*$\chi^2_{95\%}(2) = 5.99$, instead of the reduced rank $\chi^2_{95\%}(1) = 3.84$ critical point. In practice, the econometrician does not know a priori which one to use; he is better off using the same full-rank $\chi^2_{95\%}(2) = 5.99$ associated with the full-rank regularized statistic, W_3. If he uses the full-rank critical point given by $\chi^2_{95\%}(2) = 5.99$ associated with the modified Moore-Penrose statistic, he will converge to the right nominal size, but if he picks up the wrong reduced one given by $\chi^2_{95\%}(1) = 3.84$, the size distortion increases. Indeed, we report evidence on this claim in Table 1, panel C, where the frequencies shown in parentheses correspond to the wrong reduced critical point ($\chi^2_{95\%}(1) = 3.84$) in a regular setup. Note also that we have tried different values for the fixed threshold c, and we recommend $c = 0.1$ to control for the size, especially for the superconsistent-based estimator whose distribution is simulated. Smaller values of the fixed threshold do not guarantee a control of the size for W_4.

11.1.3. Power assessment

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

$$H_1: \quad r(\alpha) = \begin{bmatrix} \delta \\ (\alpha_{xx} + \alpha_{yy})\delta \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} ,$$

with $\alpha_{xy} = \delta$, ($\delta = 0.0632$ or $\delta = 0.1264$) whose empirical power is reported in Table 2, panels A and B. We also consider a second type of alternative for a violation of the second restriction only, while maintaining fulfilled the

Table 1. Empirical levels of tests for multistep noncausality $H_0 : r(\alpha) = 0$

$H_0 : r(\alpha) = 0$					
Panel A: irregular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.99$, $A_1 = A_{10}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.3220	0.2766	0.0052	0.0062	1.00
100	0.2550	0.2396	0.0006	0.0006	1.00
200	0.1764	0.1776	0	0	1.00
500	0.0938	0.1158	0	0	1.00
1000	0.054	0.0842	0	0	1.00
2000	0.0362	0.0664	0	0.0662	0
5000	0.0224	0.0560	0	0.0564	0
$H_0 : r(\alpha) = 0$					
Panel B: irregular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.9$, $A_1 = A_{20}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$; [0.001]					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.1046	0.1418	0.0648 [0.0944]	0.1412	1.00
100	0.0584	0.0986	0.0384 [0.0442]	0.1114	1.00
200	0.0328	0.0742	0.0236 [0.0242]	0.0834	1.00
500	0.0234	0.0560	0.0170 [0.0172]	0.0620	1.00
1000	0.0182	0.0552	0.0166 [0.0166]	0.0564	1.00
2000	0.0164	0.0512	0.0140 [0.0142]	0.0966	0
5000	0.0152	0.0530	0.0118 [0.0118]	0.0574	0
$H_0 : r(\alpha) = 0$					
Panel C: regular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = 0.3$, $A_1 = A_{11}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$; [0.001]					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0442	0.0254 (0.0656)	0.0422 [0.0442]	0.0848	0.9796
100	0.0424	0.0200(0.0624)	0.0402 [0.0424]	0.0798	0.9964
200	0.0442	0.0198(0.0562)	0.0426 [0.0442]	0.0666	0.9996
500	0.0456	0.0136 (0.0536)	0.0436 [0.0456]	0.0562	1.00
1000	0.0504	0.0160 (0.0592)	0.0484 [0.0504]	0.0588	1.00
2000	0.0432	0.0294(0.0930)	0.0426 [0.0432]	0.0498	1.00
5000	0.0478	0.0478(0.1444)	0.0476 [0.0478]	0.0540	1.00

Table 2. Locally-corrected size empirical power of tests for multistep noncausality $H_1 : r(\alpha) \neq 0$

Panel A: irregular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.99$, $A_1 = A_{10}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$;					
$H_1 : r(\alpha) \neq 0$ $\alpha_{xy} = \delta = 0.0632$, $\alpha_{xz} = 0$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.6970	0.8380	0.8496	0.8654	1.00
100	0.9764	0.9942	0.9972	0.9986	1.00
200	1.00	1.00	1.00	1.00	1.00
500	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00
2000	1.00	1.00	1.00	1.00	0
5000	1.00	1.00	1.00	1.00	0
$H_1 : r(\alpha) \neq 0$ $\alpha_{xy} = \delta = 0.1264$, $\alpha_{xz} = 0$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.9044	0.9604	0.98	0.9852	1.00
100	0.9992	0.9998	0.9998	0.9998	1.00
200	1.00	1.00	1.00	1.00	1.00
500	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00
2000	1.00	1.00	1.00	1.00	0
5000	1.00	1.00	1.00	1.00	0
Panel B: irregular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.9$, $A_1 = A_{20}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$; [0.001]					
$H_1 : r(\alpha) \neq 0$ $\alpha_{xy} = \delta = 0.1264$, $\alpha_{xz} = 0$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.4246	0.3888	0.4048 [0.4124]	0.1714 [0.0048]	1.00 [1.00]
100	0.8058	0.7892	0.7504 [0.8034]	0.5336 [0.00]	1.00 [1.00]
200	0.9830	0.9820	0.9670 [0.9782]	0.9396 [0.00]	1.00 [1.00]
500	1.00	1.00	0.9998 [1.00]	0.9998 [1.00]	1.00 [1.00]
1000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	1.00 [1.00]
2000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [1.00]
5000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [1.00]
Panel C: regular setup					
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = 0.3$, $A_1 = A_{11}$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.1$; [0.001]					
$H_1 : r(\alpha) \neq 0$ $\alpha_{xz} = \delta = 0.1264$, $\alpha_{xy} = 0$, $\alpha_{zy} = 0.4$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0918	0.028 (0.0588)	0.0854 [0.0918]	0.0952 [0.1106]	0.9840 [0.9278]
100	0.1854	0.0310 (0.0582)	0.1692 [0.1854]	0.1982 [0.2576]	0.9966 [0.9428]
200	0.4028	0.0318 (0.0662)	0.3736 [0.4028]	0.4130 [0.5186]	0.9996 [0.9336]
500	0.8312	0.0730 (0.1160)	0.8100 [0.8312]	0.8310 [0.8908]	1.00 [0.7982]
1000	0.9866	0.2576 (0.3064)	0.9854 [0.9866]	0.9874 [0.9914]	1.00 [0.3630]
2000	1.00	0.8728 (0.8780)	1.00 [1.00]	1.00 [1.00]	1.00 [0.62]
5000	1.00	1.00 (1.00)	1.00 [1.00]	1.00 [1.00]	1.00 [0.00]

first restriction, *i.e.*

$$H_1: \quad r(\alpha) = \begin{bmatrix} 0 \\ (\alpha_{xz} \times \alpha_{zy}) \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ ,$$

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

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

see Table 2, panel C. First of all, all power frequencies reported in Table 2 have been locally corrected for size distortions (only for over-rejections and *not* for under-rejections) for a fair comparison across statistics.

Strikingly, as shown in Table 2, panel A, although the full-rank regularized test statistics, W_3 and W_4 are conservative under the null hypothesis near the nonstationary region, they do not entail a loss of power under the alternative, compared to their oversized competitors W_1 and W_2. Once the latter have been corrected for size distortions, they do not over-perform the full-rank regularized statistics, W_3 and W_4 from the viewpoint of power. More importantly, the locally-level corrected statistics W_1 and W_2 are *infeasible* tests in practice, as this level correction requires to know the true value of the parameter. Note an under-performance of W_1 relative to the others, when the sample size is very small (panel A: $n = 50$). Further, the results reported in Table 2, panel B, shed light on the better finite sample power properties of the conservative bound test relative to the superconsistent estimator-based regularized statistic whose distribution is simulated. Besides being easier and faster to conduct, W_3 also exhibits better power properties in finite sample than its simulation-based competitor W_4. Also, the performance of W_3 is less sensitive to the value of the fixed threshold c compared to W_4. Finally, the most striking result is the *under-performance* of the reduced rank modified statistic proposed by Lutkepohl and Burda (1997) under the regular setup shown in panel C. As expected, 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. Violation of the null hypothesis coming from the second restriction will be missed by a statistic that underestimates the rank, which once again makes the full-rank regularized statistics more attractive. Even with a more favorable critical point given by the $\chi^2(1) = 3.84$ in parentheses, the spectral cut-off regularized statistic has trouble to reach the power performance achieved by its competitors. Indeed, it requires 2000 observations to achieve reasonable power of 87% relative to the others already at 100%. Thus, these results on power reinforce the better properties of the full-rank regularized statistics over the spectral cut-off statistic.

