Hypothesis tests when rank conditions fail: a smooth regularization approach *

Jean-Marie Dufour †
McGill University

Pascale Valéry ‡
HEC Montréal

September 25, 2009

*We are gratefull to Marine Carrasco, Russell Davidson and Raymond Kan for helpful comments as well as seminar participants in Marcel Dagenais econometric seminars at Université de Montréal. 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 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

‡ 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 considers the issue of testing hypotheses in the presence of (possibly) singular covariance matrices. We focus on the case where the finite-sample covariance matrix may not be singular but converges to an asymptotic singular population matrix. The econometrician can face this type of situations when the derivative matrix of the restrictions has a lower rank only at the true value of the parameter, or situations where there may be an identification problem asymptotically that cannot be detected in finite samples. More generally, when the econometrician suspects any possible singularity issue, the 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 as proposed by Lütkepohl and Burda (1997). To face such difficulties, we introduce *regularized* inverses in such a way that the continuity property is preserved (at least for some of them). Our modified inverses can be viewed as a generalization of the one introduced by Lütkepohl and Burda (1997). Our contributions can be summarized as follows. *First*, we introduce a class of *regularized* inverses as opposed to the class of *generalized* inverses but shares common properties with the latter. Those regularized inverses involve a *(variance) regularization function* or VRF that can be defined as a perturbation function of the small eigenvalues that stabilizes the inverse. *Second*, those *regularized inverses* are used to build *regularized* test statistics (Wald-type, score-type and LR-type test statistics) based on singular (asymptotic) covariance matrix. *Third*, using spectral decomposition based tools the inverse can be decomposed into three components, one viewed as *regular* built on large eigenvalues while the other ones which involve the small eigenvalues may not be *regular*. From this decomposition, we identify a convenient decomposition for the test statistic itself that can be exploited to set distributional bounds. *Fourth*, following Kato (1966) and Tyler (1981), we propose to work with the *total eigenprojections* in order to circumvent the discontinuity and non-uniqueness features of eigenvectors. *Fifth*, the specification of the VRF allows us to define two classes of tests whether the threshold used to disentangle the small eigenvalues from the the others is fixed or varies with the sample size.

1. Introduction

This paper considers the issue of testing hypotheses in the presence of (possibly) singular covariance matrices. We focus on the case where the finite-sample covariance matrix may not be singular but converges to a singular population matrix. The econometrician can face this type of situations when the derivative matrix of the restrictions has a lower rank only at the true value of the parameter, or situations where there may be an identification problem asymptotically that cannot be detected in finite samples. This asymptotic nonidentification can arise in a two-stage least squares framework when the coefficient matrix on the instruments is local to zero and the coefficient matrix on the regressors is fixed in the reduced form equation. In this case $\hat{Y}$ and X are asymptotically multicollinear, leading to an asymptotically singular regressor moment matrix and to nearly unidentified parameters of the structural equation [see Staiger and Stock (1997, page 569). In such cases, the typical rank convergence condition for the estimator of a singular (covariance) matrix used by Andrews (1987) does not hold. This violation will affect the asymptotic distribution of the test statistics. Another situation where we can encounter an asymptotic singularity is with *superconsistent* estimators such as the OLS estimator for the time trend. Indeed, in this model the OLS estimator for the intercept is root-n consistent while the time trend estimator is $n^{3/2}$ consistent [see Hamilton (1994, chapter 16, page 460)]. In order to come up with a nondegenerate distribution, each component has to be scaled with its suitable convergence rate. However in practice, the econometrician does not know the presence of multiple convergence rates for the different components of the parameter vector [see also Dovonon and Renault (2009) for multiple convergence rates applied to a first-order unidentified GMM setup]. Therefore a general method that can handle such difficulties can reveal very helpful in empirical applications.

More generally, there exist two different types of singularity: the most usual one addressed by Andrews (1987) deals with finite-sample singular matrices whose rank converges almost surely towards the rank of its population limit. The other type of singularity stems from a finite-sample full-rank matrix that converges to an asymptotic singular (covariance) matrix but violates the typical rank convergence condition of Andrews (1987). When dealing with problems of the first type, 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 of the type proposed by Lütkepohl 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 of interest 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 which are strictly speaking *wrong*.

To face such difficulties, we introduce *modified* inverses in such a way that the continuity property is preserved (at least for some of them). Our modified inverses can be viewed as a generalization of the one introduced by Lütkepohl and Burda (1997) and Valéry (2005). Our contributions can be summarized as follows. *First*, we introduce a class of *regularized* inverses as opposed to the class of *generalized* inverses but shares common properties with the latter. Those regularized inverses involve a *(variance) regularization function* or VRF that can be defined as a perturbation function of the small eigenvalues that stabilizes the inverse. *Second*, those *regularized inverses* are used to build *regularized* test statistics (Wald-type, score-type and LR-type test statistics) based on singular (asymptotic) covariance matrix. *Third*, using spectral decomposition based tools the inverse can be decomposed into three components, a *regular* one built on large eigenvalues while the others involving small eigenvalues may not be *regular*. From this decomposition, we identify a convenient decomposition for the test statistic itself that can be exploited to set distributional bounds. Thus, when the regularized test statistic has a nonstandard distribution (a mixture of chi-square), the standard chi-square distribution can then be used as an upper bound to conduct an *asymptotically valid* test. *Fourth*, following Kato (1966) and Tyler (1981), we propose to work with the *total eigenprojection* in order to circumvent the discontinuity and non-uniqueness features of eigenvectors. A lemma given by Tyler (1981) states the continuity property for the *total eigenprojection* (the sum of the eigenprojections over a subset of the eigenvalues). We extend those results to the case of convergence in probability. *Fifth*, the specification of the VRF allows us to define two classes of tests whether the threshold used to disentangle the small eigenvalues from the the others is fixed or varies with the sample size.

More specifically, our *regularization* approach has *good properties* in many respects. *First*, the class of *regularized* inverses introduced in this paper shares some properties with the *generalized* inverses but with the main difference that they do solve for the instability property of g-inverses (for instance, the Moore-Penrose inverse is not continuous in the elements of the initial matrix) by setting constraints on the elements of the matrix (somewhat similar to the rank condition used by Andrews (1987).

Second, following Kato (1966) and Tyler (1981), we work with the *eigenprojection* in order to circumvent 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* (the sum of the eigenprojections over a subset of the eigenvalues). In this way, the important continuity property is preserved for eigenvalues and eigenprojections even though eigenvectors are *not* continuous. By combining this total eigenprojection technique with a specific design for the perturbation function of the eigenvalues, the regularized inverses remain continuous. As a result, test statistics can be defined as a linear transformation of those total eigenprojections and the distributional theory built from them not misleading. This continuity property enjoyed by the *regularized* inverses is an important contribution to the econometric literature.

Third, regularized inverses involve a *variance regularizing function* (or VRF) which consists in inflating the small eigenvalues that fall below a certain threshold so that their inverse is well defined whereas the large eigenvalues remain unchanged. The specification of the VRF allows ones to define two classes of tests whether the threshold used to disentangle the small eigenvalues from the the others is fixed or varies with the sample size. The *first class* of tests which exploits a fixed threshold introduces regularized test statistics (Wald-type, score-type and LR-type test) with a nonstandard distribution (a mixture of chi-square). However, this class admits the standard chi-square distribution (with the degree of freedom corresponding to the full rank case) as an upper bound. This upper bound is very useful to the extent that the usual critical points can be used to provide an *asymptotically valid* test. However the econometrician can do better from the viewpoint of power by simulating the test. Indeed, usual critical points yield conservative tests which can lead to a loss of power under the alternative. The *second class* of tests allows the threshold to vary with the sample size. When the threshold declines to zero at an appropriate rate, the regularized tests are asymptotically distributed as the chi-square distribution with a reduced degree of freedom (corresponding to the number of eigenvalues greater than zero) like in Lütkepohl and Burda (1997). However this second class of tests requires more information on the asymptotic behavior of the sample eigenvalues. In this respect, a general result established by Eaton and Tyler (1994) where they generalize some nonclassical results given by a wide strand of statistical and multivariate analysis literature provide us with the convergence speed of the sample eigenvalues (in the presence of eigenvalues with multiplicity greater than 1) needed for our distributional results [see, for example Anderson (1963), Anderson (1987) and Amemya (1986)]. Some of the regularizing functions used are our own design while others are commonly used in the literature on inverse problems; see Engl, Hanke and Neubauer (2000) and Kress (1999).

Fourth, another appealing property of our *regularization* approach is that it naturally entails a decomposition of the matrix into two or three components, one viewed as *regular* built on large eigenvalues while the other ones which involve the small eigenvalues may not be *regular*. Based on this covariance matrix decomposition, we identify a convenient decomposition for the test statistic itself.

Fifth, the tests so regularized do embed three possible cases. *Case 1* corresponds to the usual case when using sample covariance matrices which converge to a full rank limiting matrix. Regularizing the matrix when unnecessary will disappear asymptotically (at least for the varying threshold case). *Case 2* corresponds to using sample covariance matrices which converge to a singular limiting matrix under the condition that the rank of the matrix estimator does converge to the rank of the limiting population matrix. 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 Lütkepohl and Burda (1997). Finally *case 3* makes use of consistent sample covariance matrices which do not satisfy the rank condition of Andrews (1987). Under such a circumstance, a regularization is required to provide a valid test but at the cost of a *fully modified* asymptotic distribution that can be obtained by simulation.

Sixth, our regularized inverses may be viewed as a generalization of Lütkepohl and Burda (1997)'s results by allowing more flexibility in selecting the threshold c for separating the eigenvalues. More generally, the estimate of the perturbation function of the eigenvalues can be viewed as a Hodges' estimator applied to the eigenvalues. With an appropriate choice of the regularization (tuning) parameter, the Hodges' estimator is a consistent estimator not only for the "small" components of the true eigenvalues but also for the large ones. This is the so-called *superefficiency* property enjoyed by the Hodges' estimator [see LeCam (1953)]. From our decomposition, the generalized Wald tests defined in Andrews (1987) and Lütkepohl and Burda (1997) can be replicated by keeping only the first *regular* component of the test statistic and setting to zero the other *nonregular* components. In discarding so the latter, their technique may result in a loss of information. Revisiting their tests as we do may yield some gains in power.

Seventh, our *regularization* approach discards any re-parametrization of the initial parameters. Finding suitable transformations of the parameters that circumvent the singularity problems can reveal tricky in highly nonlinear models for econometricians. 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. At the same time, they use, as Lütkepohl and Burda (1997) do, a generalized inverse as weighting matrix in a GMM setup to reduce the number of moment conditions accordingly. In this way, the reduced moment conditions will locally identify a subset of the initial parameter vector. To some extent, Peñaranda and Sentana (2008) do impose some restrictions on the null space of the singular limiting matrix. By reducing the number of parameter of interest, this boils down to assuming that the strongest collinearity in the design is restricted to the covariates that have no influence on the response.

In the same spirit, Knight and Fu (2000) works 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 under this nearly singular design and find that the resulting estimators have a slower rate of convergence than the usual root-n convergence rate. In this setup, the regularization has the advantage of preserving the usual root-n convergence rate of the estimators [see for instance Carrasco and Florens (2000),Carrasco, Chernov, Florens and Ghysels (2007)].

Eighth, while our main concern is testing, some authors make use of the related spectral decomposition based-tools 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. By contrast, we focus on testing issues by proposing *regularized test statistics*.

Finally, whereas Moore (1977), Andrews (1987) and Lütkepohl and Burda (1997) did focus on generalized Wald tests, we extend the regularized tests beyond the sole Wald tests, by considering $C(\alpha)$-type and LR-type tests in a GMM setup that could accommodate for identification problems. But our *regularization* approach goes beyond the GMM framework. Provided that we can find a root-n asymptotically gaussian estimator (or of the restrictions), we can build a *regularized* Wald test, whatever its origin, and proceed.

The paper is organized as follows. In Section 2 we review the well-known methodology of usual Wald-type tests, $C(\alpha)$-type tests and LR-type tests in two different frameworks, a GMM framework together with a more general setup where the only requirement is to find an asymptotically gaussian estimator or restriction functions for which the delta method applies. Then, we will introduce the *regularized* inverses together with their properties in Section 3 followed by the *regularized* test statistics in Section 4. More specifically, a decomposition of the statistic is highlighted through the corresponding decomposition of the covariance matrix. In Section 5 we review and adapt some results on eigenprojections to derive the convergence results for the regularized inverses. In particular, we highlight some (non)uniqueness and (dis)continuity properties related to eigenvectors of a given matrix and provide a solution based on eigenprojection techniques to surmount such difficulties. In Section 6, we state new asymptotic distributional results for the regularized statistics involving continuous regularizing functions. Finally an application to causality testing is provided in Section 7 followed by simulation results in Section 8. Concluding remarks follow while the proofs are gathered in the appendix.

2. Framework and test statistics

2.1. Basic assumptions

We consider an inference problem about a parameter of interest $\theta \in \Omega \in \mathbb{R}^p$. In order to identify the true value θ_0 of θ, we consider an $m \times 1$ vector score-type function $D_n(\theta, Y_n)$ where $Y_n = [y_1, y_2, \cdots, y_n]'$ is an $n \times k$ stochastic matrix satisfying the following assumptions.

Assumption 2.1 $\sqrt{n} D_n(\theta, Y_n) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} D_\infty(\theta)$, *where* $D_\infty(\theta)$ *is a random variable.*

Assumption 2.2 (Identification) $E_{\theta_0}\big[D_\infty(\theta)\big] = 0$ iff $\theta = \theta_0$, *where the expectation is taken w.r.t. the true probability distribution.*

Assumption 2.3 $D_\infty(\theta_0) \sim N\big[0, I(\theta_0)\big]$, *where the information matrix* $I(\theta_0)$ *is* $m \times m$.

Thus $\sqrt{n} D_n(\theta_0, Y_n) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N\big[0, I(\theta_0)\big]$. Assumptions **2.2** and **2.3** lead to a GMM estimator through the population orthogonality condition

$$E_{\theta_0}\big[D_\infty(\theta_0)\big] = 0 \ .$$

Typically, $D_n(\theta, Y_n)$ has the form

$$D_n(\theta, Y_n) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta, y_t)$$

where $E_{\theta_0}[h_t(\theta_0, y_t)] = 0$. Usually, Assumption **2.2** requires that $m \geq p$. A GMM estimator is obtained by minimizing w.r.t. θ an objective function of the form

$$M_n(\theta, W_n) = D_n(\theta, Y_n)' W_n D_n(\theta, Y_n) \tag{2.1}$$

where W_n is (usually) a symmetric positive definite matrix. Moreover, a standard regularity assumption which is usually made in the GMM framework is the following.

Assumption 2.4 (Nonsingular weighting matrix) *The symmetric positive definite (possibly random)* $m \times m$ *matrix* W_n *satisfies the limiting condition:*

$$\underset{n\to\infty}{\text{plim}}\, W_n = W_0 \ \textit{with} \ \det(W_0) \neq 0 \ . \tag{2.2}$$

Later on, we shall relax the above nonsingularity condition. This condition matters especially for the LR-type tests built from a GMM objective function. In this paper, we consider testing problems in the presence of singular covariance matrices. Under such circumstances, our purpose is to propose solutions which may hold irrespective of such nonsingularity assumptions. Let $\hat{\theta}_n$ be the unrestricted estimator obtained by minimizing the objective function $M_n(\theta, W_n)$, and $\hat{\theta}_n^0$ the corresponding constrained estimator under the null hypothesis:

$$H_0(\psi_0) : \psi(\theta) = \psi_0 \tag{2.3}$$

where $\psi : \mathbb{R}^p \to \mathbb{R}^q$.

Usual distributional theory critically hinges on regularity conditions such as the assumptions reviewed in this section. These assumptions define the standard regular framework upon which is built the whole usual distributional theory. We shall relax later on theses assumptions, especially those referring to rank conditions usually made on the restriction matrix or the covariance matrices of interest.

Assumption 2.5 (Local regularity) *There is a (nonempty) open neighborhood N_1 of θ_0 such that the $q \times 1$ vector function $\psi(\theta)$ is continuously differentiable with* $\text{rank}[P(\theta)] = q$ *for all $\theta \in N_1$, where*

$$P(\theta) = \partial\psi/\partial\theta' \, . \tag{2.4}$$

Let

$$H(\theta_0, Y_n) = \frac{\partial}{\partial\theta'} D_n(\theta_0, Y_n) \underset{n\to\infty}{\overset{\text{p}}{\to}} J(\theta_0) \, . \tag{2.5}$$

In the regular framework, the $m \times m$ matrix $I(\theta_0)$ and the $m \times p$ matrix $J(\theta_0)$ are (usually) assumed to be full-column rank matrices. If

$$D_n(\theta, Y_n) = \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; y_t) \, ,$$

$I(\theta_0)$ and $J(\theta_0)$ may be estimated using heteroskedastic, autocorrelation consistent (HAC) covariance matrix estimator:

$$\hat{I}_n(\theta) = \sum_{j=-(n-1)}^{n-1} \kappa(j/K_n)\hat{\Gamma}_n(j, \theta) \tag{2.6}$$

where $\kappa(\cdot)$ is a kernel function, K_n is a sample-size dependent bandwidth parameter,

$$\hat{\Gamma}_n(j, \theta) = \begin{cases} \frac{1}{n}\sum_{t=-(j-1)}^{n} h_t(\theta; y_t) h_{t-j}(\theta; y_{t-j})' & \text{if } j \geq 0, \\ \frac{1}{n}\sum_{t=-(j-1)}^{n} h_{t+j}(\theta; y_{t+j}) h_t(\theta; y_t)' & \text{if } j < 0, \end{cases} \tag{2.7}$$

and

$$\hat{J}_n(\theta) = \frac{\partial D_n}{\partial\theta'}(\theta, Y_n) = H(\theta, Y_n) \, , \tag{2.8}$$

with θ replaced by a consistent estimator.

2.2. Score and LR-type test statistics

We first focus on two types of test statistics, namely the score-type and the LR-type statistics. We will now briefly review the theory of $C(\alpha)$ tests which encompass LM-type tests. The $C(\alpha)$ test statistic is well-defined under the following assumptions; for a more comprehensive presentation of $C(\alpha)$ tests, see Dufour and Trognon (2001) and the references therein. For the statistic to be well-defined, we need consistent estimators as specified in both assumptions below.

Assumption 2.6 (Root-n consistent restricted estimator) *$\tilde{\theta}_n^0$ is a consistent restricted estimate of θ_0 such that $\psi(\tilde{\theta}_n^0) = 0$ and $\sqrt{n}(\tilde{\theta}_n^0 - \theta_0)$ is asymptotically bounded in probability.*

Assumption 2.7 (Consistent estimators)

$$\underset{n\to\infty}{\text{plim}}\ \hat{I}_n(\tilde{\theta}_n^0) = I(\theta_0), \quad \underset{n\to\infty}{\text{plim}}\ P(\tilde{\theta}_n^0) = P(\theta_0), \quad \underset{n\to\infty}{\text{plim}}\ \hat{J}_n(\tilde{\theta}_n^0) = J(\theta_0)\ .$$

For the sake of notational simplicity, let $\tilde{P}_0 = P(\tilde{\theta}_n^0)$, $\tilde{I}_0 = \hat{I}_n(\tilde{\theta}_n^0)$ and $\tilde{J}_0 = \hat{J}_n(\tilde{\theta}_n^0)$. The $C(\alpha)$ test statistic is then defined as

$$PC(\tilde{\theta}_n^0; \psi) = nD_n(\tilde{\theta}_n^0)'\tilde{W}_0 D_n(\tilde{\theta}_n^0) \tag{2.9}$$

where

$$\tilde{W}_0 \equiv \tilde{I}_0^{-1}\tilde{J}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{P}_0'\big[\tilde{P}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{P}_0'\big]^{-1}\tilde{P}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{J}_0'\tilde{I}_0^{-1}\ .$$

Under standard regularity conditions, the $C(\alpha)$ statistic $PC(\tilde{\theta}_n^0; \psi)$ has an asymptotic null distribution without nuisance parameters, namely $\chi^2(q)$.

We will also investigate a "LR-type" test initially suggested by Newey and West (1987):

$$LR(\psi) = n[M_n(\hat{\theta}_n^0, \bar{I}_n^{-1}) - M_n(\hat{\theta}_n, \bar{I}_n^{-1})] \tag{2.10}$$

where

$$M_n(\theta, \bar{I}_n^{-1}) = D_n(\theta)'\bar{I}_n^{-1}D_n(\theta)\ , \tag{2.11}$$

and $W_n = \bar{I}_n^{-1}$, where $\bar{I}_n = \hat{I}_n(\bar{\theta}_n)$, with $\bar{\theta}_n$ a consistent preliminary estimator of θ. Under standard regularity conditions, this test statistic is asymptotically pivotal with a $\chi^2(q)$ distribution.

2.3. Wald-type tests

More generally, there is no need to restrict our attention to a GMM estimator as the one introduced in equation (2.1). Indeed, a Wald-type test can be built from any root-n consistent estimator satisfying the assumption below.

Assumption 2.8 (Estimator $\sqrt{n}$ convergence) *$\hat{\theta}_n$ is a consistent estimator of θ_0 such that:*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \Sigma_{\theta_0})\ . \tag{2.12}$$

In order to conduct a Wald test, we shall restrict ourselves to restriction functions that are asymptotically gaussian. So it is important to note that Assumption **2.8** implies Assumption **2.9** under the condition that the Delta method applies, discarding degenerate distributions.

Assumption 2.9 ($\sqrt{n}$ convergence of the restrictions)

$$\sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta_0)) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \Sigma_\psi(\theta_0)) \; . \tag{2.13}$$

The usual Wald statistic for testing $H_0(\psi_0): \; \psi(\theta_0) = \psi_0$ is:

$$W_n(\psi_0) = n[\psi(\hat{\theta}_n) - \psi_0]' \hat{\Sigma}_\psi^{-1} [\psi(\hat{\theta}_n) - \psi_0] \tag{2.14}$$

provided the inverse exists, where

$$\hat{\Sigma}_\psi = P(\hat{\theta}_n) \hat{\Sigma}_\theta P(\hat{\theta}_n)' \; ,$$

with $\hat{\Sigma}_\theta$ being a consistent estimator of Σ_{θ_0} . If $H_0(\psi_0)$ holds and the $q \times q$-matrix

$$\Sigma_\psi(\theta) = P(\theta) \Sigma_\theta P(\theta)'$$

is nonsingular [*i.e.* $|\Sigma_\psi(\theta)| \neq 0$ for $\theta \in N_1$, where $|\cdot|$ stands for the determinant], then $W_n(\psi_0)$ has an asymptotic $\chi^2(q)$ distribution. This is not necessarily true, however, if $\Sigma_\psi(\theta)$ is singular. The latter may hold, if Σ_θ is singular or if $P(\theta)$ does not have full row rank q. In this case, $\Sigma_\psi(\theta)$ does not admit a usual inverse but can still be inverted by means of a generalized inverse or a *regularized* inverse. In Section 4, we introduce the family of *regularized* test statistics based on *regularized* inverses of the covariance matrix as a way to deal with such difficulties. First, let us introduce the class of *regularized* inverses as opposed to the class of *generalized* inverses.

3. 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 which consists 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 $\hat{\Sigma}$ be a consistent estimator of Σ with eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_q$ and $\hat{V}$ an orthogonal matrix such that $\hat{\Sigma} = \hat{V} \hat{\Lambda} \hat{V}'$ where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_q)$. For $c > 0$, we denote $q(\Sigma, c)$ the number of eigenvalues λ such that $\lambda > c$ and $q(\hat{\Sigma}, c)$ the number of eigenvalues $\hat{\lambda}$ such that $\hat{\lambda} > c$.

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

$$\Sigma^R = V \Lambda^{\dagger} V' \tag{3.1}$$

where

$$\Lambda^{\dagger} = \Lambda^{\dagger}[\bar{\lambda}; c] = \begin{pmatrix} g[\lambda_1; c] & & 0 \\ & \ddots & \\ 0 & & g[\lambda_q; c] \end{pmatrix} , \tag{3.2}$$

$g[\lambda; c] \geq 0$, and $g[\lambda; c]$ is non-increasing in λ. 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).

We now introduce a partition of the matrix $\Lambda^{\dagger}$ into three submatrices where c represents a threshold which may depend on the sample size and possibly on the sample itself, *i.e.* $c = c[n, Y_n]$:

$$\Lambda^{\dagger} = \begin{pmatrix} \Lambda_1^{\dagger}[\bar{\lambda}; c] & 0 & 0 \\ 0 & \Lambda_2^{\dagger}[\bar{\lambda}; c] & 0 \\ 0 & 0 & \Lambda_3^{\dagger}[\bar{\lambda}; c] \end{pmatrix} . \tag{3.3}$$

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 $q_3 = q - q_1 - q_2$. In other words, q_1 represents the number of eigenvalues $\lambda_i > c$. The three above components correspond to:

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

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

a $q_2 \times q_2$ matrix with $q_2 = m(c)$ denoting the multiplicity of the eigenvalue $\lambda = c$ (if any),

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

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.

Let $V_1(c)$ 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 of 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 $V_2(c)$ whose columns are the eigenvectors associated with the eigenvalues $\lambda = c$ and for the $q \times q_3$ matrix $V_3(c)$ whose columns are the eigenvectors associated with the eigenvalues $\lambda < c$. $\Lambda_1^{\dagger}[\hat{\lambda}; c]$, $\Lambda_2^{\dagger}[\hat{\lambda}; c]$,$\Lambda_3^{\dagger}[\hat{\lambda}; c]$, $\hat{V}_1(c)$, $\hat{V}_2(c)$ and $\hat{V}_3(c)$ denote the corresponding quantities based on the estimate $\hat{\Sigma}$, with $\dim \Lambda_1[\hat{\lambda}; c] = \hat{q}_1 = \text{card}\{i \in I : \hat{\lambda}_i > c\}$, $\dim \Lambda_2[\hat{\lambda}; c] = \hat{q}_2 = \text{card}\{i \in I : \hat{\lambda}_i = c\}$, $\dim \Lambda_3[\hat{\lambda}; c] = \hat{q}_3 = \text{card}\{i \in I : \hat{\lambda}_i < c\}$, respectively.

Using decomposition (3.3), the *regularized* inverse can be decomposed as follows:

$$\Sigma^R \;\; = \;\; V \Lambda^{\dagger} V'$$

$$
\begin{aligned}
&= [V_1 \ V_2 \ V_3] \begin{pmatrix} \Lambda_1^{\dagger}[\bar{\lambda};c] & 0 & 0 \\ 0 & \Lambda_2^{\dagger}[\bar{\lambda};c] & 0 \\ 0 & 0 & \Lambda_3^{\dagger}[\bar{\lambda};c] \end{pmatrix} \begin{bmatrix} V_1' \\ V_2' \\ V_3' \end{bmatrix} \\
&= V_1\Lambda_1^{\dagger}V_1' + V_2\Lambda_2^{\dagger}V_2' + V_3\Lambda_3^{\dagger}V_3' \ , \qquad (3.4)
\end{aligned}
$$

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

$$
\Sigma = V\Lambda V' = V_1\Lambda_1 V_1' + V_2\Lambda_2 V_2' + V_3\Lambda_3 V_3' \ . \qquad (3.5)
$$

Let Id denote the identity matrix. Let us establish some interesting properties for the regularized inverses.

Property 1 *Let Σ^R be the regularized inverse of Σ. If $g[\lambda;c]$ is non-increasing in λ, then we have*

i) $\Sigma\Sigma^R \ \leq \ Id$;

ii) $\Sigma\Sigma^R\Sigma \ \leq \ \Sigma$;

iii) $\left(\Sigma^R\right)^{-1} \ \geq \ \Sigma$;

iv) $\mathrm{rank}(\Sigma^R) \geq \mathrm{rank}(\Sigma)$

It is important to notice that any transformation of the original matrix Σ that diminishes the inverse Σ^R satisfies relation iii). It is worth noting that the generalized inverses share properties i) and iv) with the *regularized* inverses. By contrast, property ii) appears as a dominance relation for the *regularized* inverse as opposed to g-inverses for which it exactly holds [see Rao and Mitra (1971, Lemmas 2.2.1 and 2.2.3 page 20-21)]. Contrary to g-inverses, regularized inverses do solve for their instability property by setting some constraints on the elements of the matrix (somewhat similar to the rank condition of Andrews (1987)) by means of the regularization function. The VRF perturbs the small eigenvalues in order to stabilize their inverse, preventing them from exploding.

4. Regularized test statistics

In this section, we introduce the concept of regularized tests which embed three possible cases. *Case 1* corresponds to the usual case when using a sample covariance matrix which converges to a full rank limiting covariance matrix. *Case 2* corresponds to using a sample covariance matrix which converges to a singular limiting matrix under the condition that the rank of the matrix estimator does converge to the rank of the limiting population matrix. 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 Lütkepohl and Burda (1997). Finally *case 3* makes use of a consistent sample covariance matrix which does not satisfy the rank condition pined down by Andrews (1987). Under such a circumstance, a regularization is required to provide a valid test but at the cost of a *fully modified* asymptotic distribution. This is the path investigated in this paper. We consider situations when the rank of the covariance matrix is incomplete asymptotically but not in finite samples. We can face this type of situations when the derivative matrix of the restrictions has a lower rank only at the true value of the parameter, or situations where there is an identification problem asymptotically that cannot be detected in finite samples. A third situation where we can encounter such an asymptotic singularity is with *superconsistent* estimators such that the OLS estimator for the time trend. Indeed, in this model the OLS estimator for the intercept is root-n consistent while the time trend estimator is $n^{3/2}$ consistent. In order to come up with a nondegenerate distribution, each estimator has to be scaled with its suitable convergence rate. This approach based on multiple convergence rates to circumvent the difficulties raised by a singular Jacobian matrix of moment conditions is used by Dovonon and Renault (2009) to derive the asymptotic distribution (a mixture of χ^2) of the GMM overidentification J test. However in practice, the econometrician does not know the presence of multiple convergence rates for the different components of the parameters. Therefore a general method that can handle such difficulties can reveal very helpful in empirical applications. To the best of our knowledge, such cases have not been investigated in the literature. Under Assumption **4.1** and more generally for any matrix with an incomplete rank – either asymptotically or in finite samples– the usual inverse is not defined and requires either using a generalized inverse or transforming it such that it becomes invertible. Based on spectral decomposition tools, we propose to regularize the inverse of the covariance matrix by means of an eigenvalue perturbation function. The eigenvalue perturbation function will be specifically designed in order to make the inverse well-behave in a neighborhood of the true eigenvalues. Further, the variance regularization method leads to a decomposition of the regularized statistics into three components, one viewed as *regular* built on large eigenvalues while the others which involve the small eigenvalues may not be *regular*.

4.1. Regularized score-type tests

Consider the $q \times m$ matrix

$$\tilde{\mathcal{Q}}_n \equiv \tilde{\mathcal{Q}}[W_n] \equiv \tilde{P}_n \big(\tilde{J}_n' W_n \tilde{J}_n\big)^{-} \tilde{J}_n' W_n \ , \tag{4.1}$$

with $\tilde{P}_n = P(\tilde{\theta}_n^0)$, $\tilde{J}_n = \hat{J}(\tilde{\theta}_n^0)$, $P(\theta) = \partial\psi/\partial\theta'$, $\hat{J}(\theta) = \frac{\partial D_n}{\partial \theta'}(\theta) = H_n(\theta)$.

The assumptions below define the specific nonregular setup we shall be investigating throughout this paper. Specifically, we will examine to what extent relaxing the usual nonsingularity assumptions will affect the behavior of the test criterion. The proposed regularizing tools aim at curing the misbehavior of the test criterion under such circumstances.

Assumption 4.1 (Incomplete restriction rank) *There is a (non-empty) open neighborhood N_2 of θ_0 such that $\psi(\theta)$ is continuously differentiable and $rank[P(\theta)] = r_1$ for all $\theta \in N_2$ such that $0 < r_1 < q$.*

In the following, we analyze situations where the asymptotic covariance matrix does not necessarily have full rank. **We shall restrict our attention to situations where the rank is incomplete only asymptotically but not in finite samples.** To the best of our knowledge, cases where the ranks do not converge have not been investigated yet in the literature. Under Assumption **4.1** and more generally for any matrix with an incomplete rank – either asymptotically or in finite samples– the usual inverse is not defined and requires either using a generalized inverse or transforming it such that it becomes invertible. While Dufour and Trognon (2001, Proposition 3.1) have chosen the generalized inverse approach to proceed, we shall opt for the regularizing approach instead.

The following assumption is crucial for the distribution of the regularized $C(\alpha)$ test; for those interested in specific conditions under which this assumption holds in a GMM framework, see Dufour and Trognon (2001).

Assumption 4.2

$$\tilde{Z}_n \xrightarrow{\text{P}} Z_n\,, \quad Z_n \xrightarrow{\mathcal{L}} N\big[0, \mathcal{V}_\psi(\theta_0)\big]\,, \;\; \det\big(\mathcal{V}_\psi(\theta_0)\big) \geq 0 \tag{4.2}$$

where $\tilde{Z}_n = \sqrt{n}\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0)$, $Z_n = \sqrt{n}\mathcal{Q}(\theta_0)D_n(\theta_0)$.