11.2. Deviation from normality: the Delta method breaks down

We now assess the empirical level of the null hypothesis:

$$H_0(\psi_0) : \psi(\theta) = \theta'\theta = 0$$

at the nominal size of 5%. For ease of notation, we shall denote the statistics as follows; the standard Wald test is

$$W_1 = W = n\psi(\hat{\theta}_n)'\hat{\Sigma}^{-1}\psi(\hat{\theta}_n) \ ; \tag{11.2}$$

the Moore-Penrose modified Wald statistic proposed by Lutkepohl and Burda (1997) is:

$$W_2 = W^{+}(c_n) = n\psi(\hat{\theta}_n)'\hat{\Sigma}^{+}(c_n)\psi(\hat{\theta}_n) \ ; \tag{11.3}$$

the *regularized* Wald test statistic is

$$W_3 = W_{\psi}^{R}(c) = n^2\psi(\hat{\theta}_n)'\hat{\Sigma}^{R}(c)\psi(\hat{\theta}_n) \ ; \tag{11.4}$$

W_3 uses the quadratic form, ***i.e.*** $(\chi^2_{95\%}(p))^2$ as a bound. For instance for $p = 5$, the $(\chi^2_{95\%}(p))^2$ is equal to $11.07^2 = 122.55$ at 5%, while W_1 and W_2 use the $\chi^2_{95\%}(1) = 3.84$ as critical point. Finally, the *regularized* Wald test statistic, using the superconsistent estimator of the eigenvalues, uses a simulated critical point, ***i.e.***,

$$W_4 = W^R_\psi(\hat{\lambda}(c); c) = n^2\psi(\hat{\theta}_n)'\hat{\Sigma}^R(\hat{\lambda}(c); c)\psi(\hat{\theta}_n) \; ; \tag{11.5}$$

to simulate its distribution, we exploit the information that

$$n\psi(\hat{\theta}_n) = (\sqrt{n}\hat{\theta}_n)'(\sqrt{n}\hat{\theta}_n) \; \sim \chi^2(p)$$

under the null, by drawing $\chi^2(p)$ random numbers.

11.2.1. Level assessment

We can observe from those results that the standard Wald test together with the modified Moore-Penrose test proposed by Lutkepohl and Burda (1997) are never close to the nominal size of 0.05, either with an under-rejection when the dimension of θ is low or with a severe over-rejection when the dimension of θ increases to $p = 10, 14, 16$. In contrast, all the full-rank regularized Wald tests are very close to the nominal size, with an extreme precision for the simulated version of the test based on W_4. However, the choice of c is important for the *regularized* statistic W_3 that uses the bound. The message one can draw from this table is the following: first, the spectral cut-off Moore-Penrose regularized Wald test proposed by Lutkepohl and Burda (1997) is useless in this example and does not add anything to the standard Wald test when normality is violated. Second, the *regularized* Wald statistic that uses the bound control the level of the test, but is more sensitive to the threshold c in a non-Gaussian case. Note also that the bound is exact here. Third, the simulated super consistent regularized Wald test *always* controls the level of the test under this design. The simulation of the test requires only a few seconds. We also report, in the last column of the table, the frequency ($freq(\hat{\lambda}(c))$) at which the superconsistent estimator is set to the threshold c. It is mostly in small samples that the superconsistent estimator play a role.

11.2.2. Power assessment

As expected, when W_1 and W_2 under-reject under the null for low dimensions of θ , they lose power under the alternative *in small samples* compared to the full-rank regularized statistics W_3, W_4. Hence, full-rank regularized statistics based on non Gaussian distributions over-perform in term of power, in small sample sizes, when the dimension of θ is low. When the dimension of θ increases, they can match the power performance of *infeasible level-corrected* test, W_1 and W_2. The correction performed for W_1 and W_2 is locally and therefore the power shown in the table is overstated. Their power would even be lower under a global level correction.

12. Conclusion

In this paper, we introduce a new class of *regularized* inverses, as opposed to generalized inverses, that embeds the spectal cut-off and Tikhonov regularized inverses known in the literature. We propose three regularized Wald statistics for general law: the first two statistics rely on a fixed value for the threshold in the VRF $g(\lambda; c)$ while the third one lets the threshold vary with the sample size, but requires more information about the sample behavior of the eigenvalues. 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* is then derived for this first regularized statistic under general laws for the restrictions; such a bound corresponds to a χ^2 variable with *full rank* under Gaussianity. Hence, the test is *asymptotically valid*, meaning that the usual critical point (given by the χ^2 variable with *full rank*) can be used, but is conservative. The second regularized statistic

Table 3. Empirical levels of tests

$H_0 : \psi(\theta) = 0$; nominal size= 0.05					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 5$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0192	0.0192	0.0538 [0.0468]	0.0482 [0.0482]	0.2454 [0.1810]
100	0.0132	0.0132	0.0642 [0.0554]	0.0514 [0.0514]	0.0118 [0.0058]
200	0.0148	0.0148	0.0656 [0.0562]	0.0546 [0.0546]	0 [0]
500	0.0096	0.0096	0.064 [0.0540]	0.0526 [0.0526]	0 [0]
1000	0.0072	0.0072	0.0572 [0.0496]	0.0482 [0.0482]	0 [0]
2000	0.0072	0.0072	0.0568 [0.0498]	0.0460 [0.0460]	0 [0]
5000	0.0092	0.0092	0.0596 [0.0524]	0.0496 [0.0496]	0 [0]
$\dim(\theta) = 10$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.1730	0.1730	0.0114 [0.0104]	0.0510 [0.0510]	0.6138 [0.5480]
100	0.1504	0.1504	0.0698 [0.0578]	0.0546 [0.0546]	0.1458 [0.0962]
200	0.1282	0.1282	0.0712 [0.0546]	0.0498 [0.0498]	0.0008 [0.0002]
500	0.1264	0.1264	0.0610 [0.0488]	0.0444 [0.0444]	0 [0]
1000	0.1192	0.1192	0.0624 [0.0508]	0.0460 [0.0460]	0 [0]
2000	0.1184	0.1184	0.0662 [0.0532]	0.0504 [0.0504]	0 [0]
5000	0.1156	0.1156	0.0652 [0.0540]	0.0488 [0.0488]	0 [0]
$\dim(\theta) = 14$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.4130	0.4130	0.0014 [0.0012]	0.0538 [0.0538]	0.6204 [0.6148]
100	0.3778	0.3778	0.0586 [0.0494]	0.0560 [0.0560]	0.3862 [0.2820]
200	0.3606	0.3606	0.0706 [0.0560]	0.0500 [0.0500]	0.0070 [0.0022]
500	0.3648	0.3648	0.0650 [0.0486]	0.0460 [0.0460]	0 [0]
1000	0.3566	0.3566	0.0678 [0.0556]	0.0506 [0.0506]	0 [0]
2000	0.3576	0.3576	0.0672 [0.0548]	0.0484 [0.0484]	0 [0]
5000	0.3524	0.3524	0.0674 [0.0554]	0.0480 [0.0480]	0 [0]
$\dim(\theta) = 16$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.5384	0.5384	0.002 [0.0002]	0.0536 [0.0536]	0.5424 [0.5650]
100	0.5132	0.5132	0.0438 [0.0384]	0.0546 [0.0546]	0.508 [0.4074]
200	0.5068	0.5068	0.0702 [0.055]	0.0498 [0.0498]	0.015 [0.0062]
500	0.5022	0.5022	0.0676 [0.0514]	0.0488 [0.0488]	0 [0]
1000	0.4948	0.4948	0.0678 [0.0510]	0.0474 [0.0474]	0 [0]
2000	0.5058	0.5058	0.0724 [0.0540]	0.0470 [0.0470]	0 [0]
5000	0.4938	0.4938	0.0662 [0.0514]	0.0460 [0.0460]	0 [0]