We now introduce the *regularized* $C(\alpha)$ statistic:

$$\overline{PC}(\tilde{\theta}_n^0;\psi)^R = nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'\big[\tilde{\mathcal{Q}}_n\tilde{I}_0\tilde{\mathcal{Q}}_n'\big]^R\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0)\,. \tag{4.3}$$

The latter can be decomposed in three components. To do so, we denote

$$\mathcal{V}_\psi(\theta) \equiv \mathcal{Q}(\theta)I(\theta)\mathcal{Q}(\theta)'\,, \tag{4.4}$$

so that

$$\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0) \equiv \tilde{\mathcal{Q}}_n\tilde{I}_0\tilde{\mathcal{Q}}_n'\,. \tag{4.5}$$

The spectral decomposition of $\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)$ is:

$$\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0) = V(\tilde{\theta}_n^0)\Lambda(\tilde{\theta}_n^0)V(\tilde{\theta}_n^0)'\,. \tag{4.6}$$

This yields for the Moore-Penrose inverse

$$[\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)]^+ = \hat{V}_1(c)\hat{\Lambda}_1^+(c)\hat{V}_1'(c)\,, \tag{4.7}$$

and the *regularized* inverse

$$[\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)]^R = \hat{V}_1(c)\hat{\Lambda}_1^\dagger(c)\hat{V}_1'(c) + \hat{V}_2(c)\hat{\Lambda}_2^\dagger(c)\hat{V}_2'(c) + \hat{V}_3(c)\hat{\Lambda}_3^\dagger(c)\hat{V}_3'(c)\,. \tag{4.8}$$

Based on this *regularized* inverse, we introduce the *regularized* $C(\alpha)$ statistic, which can be decomposed as follows:

$$\begin{aligned}
\overline{PC}(\tilde{\theta}_n^0;\psi)^R &= \tilde{Z}_n'[\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)]^R\tilde{Z}_n \\
&= nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'[\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)]^R\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0) \\
&= nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'\begin{bmatrix}\hat{V}_1(c) & \hat{V}_2(c) & \hat{V}_3(c)\end{bmatrix}\begin{pmatrix}\hat{\Lambda}_1^\dagger(c) & 0 & 0\\ 0 & \hat{\Lambda}_2^\dagger(c) & 0\\ 0 & 0 & \hat{\Lambda}_3^\dagger(c)\end{pmatrix}\begin{bmatrix}\hat{V}_1'\\ \hat{V}_2'\\ \hat{V}_3'\end{bmatrix}\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0) \\
&= \overline{PC}_{1n}(c) + \overline{PC}_{2n}(c) + \overline{PC}_{3n}(c)\,,
\end{aligned} \tag{4.9}$$

where

$$\overline{PC}_{1n}(c) = nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'\hat{V}_1(c)\hat{\Lambda}_1^{\dagger}(c)\hat{V}_1'(c)\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0)\ , \tag{4.10}$$

$$\overline{PC}_{2n}(c) = nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'\hat{V}_2(c)\hat{\Lambda}_2^{\dagger}(c)\hat{V}_2'(c)\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0)\ , \tag{4.11}$$

$$\overline{PC}_{3n}(c) = nD_n(\tilde{\theta}_n^0)'\tilde{\mathcal{Q}}_n'\hat{V}_3(c)\hat{\Lambda}_3^{\dagger}(c)\hat{V}_3'(c)\tilde{\mathcal{Q}}_n D_n(\tilde{\theta}_n^0)\ . \tag{4.12}$$

It is important to remark that the *regularized* $C(\alpha)$ statistic given in equation (4.9) can easily be *simulated* together with its three components as soon as we have a consistent restricted estimator $\tilde{\theta}_n^0$ for θ_0. An heuristic scheme can be implemented as follows. Provided we have a consistent restricted estimator $\tilde{\theta}_n^0$ satisfying Assumption **2.6**, the regularized statistic can be simulated as follows:

1. an estimator of the covariance matrix can be built from $\tilde{\theta}_n^0$: $\tilde{\mathcal{V}}_{\psi}(\tilde{\theta}_n^0) \equiv \tilde{\mathcal{Q}}_n \tilde{I}_0 \tilde{\mathcal{Q}}_n'$;
2. the singular value decomposition of this covariance matrix can be performed;
3. the VRF to the objects obtained at stage 2 can be applied to get the three components, *i.e.* $\hat{V}_i(c)\hat{\Lambda}_i^{\dagger}(c)\hat{V}_i'(c)$ for $i = 1, 2, 3$;
4. some perturbations ϵ_n can be drawn form a $N(0, I_q)$ distribution;
5. $\tilde{Z}_n = \left(\tilde{\mathcal{V}}_{\psi}(\tilde{\theta}_n^0)\right)^{1/2}\epsilon_n$ can be formed;
6. the statistic can finally be computed by using the quadratic form and its decomposition as in equations (4.9)-(4.12).

The appealing and convenient property of being simulable enjoyed by this specific formulation widens its scope of application to general nonregular situations. Another appealing feature of this *regularized* $C(\alpha)$ statistic is that it can be decomposed into three independent components. The independence feature of its components arises from the eigenprojection onto different eigenspaces. This property greatly simplifies the proof of the distributional results of the *regularized* $C(\alpha)$ test we shall discuss later on. For the time being, let us turn our attention to the regularized LR-type tests.

4.2. Regularized LR-type tests

We now study the behavior of the LR-type statistic under the following nonregularity assumption.

Assumption 4.3 (Properties of the weighting matrix) *W_n is a positive definite matrix which converges in probability to the positive semi-definite matrix of constants W_∞,* i.e. $det(W_\infty) \geq 0$.

Let $W_{n,R}$ denote the regularized inverse of W_n and let us introduce the regularized GMM criterion

$$M_{n,R}(\theta) \equiv M_n(\theta, W_{n,R}) = D_n(\theta)'W_{n,R}D_n(\theta)\ .$$

Let $\hat{\theta}_{n,R}$ be the unrestricted regularized GMM estimator while $\hat{\theta}^0_{n,R}$ represents its restricted counterpart.

We propose to *regularize* the "LR-type " test statistic by taking a regularized inverse of the weighting matrix denoted $W_{n,R}$:

$$LR(\psi)^R = n[M_n(\hat{\theta}^0_n, W_{n,R}) - M_n(\hat{\theta}_n, W_{n,R})] \tag{4.13}$$

where

$$M_n(\theta, W_{n,R}) = D_n(\theta)'W_{n,R}D_n(\theta) = D_n(\theta)' \begin{bmatrix} \hat{V}_1(c) & \hat{V}_2(c) & \hat{V}_3(c) \end{bmatrix} \begin{pmatrix} \hat{\Lambda}^\dagger_1(c) & 0 & 0 \\ 0 & \hat{\Lambda}^\dagger_2(c) & 0 \\ 0 & 0 & \hat{\Lambda}^\dagger_3(c) \end{pmatrix} \begin{bmatrix} \hat{V}_1' \\ \hat{V}_2' \\ \hat{V}_3' \end{bmatrix} D_n(\theta)$$

$$= D_n(\theta)'\hat{V}_1(c)\Lambda^\dagger_1(c)\hat{V}_1'(c)D_n(\theta) + D_n(\theta)'\hat{V}_2(c)\Lambda^\dagger_2(c)\hat{V}_2'(c)D_n(\theta) + D_n(\theta)'\hat{V}_3(c)\Lambda^\dagger_3(c)\hat{V}_3'(c)D_n(\theta)$$

$$= D_n(\theta)'W_{11,nR}D_n(\theta) + D_n(\theta)'W_{22,nR}D_n(\theta)D_n(\theta)'W_{33,nR}D_n(\theta) \ , \tag{4.14}$$

and

$$W_{ii,nR} = \hat{V}_i(c)\Lambda^\dagger_i(c)\hat{V}_i'(c) \quad i = 1, 2, 3 \ . \tag{4.15}$$

Thus the regularized GMM criterion can be decomposed as follows:

$$M_n(\theta, W_{n,R}) = M_{1n,R}(\theta) + M_{2n,R}(\theta) + M_{3n,R}(\theta) \ , \tag{4.16}$$

with

$$M_{in,R}(\theta) = D_n(\theta)'W_{ii,nR}D_n(\theta) \text{ for } i = 1, 2, 3 \ .$$

A typical choice for the weighting matrix is to take the inverse of the covariance matrix of the moment conditions which corresponds in our framework to take the inverse of the information matrix associated with the score function, *i.e.* $W_\infty = I(\theta_0)$ and $W_{\infty,R} = I^R(\theta_0)$. It is worth noting that the decomposition of the regularized inverse carries an important linearity property that is transmitted to the regularized GMM criterion together with the tests statistics themselves as displayed below. This linearity property will greatly simplify the distributional proofs later on.

This yields a decomposition of the LR statistic into three components

$$\begin{aligned} LR(\psi)^R &= n[M_n(\hat{\theta}^0_n, W_{n,R}) - M_n(\hat{\theta}_n, W_{n,R})] \\ &= n\Big[\big[M_{1n,R}(\hat{\theta}^0_n) + M_{2n,R}(\hat{\theta}^0_n) + M_{3n,R}(\hat{\theta}^0_n)\big] - \big[M_{1n,R}(\hat{\theta}_n) + M_{2n,R}(\hat{\theta}_n) + M_{3n,R}(\hat{\theta}_n)\big]\Big] \\ &= n[M_{1n,R}(\hat{\theta}^0_n) - M_{1n,R}(\hat{\theta}_n)] + n[M_{2n,R}(\hat{\theta}^0_n) - M_{2n,R}(\hat{\theta}_n)] + n[M_{3n,R}(\hat{\theta}^0_n) - M_{3n,R}(\hat{\theta}_n)] \\ &= LR_{1,n} + LR_{2,n} + LR_{3,n} \end{aligned} \tag{4.17}$$

4.3. Regularized Wald tests

Similarly, we can decompose the original Wald statistic $W_n(\psi_0)$ in (2.14) as

$$W_n(\psi_0) = W_{1n} + W_{2n} + W_{3n} \tag{4.18}$$

where $W_{in} = n\ \psi(\hat{\theta})'\hat{V}_i\hat{\Lambda}_i^{-1}\hat{V}_i'\psi(\hat{\theta})$, for $i = 1, ..., 3$. The *regularized* counterpart of the above Wald statistics can be defined as follows:

$$\begin{aligned} W_n^R(\psi_0) &= n\ \psi(\hat{\theta})'\hat{\Sigma}^R\psi(\hat{\theta}) = n\ \psi(\hat{\theta})'\hat{V}\hat{\Lambda}^\dagger\ \hat{V}'\psi(\hat{\theta}) \\ &= n\ \psi(\hat{\theta})'\hat{V}\begin{pmatrix} g(\hat{\lambda}_1; c) & & 0 \\ & \ddots & \\ 0 & & g(\hat{\lambda}_q; c)\end{pmatrix}\hat{V}'\psi(\hat{\theta}) \\ &= n\ \psi(\hat{\theta})'\begin{bmatrix}\hat{V}_1(c) & \hat{V}_2(c) & \hat{V}_3(c)\end{bmatrix}\begin{pmatrix}\hat{\Lambda}_1^\dagger(c) & 0 & 0 \\ 0 & \hat{\Lambda}_2^\dagger(c) & 0 \\ 0 & 0 & \hat{\Lambda}_3^\dagger(c)\end{pmatrix}\begin{bmatrix}\hat{V}_1' \\ \hat{V}_2' \\ \hat{V}_3'\end{bmatrix}\psi(\hat{\theta}) \end{aligned} \tag{4.19}$$

$$\begin{aligned} W_n^R(\psi_0) &= n\ \psi(\hat{\theta})'\hat{V}_1(c)\hat{\Lambda}_1^\dagger(c)\hat{V}_1'(c)\psi(\hat{\theta}) + n\psi(\hat{\theta})'\hat{V}_2(c)\hat{\Lambda}_2^\dagger(c)\hat{V}_2'(c)\psi(\hat{\theta}) + n\psi(\hat{\theta})'\hat{V}_3(c)\hat{\Lambda}_3^\dagger(c)\hat{V}_3'(c)\psi(\hat{\theta}) \\ &= W_{1n}^R(c) + W_{2n}^R(c) + W_{3n}^R(c)\ . \end{aligned} \tag{4.20}$$

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, a second component gathering the eigenvalues exactly equal to the threshold c and a third component which incorporates the small remaining eigenvalues. As we will see below, the first component can be viewed as "regular", while the other ones 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 will replace the latter with eigenvalues bounded away from zero. Instead, Lütkepohl and Burda (1997) keep only the eigenvalues which are greater than the threshold c, which boils down to setting $W_{2n}^R(c) = 0$ and $W_{3n}^R(c) = 0$. Discarding the small eigenvalues in such a way may result in a loss of information. However, as Lütkepohl and Burda (1997) use a χ^2 distribution with a reduced degree of freedom, a deeper investigation must be conducted to assess the power properties. It is important to note that in finite samples it will be difficult to disentangle between the estimates which really correspond to $\lambda = c$ from those which are close to c but are 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 the threshold c, $m(c) = 0$, and the decomposition involves only two components.

5. Results on Eigenprojections

In this section, we discuss some non-uniqueness and discontinuity issues regarding the eigenvectors of a given matrix. While it is well-known in spectral theory that the 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 always cautious about such considerations that could entail some 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 some 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 the discontinuity of the eigenvectors, we consider the following simple counter-example.

Example 5.1 Let $A(x)$ be the matrix function defined as:

$$A(x) = \begin{cases} \begin{bmatrix} 1+x & 0 \\ 0 & 1-x \end{bmatrix} & \text{if } x < 0 \\ \\ \begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix} & \text{if } x \geq 0 \ . \end{cases} \tag{5.1}$$

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:

$$A(x) = (1+x) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + (1-x) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{5.2}$$

[with $(1+x)$ and $(1-x)$ being the eigenvalues and $(1,0)'$ and $(0,1)'$ the eigenvectors], while for $x > 0$, it is

$$A(x) = \frac{1}{\sqrt{2}}(1+x) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{1}{\sqrt{2}}(1-x) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{5.3}$$

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

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 the eigenvalues of Σ. The *eigenspace* of Σ associated with λ is defined as all the linear combinations from a basis of eigenvectors $\mathbf{x}_i(\lambda)$, $i = 1, \ldots, m(\lambda)$, *i.e.*

$$V(\lambda) = \{\mathbf{x}_i(\lambda) \in \mathbb{R}^q | \Sigma \mathbf{x}_i(\lambda) = \lambda \mathbf{x}_i(\lambda)\} . \tag{5.4}$$

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 $\mathbf{T}$ (*i.e.* $\mathbf{T}\Sigma$ is symmetric), we have:

$$\mathbb{R}^q = \sum_{\lambda \in \mathcal{S}(\Sigma)} V(\lambda) . \tag{5.5}$$

The *eigenprojection* of Σ associated with λ, denoted $P(\lambda)$, is the projection operator onto $V(\lambda)$ w.r.t. decomposition (5.5) of $\mathbb{R}^q$. For any set of vectors $\mathbf{x}_i(\lambda)$ in $V(\lambda)$ such that $\mathbf{x}_i(\lambda)'\mathbf{T}\mathbf{x}_j(\lambda) = \delta_{ij}$, where δ_{ij} denotes the Kronecker' s delta, $P(\lambda)$ has the representation

$$P(\lambda) = \sum_{j=1}^{m(\lambda)} \mathbf{x}_j(\lambda)\mathbf{x}_j(\lambda)'\mathbf{T} . \tag{5.6}$$

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

$$\Sigma = \sum_{\lambda \in \mathcal{S}(\Sigma)} \lambda P(\lambda) , \qquad \hat{\Sigma} = \sum_{\hat{\lambda} \in \mathcal{S}(\hat{\Sigma})} \hat{\lambda} P(\hat{\lambda}) . \tag{5.7}$$

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 $\sum_{\lambda \in v} P(\lambda)$. More specifically, let $P_{k,t}(\lambda) = \sum_{l=k}^{t} P(\lambda_l)$. We report below a lemma given by Tyler (1981, Lemma 2.1, p. 726) which states an important continuity property for eigenvalues and eigenprojections on eigenspaces for non-random symmetric matrices which will be useful to establish the consistency for the sample regularized inverses toward their population quantity.

Lemma 5.2 (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 tell 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 relation relating the operator to the matrix with respect to the given bases] associated with the eigenvalues and transmitted to their sum. No matter what the multiplicity of the eigenvalues involved in the total eigenprojection [$P_{k,t}(\lambda) = \sum_{l=k}^{t} P(\lambda_l)$], this continuity property holds provided that we can find one eigenvalue just before and one just after that are distinct.

It will be useful to extend Lemma **5.2** to random symmetric matrices. To do so, we first consider almost sure (a.s.) convergence and then convergence in probability (in symbol $\xrightarrow{\mathbf{P}}$). To the best of our knowledge, these results are not 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 5.3 *Let $\hat{\Sigma}_n$ be a $q \times q$ real random matrix symmetric in the metric of a real positive definite symmetric random matrix T_n and with eigenvalues $\lambda_1(\hat{\Sigma}_n) \geq \lambda_2(\hat{\Sigma}_n) \geq \ldots \geq \lambda_q(\hat{\Sigma}_n)$. Let $P_{k,t}(\hat{\Sigma}_n)$ represent the total eigenprojection for $\hat{\Sigma}_n$ associated with $\lambda_k(\hat{\Sigma}_n) \ldots \lambda_t(\hat{\Sigma}_n)$ for $t \geq k$. If $\hat{\Sigma}_n \xrightarrow{a.s.} \Sigma$ as $n \to \infty$, then:*

i) $\lambda_k(\hat{\Sigma}_n) \xrightarrow{a.s.} \lambda_k(\Sigma)$, *and*

ii) $P_{k,t}(\hat{\Sigma}_n) \xrightarrow{a.s.} P_{k,t}(\Sigma)$ *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 5.4 *Let $\hat{\Sigma}_n$ be a $q \times q$ real random matrix symmetric in the metric of a real positive definite symmetric random matrix T_n with eigenvalues $\lambda_1(\hat{\Sigma}_n) \geq \lambda_2(\hat{\Sigma}_n) \geq \ldots \geq \lambda_q(\hat{\Sigma}_n)$. Let $P_{k,t}(\hat{\Sigma}_n)$ represent the total eigenprojection for $\hat{\Sigma}_n$ associated with $\lambda_k(\hat{\Sigma}_n), \ldots, \lambda_t(\hat{\Sigma}_n)$ for $t \geq k$. If $\hat{\Sigma}_n \xrightarrow{\mathbf{P}} \Sigma$ as $n \to \infty$, then:*

i) $\lambda_k(\hat{\Sigma}_n) \xrightarrow{\mathbf{P}} \lambda_k(\Sigma)$, *and*

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

6. Asymptotic theory

In this section, we state some asymptotic results for the *regularized* statistics which rely on different specifications for the regularized inverses. We will first study the simple case of a continuous *variance regularization function* (VRF) before considering the monotonic case. The monotonicity property can be exploited to yield some bounds on the limiting distribution of the modified statistics.

Based on the spectral decomposition defined in equation (5.7), we immediately deduce a spectral decomposition for the regularized inverses:

$$\Sigma^R = \sum_{\lambda \in \mathcal{S}(\Sigma)} g(\lambda; c) P(\lambda) , \qquad \hat{\Sigma}^R = \sum_{\hat{\lambda} \in \mathcal{S}(\hat{\Sigma})} g(\hat{\lambda}; c) P(\hat{\lambda}) \ . \tag{6.1}$$

We will provide below various regularized inverses based on different specifications for the VRF $g(\lambda; c)$. We will consider different specifications for the VRF that may depend on the sample size n. This path dependency may be introduced through the threshold c_n as

$$g(\lambda, n) = g(\lambda, c_n) \; . \tag{6.2}$$

6.1. Distributional theory for regularizing functions continuous almost everywhere

Let the general VRF $g(\lambda; c)$ be defined as:

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda} & \text{if } \; \lambda > c \; , \\ \frac{1}{\epsilon+\gamma(c-\lambda)} & \text{if } \; \lambda \leq c \; . \end{cases} \tag{6.3}$$

Interesting special cases include:

i) $\gamma = \infty$, hence

$$g(\lambda; c) = \begin{cases} \frac{1}{\lambda} & \text{if } \; \lambda > c \\ 0 & \text{if } \; \lambda \leq c \end{cases}$$

and therefore $\varLambda^{\dagger} = \varLambda_c^{+}$, where

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

corresponds to a spectral cut-off regularization scheme [see Carrasco (2007) and the references therein]; $I(s)$ is equal to 1 if the relation s is satisfied; in particular $\varLambda_c^{+}$ is a *modified version* of the Moore-Penrose inverse of

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

used by Lütkepohl and Burda (1997, henceforth LB). Below the eigenvalues λ_{q_1+1} to λ_q are smaller than the threshold c but not necessarily equal to zero;

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

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

henceforth called DV1.

The following VRF will turn out particularly convenient.

$$g(\hat{\lambda}_{n,j}; c) = \begin{cases} \frac{1}{\hat{\lambda}_{n,j}} & \text{if } \hat{\lambda}_{n,j} > c \\ \frac{1}{\epsilon+\gamma(c-\hat{\lambda}_{n,j})} & \text{if } \hat{\lambda}_{n,j} \leq c \end{cases} \; , \tag{6.4}$$

where $\gamma > 0$. It can be viewed as a *modified* Hodges' estimator applied to the eigenvalues. See LeCam (1953).

Again all the VRF $g(\lambda; c)$ above allow for possibly sample-size dependent thresholds c_n.

Proposition 6.1 (Almost sure convergence of the regularized inverses) *Let $g(\lambda; c)$ be a continuous function at each eigenvalue of Σ. If $\hat{\Sigma}_n \overset{a.s.}{\to} \Sigma$ and $\lambda_j(\Sigma) \neq c$ for all j, then*

$$\hat{\Sigma}_n^R \overset{a.s.}{\to} \Sigma^R \,. \tag{6.5}$$

Before establishing the main convergence result for the regularized covariance matrices, we must 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 some nonclassical results due to Anderson (1963) Anderson (1987) on the behavior of sample roots (of a determinantal equation). Specifically, under fairly weak conditions they show that if the sequence of random matrices of interest X_n is such that $Q_n = b_n(X_n - B_0) \overset{\mathcal{L}}{\to} Q$ with the limit B_0 nonstochastic, then the sample eigenvalues will have the same convergence rate, with $b_n[\Psi(X_n) - \Psi(B_0)] \overset{\mathcal{L}}{\to} \Psi(Q)$, where $\Psi(B)$ denotes the $q-$vector of singular values of an arbitrary matrix B. This convergence rate of the sample eigenvalues will be useful in establishing the asymptotic distribution of the regularized tests.

Let $d_1 > d_2 > \cdots > d_k$ denote the distinct eigenvalues of C and let m_i be their multiplicities, $i = 1, \ldots, k$. Given the eigenvalue multiplicities of C, a function H on $q \times q$ symmetric matrices can be defined by

$$H(C) = \begin{pmatrix} \rho(C_{11}) \\ \rho(C_{22}) \\ \vdots \\ \rho(C_{kk}) \end{pmatrix} \tag{6.6}$$

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

$$\Gamma A \Gamma' = D,$$

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 6.2 *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*

$$W_n = b_n(S_n - A) \overset{\mathcal{L}}{\to} W \,. \tag{6.7}$$

Then

$$b_n\big(\rho(S_n) - \rho(A)\big) \overset{\mathcal{L}}{\to} H\big(\Gamma W \Gamma'\big) \,. \tag{6.8}$$

For any $p \times q$ real matrix B, let

$$\Psi(B) = f(\rho(B'B))$$

where

$$f(x) = \begin{pmatrix} \sqrt{x_1} \\ \vdots \\ \sqrt{x_q} \end{pmatrix}$$

and let

$$T = \big(df(\xi)\big) = \frac{1}{2}\text{diag}(\xi_1^{-1/2}, \ldots, \xi_q^{-1/2}) \;, \tag{6.9}$$

where $\xi_1 \geq \cdots \geq \xi_q > 0$ are the eigenvalues of $B'B$. In the lemma below, we gather the special cases where the matrix B has rank $r = 0$ and $r = q$. Let X_n a sequence of $p \times q$ real random matrices, ($p \geq q$) and there exists a fixed $p \times q$ matrix B such that

$$Q_n = b_n\big(X_n - B\big) \xrightarrow{\mathcal{L}} Q \;. \tag{6.10}$$

Lemma 6.3 *When $B = 0$, then*

$$b_n\big(\Psi(X_n) - \Psi(B)\big) \xrightarrow{\mathcal{L}} \Psi(Q) \;. \tag{6.11}$$

When B has full rank q, then

$$b_n\big(\Psi(X_n) - \Psi(B)\big) \xrightarrow{\mathcal{L}} TH\big(\Gamma\big[B'Q + Q'B\big]\Gamma'\big) \;. \tag{6.12}$$

See Eaton and Tyler (1994, Propositions 3.1 and 3.4) for a proof. Write the $p \times q$ matrix Σ in the form

$$\Sigma = \Gamma_1' \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Gamma_2' \tag{6.13}$$

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

$$\Sigma_n = \begin{pmatrix} \Sigma_{n11} & \Sigma_{n12} \\ \Sigma_{n21} & \Sigma_{n22} \end{pmatrix} \tag{6.14}$$

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 $D = \text{diag}(\xi_1^{1/2}, \ldots, \xi_r^{1/2})$ defines a function H_D on $r \times r$ symmetric matrices. Let $T_D = \frac{1}{2}\text{diag}(\xi_1^{-1/2}, \ldots, \xi_r^{-1/2})$. The general case $1 \leq r < q$ can be thought as gluing together the two special cases $r = 0$ and $r = q$ given in Lemma **6.3** above.

Lemma 6.4 *Let Σ_n be a sequence of $p \times q$ real random matrix and Σ be a $p \times q$ real nonstochastic matrix such that $rank(\Sigma) = r$, $1 \leq r < q$. Let also b_n be a sequence of real constants such that*

$$Q_n = b_n(\Sigma_n - \Sigma) \xrightarrow{\mathcal{L}} Q \tag{6.15}$$

where $b_n \to +\infty$ and Q is a random matrix. Then,

$$E_n = b_n\big[\Psi(\Sigma_n) - \Psi(\Sigma)\big] \xrightarrow{\mathcal{L}} \begin{bmatrix} H_D\big(\frac{1}{2}[Q_{11}' + Q_{11}]\big) \\ \Psi(Q_{22}) \end{bmatrix} \tag{6.16}$$

where Q_{11} and Q_{22} are well-defined random elements.

For a proof, see Eaton and Tyler (1994, Theorem 4.2).

We state now the convergence result for the regularized inverses which is fundamental to obtain well-behaved regularized test statistics. Let $\hat{\lambda}_{n,i} = \lambda_i(\Sigma_n)$ and $\lambda_i = \lambda_i(\Sigma)$. First when designing the VRF $g(\lambda; c(n))$, the threshold $c(n)$ must be selected so that

$$Pr\big[|\hat{\lambda}_{n,i} - \lambda_i| > c(n)\big] = Pr\big[|b_n(\hat{\lambda}_{n,i} - \lambda_i)| > b_n c(n)\big] \underset{n\to\infty}{\to} 0$$

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 guarantees that the nonzero eigenvalues of the covariance matrix will eventually be greater than the threshold, while the true zero eigenvalues are still smaller than 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 the threshold is to set $c(n) = n^{-1/3}$. This convergence rate plays a crucial role in Proposition **6.5** below.

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.

Proposition 6.5 *Let Σ be a $q \times q$ real positive semidefinite nonstochastic matrix. Let Σ_n be a $q \times q$ real symmetric random matrix and suppose the assumptions of Lemma* **6.4** *hold. Let $g(\lambda; c(n))$ be a function continuous almost everywhere w.r.t. λ, with the set consisted of the discontinuities of g being of measure zero. Suppose further that $c(n) \to 0$ and $b_n c(n) \to \infty$ as $n \to \infty$. Then*

$$(\Sigma_n)^R \xrightarrow{\text{p}} (\Sigma)^R .$$

See appendix for a proof.

In the following, we establish a *characterization* of the asymptotic distribution of the *regularized* test statistics in the general case. This characterization makes use of a decomposition of the *modified* statistic into a regular component and a regularized component. We first formulate the proposition under the following generic assumption before considering specific statistics as $C(\alpha)$ and Wald test statistics.

Assumption 6.6 *Let Z_n and Z_∞ be $q \times 1$ random vectors defined on a common probability space* *such that:*

$$Z_n \xrightarrow{\mathcal{L}} Z_\infty \sim N[0, \Sigma] \text{ , with } \operatorname{rank}(\Sigma) = q_1 \le q \ .$$

Let $v_1 = \{\lambda : \lambda > 0\}$ and $v_2 = \{\lambda : \lambda \le 0\}$ with $v_1 \cup v_2 = \mathcal{S}(\Sigma)$. Similarly, let $\hat{v}_1 \equiv \hat{v}_1(\omega, c_n) = \{\hat{\lambda}_n(\omega) : \hat{\lambda}_n(\omega) > c_n\}$ and $\hat{v}_2 \equiv \hat{v}_2(\omega, c_n) = \{\hat{\lambda}_n(\omega) : \hat{\lambda}_n(\omega) \le c_n\}$ with $\hat{v}_1 \cup \hat{v}_2 = \mathcal{S}(\Sigma_n)$. Let $B_{i,n}(c_n) = \{\omega \in \Omega : \hat{v}_i(\omega, c_n) = v_i\}$ such that $Pr[B_{in}(c_n)] \underset{n\to\infty}{\to} 1$.

Proposition 6.7 (Asymptotic distribution of *regularized* statistics when the threshold goes to zero) *Suppose that the assumptions of Proposition* **6.5** *are satisfied. Suppose further that Assumption* **6.6** *holds. The regularized statistic:*

$$Z_n'[\Sigma_n]^R Z_n = Z_{1n}'\hat{\Lambda}_{1n}^\dagger Z_{1n} + Z_{2n}'\hat{\Lambda}_{2n}^\dagger Z_{2n} \xrightarrow{\mathcal{L}} Z_1'\Lambda_1^\dagger Z_1$$

with $Z_1'\Lambda_1^\dagger Z_1 \ \sim \chi^2(q_1)$.

It is worth noting that when the threshold $c(n)$ converges to zero at an appropriate rate based on the sample eigenvalues convergence rate, the statistic thus regularized is asymptotically distributed as a $\chi^2(q_1)$ variable with q_1 denoting the number of nonzero eigenvalues. Under this circumstance, we find the asymptotic result obtained by Lütkepohl and Burda (1997) but at the difference that the above regularized statistic is bounded from below by that of Lütkepohl and Burda (1997) *in finite samples*. While a simulation-based approach is required to control the level of the regularized test under the null hypothesis in finite samples, this feature may entail some gains in power under the alternative.

In the following we apply the above distributional result to specific test statistics such as the regularized $C(\alpha)$, the regularized Wald test. To do so, we need to check that the required assumptions for Proposition **6.7** to hold are satisfied for those specific statistics. Also, under the assumptions **2.4** and **2.7**, we have:

$$\underset{n\to\infty}{\text{plim}}\ \tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0) = \underset{n\to\infty}{\text{plim}}\ \tilde{\mathcal{Q}}_n \tilde{I}_0 \tilde{\mathcal{Q}}_n' = \mathcal{Q}(\theta_0) I(\theta_0) \mathcal{Q}(\theta_0)' = \mathcal{V}_\psi(\theta_0) \ .$$

Corollary 6.8 *Under the assumptions* **2.4** *and* **2.7***, and the assumptions of Proposition* **6.5***, we have:*

$$[\tilde{\mathcal{V}}_\psi(\tilde{\theta}_n^0)]^R \xrightarrow{\text{p}} [\mathcal{V}_\psi(\theta_0)]^R \ . \tag{6.17}$$

The results stated in Proposition **6.7** when applied to $Z_n = \mathcal{Q}_n D_n(\tilde{\theta}_n^0)$ yield a χ^2 distribution for the *regularized* $C(\alpha)$ test.

Corollary 6.9 (Asymptotic distribution of *regularized* $C(\alpha)$ statistic) *Suppose that the assumptions of Proposition* **6.7** *hold. Suppose further that the assumptions of Corollary* **6.8** *are satisfied together with Assumption* **4.2***. Then*

$$\overline{PC}_R(\tilde{\theta}_n^0; \psi) \xrightarrow{\mathcal{L}} \overline{PC}_R(\theta_0; \psi) = D_\infty(\theta_0)'\mathcal{Q}(\theta_0)'[\mathcal{V}_\psi(\theta_0)]^R \mathcal{Q}(\theta_0) D_\infty(\theta_0) \ \sim \ \chi^2(q_1) \tag{6.18}$$

The results established above deserve some discussion. Thus, it is important to notice that both components are asymptotically independent. Whereas the first component is asymptotically χ^2 distributed, the second one collapses to zero. To the extent that the regularized $C(\alpha)$ statistic can be viewed as the sum of two "small" independent $C(\alpha)$ statistics, it can be easily simulated to provide a feasible test under singularity conditions.

Similarly when taking $Z_n = \sqrt{n}[\psi(\hat{\theta})-\psi_0]$, we can deduce from Proposition **6.7** the asymptotic distribution of the *regularized* Wald test under the null. Recall that we want to test the null:

$$H_0(\psi_0) : \psi(\theta_0) = \psi_0 \; . \tag{6.19}$$

Let us reformulate Assumption **2.9** under possible singularity issue.

Assumption 6.10 ($\sqrt{n}$ convergence of the restrictions)

$$\sqrt{n}(\psi(\hat{\theta}_n) - \psi(\theta_0)) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N(0, \Sigma_\psi(\theta_0)) \;\; \det\bigl(\Sigma_\psi(\theta_0)\bigr) \geq \; 0 \tag{6.20}$$

Corollary 6.11 (Asymptotic distribution of *regularized* Wald statistics) *Suppose that the assumptions of Proposition* **6.7** *hold. Suppose further that Assumption* **6.10** *holds. Let*

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

with

$$W_{1n}^R = n[\psi(\hat{\theta}) - \psi_0]'\hat{V}_1(c(n))\hat{\Lambda}_1^\dagger(c(n))\hat{V}_1'(c(n))[\psi(\hat{\theta}) - \psi_0] \tag{6.21}$$

and

$$W_{2n}^R = n[\psi(\hat{\theta}) - \psi_0]'\hat{V}_2(c(n))\hat{\Lambda}_2^\dagger(c(n))\hat{V}_2'(c(n))[\psi(\hat{\theta}) - \psi_0] \; . \tag{6.22}$$

Then, under 6.19*:*

$$W_{1n}^R \overset{\mathcal{L}}{\to} W_1 \; , \;\; \textit{with}\; W_1 \;\sim\; \chi^2(q_1) \;\; \textit{and} \;\; W_{2n} \overset{\text{p}}{\to} \; 0. \tag{6.23}$$

Notice that the test is consistent against fixed alternatives:

$$H_1(\psi_0) : \psi(\theta_0) \neq \psi_0 \; . \tag{6.24}$$

Let $\psi(\theta_0) - \psi_0 \; = \; \psi_1 \; \neq \; 0$ with ψ_1 a q_1-vector. The *regularized* test has local power against alternatives:

$$H_1(\psi_0) : \psi(\theta_0) - \psi_0 = \frac{\psi_1}{\sqrt{n}} \; . \tag{6.25}$$

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

$$W_n^R \overset{\mathcal{L}}{\to} \; \chi^2\bigl(q_1, \psi_1'[\Sigma]_{11}^R\psi_1\bigr) \; . \tag{6.26}$$

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

$$W_n^R = W_{1n}^R + W_{2n}^R = n\psi(\hat{\theta})'\hat{V}_1(c(n))\hat{\Lambda}_1^{\dagger}(c(n))\hat{V}_1'(c(n))\psi(\hat{\theta}) + 0 = n\psi(\hat{\theta})'\hat{V}(c(n))\hat{\Lambda}^{+}(c(n))\hat{V}'(c(n))\psi(\hat{\theta}) = W_n^{+}(c(n)) \,, \quad (6.27)$$

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 $\Lambda = diag(\lambda_1, \ldots, \lambda_{q_1}, \lambda_{q_1+1}, \ldots, \lambda_q)$. Therefore $W_n^{+}(c(n)) = W_{1n}^R(c(n))$.