Table 4. Empirical power of tests

$H_1 : \psi(\theta) = 0.075916 \ \neq 0$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 5$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.1384	0.1384	0.2884 [0.2724]	0.2810 [0.2810]	0.4994 [0.4388]
100	0.3124	0.3124	0.5476 [0.5366]	0.5398 [0.5398]	0.3030 [0.2318]
200	0.6990	0.6990	0.8712 [0.8686]	0.8730 [0.8730]	0.1004 [0.0528]
500	0.9948	0.9948	0.9998 [0.9998]	0.9998 [0.9998]	0.0022 [0.0006]
1000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
2000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
5000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
$H_1 : \psi(\theta) = 0.05223916 \ \neq 0$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 5$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0912	0.0912	0.2044 [0.1928]	0.2008 [0.2008]	0.4322 [0.3670]
100	0.1882	0.1882	0.3804 [0.3726]	0.3736 [0.3736]	0.1898 [0.1332]
200	0.4596	0.4596	0.7006 [0.6952]	0.7058 [0.7058]	0.0292 [0.0142]
500	0.9546	0.9546	0.9888 [0.9882]	0.9886 [0.9886]	0 [0]
1000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
2000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
5000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]
$H_1 : \psi(\theta) = 0.00355332 \ \neq 0$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 10$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.052	0.052	0.0112 [0.011]	0.0568 [0.0568]	0.6166 [0.5546]
100	0.0614	0.0614	0.0658 [0.060]	0.0656 [0.0656]	0.1692 [0.1104]
200	0.0708	0.0708	0.0738 [0.0738]	0.0744 [0.0744]	0.0022 [0.0004]
500	0.1216	0.1216	0.1260 [0.1240]	0.1164 [0.1164]	0 [0]
1000	0.2090	0.2090	0.1962 [0.1998]	0.1954 [0.1954]	0 [0]
2000	0.3884	0.3884	0.3912 [0.3930]	0.3908 [0.3908]	0 [0]
5000	0.8454	0.8454	0.8490 [0.8490]	0.8454 [0.8454]	0 [0]
$H_1 : \psi(\theta) = 0.01056748 \ \neq 0$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 14$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0678	0.0678	0.0682	0.0666	0.5994 [0.6052]
100	0.0792	0.0792	0.0784	0.0808	0.4450 [0.3412]
200	0.1072	0.1072	0.1180	0.1166	0.0208 [0.0110]
500	0.2416	0.2416	0.2490	0.2382	0 [0]
1000	0.5082	0.5082	0.5074	0.5068	0 [0]
2000	0.8582	0.8582	0.8628	0.8612	0 [0]
5000	0.9996	0.9996	0.9998	0.9998	0 [0]
$H_1 : \psi(\theta) = 0.01733792 \ \neq 0$					
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.9$; [0.97]					
$\dim(\theta) = 16$					
n	W_1	W_2	W_3	W_4	$freq(\hat{\lambda}(c))$
50	0.0672	0.0672	0.0006 [0.0006]	0.0716 [0.0716]	0.4994 [0.5250]
100	0.0896	0.0896	0.0810 [0.0726]	0.1002 [0.1002]	0.5960 [0.4928]
200	0.1436	0.1436	0.1598 [0.1598]	0.1580 [0.1580]	0.0556 [0.0290]
500	0.3672	0.3672	0.3934 [0.3934]	0.3822 [0.3822]	0 [0]
1000	0.7430	0.7430	0.7532 [0.7532]	0.7444 [0.7444]	0 [0]
2000	0.9840	0.9840	0.9846 [0.9846]	0.9836 [0.9836]	0 [0]
5000	1.00	1.00	1.00 [1.00]	1.00 [1.00]	0 [0]

relies on a *superconsistent* estimator of the eigenvalues at the threshold c whose distribution can be simulated. Finally, when the threshold goes to zero with the sample size, we obtain the spectral cut-off modified Wald statistic of Lutkepohl and Burda (1997) as a special case. Under normality, the test has the asymptotic χ^2 distribution with a reduced rank, *i.e.* the number of eigenvalues greater than zero. Note that Lutkepohl and Burda (1997) result only holds for distinct eigenvalues whereas our result accounts for eigenvalues with multiplicity larger than one. *Seventh*, we also show that the regularized statistics are consistent against global alternatives, but the spectral cut-off Wald test used by Lutkepohl and Burda (1997) has reduced power in some directions of the alternative. In brief, our regularization approach, especially when implemented through the full-rank regularized bound test, has better size and power properties than the reduced rank statistic and the conventional Wald test. Besides our approach is systematic, and robust to regular setups unlike Lutkepohl and Burda (1997) 's reduced rank procedure.

A. Appendix: Proofs

Proof of Property 1 Using decomposition (4.1) and (4.9), we have:

$$\Sigma\Sigma^R(c) = V\Lambda V' V\Lambda^\dagger(c)V' = V\Lambda\Lambda^\dagger(c)V'$$

where we use the fact that the V_i's are orthogonal matrices. For all λ, $0 \le \lambda g(\lambda; c) \le 1$, so that:

$$\Sigma\Sigma^R(c) = V\mathrm{diag}\big[\lambda_j g(\lambda_j; c)\big]_{j=1,\cdots,q} V' \;\le\; I_q \ .$$

Regarding $ii)$, we have:

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

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

$$\Sigma - \Sigma\Sigma^R(c)\Sigma \ge \ 0 \ \Leftrightarrow \ \Sigma\big(I_q - \Sigma^R(c)\Sigma\big) \ge \ 0 \ \Rightarrow \ I_q - \Sigma^R(c)\Sigma \ \ge \ 0 \tag{A.1}$$

since Σ is semi definite positive. The last implication holds by $i)$.

Regarding $iv)$, for all $\lambda \ge 0$, $g(\lambda; c)$ bounded, and if $g(\lambda; c) > 0$, we have:

$$\lambda g(\lambda; c) \le 1 \ \Rightarrow 0 < g(\lambda; c) \le \ \frac{1}{\lambda} \le \infty \quad \text{hence} \quad \Big(g(\lambda; c)\Big)^{-1} - \lambda \ \ge \ 0 \ .$$

Hence,

$$\big(\Sigma^R(c)\big)^{-1} - \Sigma = V\mathrm{diag}\big[\big(g(\lambda_j; c)\big)^{-1} - \lambda_j\big]_{j=1,\cdots,q}V' \ \ge 0 \ .$$

Finally for $v)$, the rank is given by the number of eigenvalues greater than zero. As $\Sigma^R(c) = V g(\lambda_j; c)_{j=1,\cdots,q}V'$, hence

$$\big(\lambda > 0 \ \Rightarrow \ g(\lambda; c) > 0\big) \quad \Rightarrow \quad \big(\mathrm{rank}\big(\Sigma^R(c)\big) \ge \mathrm{rank}(\Sigma)\big) \ .$$

□

PROOF of Lemma 6.3 If $\Sigma_n \overset{a.s.}{\to} \Sigma$, then the event $A = \{\omega : \ \Sigma_n(\omega) \underset{n\to\infty}{\to} \Sigma\}$ has probability one, *i.e.* $P(A) = 1$. For any $\omega \in A$, we have by Lemma **6.2**:

$$[\Sigma_n(\omega) \underset{n\to\infty}{\to} \Sigma] \ \Rightarrow [\lambda_j(\Sigma_n(\omega)) \to \lambda_j(\Sigma), \ \ j = 1, \dots, J] \ .$$

Denoting $B = \{\omega : \ \lambda_j(\Sigma_n(\omega)) \underset{n\to\infty}{\to} \lambda_j(\Sigma)\}$, 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 6.4

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

i) $\lambda_j(\Sigma_{m_l}) \overset{a.s.}{\to} \lambda_j(\Sigma)$,

ii) $P_{j,t}(\Sigma_{m_l}) \overset{a.s.}{\to} P_{j,t}(\Sigma)$ provided $\lambda_{j-1}(\Sigma) \ne \lambda_j(\Sigma)$ and $\lambda_t(\Sigma) \ne \lambda_{t+1}(\Sigma)$.