When the threshold goes to zero at the appropriate speed, the regularized test statistics have a standard chi-square limiting distribution, with the nonregular component collapsing to zero. A more interesting case producing a nonstandard asymptotic distribution would be the case that the threshold is fixed putting a positive probability mass on the eigenvalues exactly equal to the threshold c. For instance, in the DV1 case (*i.e.* with $\Lambda_2^{\dagger}(c) = \frac{1}{c} I_{(q-q_1)}$), the modified statistic corresponds to:

$$\begin{aligned} W_n^R &= W_{1n}^R(c) + W_{2n}^R(c) \\ &= n\psi(\hat{\theta})'\hat{V}_1(c)\hat{\Lambda}_1^{\dagger}(c)\hat{V}_1'(c)\psi(\hat{\theta}) + n\psi(\hat{\theta})'\hat{V}_2(c)\hat{\Lambda}_2^{\dagger}(c)\hat{V}_2'(c)\psi(\hat{\theta}) \\ &= n\psi(\hat{\theta})'\hat{V}_1(c)\hat{\Lambda}_1^{\dagger}(c)\hat{V}_1'(c)\psi(\hat{\theta}) + \frac{n}{c}\psi(\hat{\theta})'\hat{V}_2(c)\hat{V}_2'(c)\psi(\hat{\theta}) \,. \end{aligned} \quad (6.28)$$

Let us now study this case.

Proposition 6.12 (Asymptotic distribution of *regularized* statistics when the threshold is fixed) *Suppose Z_n and Z to be $q \times 1$ random vectors defined on a common probability space such that: $Z_n \xrightarrow{\mathcal{L}} Z \sim N[0, \Sigma]$, with $\mathrm{rank}(\Sigma) = r \leq q$. Suppose the assumptions of Lemma* **5.4** *are satisfied. Let the VRF $g(., c)$ be a function continuous a.e. and bounded. Let the regularized statistic satisfy the following decomposition:*

$$Z_n'[\Sigma_n]^R Z_n = Z_{1n}'\hat{\Lambda}_{1n}^{\dagger} Z_{1n} + Z_{2n}'\hat{\Lambda}_{2n}^{\dagger} Z_{2n} + Z_{3n}'\hat{\Lambda}_{3n}^{\dagger} Z_{3n},$$

then

$$Z_{1n}'\hat{\Lambda}_{1n}^{\dagger} Z_{1n} = \sum_{j=1}^{q_1} z_{1nj} g(\lambda_{1nj}; c) z_{1nj} \xrightarrow{\mathcal{L}} \chi^2(q_1)$$

$$Z_{2n}'\hat{\Lambda}_{2n}^{\dagger} Z_{2n} = \sum_{j=q_1+1}^{q_2} z_{2nj} g(\lambda_{2nj}; c) z_{2nj} \xrightarrow{\mathcal{L}} \sum_{j=q_1+1}^{q_2} c g(c; c) I(\lambda_{2j} = c) v_j^2 \;\leq\; \chi^2(q_2)$$

$$Z_{3n}'\hat{\Lambda}_{3n}^{\dagger} Z_{3n} = \sum_{j=q_2+1}^{q} z_{3nj} g(\lambda_{3nj}; c) z_{3nj} \xrightarrow{\mathcal{L}} \sum_{j=q_2+1}^{q} \lambda_{3j} g(\lambda_{3j}; c) I(\lambda_{3j} < c) v_j^2 \;\leq\; \chi^2(q_3)$$

where $v_j \overset{i.i.d.}{\sim} N(0, 1)$.

Let $P_0 = P(\theta_0)$.

Proposition 6.13 (Asymptotic distribution of the *regularized* LR statistic when the threshold is fixed) *Suppose Assumptions* **4.3**, **B.1**-**B.7** *listed in Appendix hold. Then,*

$$LR(\psi)^R = n[M_n(\hat{\theta}_n^0, W_n^R) - M_n(\hat{\theta}_n, W_n^R)] \overset{\mathcal{L}}{\to} LR(\psi)_\infty^R \le \chi^2(q) \;, \tag{6.29}$$

where

$$LR(\psi)_\infty^R = \frac{1}{2} u' U_0^{1/2} V_0^{-1} P_0' \big(P_0 V_0^{-1} P_0'\big)^{-1} P_0 V_0^{-1} U_0^{1/2} u \;, \tag{6.30}$$

with $u \sim N(0, I)$

$$U_0 = 4J(\theta_0)' I^R(\theta_0) I(\theta_0) I^R(\theta_0) J(\theta_0) \tag{6.31}$$

and

$$V_0 = 2J(\theta_0)' I^R(\theta_0) J(\theta_0) \;. \tag{6.32}$$

Note that as soon as we have consistent estimators of the quantities above the limiting random variable $LR(\psi)_\infty^R$ can be simulated. We also provide below an asymptotic characterization of the regularized LR statistic component by component. An appealing feature of the asymptotic characterization is that each component can be written as a quadratic form in Gaussian variables.

Proposition 6.14 (Asymptotic characterization of the regularized LR components) *Let*

$$AV_{i0}A - V_0^{-1} V_{i0} V_0^{-1} = -V_0^{-1} V_{i0} V_0^{-1} P_0' \big(P_0 V_0^{-1} P_0'\big)^{-1} P_0 V_0^{-1} \;,$$
$$V_{i0} = 2J_0' I_{ii}^R(\theta_0) J_0$$
$$V_0^{-1} - A = V_0^{-1} P_0' \big(P_0 V_0^{-1} P_0'\big)^{-1} P_0 V_0^{-1} \;.$$

Under the assumptions of Proposition **6.13***, we have:*

$$n\xi_n = \sum_{i=1}^{3} \Big\{ \sqrt{n} X_{in,R}' \big(-A + V_0^{-1}\big) \sqrt{n} X_{n,R} + \frac{1}{2} \sqrt{n} X_{n,R}' \big(AV_{i0}A - V_0^{-1} V_{i0} V_0^{-1}\big) \sqrt{n} X_{n,R} \Big\} + ne_n, \;\; ne_n = o_p(1) \;, \tag{6.33}$$

where

$$\sqrt{n} X_{in,R} = \sqrt{n} \frac{\partial M_{in,R}}{\partial \theta}(\theta_0) \sim \; N[0, U_{i0}] \tag{6.34}$$

$$U_{i0} = 4J_0' I_{ii}^R(\theta_0) I(\theta_0) I_{ii}^R(\theta_0) J_0 \;\; \textit{for } i = 1, 2, 3, \tag{6.35}$$

and

$$\sqrt{n} X_{n,R} = \sqrt{n} \frac{\partial M_{n,R}}{\partial \theta}(\theta_0) \sim , N[0, U_0] \;.$$

See Appendix for a proof.

7. Examples

7.1. Multistep noncausality

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

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

Let consider

$$\begin{gathered} Y \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_T) \quad (3 \times T) \\ B \equiv (A_1) \quad (3 \times 3) \\ Z_t \equiv [\mathbf{y}_t] \;\; (3 \times 1) \quad Z \equiv (Z_0, \ldots, Z_{T-1}) \quad (3 \times T) \\ U = (u_1, \ldots, u_T) \quad (3 \times T) \end{gathered}$$

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 $\alpha = \text{vec}(A_1) = \text{vec}(B)$. Testing $H_0 : \; y_t \overset{(\infty)}{\nrightarrow} x_t$ 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

$$\begin{gathered} \alpha_{xy} = \alpha_{xz} = 0, \;\; \alpha_{zy} \neq 0 \\ \alpha_{xy} = \alpha_{zy} = 0, \;\; \alpha_{xz} \neq 0 \\ \alpha_{xy} = \alpha_{xz} = \alpha_{zy} = 0 \end{gathered} \tag{7.1}$$

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

$$\frac{\partial r}{\partial \alpha'} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{xy} & 0 & 0 & \alpha_{xx} + \alpha_{yy} & \alpha_{xy} & \alpha_{xz} & \alpha_{zy} & 0 & 0 \end{bmatrix} \tag{7.2}$$

if (7.1) 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 $\alpha = \text{vec}(A_1) = \text{vec}(B)$. Using the column stacking operator vec we have:

$$Y = BZ + U \tag{7.3}$$

or

$$\text{vec}(Y) = \text{vec}(BZ) + \text{vec}(U) \tag{7.4}$$

$$y = \big(Z' \otimes I_3\big)\text{vec}(B) + \text{vec}(U) \tag{7.5}$$

$$y = \big(Z' \otimes I_3\big)\alpha + u \tag{7.6}$$

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

$$\hat{\alpha} = \Bigg((ZZ')^{-1}Z \otimes I_3\Bigg)y \ . \tag{7.7}$$

The asymptotic distribution of the multivariate LS estimator:

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N\big(0, \varGamma^{-1} \otimes \varSigma_u)\big) \tag{7.8}$$

implies the asymptotic distribution for the restrictions:

$$\sqrt{T}(r(\hat{\alpha}) - r(\alpha)) \xrightarrow{\mathcal{L}} N\big(0, \varSigma_{r(\alpha)})\big) \tag{7.9}$$

where

$$\hat{\varSigma}_{r(\alpha)} = \frac{\partial r}{\partial \alpha'}(\hat{\alpha})\hat{\varSigma}_{\alpha}\frac{\partial r'}{\partial \alpha}(\hat{\alpha}) \tag{7.10}$$

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

$$\hat{\varSigma}_{\alpha} = \hat{\varGamma}^{-1} \otimes \hat{\varSigma}_u \tag{7.11}$$

is a consistent estimator for $\varSigma_{\alpha}$ with

$$\hat{\varGamma} = \frac{1}{T}ZZ' \tag{7.12}$$

and

$$\hat{\varSigma}_u = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t' = \frac{1}{T}Y\big[I_T - Z'(ZZ')^{-1}Z\big]Y' \ . \tag{7.13}$$

From the asymptotic distribution (7.9), a Wald-type test is easily obtained to test the null H_0 : $r(\alpha) = 0$, *i.e.*

$$W_{\psi} = Tr(\hat{\alpha})'\hat{\varSigma}^{R}_{r(\alpha)}r(\hat{\alpha}) \tag{7.14}$$

where a regularization is required under parameter setting (7.1).

8. Simulation results

In this section we perform a limited Monte Carlo experiment to assess the empirical behavior of the regularized Wald tests for testing multi-step noncausality. The study of the empirical properties of the other tests ($C(\alpha)$ and LR-type tests) are still pending. The results are very preliminary. The simulation study need to be completed.

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

$$W_{\psi} = Tr(\hat{\alpha})'\hat{\varSigma}^{R}_{r(\alpha)}r(\hat{\alpha}) \tag{8.15}$$

where a regularization is required under parameter setting (7.1). To fix notations, let W denote the (unmodified) standard Wald test statistic with $\hat{\Sigma}^{R}_{r(\alpha)} = \hat{\Sigma}^{-1}_{r(\alpha)}$; let W_{DV} denote the *regularized* Wald test based on fixed threshold c whereas W_{LB} stands for the *modified* Moore-Penrose Wald test using a sample size dependent threshold c_n.

8.1. **Simulation design**

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

$$A_1 = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ 0 & \alpha_{yy} & \alpha_{yz} \\ 0 & 0 & \alpha_{zz} \end{bmatrix}$$

$$1)\ \ \alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.9$$

$$2)\ \ \alpha_{xx} = \alpha_{yy} = \alpha_{zz} = 0.3$$

Because of its triangular form, the diagonal elements of A_1 represent its eigenvalues. The first parameter setting involve parameters close to the nonstationary region whereas the second one fall 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

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

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 key parameter to disentangle between the regularity point and singularity point is α_{xz} ($\alpha_{xz} = 0$ corresponds to a singularity point and $\alpha_{xz} \neq 0$ to a regularity point). 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 $\chi^2_{95\%}(1) = 3.81$.

We first study empirical level for testing

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

with $\alpha_{xy} = \alpha_{xz} = 0$, using two different approaches. The first one is based on the asymptotic χ^2-approximation while the second one resort to simulation-based bootstrap tests. The bootstrap tests are implemented through a consistent point estimate of the possible nuisance parameters. For more details on Monte Carlo and bootstrap tests see Dufour (2006).

The nominal size has been fixed to 5%.

$H_0 : r(\alpha) = 0$						
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.9$						
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.001$; [0.1]						
	$n = 50$		$n = 100$		$n = 200$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	10.46	6.86	5.84	5.48	3.28	4.62
W_{DV}	9.44 [6.48]	6.44 [6.18]	4.42 [3.84]	5.14 [5.40]	2.42 [2.36]	4.36 [4.5]
W_{LB}	14.28	5.52	9.92	5.08	7.5	4.38
	$n = 500$		$n = 1000$		$n = 2000$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	2.34	4.26	1.82	4.74	1.64	4.08
W_{DV}	1.72 [1.7]	4.26 [4.28]	1.66 [1.66]	5.04 [5.06]	1.42 [1.40]	4.02 [4.02]
W_{LB}	5.66	4.22	5.58	5.06	5.24	4.02

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 $\alpha_{xy} = \delta = 0.1264$ and $\alpha_{xz} = 0$.

The empirical power for locally size-corrected tests is reported in the table below.

$H_1 : r(\alpha) \neq 0 \;\; \alpha_{xy} = \delta = 0.1269$						
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = -0.9$						
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.001$; [0.1]						
	$n = 50$		$n = 100$		$n = 200$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	42.06	32.92	81.42	77.56	98.3	98.28
W_{DV}	40.08 [39.42]	31.54 [32.04]	80.36 [75.04]	77.84 [75.18]	97.82 [96.7]	98.44 [97.68]
W_{LB}	37.84	29.10	79.82	74.9	98.34	97.68
	$n = 500$		$n = 1000$		$n = 2000$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	100	100	100	100	100	100
W_{DV}	100 [99.98]	100 [100]	100 [100]	100 [100]	100 [100]	100 [100]
W_{LB}	100	100	100	100	100	100

We also consider a second alternative for a violation of the second restriction only while maintaining fulfilled the 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 this design:

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

Under this design, the LB Wald statistic based on a modified version of the Moore-Penrose inverse is expected to display reduced power. Indeed, by underestimating the true rank of the covariance matrix, this statistic put 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. The empirical power for locally size-corrected tests is reported below.

$H_1 : r(\alpha) \neq 0\ \ \alpha_{xz} = \delta = 0.1269$						
$\alpha_{xx} = \alpha_{yy} = \alpha_{zz} = 0.3$						
$c_n = \hat{\lambda}_1 n^{-1/3}$, $c = 0.001;\ [0.1]$						
	$n = 50$		$n = 100$		$n = 200$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	9.18	8.98	18.54	18	40.28	38.46
W_{DV}	[8.54]	[8.50]	[16.92]	[16.72]	[37.36]	[36.46]
W_{LB}	5.20	3.58	5.20	3.86	5.78	4.96
	$n = 500$		$n = 1000$		$n = 2000$	
	Asy.	Boot.	Asy.	Boot.	Asy.	Boot.
W	83.12	78.98	98.66	97.76	100	100
W_{DV}	[81.0]	[77.66]	[98.54]	[97.62]	[100]	[100]
W_{LB}	6.86	4.84	8.88	6.26	12.10	10.28

8.2. Remarks on simulation results

- The unmodified standard Wald test is shown to be conservative *when it works*.
- Our regularized test based on a fixed threshold c is always implementable, displays good size properties when conducted in its simulated version, has good power properties under different types of alternatives unlike the Lütkepohl and Burda modified Moore-Penrose Wald test.
- A *strong* regularization scheme ($c = 0.1$) against a weak regularization ($c = 0.001$) has better size properties in small samples without a substantial loss of power relative to the infeasible size-corrected tests when we are close to the nonstationary region.

9. Conclusion

In this paper, we introduce a new class of inverses as opposed to generalized inverses with the difference that the regularized inverses enjoy better stability properties. Exploiting the continuity property of the total eigenprojections stated in Tyler (1981) enables us to built continuous regularized inverses that greatly simplifies the derivation of distributional results for the regularized tests statistics. Two classes of tests can be defined depending on the choice of the threshold in the specification of the VRF. Further, the decomposition property enjoyed by the regularized inverses makes it easier to find some distributional bounds for the nonstandard case. Thus, the standard distributional results can be used to provide asymptotically valid tests in a convenient way. However, the usual critical point may be conservative. In this respect, a simulation-based approach can improve the performance of the procedure under the alternative. Finally, the tuning parameter requires a deeper examination together with the study of monotonic variance regularization function.

A. Appendix: Proofs

PROOF of Property 1 Recall the spectral decomposition

$$\Sigma = V \Lambda V' = V \mathrm{diag}\big[\lambda_1, \lambda_2, \ldots, \lambda_q\big] V'$$

with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q \geq 0$.