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

PROOF of Proposition 7.3 If $\Sigma_n \stackrel{a.s.}{\to} \Sigma$, then by lemma **6.3** i), we have $\hat{\lambda}_i \stackrel{a.s.}{\to} d_j,\ \forall i \in I_j$, where $I_j = \{i \in I : \lambda_i = d_j\}$. Under the additional Assumption **7.2**, and the a.e. continuity of $g(.,c)$, we have $g(\hat{\lambda}_i; c) \stackrel{a.s.}{\to} g(d_j; c)\ \forall i \in I_j$. Moreover, by lemma **6.3** ii), we have $P_{I_j}(\Sigma_n) \stackrel{a.s.}{\to} P_j(\Sigma)$. Hence,

$$\Sigma_n^R(c) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\left[g(d_j;c) - g(d_j;c) + \frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i;c)\right]$$

$$= \sum_{j=1}^{k} P_{I_j}(\Sigma_n) g(d_j;c) + \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i;c) - g(d_j;c)\right] \stackrel{a.s.}{\to} \sum_{j=1}^{k} P_j(\Sigma) g(d_j;c) \quad \text{(A.2)}$$

since $g(d_j; c) = \frac{1}{m(d_j)} \times m(d_j) g(d_j; c) = \frac{1}{m(d_j)} \sum_{i\in I_j} g(d_j; c)$.

□

PROOF of Proposition 7.4 Using decomposition (4.7)-(4.8), and from equation (7.14), we have:

$$\Sigma_n^R(c) = \sum_{i=1}^{3} \Sigma_{i,n}^R(c) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c) \quad \text{(A.3)}$$

where $\Sigma_{1,n}^R(c) = \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c)$, $\quad \Sigma_{2,n}^R(c) = P_{I(c)}(\Sigma_n)\frac{1}{m(c)}\sum_{i\in I(c)} g(\hat{\lambda}_i, c)$

and $\Sigma_{3,n}^R(c) = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c)$.

By Lemma **6.4** i) and ii), eigenvalues and total eigenprojections are continuous. Under Assumption **7.2**, we have:

$$\forall\ i \in I_j\ , g(\hat{\lambda}_i, c) \stackrel{p}{\to} g(d_j; c), \quad \text{and} \quad P_{I_j}(\Sigma_n) \stackrel{p}{\to} P_j(\Sigma) .$$

For

$$\Sigma_{1,n}^R(c) = \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i; c) = \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)\left[g(d_j;c) - g(d_j;c) + \frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i;c)\right]$$

$$= \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n) g(d_j;c) + \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i;c) - g(d_j;c)\right] \stackrel{p}{\to} \sum_{j=1}^{k_1} g(d_j;c) P_j(\Sigma)$$

since $g(d_j; c) = \frac{1}{m(d_j)} \times m(d_j) g(d_j; c) = \frac{1}{m(d_j)} \sum_{i\in I_j} g(d_j; c)$. Hence,

$$\Sigma_{1,n}^R(c) = \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c) \stackrel{p}{\to} \sum_{j=1}^{k_1} g(d_j;c) P_j(\Sigma) \equiv \Sigma_1^R(c)$$

$$\Sigma_{2,n}^R(c) = P_{I(c)}(\Sigma_n)\frac{1}{m(c)}\sum_{i\in I(c)} g(\hat{\lambda}_i, c) \ \xrightarrow{p} \ g(c;c)1_{\{d_j=c\}}P_{j(c)}(\Sigma) \equiv \ \Sigma_2^R(c) \ .$$

The proof for $\Sigma_{3,n}^R(c)$ is similar to that of $\Sigma_{1,n}^R(c)$. Hence,

$$\Sigma_{3,n}^R(c) = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c) \ \xrightarrow{p} \ \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)P_j(\Sigma) \equiv \ \Sigma_3^R(c) \ .$$

Therefore, $\Sigma_n^R(c) \ \xrightarrow{p} \ \Sigma^R(c)$.

□

PROOF of Proposition 8.1 By Proposition **7.4**, we have $\Sigma_n^R(c) \ \xrightarrow{p} \ \Sigma^R(c)$ and under Assumption **2.1**, $X_n \ \xrightarrow{\mathcal{L}} \ X$, hence $W_n^R(c) = X_n'\Sigma_n^R(c)X_n \ \xrightarrow{\mathcal{L}} \ X'\Sigma^R(c)X = W^R(c)$. Using representation (7.13) for $\Sigma^R(c)$, and (2.7), we can write:

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

We can further decompose the overall statistic into three blocks depending on the ranking of the eigenvalues w.r.t. c, with $k_1 = \sum_{j=1}^{k} 1_{\{d_j>c\}}$, *i.e.*,

$$W_1^R(c) \ = \ X'\Sigma_{11}^R(c)X = \sum_{j=1}^{k} g(d_j;c)1_{\{d_j>c\}}X'P_j(\Sigma)X = \sum_{j=1}^{k_1} g(d_j;c)X'P_j(\Sigma)X = \sum_{j=1}^{k_1} g(d_j;c)X'B(d_j)B(d_j)'X \ .$$

Similarly, $W_2^R(c) = X'\Sigma_{22}^R(c)X = g(c;c)1_{\{d_j=c\}}X'P_{j(c)}(\Sigma)X = g(c;c)1_{\{d_j=c\}}X'B(c)B(c)'X$. Finally,

$$W_3^R(c) = X'\Sigma_{33}^R(c)X = \sum_{j=1}^{k} g(d_j;c)1_{\{d_j<c\}}X'P_j(\Sigma)X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X \ .$$

□

PROOF of Corollary 8.2 In the Gaussian case, we have: $B(d_j)'X = \sqrt{d_j}x_j$, where $x_j = N\left[0, I_{m(d_j)}\right]$, hence

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

with the three blocks corresponding to

$$W_1^R(c) = X'\Sigma_{11}^R(c)X = \sum_{j=1}^{k_1} g(d_j;c)X'B(d_j)B(d_j)'X = \sum_{j=1}^{k_1} g(d_j;c)d_jx_j'x_j \ ,$$

$$W_2^R(c) = X'\Sigma_{22}^R(c)X = g(c;c)1_{\{d_j=c\}}X'B(c)B(c)'X = g(c;c)1_{\{d_j=c\}}cx_j'x_j \ ,$$

and $W_3^R(c) = X'\Sigma_{33}^R(c)X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)X'B(d_j)B(d_j)'X = \sum_{j=k_1+1_{\{d_j=c\}}+1}^{k} g(d_j;c)d_j x_j' x_j$.

□

PROOF of Proposition 8.5

The quantity $a_n\big[\hat{\psi}_n - \psi_0\big]$ can be written as:

$$a_n\big[\hat{\psi}_n - \psi_0\big] = a_n\big[\hat{\psi}_n - \psi_1 + \psi_1 - \psi_0\big] = a_n[\hat{\psi}_n - \psi_1\big] + a_n\big[\psi_1 - \psi_0\big]\ . \tag{A.4}$$

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

$$\begin{aligned} W_n^R(c) &= \{a_n[\hat{\psi}_n - \psi_1\big] + a_n\big[\psi_1 - \psi_0\big]\}'\Sigma_n^R(c)\{a_n[\hat{\psi}_n - \psi_1\big] + a_n\big[\psi_1 - \psi_0\big]\} \\ &= a_n[\hat{\psi}_n - \psi_1\big]'\Sigma_n^R(c)a_n[\hat{\psi}_n - \psi_1\big] + 2a_n[\hat{\psi}_n - \psi_1\big]'\Sigma_n^R(c)a_n\big[\psi_1 - \psi_0\big] \\ &\quad + a_n\big[\psi_1 - \psi_0\big]'\Sigma_n^R(c)a_n\big[\psi_1 - \psi_0\big] \\ &= X_n'\Sigma_n^R(c)X_n + 2X_n'\Sigma_n^R(c)a_n\Delta + a_n^2\Delta'\Sigma_n^R(c)\Delta \\ &\overset{\mathcal{L}}{\to} X'\Sigma^R(c)X + 2X'\Sigma^R(c)a_n\Delta + a_n^2\Delta'\Sigma^R(c)\Delta \ \to\ \infty \end{aligned} \tag{A.5}$$

since $X_n \overset{\mathcal{L}}{\to} X$, $\Sigma_n^R(c) \overset{p}{\to} \Sigma^R(c)$, but $a_n(\psi_1 - \psi_0) = a_n\Delta \ \to\ \infty$, as a_n grows to infinity. Hence $W_n^R(c)$ converges to infinity with probability 1. The quantity

$$X'\Sigma^R(c)X + 2X'\Sigma^R(c)a_n\Delta + a_n^2\Delta'\Sigma^R(c)\Delta$$

is asymptotically equivalent to

$$X'\Sigma^R(c)X + a_n^2\Delta'\Sigma^R(c)\Delta \tag{A.6}$$

due to the dominance principle of $a_n\Delta'\Sigma^R(c)\Delta$ over $2X'\Sigma^R(c)\Delta$, *i.e.*,

$$X'\Sigma^R(c)X + 2X'\Sigma^R(c)a_n\Delta + a_n^2\Delta'\Sigma^R(c)\Delta = X'\Sigma^R(c)X + a_n\big[2X'\Sigma^R(c)\Delta + a_n\Delta'\Sigma^R(c)\Delta\big]\ .$$