Let us decompose the matrix Σ as:

$$\Sigma = V(c)\Lambda(c)V(c)' = V_1(c)\Lambda_1(c)V_1(c)' + V_2(c)\Lambda_2(c)V_2(c)' + V_3(c)\Lambda_3(c)V_3(c)'$$

with

$$\Lambda_1(c) = \mathrm{diag}\big[\lambda_1, \ldots, \lambda q_1\big] \text{ such that } \lambda > c$$

$$\Lambda_2(c) = \mathrm{diag}\big[c, \ldots, c\big] \text{ such that } \lambda = c \text{ with multiplicity } m(c)$$

and

$$\Lambda_3(c) = \mathrm{diag}\big[\lambda_{q_1+m(c)+1}, \ldots, \lambda q\big] \text{ such that } \lambda < c\ .$$

Let us decompose its regularized inverse likewise:

$$\Sigma^R = V(c)\Lambda^\dagger(c)V(c)' = V_1(c)\Lambda_1^\dagger[\bar{\lambda}; c]V_1'(c) + V_2(c)\Lambda_2^\dagger[\bar{\lambda}; c]V_2'(c) + V_3(c)\Lambda_3^\dagger[\bar{\lambda}; c]V_3'(c)\ .$$

We have:

$$\Sigma\Sigma^R = V(c)\Lambda(c)\Lambda^\dagger(c)V(c)' = V_1(c)\Lambda_1(c)\Lambda_1^\dagger[\bar{\lambda}; c]V_1'(c) + V_2(c)\Lambda_2(c)\Lambda_2^\dagger[\bar{\lambda}; c]V_2'(c) + V_3(c)\Lambda_3(c)\Lambda_3^\dagger[\bar{\lambda}; c]V_3'(c)$$

because the V_i's are orthogonal matrices. As $g[\lambda; c]$ is a non-increasing function in λ, with

$$\Lambda_1^\dagger[\bar{\lambda}; c] = \mathrm{diag}\big[g[\lambda_1; c], \ldots, g[\lambda_{q_1}; c]\big] = \Lambda_1(c)^{-1} \text{ for } \lambda > c\ ,\ \Rightarrow\ \Lambda_1(c)\Lambda_1^\dagger = \Lambda_1(c)\Lambda_1(c)^{-1} = Id\ .$$

To alleviate the notations $\Lambda_i^\dagger$ stands for $\Lambda_i^\dagger[\bar{\lambda}; c]$, for all i and for the whole matrix as well. As for $i = 2, 3$:

$$\Lambda_i^\dagger[\bar{\lambda}; c] = \mathrm{diag}\big[g[\lambda_j; c], \ldots, g[\lambda_q; c]\big] \text{ for } \lambda \leq c$$

$$\Rightarrow\ \Lambda_i(c)\Lambda_i^\dagger = \begin{pmatrix} \lambda_j g[\lambda_j; c] & 0 & 0\ldots0 & 0 \\ 0 & \lambda_{j+1} g[\lambda_{j+1}; c] & 0\ldots0 & 0 \\ \vdots & 0 & 0\ldots0 & \lambda_{j'} g[\lambda_{j'}; c] \end{pmatrix} \leq Id$$

since $\lambda_j g[\lambda_j; c] \leq 1$ for $\lambda_j \leq c$. This means that for all nonzero vector x:

$$x'\big[Id - \Lambda_i(c)\Lambda_i^\dagger\big]x \quad = \quad \begin{bmatrix} x_j & \ldots & x_{j'} \end{bmatrix} \begin{pmatrix} 1 - \lambda_j g[\lambda_j; c] & 0 & 0\ldots0 & 0 \\ 0 & 1 - \lambda_{j+1} g[\lambda_{j+1}; c] & 0\ldots0 & \vdots \\ \vdots & 0 & 0\ldots0 & 1 - \lambda_{j'} g[\lambda_{j'}; c] \end{pmatrix} \begin{bmatrix} x_j \\ \ldots \\ x_{j'} \end{bmatrix}$$

$$= \sum_{k=j}^{j'} x_k \big[1 - \lambda_k g[\lambda_k; c]\big] x_k \geq 0 . \tag{A.1}$$

Since

$$\Lambda_i(c)\Lambda_i^{\dagger} \leq Id \Rightarrow V_i(c)\Lambda_i(c)\Lambda_i^{\dagger}V_i(c)' \leq V_i(c)V_i(c)' = Id .$$

Since $\lambda g[\lambda; c] \leq 1$ all λ then we also have for all nonzero vector x:

$$\begin{aligned} x'\big[Id - \Lambda(c)\Lambda^{\dagger}\big]x &= \begin{bmatrix} x_1 & \dots & x_q \end{bmatrix} \begin{pmatrix} 1-\lambda_1 g[\lambda_1;c] & 0 & 0\dots0 & 0 \\ 0 & 1-\lambda_2 g[\lambda_2;c] & 0\dots0 & \vdots \\ \vdots & 0 & 0\dots0 & 1-\lambda_q g[\lambda_q;c] \end{pmatrix} \begin{bmatrix} x_1 \\ \dots \\ x_q \end{bmatrix} \\ &= \sum_{k=1}^{q} x_k \big[1 - \lambda_k g[\lambda_k; c]\big] x_k \geq 0 . \end{aligned} \tag{A.2}$$

$$\Lambda(c)\Lambda^{\dagger}(c) \leq Id \Rightarrow V(c)\Lambda(c)\Lambda^{\dagger}(c)V(c)' \leq Id$$

hence

$$\Sigma\Sigma^R \leq Id .$$

Obviously, we have

$$\Lambda^{\dagger}(c)\Lambda(c) \leq Id \Rightarrow V(c)\Lambda^{\dagger}(c)\Lambda(c)V(c)' \leq Id$$

hence

$$\Sigma^R\Sigma \leq Id .$$

Postmultiplying $\Sigma\Sigma^R$ with Σ we have:

$$\begin{aligned} \Sigma\Sigma^R\Sigma &= V_1 \underbrace{\Lambda_1\Lambda_1^{-1}}_{=Id} \Lambda_1 V_1' + V_2 \underbrace{\Lambda_2\Lambda_2^{\dagger}}_{\leq Id} \Lambda_2 V_2' + V_3 \underbrace{\Lambda_3\Lambda_3^{\dagger}}_{\leq Id} \Lambda_3 V_3' \\ &\leq V_1\Lambda_1V_1' + V_2\Lambda_2V_2' + V_3\Lambda_3V_3' = \Sigma , \end{aligned} \tag{A.3}$$

where the dependance on c has been intentionally dropped for clarity purposes. Concerning result iii), we have:

$$\big(\Sigma^R\big)^{-1} = V \begin{pmatrix} \big(\Lambda_1^{\dagger}\big)^{-1} & 0 & 0 \\ 0 & \big(\Lambda_2^{\dagger}\big)^{-1} & 0 \\ 0 & 0 & \big(\Lambda_3^{\dagger}\big)^{-1} \end{pmatrix} V' .$$

Recalling that $g[\lambda; c]$ is a non-increasing function of λ such that

$$g(\lambda; c) \leq \frac{1}{\lambda} \text{ for } \lambda \leq c$$

hence

$$\Big(g(\lambda;c)\Big)^{-1} \geq \lambda \ \text{ for } \ \lambda \leq c \ \Leftrightarrow \ \Big(g(\lambda;c)\Big)^{-1} - \lambda \ \geq \ 0 \ .$$

Further,

$$g(\lambda;c) = \frac{1}{\lambda} \ \text{ for } \lambda > c \ .$$

As this holds component by component, we have

$$\text{diag}[\big(g(\lambda;c)\big)^{-1} - \lambda] \geq \ 0 \ \ \forall \ \lambda \ .$$

Hence,

$$\begin{aligned} \big(\Sigma^R\big)^{-1} - \Sigma &= V \begin{pmatrix} \big(\Lambda_1^\dagger\big)^{-1} - \Lambda_1 & 0 & 0 \\ 0 & \big(\Lambda_2^\dagger\big)^{-1} - \Lambda_2 & 0 \\ 0 & 0 & \big(\Lambda_3^\dagger\big)^{-1} - \Lambda_3 \end{pmatrix} V' \\ &= V_1\big(\Lambda_1 - \Lambda_1\big)V_1' + V_2\Big(\big(\Lambda_2^\dagger\big)^{-1} - \Lambda_2\Big)V_2' + V_3\Big(\big(\Lambda_3^\dagger\big)^{-1} - \Lambda_3\Big)V_3' \geq 0 \end{aligned} \tag{A.4}$$

Finally, to prove result $iv)$ let $\mathcal{M}(\Sigma)$ (and $\mathcal{M}(\Sigma^R)$) denote the linear space spanned by the columns of Σ (and Σ^R respectively). Recall that $\text{rank}(\Sigma)$ is defined as the number of independent columns or rows of Σ. As $\mathcal{M}(\Sigma) \subseteq \mathcal{M}(\Sigma^R)$, then $\text{rank}(\Sigma^R) \geq \text{rank}(\Sigma)$. □

PROOF of Lemma 5.3 If $\hat{\Sigma}_n \overset{a.s.}{\to} \Sigma$ then the event $A = \{\omega : \ \hat{\Sigma}_n(\omega) \underset{n\to\infty}{\to} \Sigma\}$ has probability one: $P(A) = 1$. For any $\omega \in A$, we have by Lemma **5.2**:

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

By denoting $B = \{\omega : \ \lambda_j(\hat{\Sigma}_n(\omega)) \underset{n\to\infty}{\to} \lambda_j(\Sigma)\}$, we have $A \subseteq B$, hence we have with probability one the result i) stated in the Lemma. By using the same type of argument, we have the result ii) for the eigenprojections.

□

PROOF of Lemma 5.4

If $\hat{\Sigma}_n \overset{p}{\to} \Sigma$ with eigenvalues $\{\lambda_j(\hat{\Sigma}_n)\}$, then every subsequence $\{\hat{\Sigma}_{n_k}\}$ with eigenvalues $\{\lambda(\hat{\Sigma}_{n_k})\}$, also satisfies $\hat{\Sigma}_{n_k} \overset{p}{\to} \Sigma$. By Lukacs (1975, theorem 2.4.3, page 48), there exists $\{\hat{\Sigma}_{m_l}\} \subseteq \{\hat{\Sigma}_{n_k}\}$ such that $\hat{\Sigma}_{m_l} \overset{a.s.}{\to} \Sigma$. Hence by Lemma **5.3**, we have

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

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

As we have $\{\hat{\Sigma}_{m_l}\} \subseteq \{\hat{\Sigma}_{n_k}\} \subseteq \{\hat{\Sigma}_n\}$ with the corresponding eigenvalues $\{\lambda_j(\hat{\Sigma}_{m_l})\} \subseteq$

$\{\lambda_j(\hat{\Sigma}_{n_k})\} \subseteq \{\lambda_j(\hat{\Sigma}_n)\}$ by Lukacs (1975, theorem 2.4.4 page 49), it suffices that every subsequence $\{\lambda_j(\hat{\Sigma}_{n_k})\}$ of $\{\lambda_j(\hat{\Sigma}_n)\}$ contains a subsequence $\{\lambda_j(\hat{\Sigma}_{m_l})\}$ which converges almost surely, to get the result of convergence in probability: $\lambda_j(\hat{\Sigma}_n) \xrightarrow{p} \lambda_j(\Sigma)$. By applying the same type of argument, we get $P_{j,t}(\hat{\Sigma}_n) \xrightarrow{p} P_{j,t}(\Sigma)$. □

PROOF of Proposition 6.1

We shall first focus on the case where all the eigenvalues are distinct before considering multiple eigenvalues. Let abbreviate $\hat{\lambda}_j = \lambda_j(\hat{\Sigma}_n(\omega))$, $P(\hat{\lambda}_j) = P_{j,j}[\hat{\Sigma}_n(\omega)]$, $\lambda_j = \lambda_j(\Sigma)$ and $P(\lambda_j) = P_{j,j}(\Sigma)$. Recalling equation (6.1) we have:

$$\hat{\Sigma}_n^R(\omega) = \sum_{\hat{\lambda}_j \in \mathcal{S}(\hat{\Sigma}_n)} g(\hat{\lambda}_j;c)P(\hat{\lambda}_j) \quad and \quad \Sigma^R = \sum_{\lambda_j \in \mathcal{S}(\Sigma)} g(\lambda_j;c)P(\lambda_j) \tag{A.5}$$

If $\hat{\Sigma}_n \xrightarrow{a.s.} \Sigma$ then by lemma **5.3** i), we have $\hat{\lambda}_j \xrightarrow{a.s.} \lambda_j$ for all j, hence by continuity of $g(.,c)$ we have $g(\hat{\lambda}_j;c) \xrightarrow{a.s.} g(\lambda_j;c)\ \forall\ j$. By lemma **5.3** ii) and the continuity of the product, we have $g(\hat{\lambda}_j;c)P(\hat{\lambda}_j) \xrightarrow{a.s.} g(\lambda_j;c)P(\lambda_j)\ \forall\ j$. The convergence for the sum follows from the component-wise convergence. Hence $\hat{\Sigma}_n^R \xrightarrow{a.s.} \Sigma^R$ for $m(\lambda) = 1$.

Now, let us establish the consistency property in the case of multiple eigenvalues.Suppose that λ_j has multiplicity $m(\lambda_j)$. Let $I_j = \{i \in I : \lambda_i = \lambda_j\}$ so that $\mathrm{card}(I_j) = m(\lambda_j)$.

For simplicity let us focus on the corresponding part of the spectral decomposition of $\hat{\Sigma}_n$, *i.e.*

$$\begin{aligned}
\sum_{i\in I_j} g(\hat{\lambda}_i;c)P(\hat{\lambda}_i) &= g(\hat{\lambda}_j;c)P(\hat{\lambda}_j) + \sum_{i\in I_j\setminus\{j\}} g(\hat{\lambda}_i;c)P(\hat{\lambda}_i) \\
&= g(\hat{\lambda}_j;c)P(\hat{\lambda}_j) + \sum_{i\in I_j\setminus\{j\}} [g(\hat{\lambda}_i;c) - g(\hat{\lambda}_j;c) + g(\hat{\lambda}_j;c)]P(\hat{\lambda}_i) \\
&= g(\hat{\lambda}_j;c)P(\hat{\lambda}_j) + \sum_{i\in I_j\setminus\{j\}} g(\hat{\lambda}_j;c)P(\hat{\lambda}_i) + \sum_{i\in I_j\setminus\{j\}} [g(\hat{\lambda}_i;c) - g(\hat{\lambda}_j;c)]P(\hat{\lambda}_i) \\
&= g(\hat{\lambda}_j;c)\sum_{i\in I_j} P(\hat{\lambda}_i) + \sum_{i\in I_j\setminus\{j\}} [g(\hat{\lambda}_i;c) - g(\hat{\lambda}_j;c)]P(\hat{\lambda}_i) \\
&\xrightarrow{a.s.} g(\lambda_j;c)\sum_{i\in I_j} P(\lambda_i) + \sum_{i\in I_j\setminus\{j\}} [g(\lambda_j;c) - g(\lambda_j;c)]P(\lambda_i)
\end{aligned} \tag{A.6}$$

by Lemma **5.3** and by continuity of $g(.;c)$. Indeed, $\forall i,j \in I_j$, $\lim_{n\to\infty} g(\hat{\lambda}_i;c) = \lim_{n\to\infty} g(\hat{\lambda}_j;c) = g(\lambda_j;c)$. For the projection operator P is a continuous linear operator; following Kolmogorov and Fomin (1970, theorem 1, page 223), a necessary condition for a linear operator P to be continuous on a topological linear space is that P be *bounded*. Therefore, equation (A.6) becomes

$$\sum_{i\in I_j} g(\hat{\lambda}_i;c)P(\hat{\lambda}_i) \xrightarrow{a.s.} g(\lambda_j;c)\sum_{i\in I_j} P(\lambda_i) + \sum_{i\in I_j\setminus\{j\}} \underbrace{[g(\lambda_j;c) - g(\lambda_j;c)]}_{0}\underbrace{P(\lambda_i)}_{O(1)}$$

$$\stackrel{a.s.}{\to} \quad g(\lambda_j; c) \sum_{i \in I_j} P(\lambda_i) \tag{A.7}$$

As the result holds for any λ_j with an arbitrary multiplicity $m(\lambda_j) > 1$, hence

$$\hat{\Sigma}^R = \sum_{\hat{\lambda} \in \mathcal{S}(\hat{\Sigma})} g(\hat{\lambda}; c) P(\hat{\lambda}) \stackrel{a.s.}{\to} \sum_{\lambda \in \mathcal{S}(\Sigma)} g(\lambda; c) P(\lambda) = \Sigma^R \quad \forall \;\; m(\lambda) \geq \; 1 \; .$$

□

Proof of Proposition 6.5 We wish to show that $lim_{n\to\infty} Pr\big[\|(\Sigma_n)^R - (\Sigma)^R\| > \epsilon\big] \to 0$ for every $\epsilon > 0$. Let $I = \{1, 2, \ldots, q\}$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q$, and $J = \{1, 2, \ldots, k\}$ the subset of I corresponding to the distinct eigenvalues of Σ: $\lambda_1 > \lambda_2 > \ldots > \lambda_j > \ldots > \lambda_k$, with $\sum_{j=1}^{k} m(\lambda_j) = q$, and $q \geq 1$ and $k \geq 1$. For $j \in J$, let $I_j = \{i \in I \;\; : \lambda_i = \lambda_j\}$, so $\text{card}(I_j) = m(\lambda_j)$ and $\sum_{i=1}^{q} \lambda_i = \sum_{i \in I} \lambda_i = \sum_{j \in J} \sum_{i \in I_j} \lambda_i$. Let $P(\lambda_j)$ represent the eigenprojection onto the eigenspace $\mathcal{V}(\lambda_j)$ associated with λ_j. $P(\lambda_j)$ has the representation $P(\lambda_j) = \sum_{i=1}^{m(\lambda_j)} x_i x_i' = \sum_{i \in I_j} x_i x_i'$, where $\{x_i\}_{i \in I_j}$ is an orthonormal basis for $\mathcal{V}(\lambda_j)$. Let $\nu_{(j)} = \min\{i : i \in I_j\}$ and $\hat{\lambda}_{n,\nu(j)}$ an estimate of λ_j. Similarly, we define $P_j(\hat{\lambda}_{n,\nu(j)}) = \sum_{i \in I_j} \hat{x}_i \hat{x}_i'$ the eigenprojection onto the eigenspace $\mathcal{V}(\hat{\lambda}_{n,\nu(j)})$ associated with $\hat{\lambda}_{n,\nu(j)}$ with $\{\hat{x}_i\}_{i \in I_j}$ being an orthonormal basis for $\mathcal{V}(\hat{\lambda}_{n,\nu(j)})$. We also recall that the total eigenprojection for Σ is defined as $\sum_{\lambda \in v} P(\lambda)$ for any subset v of the spectral set $\mathcal{S} = \{\lambda_1, \ldots, \lambda_q\}$. As Σ is positive semidefinite, $\lambda_i \geq 0 \;\; \forall \; i \in I$. If there exist some eigenvalues $\lambda_i = 0$, then the smallest distinct eigenvalue is such that $\lambda_k = 0$. Let J_{-k} contain the nonzero distinct eigenvalues of Σ, *i.e.*

$$J_{-k} = \begin{cases} J & \text{if } \lambda_k \neq 0 \\ \{1, \ldots, k-1\} & \text{if } \lambda_k = 0 \\ \emptyset & \text{if } \lambda_1 = 0 \; . \end{cases} \tag{A.8}$$