□

PROOF of Proposition 8.6

Under the local alternative $a_n(\psi_{1n} - \psi_0) \to\ \Delta \neq 0$, then

$$\begin{aligned} W_n^R(c) &= a_n[\hat{\psi}_n - \psi_{1n}\big]'\Sigma_n^R(c)a_n[\hat{\psi}_n - \psi_{1n}\big] + 2a_n[\hat{\psi}_n - \psi_{1n}\big]'\Sigma_n^R(c)a_n\big[\psi_{1n} - \psi_0\big] \\ &\quad + a_n\big[\psi_{1n} - \psi_0\big]'\Sigma_n^R(c)a_n\big[\psi_{1n} - \psi_0\big] \\ &= X_n'\Sigma_n^R(c)X_n + 2X_n'\Sigma_n^R(c)a_n\big[\psi_{1n} - \psi_0\big] + a_n\big[\psi_{1n} - \psi_0\big]'\Sigma_n^R(c)a_n\big[\psi_{1n} - \psi_0\big] \\ &\underset{n\to\infty}{\overset{\mathcal{L}}{\to}} X'\Sigma^R(c)X + 2X'\Sigma^R(c)\Delta + \Delta'\Sigma^R(c)\Delta \end{aligned} \tag{A.7}$$

as $X_n \overset{\mathcal{L}}{\to} X$, $\Sigma_n^R(c) \overset{p}{\to} \Sigma^R(c)$.

□

PROOF of corollary 8.7 From Proposition **8.6**, we have:

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

As $\Delta \in \mathcal{V}(0)$, $P(0)(\Sigma)\Delta = \Delta$, and we have:

$$\Sigma^R(c)\Delta = \sum_{d_j} g(d_j; c)P_j(\Sigma)\Delta = g(0; c)P(0)(\Sigma)\Delta = g(0; c)\Delta$$

since $P_j(\Sigma)\Delta = 0$, for all eigenprojections on the eigenspaces different from $\mathcal{V}(0)$. Hence,

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

□

PROOF of Property 3 If $\lambda_i = c$, then $|\hat{\lambda}_i - c| = |\hat{\lambda}_i - \lambda_i|$ and

$$P\big[\hat{\lambda}_i(c) = c\big] = P\big[b_n|\hat{\lambda}_i - c| \le \nu e_n\big] = P\big[b_n|\hat{\lambda}_i - \lambda_i| \le \nu e_n\big] \underset{n\to\infty}{\to} 1 \ , \tag{A.8}$$

since $b_n(\hat{\lambda}_i - \lambda_i) = O_p(1)$, and $\nu e_n \to \infty$. Hence

$$P\big\{s[\hat{\lambda}_i(c) - c] = s[\lambda_i - c]\big\} \underset{n\to\infty}{\to} 1 \ , \ \forall\, i \in I(c) = \{i \in I : \ \lambda_i = c\} \ . \tag{A.9}$$

On the other hand, the modified estimator $\hat{\lambda}_i(c)$ is designed such that:

$$|\hat{\lambda}_i(c) - \lambda_i| = |\hat{\lambda}_i(c) - \hat{\lambda}_i + \hat{\lambda}_i - \lambda_i| = |\hat{\lambda}_i - \lambda_i| \text{ if } |\hat{\lambda}_i - c| > \nu\frac{e_n}{b_n} \ .$$

Hence, $\forall\ \epsilon > 0$,

$$P\big[|\hat{\lambda}_i(c) - \lambda_i| \le \epsilon\big] = P\big[|\hat{\lambda}_i - c| > \nu\frac{e_n}{b_n}\big] = \underset{n\to\infty}{\to} 1 \text{ , if } \lambda_i \neq c \tag{A.10}$$

since $\frac{e_n}{b_n} \underset{n\to\infty}{\to} 0$. Thus, if $\lambda_i > c$, we have $P\big[(\hat{\lambda}_i - c) > \nu\frac{e_n}{b_n}\big] \underset{n\to\infty}{\to} 1$, hence

$$P\big\{s[\hat{\lambda}_i(c) - c] = s[\lambda_i - c]\big\} \underset{n\to\infty}{\to} 1 \ . \tag{A.11}$$

Also, if $\lambda_i < c$, we have $P\big[(c - \hat{\lambda}_i) > \nu\frac{e_n}{b_n}\big] \underset{n\to\infty}{\to} 1$, hence

$$P\big\{s[\hat{\lambda}_i(c) - c] = s[\lambda_i - c]\big\} \underset{n\to\infty}{\to} 1 \ . \tag{A.12}$$

□

PROOF of Proposition 9.1 As $W_n^R(c) = X_n'\Sigma_n^R(c)X_n$, and $\tilde{W}_n^R(\hat{\lambda}(c); c) = X_n'\tilde{\Sigma}_n^R(\hat{\lambda}(c); c)X_n$, it is sufficient to show that $\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) \overset{p}{\to} \Sigma^R(c)$ to have $\tilde{W}_n^R(\hat{\lambda}(c); c) \overset{a}{\sim} W_n^R(c)$, where $\overset{a}{\sim}$ denotes the asymptotic equivalence. We want to show that $\forall \epsilon > 0$

$$p\{|\tilde{W}_n^R(\hat{\lambda}(c); c) - W_n^R(c)| > \epsilon\} \underset{n\to\infty}{\to} 0 \ .$$

As $|\tilde{W}_n^R(\hat{\lambda}(c); c) - W_n^R(c)| = |X_n'\tilde{\Sigma}_n^R(\hat{\lambda}(c); c)X_n - X_n'\Sigma_n^R(c)X_n| = |X_n'\big(\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) - \Sigma_n^R(c)\big)X_n|$ it is equivalent to show, $\forall\ \epsilon > 0$, $p\{\|\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) - \Sigma_n^R(c)\| > \epsilon\} \underset{n\to\infty}{\to} 0$. More specifically,

$$\|\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) - \Sigma_n^R(c)\| = \|\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) - \Sigma^R(c) + \Sigma^R(c) - \Sigma_n^R(c)\| \le \|\tilde{\Sigma}_n^R(\hat{\lambda}(c); c) - \Sigma^R(c)\| + \|\Sigma^R(c) - \Sigma_n^R(c)\|$$

but $p\{\|\Sigma^R(c) - \Sigma_n^R(c)\| > \epsilon\} = p\{\|\Sigma_n^R(c) - \Sigma^R(c)\| > \epsilon\} \to 0$ by Proposition **7.4**. Hence, it is sufficient to show that $\tilde{\Sigma}_n^R(\hat{\lambda}(c);c) \xrightarrow{p} \Sigma^R(c)$. To do so, let us study

$$\begin{aligned}
\|\tilde{\Sigma}_n^R(\hat{\lambda}(c);c) - \Sigma^R(c)\| &= \|\big[\tilde{\Sigma}_{11,n}^R\big(\hat{\lambda}(c);c\big) + \tilde{\Sigma}_{22,n}^R\big(\hat{\lambda}(c);c\big) + \tilde{\Sigma}_{33,n}^R\big(\hat{\lambda}(c);c\big)\big] - \big[\Sigma_{11}^R(c) + \Sigma_{22}^R(c) + \Sigma_{33}^R(c)\big]\| \\
&= \|\big[\tilde{\Sigma}_{11,n}^R\big(\hat{\lambda}(c);c\big) - \Sigma_{11}^R(c)\big] + \big[\tilde{\Sigma}_{22,n}^R\big(\hat{\lambda}(c);c\big) - \Sigma_{22}^R(c)\big] + \big[\tilde{\Sigma}_{33,n}^R\big(\hat{\lambda}(c);c\big) - \Sigma_{33}^R(c)\big]\| \\
&\leq \|\Delta\Sigma_{11,n}\| + \|\Delta\Sigma_{22,n}\| + \|\Delta\Sigma_{33,n}\| \qquad \text{(A.13)}
\end{aligned}$$

where $\Delta\Sigma_{ii,n} = \big[\tilde{\Sigma}_{ii,n}^R\big(\hat{\lambda}(c);c\big) - \Sigma_{ii}^R(c)\big]$ for $i = 1,2,3$. Consider first:

$$\|\Delta\Sigma_{11,n}\| = \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)>0\}} - \sum_{j=1}^{k} P_j(\Sigma)g(d_j;c)1_{\{(d_j-c)>0\}}\| \,. \qquad \text{(A.14)}$$

By adding and substracting simultaneously, $\sum_{j=1}^{k} P_{I_j}(\Sigma_n)g(d_j;c)1_{\{(d_j-c)>0\}}$ we have:

$$\begin{aligned}
\|\Delta\Sigma_{11,n}\| &= \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)g(d_j;c)1_{\{(d_j-c)>0\}} - \sum_{j=1}^{k} P_{I_j}(\Sigma_n)g(d_j;c)1_{\{(d_j-c)>0\}} \\
&+ \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)>0\}} - \sum_{j=1}^{k} P_j(\Sigma)g(d_j;c)1_{\{(d_j-c)>0\}}\| \\
&= \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)g(d_j;c)1_{\{(d_j-c)>0\}} - \sum_{j=1}^{k} P_j(\Sigma)g(d_j;c)1_{\{(d_j-c)>0\}} \\
&\quad + \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)>0\}} - g(d_j;c)1_{\{(d_j-c)>0\}}\big]\|
\end{aligned}$$

$$\begin{aligned}
\|\Delta\Sigma_{11,n}\| &= \|\sum_{j=1}^{k}\big[P_{I_j}(\Sigma_n) - P_j(\Sigma)\big]g(d_j;c)1_{\{(d_j-c)>0\}} \\
&\quad + \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)>0\}} - g(d_j;c)1_{\{(d_j-c)>0\}}\big]\| \\
&\leq \|\sum_{j=1}^{k}\big[P_{I_j}(\Sigma_n) - P_j(\Sigma)\big]g(d_j;c)1_{\{(d_j-c)>0\}}\| \\
&\quad + \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)>0\}} - g(d_j;c)1_{\{(d_j-c)>0\}}\big]\|
\end{aligned}$$

$$\|\Delta\Sigma_{11,n}\| \leq \sum_{j=1}^{k}\|P_{I_j}(\Sigma_n) - P_j(\Sigma)\||g(d_j;c)1_{\{(d_j-c)>0\}}|$$

$$+\sum_{j=1}^{k}\|P_{I_j}(\Sigma_n)\||\frac{1}{m(d_j)}\sum_{i\in I_j}\big[g(\hat\lambda_i(c);c)1_{\{(\hat\lambda_i(c)-c)>0\}}-g(d_j;c)1_{\{(d_j-c)>0\}}\big]| \tag{A.15}$$

$$\|\Delta\Sigma_{11,n}\|\leq\sum_{j=1}^{k}\|P_{I_j}(\Sigma_n)-P_j(\Sigma)\||g(d_j;c)1_{\{(d_j-c)>0\}}|$$
$$+\sum_{j=1}^{k}\|P_{I_j}(\Sigma_n)\|\tfrac{1}{m(d_j)}\sum_{i\in I_j}|g(\hat\lambda_i(c);c)1_{\{(\hat\lambda_i(c)-c)>0\}}-g(d_j;c)1_{\{(d_j-c)>0\}}|\ .$$

$\forall i\in I_j=\{i\in I:\ \lambda_i=d_j\}$, for the j indices such that $d_j\neq c$, we have by (A.10) and (A.11)

$$\forall\,\epsilon>0,\ P\big[|\hat\lambda_i(c)-d_j|\leq\epsilon\big]=P\big[b_n|\hat\lambda_i-c|>\nu e_n\big]\underset{n\to\infty}{\to}1\text{ if, }d_j\neq c \tag{A.16}$$

and $\forall\,\epsilon>0$,

$$P\big[|1_{\{(\hat\lambda_i(c)-c)>0\}}-1_{\{(d_j-c)>0\}}|\leq\epsilon\big]=P\big\{s[\hat\lambda_i(c)-c]=s[\lambda_i-c]\big\}\underset{n\to\infty}{\to}1\ . \tag{A.17}$$

We can write the quantity $|g(\hat\lambda_i(c);c)1_{\{(\hat\lambda_i(c)-c)>0\}}-g(d_j;c)1_{\{(d_j-c)>0\}}|=|\Delta_g|$. Also

$$\begin{aligned}|\Delta_g| &= |\big[g(\hat\lambda_i(c);c)-g(d_j;c)+g(d_j;c)\big]1_{\{(\hat\lambda_i(c)-c)>0\}}-g(d_j;c)1_{\{(d_j-c)>0\}}|\\ &= |\big[g(\hat\lambda_i(c);c)-g(d_j;c)\big]1_{\{(\hat\lambda_i(c)-c)>0\}}+g(d_j;c)\big[1_{\{(\hat\lambda_i(c)-c)>0\}}-1_{\{(d_j-c)>0\}}\big]|\\ &\leq |g(\hat\lambda_i(c);c)-g(d_j;c)|1_{\{(\hat\lambda_i(c)-c)>0\}}+g(d_j;c)|1_{\{(\hat\lambda_i(c)-c)>0\}}-1_{\{(d_j-c)>0\}}|\end{aligned} \tag{A.18}$$

By Property **??**i), $\forall\ i\in I_j:\hat\lambda_i(c)\overset{p}{\to}d_j$, $d_j\neq c$, and $g\in\mathcal{G}_c$ is such that g is continuous a.e., except possibly at c, hence $g(\hat\lambda_i(c);c)\overset{p}{\to}g(d_j;c)$. As $1_{\{(\hat\lambda_i(c)-c)>0\}}=O_p(1)$, we have $|g(\hat\lambda_i(c);c)-g(d_j;c)|1_{\{(\hat\lambda_i(c)-c)>0\}}\overset{p}{\to}0$. By equation (A.17), $|1_{\{(\hat\lambda_i(c)-c)>0\}}-1_{\{(d_j-c)>0\}}|\overset{p}{\to}0$ and $g(d_j;c)=O(1)$. Hence, $|\Delta_g|\overset{p}{\to}0\ \ \forall\ i\in I_j$, and the j's are such that $d_j\neq c$. Besides, the projection operator $P_{I_j}(\Sigma_n)=O_p(1)$, and $\underset{n\to\infty}{\text{plim}}P_{I_j}(\Sigma_n)=P_j(\Sigma)$ by Lemma **6.4** ii). Hence, we have $\underset{n\to\infty}{\text{plim}}\big[\|\Delta\Sigma_{11,n}\|>\epsilon\big]=0$. Note that $g(c;c)=\frac{1}{\hat m(c)}\hat m(c)g(c;c)=\frac{1}{\hat m(c)}\sum_{i\in\hat I(c)}g(c;c)$. Thus for the second component, we have:

$$\|\Delta\Sigma_{22,n}\| = \|P_{\hat I(c)}(\Sigma_n)\frac{1}{\hat m(c)}\sum_{i\in\hat I(c)}g(c;c)1_{\{\hat\lambda_i(c)=c\}}-g(c;c)1_{\{d_j=c\}}P_{j(c)}(\Sigma)\|\ .$$

Similarly to $\|\Delta\Sigma_{11,n}\|$, we add and substract $P_{\hat I(c)}(\Sigma_n)g(c;c)1_{\{d_j=c\}}$ and by gathering the terms together, we get:

$$\begin{aligned}\|\Delta\Sigma_{22,n}\| &= \|\big[P_{\hat I(c)}(\Sigma_n)-P_{j(c)}(\Sigma)\big]g(c;c)1_{\{d_j=c\}}+P_{\hat I(c)}(\Sigma_n)\frac{1}{\hat m(c)}\sum_{i\in\hat I(c)}g(c;c)\big[1_{\{\hat\lambda_i(c)=c\}}-1_{\{d_j=c\}}\big]\|\\ &\leq \|\big[P_{\hat I(c)}(\Sigma_n)-P_{j(c)}(\Sigma)\big]g(c;c)1_{\{d_j=c\}}\|+\|P_{\hat I(c)}(\Sigma_n)\frac{1}{\hat m(c)}\sum_{i\in\hat I(c)}g(c;c)\big[1_{\{\hat\lambda_i(c)=c\}}-1_{\{d_j=c\}}\big]\|\\ &= \|\big[P_{\hat I(c)}(\Sigma_n)-P_{I(c)}(\Sigma_n)+P_{I(c)}(\Sigma_n)-P_{j(c)}(\Sigma)\big]g(c;c)1_{\{d_j=c\}}\|\\ &+ \|P_{\hat I(c)}(\Sigma_n)\||\frac{1}{\hat m(c)}\sum_{i\in\hat I(c)}g(c;c)\big[1_{\{\hat\lambda_i(c)=c\}}-1_{\{d_j=c\}}\big]|\end{aligned}$$

$$\leq \|P_{\hat{I}(c)}(\Sigma_n) - P_{I(c)}(\Sigma_n)\| |g(c;c)1_{\{d_j=c\}}| + \|P_{I(c)}(\Sigma_n) - P_{j(c)}(\Sigma)\| |g(c;c)1_{\{d_j=c\}}|$$
$$+ \|P_{\hat{I}(c)}(\Sigma_n)\| |\frac{1}{\hat{m}(c)} \sum_{i\in \hat{I}(c)} g(c;c)\left[1_{\{\hat{\lambda}_i(c)=c\}} - 1_{\{d_j=c\}}\right]|$$