First, let us consider the case that $\lambda_1 = 0$ with multiplicity $m(0) = q$. In this case, $\Sigma_n \stackrel{P}{\to} \Sigma = 0$, *i.e.* Σ_n converges to a degenerate 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 $P_0(\Sigma_n)$ denote the eigenprojection of Σ_n associated with its zero eigenvalue which projects onto the corresponding eigenspace $\mathcal{V}_0(\hat{\lambda}_n)$. Following Tyler (1981), $P_0(\Sigma_n) \;\; \stackrel{P}{\to} \;\; P_0(\Sigma)$, or equivalently the eigenspace $\mathcal{V}_0(\hat{\lambda}_n)$ converges to $\mathcal{V}_0$, then

$$\Sigma_n^R = \frac{1}{\epsilon} P_0(\Sigma_n) \;\; \stackrel{P}{\to} \frac{1}{\epsilon} P_0(\Sigma) = \Sigma^R \; .$$

Let us now study the case that $\lambda_k = 0$.

$$
\begin{aligned}
\|(\Sigma_n)^R - (\Sigma)^R\| &= \|\sum_{i\in I} g(\hat{\lambda}_{n,i}; c(n))P(\hat{\lambda}_{n,i}) - \sum_{i\in I} g(\lambda_i; c(n))P(\lambda_i)\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} g(\hat{\lambda}_{n,i}; c(n))P(\hat{\lambda}_{n,i}) - \sum_{j\in J}\sum_{i\in I_j} g(\lambda_i; c(n))P(\lambda_i)\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} g(\hat{\lambda}_{n,i}; c(n))P(\hat{\lambda}_{n,i}) - \sum_{j\in J} g(\lambda_j; c(n))\sum_{i\in I_j} P(\lambda_i)\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} g(\hat{\lambda}_{n,i}; c(n))P(\hat{\lambda}_{n,i}) - \sum_{j\in J} g(\lambda_j; c(n))P(\lambda_j)\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} g(\hat{\lambda}_{n,i}; c(n))P(\hat{\lambda}_{n,i}) - \sum_{j\in J} g(\lambda_j; c(n))\sum_{i\in I_j} x_i x_i'\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} \big[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))\big]P(\hat{\lambda}_{n,i}) + \sum_{j\in J} g(\hat{\lambda}_{n,\nu(j)}; c(n))\sum_{i\in I_j} P(\hat{\lambda}_{n,i}) \\
&\quad - \sum_{j\in J} g(\lambda_j; c(n))\sum_{i\in I_j} x_i x_i'\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} \big[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))\big]P(\hat{\lambda}_{n,i}) + \sum_{j\in J} g(\hat{\lambda}_{n,\nu(j)}; c(n))\big[P(\hat{\lambda}_{n,\nu(j)}) + o_p(1)\big] \\
&\quad - \sum_{j\in J} g(\lambda_j; c(n))P(\lambda_j)\| \\
&= \|\sum_{j\in J}\sum_{i\in I_j} \big[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))\big]P(\hat{\lambda}_{n,i}) + \sum_{j\in J_{-k}} g(\hat{\lambda}_{n,\nu(j)}; c(n))\big[P(\hat{\lambda}_{n,\nu(j)}) + o_p(1 \\
&\quad + g(\hat{\lambda}_{n,\nu(k)}; c(n))\big[P(\hat{\lambda}_{n,\nu(k)}) + o_p(1)\big] - \sum_{j\in J_{-k}} g(\lambda_j; c(n))P(\lambda_j) - g(\lambda_k; c(n))P(\lambda_k)\| \\
&= \|\sum_{j\in J_{-k}} g(\hat{\lambda}_{n,\nu(j)}; c(n))\big[P(\hat{\lambda}_{n,\nu(j)}) + o_p(1)\big] - \sum_{j\in J_{-k}} g(\lambda_j; c(n))P(\lambda_j) \\
&+ g(\hat{\lambda}_{n,\nu(k)}; c(n))\big[P(\hat{\lambda}_{n,\nu(k)}) + o_p(1)\big] - g(\lambda_k; c(n))P(\lambda_k) \\
&+ \sum_{j\in J}\sum_{i\in I_j} \big[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))\big]P(\hat{\lambda}_{n,i})\| \\
&\leq A + B + C \qquad \text{(A.}
\end{aligned}
$$

using the triangle inequality property for the norm, where

$$
A = \|\sum_{j\in J_{-k}} g(\hat{\lambda}_{n,\nu(j)}; c(n))P(\hat{\lambda}_{n,\nu(j)}) - \sum_{j\in J_{-k}} g(\lambda_j; c(n))P(\lambda_j)\|
$$

where, without loss of generality, we have dropped the $o_p(1)$ terms,

$$B = \|g(\hat{\lambda}_{n,\nu(k)}; c(n))P(\hat{\lambda}_{n,\nu(k)}) - g(\lambda_k; c(n))P(\lambda_k)\|$$

and

$$C = \| \sum_{j \in J} \sum_{i \in I_j} \big[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))\big] P(\hat{\lambda}_{n,i})\| \ .$$

Let us first focus on A:

$$\begin{aligned}
A &= \| \sum_{j \in J_{-k}} g(\hat{\lambda}_{n,\nu(j)}; c(n))P(\hat{\lambda}_{n,\nu(j)}) - \sum_{j \in J_{-k}} g(\lambda_j; c(n))P(\lambda_j)\| \\
&= \| \sum_{j \in J_{-k}} \big[g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))\big] P(\hat{\lambda}_{n,\nu(j)}) \\
&+ \sum_{j \in J_{-k}} g(\lambda_j; c(n))P(\hat{\lambda}_{n,\nu(j)}) - \sum_{j \in J_{-k}} g(\lambda_j; c(n))P(\lambda_j)\| \\
&= \| \sum_{j \in J_{-k}} \big[g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))\big] P(\hat{\lambda}_{n,\nu(j)}) \\
&+ \sum_{j \in J_{-k}} g(\lambda_j; c(n)) \Big[P(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\Big] \| \\
&\leq \| \sum_{j \in J_{-k}} \big[g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))\big] P(\hat{\lambda}_{n,\nu(j)})\| \\
&+ \| \sum_{j \in J_{-k}} g(\lambda_j; c(n)) \big[P(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\big] \| \\
&\leq \sum_{j \in J_{-k}} |g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))| \ \|P(\hat{\lambda}_{n,\nu(j)})\| \\
&+ \| \sum_{j \in J_{-k}} g(\lambda_j; c(n)) \big[P_j(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\big] \|
\end{aligned} \tag{A.10}$$

Let $A_1 = \sum_{j \in J_{-k}} |g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))| \|P(\hat{\lambda}_{n,\nu(j)})\|$ and $A_2 = \| \sum_{j \in J_{-k}} g(\lambda_j; c(n)) \big[P(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\big] \|$.

$$A_1 \leq \max_{j \in J_{-k}} |g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))| \sum_{j \in J_{-k}} \|P(\hat{\lambda}_{n,\nu(j)})\| \tag{A.11}$$

As $P(\hat{\lambda}_{n,\nu(j)})$ is bounded in probability, then the quantity $\sum_{j \in J_{-k}} \|P(\hat{\lambda}_{n,\nu(j)})\|$ is bounded. let us study now the quantity $lim_{n \to \infty} Pr[\max_{j \in J_{-k}} |g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))| > \epsilon]$.

As $\hat{\Sigma}_n \xrightarrow{\text{p}} \Sigma$, then by Lemma **5.4** i), we have $\hat{\lambda}_{n,i} \xrightarrow{\text{p}} \lambda_j, \ \ \forall \ i \in I_j$. The result still holds for

$\min_{i \in I_j} \hat{\lambda}_{n,i} = \hat{\lambda}_{n,\nu(j)} \overset{\text{p}}{\to} \lambda_j$. Hence, with $c(n) \underset{n\to\infty}{\to} 0$

$$Pr\big[|\hat{\lambda}_{n,\nu(j)} - \lambda_j| > c(n)\big] = Pr\big[|b_n(\hat{\lambda}_{n,\nu(j)} - \lambda_j)| > b_n c(n)\big] \underset{n\to\infty}{\to} 0 \qquad \text{(A.12)}$$

since $b_n c(n) \to \infty$ slower than b_n does, and $b_n\big(\hat{\lambda}_{n,\nu(j)} - \lambda_j\big)$ converges in distribution by Lemma **6.4**. As $\hat{\lambda}_{n,\nu(j)} \to \lambda_j$ in probability and g is continuous a.e. except on a set with measure zero, then $lim_{n\to\infty} Pr[\max_{j\in J_{-k}} |g(\hat{\lambda}_{n,\nu(j)}; c(n)) - g(\lambda_j; c(n))| > \epsilon] = 0$. Hence $lim_{n\to\infty} Pr[A_1 > \epsilon] = 0$. Similarly,

$$\begin{aligned} A_2 = \| \sum_{j\in J_{-k}} g(\lambda_j; c(n)) \big[P(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\big] \| &\leq \| \max_{j\in J_{-k}} g(\lambda_j; c(n)) \sum_{j\in J_{-k}} \big[P(\hat{\lambda}_{n,\nu(j)}) - P(\lambda_j)\big] \| \\ &\leq |\max_{j\in J_{-k}} g(\lambda_j; c(n))| \, \| \sum_{j\in J_{-k}} P(\hat{\lambda}_{n,\nu(j)}) - \sum_{j\in J_{-k}} P(\lambda_j) \| \,. \qquad \text{(A.13)} \end{aligned}$$

Following Tyler (1981) and Lemma **5.4** ii), $\| \sum_{j\in J_{-k}} P_j(\hat{\lambda}_{n,\nu(j)}) - \sum_{j\in J_{-k}} P(\lambda_j) \|$ converges to zero in probability, provided that $\lambda_1 \neq \lambda_2$ and $\lambda_{k-1} \neq \lambda_k$. As we work on the set consisted of the distinct eigenvalues, we can select one eigenvalue before and one after such that the condition is satisfied. Hence, $lim_{n\to\infty} Pr[A_2 > \epsilon] = 0$. As $A \leq A_1 + A_2$, with each quantity converging to zero in probability, then $lim_{n\to\infty} Pr[A > \epsilon] = 0$.

Concerning B, we have:

$$\begin{aligned} B &= \|g(\hat{\lambda}_{n,\nu(k)}; c(n)) P(\hat{\lambda}_{n,\nu(k)}) - g(\lambda_k; c(n)) P(\lambda_k)\| \\ &= \| \big[g(\hat{\lambda}_{n,\nu(k)}; c(n)) - g(\lambda_k; c(n))\big] P(\hat{\lambda}_{n,\nu(k)}) + g(\lambda_k; c(n)) P(\hat{\lambda}_{n,\nu(k)}) - g(\lambda_k; c(n)) P(\lambda_k)\| \\ &\leq \| \big[g(\hat{\lambda}_{n,\nu(k)}; c(n)) - g(\lambda_k; c(n))\big] P(\hat{\lambda}_{n,\nu(k)})\| + \|g(\lambda_k; c(n)) \big[P(\hat{\lambda}_{n,\nu(k)}) - P(\lambda_k)\big]\| \\ &= |g(\hat{\lambda}_{n,\nu(k)}; c(n)) - g(\lambda_k; c(n))| \, \|P(\hat{\lambda}_{n,\nu(k)})\| + \|g(\lambda_k; c(n)) \big[P(\hat{\lambda}_{n,\nu(k)}) - P(\lambda_k)\big]\| \\ &= |g(\hat{\lambda}_{n,\nu(k)}; c(n)) - g(\lambda_k; c(n))| \, \underbrace{\|P(\hat{\lambda}_{n,\nu(k)})\|}_{O_p(1)} + |g(\lambda_k; c(n))| \, \|P(\hat{\lambda}_{n,\nu(k)}) - P(\lambda_k)\| \end{aligned}$$

$|g(\lambda_k; c(n))| = O(1)$ by definition of the regularizing function $g(.)$. By Lemma **5.4** ii), we have $\|P(\hat{\lambda}_{n,\nu(k)}) - P(\lambda_k)\| \overset{\text{p}}{\to} 0$. Again, from equation (A.12) and with $\hat{\lambda}_{n,\nu(k)}) \overset{\text{p}}{\to} \lambda_k$, g being continuous almost everywhere (a.e.) except on a set with measure zero, then

$$lim_{n\to\infty} Pr[|g(\hat{\lambda}_{n,\nu(k)}; c(n)) - g(\lambda_k; c(n))| > \epsilon] = 0 \,.$$

Hence, B converges in probability to zero.

As for $C = \| \sum_{j \in J} \sum_{i \in I_j} \left[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n)) \right] P(\hat{\lambda}_{n,i}) \|$, we have:

$$C \leq \sum_{j \in J} \| \sum_{i \in I_j} \left[g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n)) \right] P(\hat{\lambda}_{n,i}) \| \leq \sum_{j \in J} \sum_{i \in I_j} |g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n))| \underbrace{\|P(\hat{\lambda}_{n,i})\|}_{O_p(1)} . \tag{A.14}$$

As i and $\nu(j)$ both belong to I_j, $\hat{\lambda}_{n,i}$ and $\hat{\lambda}_{n,\nu(j)}$ both converge to λ_j. As g is continuous a.e. with the set of discontinuities of g having measure zero, hence $g(\hat{\lambda}_{n,i}; c(n)) - g(\hat{\lambda}_{n,\nu(j)}; c(n)) \xrightarrow{p} 0 \; \forall \, i \in I_j$. As this holds for an arbitrary $j \in J$, it still holds for any j. Hence, the result still holds for the sum over all j in J. Hence $C \xrightarrow{p} 0$.

Since A, B and C converge to zero in probability, we can finally conclude that:

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

□

PROOF of Proposition 6.7 Under Assumption **6.6**, we have $Z_n \xrightarrow{\mathcal{L}} Z_\infty \sim N[0, \Sigma]$ and by Proposition **6.5**, we have $[\Sigma_n]^R \xrightarrow{p} [\Sigma]^R$. Hence,

$$Z_n' [\Sigma_n]^R Z_n \xrightarrow{\mathcal{L}} Z_\infty' [\Sigma]^R Z_\infty \quad . \tag{A.15}$$

Further, let

$$\begin{aligned} X_n &= Z_n' [\Sigma_n]^R Z_n = Z_n' \Big[\sum_{\hat{\lambda}_n \in \mathcal{S}(\Sigma_n)} g(\hat{\lambda}_n, c_n) P(\hat{\lambda}_n) \Big] Z_n \\ &= Z_n' \Big[\sum_{\hat{\lambda}_n \in \hat{v}_1} g(\hat{\lambda}_n, c_n) P(\hat{\lambda}_n) \Big] Z_n + Z_n' \Big[\sum_{\hat{\lambda}_n \in \hat{v}_2} g(\hat{\lambda}_n, c_n) P(\hat{\lambda}_n) \Big] Z_n \end{aligned}$$

Let

$$X_{1n} = Z_n' \Big[\sum_{\hat{\lambda}_n \in \hat{v}_1} g(\hat{\lambda}_n, c_n) P(\hat{\lambda}_n) \Big] Z_n$$

and

$$X_{2n} = Z_n' \Big[\sum_{\hat{\lambda}_n \in \hat{v}_2} g(\hat{\lambda}_n, c_n) P(\hat{\lambda}_n) \Big] Z_n \; .$$

For the sake of simplicity, let

$$\begin{aligned} \Sigma_{11,n}^R &= \sum_{\hat{\lambda} \in \hat{v}_1} g(\hat{\lambda}, c_n) P(\hat{\lambda}) = \sum_{j \in v_1} \sum_{i \in I_j} g(\hat{\lambda}_i, c_n) P(\hat{\lambda}_i) \\ &= \sum_{j \in v_1} \Big[g(\hat{\lambda}_j, c_n) P(\hat{\lambda}_j) + \sum_{i \in I_j \setminus \{j\}} g(\hat{\lambda}_i, c_n) P(\hat{\lambda}_i) \Big] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\in v_1}\Big[g(\hat\lambda_j,c_n)P(\hat\lambda_j)+\sum_{i\in I_j\backslash\{j\}}[g(\hat\lambda_i,c_n)-g(\hat\lambda_j,c_n)+g(\hat\lambda_j,c_n)]P(\hat\lambda_i)\Big]\\
&= \sum_{j\in v_1}\Big[g(\hat\lambda_j,c_n)P(\hat\lambda_j)+\sum_{i\in I_j\backslash\{j\}}g(\hat\lambda_j,c_n)P(\hat\lambda_i)+\sum_{i\in I_j\backslash\{j\}}[g(\hat\lambda_i,c_n)-g(\hat\lambda_j,c_n)]P(\hat\lambda_i)\Big]\\
&= \sum_{j\in v_1}\Big[g(\hat\lambda_j,c_n)\sum_{i\in I_j}P(\hat\lambda_i)+\sum_{i\in I_j\backslash\{j\}}[g(\hat\lambda_i,c_n)-g(\hat\lambda_j,c_n)]P(\hat\lambda_i)\Big]
\end{aligned}
$$