If $\lambda_i = d_j = c$, by equation (A.8), $P\left[\hat{\lambda}_i(c) = c\right] = P\left[b_n|\hat{\lambda}_i - \lambda_i| \leq \nu e_n\right] \underset{n\to\infty}{\to} 1$. By equation (9.4), $P\left[\hat{I}(c) = I(c)\right] = P\left[\bigcap_{i\in I(c)} \left\{\hat{\lambda}_i(c) = c\right\}\right] \to 1$. Hence, by (A.9) we have:

$$|1_{\{\hat{\lambda}_i(c)=c\}} - 1_{\{d_j=c\}}| \underset{n\to\infty}{\overset{p}{\to}} 0 \ \forall\, i \in \hat{I}(c) .$$

Moreover, $P\left[\hat{I}(c) = I(c)\right] \to 1$ implies $\underset{n\to\infty}{\text{plim}} P_{\hat{I}(c)}(\Sigma_n) = \underset{n\to\infty}{\text{plim}} P_{I(c)}(\Sigma_n) = P_{j(c)}(\Sigma)$ by Lemma **6.4** ii). Besides, $P_{\hat{I}(c)}(\Sigma_n) = O_p(1)$. Hence, $\underset{n\to\infty}{\text{plim}}\left[\|\Delta\Sigma_{22,n}\| > \epsilon\right] = 0$. Finally, the proof of $\underset{n\to\infty}{\text{plim}}\left[\|\Delta\Sigma_{33,n}\| > \epsilon\right] = 0$, with

$$\|\Delta\Sigma_{33,n}\| = \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i(c);c)1_{\{(\hat{\lambda}_i(c)-c)<0\}} - \sum_{j=1}^{k} P_j(\Sigma)g(d_j;c)1_{\{(d_j-c)<0\}}\|$$

is similar to $\Delta\Sigma_{11,n}$. Also, the result follows.

□

Proof of Proposition 10.1 We need to show that $lim_{n\to\infty} Pr\left[\|\Sigma_n^R(c_n) - \Sigma^R(0)\| > \epsilon\right] = 0$ for every $\epsilon > 0$. Let r denote the rank of the matrix of interest Σ. 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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_i \geq \ldots \geq \lambda_q \geq 0$, and $J = \{1, 2, \ldots, k\}$ the subset of I corresponding to the indices of the distinct eigenvalues of Σ: $d_1 > d_2 > \ldots > d_j > \ldots > d_k$, 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 there exist some eigenvalues $\lambda_i = 0$, then $d_k = 0$. Let $P(d_j)$ represent the eigenprojection operator projecting onto the eigenspace $\mathcal{V}(d_j)$ associated with d_j.

First let us show that

$$\underset{n\to\infty}{lim} Pr[\sup_{i\in I_j}|g(\hat{\lambda}_i; c_n) - g(d_j; 0)| > \epsilon] = 0 \ \ \forall\, \epsilon > 0 \tag{A.19}$$

as it will be used extensively throughout the proof. By Lemma **6.4** i), we have for all $i \in I_j$, $\hat{\lambda}_i \overset{\text{p}}{\to} d_j$. Besides, as $c_n \underset{n\to\infty}{\to} 0$, we have

$$Pr\left[|\hat{\lambda}_i - d_j| > c_n\right] = Pr\left[|b_n(\hat{\lambda}_i - d_j)| > b_n c_n\right] \underset{n\to\infty}{\to} 0 \tag{A.20}$$

since $b_n c_n \to \infty$ slower than b_n does, and $b_n\left(\hat{\lambda}_i - d_j\right)$ converges in distribution by Lemma **6.6**. As $\hat{\lambda}_i \overset{\text{p}}{\to} d_j, \ \ \forall\, i \in I_j$ and g is continuous a.e., then $\underset{n\to\infty}{lim} Pr[\sup_{i\in I_j}|g(\hat{\lambda}_i; c_n) - g(d_j; 0)| > \epsilon] = 0, \forall\, \epsilon > 0$. As j was taken arbitrary, it holds for any j.

Consider first the case where the limiting matrix Σ has full rank, *i.e.* $\text{rank}(\Sigma) = r = q$. For all $j \in J : \ d_j > 0$, since $r = q$, then by Lemma **6.4** i) and ii), and the continuity of g a.e., we have:

$$g(\hat{\lambda}_i; c_n) \overset{p}{\to} g(d_j; 0) \ and \ P_{I_j}(\Sigma_n) \overset{p}{\to} P_j(\Sigma) .$$

Since

$$g(d_j;0) = \frac{1}{m(d_j)} \times m(d_j) g(d_j;0) = \frac{1}{m(d_j)} g(d_j;0) \underbrace{\sum_{i \in I_j} 1}_{m(d_j)} = \frac{1}{m(d_j)} \sum_{i \in I_j} g(d_j;0) \ ,$$

we have:

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

since $P_{I_j}(\Sigma_n) \xrightarrow{p} P_j(\Sigma)$ and $|g(\hat{\lambda}_i; c_n) - g(d_j;0)| \xrightarrow{p} 0$ by (A.19).

Second, consider the case where $d_1 = 0$ with multiplicity $m(0) = q$. In this case, $\Sigma_n \xrightarrow{p} \Sigma = 0$, *i.e.* Σ_n converges to a zero matrix so that the range of the mapping A_Σ corresponding to Σ is $\mathcal{R}(A_\Sigma) = \{0\}$ and its nullspace is $\mathcal{N}(A_\Sigma) = \mathbb{R}^q$. Let $P_1(\Sigma) = P(d_1)(\Sigma)$ denote the eigenprojection operator of Σ associated with its zero eigenvalue which projects onto the corresponding eigenspace $\mathcal{V}(0)$. By Lemma **6.4** i) and ii), and the continuity of g a.e., we have:

$$g(\hat{\lambda}_i; c_n) \xrightarrow{p} g(d_1;0) = g(0;0) \ , \forall \ i \in I_1$$

$$P_{I_1}(\Sigma_n) \xrightarrow{p} P_1(\Sigma) \ ,$$

hence

$$\begin{aligned}
\Sigma_n^R(c_n) &= P_{I_1}(\Sigma_n) \frac{1}{m(d_1)} \sum_{i \in I_1} g(\hat{\lambda}_i; c_n) = P_{I_1}(\Sigma_n) \big[g(0;0) - g(0;0) + \frac{1}{m(0))} \sum_{i \in I_1} g(\hat{\lambda}_i; c_n) \big] \\
&= P_{I_1}(\Sigma_n) g(0;0) + P_{I_1}(\Sigma_n) \frac{1}{m(0)} \sum_{i \in I_1} \big[g(\hat{\lambda}_i; c_n) - g(0;0) \big] \\
&\xrightarrow{p} g(0;0) P_1(\Sigma) = \Sigma^R(0) \ ,
\end{aligned} \tag{A.21}$$

since $P_{I_1}(\Sigma_n) \xrightarrow{p} P_1(\Sigma)$, $P_{I_1}(\Sigma_n) = O_p(1)$ and $|g(\hat{\lambda}_i; c_n) - g(0;0)| \xrightarrow{p} 0$, by (A.19).

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

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

$$
\begin{aligned}
&\leq \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i;c_n)-g(d_j;0)\right]\| + \|\sum_{j=1}^{k} g(d_j;0)\left[P_{I_j}(\Sigma_n)-P_j(\Sigma)\right]\| \\
&\leq \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i;c_n)-g(d_j;0)\right]\| \\
&\quad + \sum_{j=1}^{k} |g(d_j;0)|\|P_{I_j}(\Sigma_n)-P_j(\Sigma)\| \\
&\leq \|\sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j}\left[g(\hat{\lambda}_i;c_n)-g(d_j;0)\right]\| + \sum_{j=1}^{k} |g(d_j;0)|\|P_{I_j}(\Sigma_n)-P_j(\Sigma)\|
\end{aligned}
\tag{A.22}
$$

since $P_{I_j}(\Sigma_n) = O_p(1)$, $|g(\hat{\lambda}_i;c_n) - g(0;0)| \xrightarrow{p} 0$ by (A.19), $g(d_j;0) = O(1)$ and $\|P_{I_j}(\Sigma_n) - P_j(\Sigma)\| \xrightarrow{p} 0$, by Lemma **6.4** ii).

We can finally conclude that:

$$
\lim_{n\to\infty} Pr\left[\|\Sigma_n^R(c_n) - \Sigma^R(0)\| \geq \epsilon\right] = 0 \ .
$$

□

PROOF of Proposition 10.2

By Proposition **10.1**, we have $\Sigma_n^R(c_n) \xrightarrow{p} \Sigma^R(0)$. Then by Assumption **2.1**, $X_n \xrightarrow{\mathcal{L}} X$, hence

$$
X_n' \Sigma_n^R(c_n) X_n \xrightarrow{\mathcal{L}} X' \Sigma^R(0) X \ . \tag{A.23}
$$

Let us project

$$
W_n^R(c_n) = X_n' \Sigma_n^R(c_n) X_n \ ,
$$

where

$$
\Sigma_n^R(c_n) = \sum_{j=1}^{k} P_{I_j}(\Sigma_n)\frac{1}{m(d_j)}\sum_{i\in I_j} g(\hat{\lambda}_i, c_n)
$$

onto the two orthogonal eigenspaces, $\mathcal{V}_1(\Sigma_n)$ and $\mathcal{V}_2(\Sigma_n)$, such that:

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

where

$$
\begin{aligned}
W_{1n}^R(c_n) &= X_n' \Sigma_{11,n}^R(c_n) X_n \ , \\
W_{2n}^R(c_n) &= X_n' \Sigma_{22,n}^R(c_n) X_n \ ,
\end{aligned}
$$

with

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

$$= \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n) g(d_j;0) + \sum_{j=1}^{k_1} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i \in I_j} \big[g(\hat{\lambda}_i;c_n) - g(d_j;0)\big] \quad \text{(A.24)}$$

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 **6.4** i) and ii), and under the assumption that $g(.,(c_n))$ is continuous a.e., we have $P_{I_j}(\Sigma_n) \xrightarrow{p} P_j(\Sigma)$ and $\forall \epsilon > 0$, ϵ small, $\lim_{n\to\infty} Pr\big[\sup_{i \in I_j} |g(\hat{\lambda}_i;c_n) - g(d_j;0)| > \epsilon\big] = 0$, with $c_n \to 0$ by (A.19). Besides, we know that projection operators are bounded in probability. Hence,

$$\Sigma^R_{11,n}(c_n) \xrightarrow{p} \sum_{j=1}^{k_1} g(d_j;0) P_j(\Sigma) \equiv \Sigma^R_{11}(0) \ . \quad \text{(A.25)}$$

Therefore, we have:

$$W^R_{1n}(c_n) = X_n' \Sigma^R_{11,n}(c_n) X_n \xrightarrow{\mathcal{L}} X' \Sigma^R_{11}(0) X \equiv W^R_1(0) \ .$$

Similarly, we have

$$\Sigma^R_{22,n}(c_n) = \sum_{j=k_1+1}^{k} P_{I_j}(\Sigma_n) \frac{1}{m(d_j)} \sum_{i \in I_j} g(\hat{\lambda}_i, c_n) \xrightarrow{p} \sum_{j=k_1+1}^{k} g(d_j;0) P_j(\Sigma) \equiv \Sigma^R_{22}(0) \equiv 0$$

since $g(d_j;0) \equiv 0$ for all $j = k_1+1, \ldots, k$. As $X_n \xrightarrow{\mathcal{L}} X$, and $\Sigma^R_{22,n}(c_n) \xrightarrow{p} \Sigma^R_{22}(0) \equiv 0$, we have:

$$W^R_{2n}(c_n) = X_n' \Sigma^R_{22,n}(c_n) X_n \xrightarrow{\mathcal{L}} X' \Sigma^R_{22}(0) X \equiv 0 = W^R_2(0) \ .$$

□

PROOF of Corollary 10.3

Apply the results of Proposition **10.2** with

$$X_n = \sqrt{n}\big[\psi(\hat{\theta}_n) - \psi_0\big] \xrightarrow{\mathcal{L}} N\big[0, \Sigma\big] = X \ \text{, and } \ x_j = N\big[0, I_{m(d_j)}\big] \ .$$

$$\begin{aligned} W^R_1(0) &= X' \Sigma^R_{11}(0) X = X' \Big(\sum_{j=1}^{k_1} g(d_j;c) P_j(\Sigma)\Big) X = \sum_{j=1}^{k_1} g(d_j;c) X' P_j(\Sigma) X \\ &= \sum_{j=1}^{k_1} g(d_j;c) X' V(d_j) V(d_j)' X = \sum_{j=1}^{k_1} g(d_j;c) d_j x_j' x_j \\ &= \sum_{j=1}^{k_1} \frac{1}{d_j} d_j x_j' x_j = \sum_{j=1}^{k_1} x_j' x_j \ , \end{aligned}$$

where $x_j = N\big[0, I_{m(d_j)}\big]$. As $\sum_{j=1}^{k_1} m(d_j) = q_1$, $W^R_1(0) \sim \chi(q_1)$.

□

References

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

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.

Antoine, B. and Renault, E. (2010*a*), 'Efficient Minimum Distance Estimation with Multiple Rates of Convergence', *Journal of Econometrics, forthcoming* .

Antoine, B. and Renault, E. (2010*b*), Specification tests for strong identification. Simon Fraser University.

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.

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**.

Carrasco, M. (2007), A regularizing approach to the many instrument problem. CIREQ working paper, University of Montreal.

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, *in* J. Heckman and E. Leamer, eds, 'Handbook of econometrics', Vol. 6, pNH, Amsterdam.

Chen, X. and Ludvigson, S. (2009), 'Land of addicts ? An empirical investigation of habit-based asset pricing behavior', *Journal of Applied Econometrics* **24**, 1057–1093.

Cochrane, J. H. and Piazzesi, M. (2005), 'Bond risk premia', *American Economic Review* **95**, 138–160.

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.

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

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

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. 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.

Engl, H. W., Hanke, M. and Neubauer, A. (2000), *Regularization of inverse problems*, Kluwer Academic Publishers.

Fan, J., Fan, Y. and Lv, J. (2006), High Dimensional Covariance Matrix Estimation Using a Factor Model, Technical report, Princeton University.

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.

Hansen, L. P. and Jagannathan, R. (1991), 'Implications of security market data for models of dynamic economies', *Journal of Political Economy* **99**, 225–262.

Hansen, L. P. and Jagannathan, R. (1997), 'Assessing Specification Errors in Stochastic Discount Factor Models', *Journal of Finance* **52**, 557–590.

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

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

Kan, R. and Zhou, G. (2006), 'A new variance bound on the stochastic discount factor', *Journal of Business* **79**, 941–961.

Kato, T. (1966), *Perturbation theory for linear operators*, Springer.

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

Kress, R. (1999), *Linear integral equations*, Applied mathematical sciences, second edn, Springer, New York.

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.

Leeb, H. and Potscher, 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*, second edn, Springer.

Ludvigson, S. C. and Ng, S. (2009), 'Macro factors in bond risk premia', *Review of Financial Studies* **22**, 5027–5067.

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

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

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

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*, Engle, R. F., McFadden, D. L., North Holland, Amsterdam.

Peñaranda, F. and Sentana, E. (2008), Spanning tests in return and stochastic discount factor mean-variance frontiers: a unifying approach. CEMFI working paper.

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

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.

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

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

Valéry, P. (2005), Simulation-based inference and nonlinear canonical analysis in financial econometrics, PhD thesis, University of Montreal.