Using the continuity property of the eigenvalues and total eigenprojections given in **5.4** i) and ii), and under the assumption that $g(.,c)$ is continuous a.e., we have

$$
\begin{aligned}
\Sigma^R_{11,n} &= \sum_{j\in v_1}\Big[g(\hat\lambda_j,c_n)\sum_{i\in I_j}P(\hat\lambda_i)+\sum_{i\in I_j\backslash\{j\}}[g(\hat\lambda_i,c_n)-g(\hat\lambda_j,c_n)]P(\hat\lambda_i)\Big]\\
&\xrightarrow{\mathbf{P}} \sum_{j\in v_1}\Big[g(\lambda_j,0)P(\lambda_j)+\underbrace{[g(\lambda_j,0)-g(\lambda_j,0)]}_{0}\sum_{i\in I_j\backslash\{j\}}P(\lambda_i)\Big]
\end{aligned}
$$

since $\forall i,j\in I_j$, $\lim_{n\to\infty}g(\hat\lambda_i;c_n)=\lim_{n\to\infty}g(\hat\lambda_j;c_n)=g(\lambda_j;0)$, with $c_n\to 0$ and by multiplicity of λ_j, $\sum_{i\in I_j}P(\hat\lambda_i)\xrightarrow{\mathbf{P}}P(\lambda_j)$. Hence,

$$
\Sigma^R_{11,n} \xrightarrow{\mathbf{P}} \sum_{j\in v_1}g(\lambda_j,0)P(\lambda_j)\equiv\Sigma^R_{11}
$$

Therefore we have:

$$
X_{1n}=Z_n'\Sigma^R_{11,n}Z_n\xrightarrow{\mathcal{L}}Z_\infty'\Sigma^R_{11}Z_\infty\equiv X_1\ .
$$

Write X_1 in the form

$$
\begin{aligned}
X_1 &= Z_\infty'\Big[\sum_{j\in v_1}g(\lambda_j,0)\underbrace{\sum_{l=1}^{m(\lambda_j)}x_l(\lambda_j)x_l(\lambda_j)'}_{P(\lambda_j)}\Big]Z_\infty\\
&= \sum_{j\in v_1}g(\lambda_j,0)\sum_{l=1}^{m(\lambda_j)}z_l(\lambda_j)z_l(\lambda_j)'
\end{aligned}
$$

where $Z_\infty'x_l(\lambda_j)=z_l(\lambda_j)=\sqrt{\lambda_j}v_l$ where $v_l\sim\ N(0,1)$. Further, we have $\forall\lambda\in v_1: g(\lambda,0)=\frac{1}{\lambda}$.

$$
X_1 = \sum_{j\in v_1}\sum_{l=1}^{m(\lambda_j)}\frac{1}{\lambda_j}\sqrt{\lambda_j}v_l\sqrt{\lambda_j}v_l=\sum_{j\in v_1}\sum_{l=1}^{m(\lambda_j)}v_l^2
$$

Denote $q_1 = \sum_{j \in v_1} m(\lambda_j)$ where the dimension of the eigenspace is given by the multiplicity of λ_j, *i.e.* $\dim \mathcal{V}(\lambda_j) = m(\lambda_j)$. Hence we have

$$X_1 \sim \chi^2(q_1) \; .$$

Using the orthogonality property of the eigenspaces and the same arguments as above, we have:

$$X_{2n} \xrightarrow{\mathcal{L}} X_2 \equiv Z_\infty' \Sigma_{22}^R Z_\infty = \sum_{j \in v_2} \sum_{l=1}^{m(\lambda_j)} g(\lambda_j, 0) \lambda_j v_l^2 \; .$$

For $\forall \lambda \in v_2 : \; \lambda g(\lambda, 0) = 0$, hence $X_2 = 0$.

□

PROOF of Proposition 6.12 We have $Z_n \xrightarrow{\mathcal{L}} Z \sim N[0, \Sigma]$ with $rank(\Sigma) = r < q$. Let the spectral set $\mathcal{S}(\Sigma) = \sum_{i=1}^{3} v_i$ with $v_1 = \{\lambda : \lambda > c\}$, $v_2 = \{\lambda : \lambda = c\}$ and $v_3 = \{\lambda : \lambda < c\}$. Let the columns in V_1 (V_2, or V_3 respectively) be a basis for the eigenspace $\mathcal{V}_1$ ($\mathcal{V}_2$, or $\mathcal{V}_3$ respectively) corresponding to the eigenvalues in the subsets v_1 (v_2, v_3, respectively). Let $P_i(\lambda) = \sum_{\lambda \in v_i} P(\lambda)$ denote the total eigenprojection onto the corresponding eigenspace $\mathcal{V}_i$, for $i = 1, 2, 3$. Moreover, $\dim\big(\mathcal{V}_i\big) = q_i$ with $q = \sum_{i=1}^{3} q_i$. Let $\hat{V}_{in}$, $\mathcal{V}_i(\hat{\lambda}_n)$, $P_i(\hat{\lambda}_n)$ and $v_i(\hat{\lambda}_n)$ denote the sample analogs based on the estimates. Let $B_i = \{\omega \in \Omega : \; v_i(\hat{\lambda}_n) = v_i\}$ with $Pr(B_i) \underset{n \to \infty}{\rightarrow} 1$. Given that $P(\lambda)$ has the following representation with $\mathbf{x}_j(\lambda)$ vectors in $\mathbb{R}^q$,

$$P(\lambda) = \sum_{j=1}^{m(\lambda)} \mathbf{x}_j(\lambda) \mathbf{x}_j(\lambda)' \; ,$$

we have:

$$\begin{aligned} \Sigma_n^R = \sum_{\lambda \in \mathcal{S}(\Sigma_n)} g(\hat{\lambda}_n; c) P(\hat{\lambda}_n) = \sum_{i=1}^{3} \sum_{\hat{\lambda}_n \in v_i(\hat{\lambda}_n)} g(\hat{\lambda}_n; c) P_i(\hat{\lambda}_n) \xrightarrow{\mathbf{P}} \sum_{i=1}^{3} \sum_{\lambda \in v_i} g(\lambda; c) P_i(\lambda) \; , \forall \; \omega \; \in \; \bigcap_i B_i \\ = \sum_{\lambda \in \mathcal{S}(\Sigma)} g(\lambda; c) P(\lambda) = \Sigma^R \quad \text{(A.16)} \end{aligned}$$

by continuity of the eigenvalues and total eigenprojections given in **5.4** i) and ii), and under the assumption that $g(., c)$ is continuous a.e..

$$\begin{aligned} Z_n' \Sigma_n^R Z_n &\xrightarrow{\mathcal{L}} Z' \Sigma^R Z \\ &= Z' \Big[\sum_{i=1}^{3} \sum_{\lambda \in v_i} g(\lambda; c) P_i(\lambda) \Big] Z \end{aligned}$$

$$= \quad Z'\Big[\sum_{i=1}^{3}\sum_{\lambda\in v_i} g(\lambda;c)\sum_{j=1}^{m(\lambda)} \mathbf{x}_{ij}(\lambda)\mathbf{x}_{ij}(\lambda)'\Big]Z$$
$$= \quad \sum_{i=1}^{3}\sum_{\lambda\in v_i}\sum_{j=1}^{m(\lambda)} Z'\mathbf{x}_{ij}(\lambda)g(\lambda;c)\mathbf{x}_{ij}(\lambda)'Z \tag{A.17}$$

As $Var(Z'\mathbf{x}_{ij}(\lambda)) = \lambda$ with $\lambda \in v_i$ with a possible multiplicity $m(\lambda)$, then in order to simplify notations, we drop the explicit multiplicity formulation and obtain

$$\begin{aligned} Z_n'\Sigma_n^R Z_n &\overset{\mathcal{L}}{\to} Z'\Sigma^R Z \\ &= \sum_{i=1}^{3}\sum_{\lambda\in v_i} Z'\mathbf{x}_i(\lambda)g(\lambda;c)\mathbf{x}_i(\lambda)'Z \end{aligned} \tag{A.18}$$

Let $z_i = Z'\mathbf{x}_i(\lambda)$ for $i = 1, 2, 3$. Hence,

$$\begin{aligned} Z_n'\Sigma_n^R Z_n &\overset{\mathcal{L}}{\to} Z'\Sigma^R Z \\ &= \sum_{i=1}^{3}\sum_{\lambda\in v_i} z_i(\lambda)g(\lambda;c)z_i' \\ &= \sum_{\lambda\in\, v_1} z_1(\lambda)g(\lambda;c)z_1' + \sum_{\lambda\in\, v_2} z_2(\lambda)g(\lambda;c)z_2' + \sum_{\lambda\in\, v_3} z_3(\lambda)g(\lambda;c)z_3' \end{aligned} \tag{A.19}$$

As $z_i \sim N(0,\lambda)$ with $\lambda \in v_i$ for $i = 1, 2, 3$. For $\lambda \in v_1$, $g(\lambda;c) = 1/\lambda$, hence

$$\sum_{\lambda\in\, v_1} z_1(\lambda)g(\lambda;c)z_1' \sim \chi^2(q_1)\ .$$

For $\lambda \in v_2$, $\lambda = c$, *i.e.* $cg(c;c) \leq 1$, hence

$$\sum_{\lambda\in\, v_2} z_2(\lambda)g(c;c)z_2' = \sum_{\lambda\in\, v_2} cg(c;c)I(\lambda = c)v^2 \leq \chi^2(q_2)\ .$$

where $v \sim N(0,1)$, and $I(\lambda)$ is an indicator function. For $\lambda \in v_3$, $\lambda < c$, hence we have $\lambda g(\lambda;c) < 1$

$$\sum_{\lambda\in\, v_3} z_3(\lambda)g(\lambda;c)z_3' = \sum_{\lambda\in\, v_3} \lambda g(\lambda;c)I(\lambda < c)v^2 \leq \chi^2(q_3)\ .$$

where $v \sim N(0,1)$, and $I(\lambda)$ is an indicator function. □

B. Appendix:

PROOF of Proposition 6.13

The sketch of the proof is very similar to the proof of Dufour and Trognon (2001, Proposition 5.1).

Assumption B.1 (Compact parameter space) *$\theta \in \Theta$, where Θ is a compact set of $\mathbb{R}^p$.*

Assumption B.2 (Objective function continuity) *$M_{n,R}(\theta) = M_{n,R}(\theta, \omega)$ is a real function on $\Theta \times \Omega$, such that $M_{n,R}(\theta, \omega)$ is a continuous function of θ for all $\omega \in \Omega$.*

Assumption B.3 (Objective function uniform convergence) *There is a fixed (non-random) function $M_{\infty,R}(\theta)$ such that*

$$P\Big(\{\omega : \max_{\theta} |M_{n,R}(\theta, \omega) - M_{\infty,R}(\theta)| \to 0\}\Big) = 1 .$$

Assumption B.4 (Asymptotic identification) *$M_{\infty,R}(\theta)$ has a unique minimum at $\theta = \theta_0$, where θ_0 is an interior point of Θ;* i.e.

$$M_{\infty,R}(\theta) = D_\infty(\theta)' W_{\infty,R} D_\infty(\theta) = 0 \;\; \Leftrightarrow \;\theta = \theta_0 \tag{B.20}$$

Assumption B.5 (Uniform convergence of second derivatives) *$M_{n,R}(\theta, \omega)$ is a twice continuously differentiable function in θ and there is a fixed (non-random) function $V(\theta)$ such that:*

$$P\Big(\{\omega : \sup_{\theta} \| \frac{\partial^2}{\partial\theta\partial\theta'} M_{n,R}(\theta, \omega) - \frac{\partial^2}{\partial\theta\partial\theta'} M_{\infty,R}(\theta) \|_F \underset{n\to\infty}{\to} 0\}\Big) = 1 .$$

where $\|\cdot\|_F$ is the Frobenius norm, i.e. *$\|\Sigma\|_F^2 = tr[\Sigma'\Sigma]$.*

Assumption B.6 (Boundedness of the regularized weighting matrix)

$$W_{n,R} \overset{\text{p}}{\to} W_{\infty,R}$$

where $W_{\infty,R}$ is a non-random definite positive matrix.

Assumption B.7 (Asymptotic normality of the Score function)

$$\sqrt{n} D_n(\theta_0, \omega) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} D_\infty(\theta_0), \;\; D_\infty(\theta_0) \sim N[0, I(\theta_0)]$$

Assumption B.8 (Consistent estimators)

$$\frac{\partial D_n}{\partial\theta'}(\theta_0, \omega) \overset{\text{p}}{\to} J(\theta_0)$$

and

$$I_n(\theta_0) \overset{\text{p}}{\to} I(\theta_0) .$$

The proof rests upon the three first order conditions (FOC) below. Both estimators satisfy their respective FOC, *i.e.*:

$$\frac{\partial M_{n,R}}{\partial \theta}(\hat{\theta}_{n,R}) = 0 \tag{B.21}$$

$$\frac{\partial M_{n,R}}{\partial \theta}(\hat{\theta}^0_{n,R}) - \frac{\partial \psi'}{\partial \theta}(\hat{\theta}^0_{n,R})\hat{\lambda}^{\mathcal{L}}_n = 0 \tag{B.22}$$

$$\psi(\hat{\theta}^0_{n,R}) = 0 \tag{B.23}$$

where $\hat{\lambda}^{\mathcal{L}}_n$ denotes the Lagrange multiplier associated with the constraints. Set $W_{n,R} = \bar{I}^R_n$, the regularized inverse of the information matrix in the regularized GMM criterion. Taking its derivative w.r.t. θ yields:

$$\sqrt{n}\frac{\partial M_{n,R}}{\partial \theta}(\theta_0) = 2\frac{\partial D_n}{\partial \theta}(\theta_0)'\bar{I}^R_n\sqrt{n}D_n(\theta_0) \ .$$

Under Assumptions **B.1**-**B.8**:

$$\sqrt{n}\frac{\partial M_{n,R}}{\partial \theta}(\theta_0) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} 2J(\theta_0)'I^R(\theta_0)D_\infty(\theta_0)$$

hence

$$\sqrt{n}\frac{\partial M_{n,R}}{\partial \theta}(\theta_0) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} N\big[0, U_0\big] \ , \tag{B.24}$$

where

$$U_0 = 4J_0'I^R(\theta_0)I(\theta_0)I^R(\theta_0)J_0 \ . \tag{B.25}$$

Let

$$V_0 = 2J_0'I^R(\theta_0)J_0 \ . \tag{B.26}$$

Then we have the following equivalence:

$$\begin{aligned} 2V_0 - U_0 &= 4J_0'I^R(\theta_0)J_0 - 4J_0'I^R(\theta_0)I(\theta_0)I^R(\theta_0)J_0 \geq 0 \\ &\Leftrightarrow J_0'\Big[I^R(\theta_0) - I^R(\theta_0)I(\theta_0)I^R(\theta_0)\Big]J_0 \geq 0 \\ &\Leftrightarrow J_0'I^R(\theta_0)\Big[\big(I^R(\theta_0)\big)^{-1} - I(\theta_0)\Big]I^R(\theta_0)J_0 \geq 0 \\ &\Leftrightarrow \big(I^R(\theta_0)\big)^{-1} - I(\theta_0) \geq 0 \ . \end{aligned}$$

By Property 1 $iii)$, $\big(I^R(\theta_0)\big)^{-1} - I(\theta_0) \geq 0$, hence

$$U_0 \leq \ 2V_0 \ . \tag{B.27}$$

The matrix of second derivatives is given by

$$\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0) = 2\frac{\partial D_n(\theta_0)}{\partial\theta}' \bar{I}_n^R \frac{\partial D_n(\theta_)}{\partial\theta} + 2\Sigma_n$$

with

$$\Sigma_n = \left[\Sigma_{n,ij}\right], \quad \Sigma_{n,ij} = \frac{\partial^2 D_n(\theta_0)}{\partial\theta_i\partial\theta_j}\bar{I}_n^R D_n(\theta_0) \ .$$

By Assumptions **B.4** and **B.6-B.8**, $D_n(\theta_0) = o_p(1)$, hence $\Sigma_n \underset{n\to\infty}{\overset{\text{p}}{\to}} 0$. Therefore, we have:

$$\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0) \underset{n\to\infty}{\overset{\text{p}}{\to}} 2J(\theta_0)' I^R(\theta_0) J(\theta_0) = V_0 \ . \tag{B.28}$$

Here V_0 is not singular, hence it is invertible in the usual sense, *i.e.* $V_0^{-1}V_0 = Id$. We shall use Taylor's theorem for multivariable function [see Khuri (2003, Theorem 7.5.1, page 278) for all x in a neighborhood of x_0, *i.e.* $x \in N_\delta(x_0)$:

$$f(x) = f(x_0) + \frac{[(x-x_0)'\nabla]^{(1)}}{1!} f(x_0) + \frac{[(x-x_0)'\nabla]^{(2)}}{2!} f(z_0) \tag{B.29}$$

where $z_0 = x_0 + \xi h = x_0 + \xi(x - x_0)$, $0 < \xi < 1$, is a point on the line segment from x_0 to x. The differential operator $(x'\nabla)^m$ applied to a function $f(x)$ results in:

$$(x'\nabla)^m f(x) = \sum_{\substack{k_1,k_2,\ldots,k_p \\ k_1+k_2+\ldots+k_p=m}} \frac{m!}{k_1!k_2!\ldots k_p!} x_1^{k_1} x_2^{k_2} \cdots x_p^{k_p} \times \frac{\partial^m f(x)}{\partial x_1^{k_1}\partial x_2^{k_2}\cdots\partial x_p^{k_p}} \tag{B.30}$$

As $\left[\frac{\partial M_{n,R}}{\partial\theta}\right](\theta)$ is a vector, we apply the theorem on each component of the vector, *i.e.* $f(x) = \frac{\partial M_{n,R}}{\partial\theta_j}$ in equation (B.29), with $x = \hat{\theta}_{n,R}$ and $x_0 = \theta_0$. This yields:

$$\begin{aligned}
\underbrace{\frac{\partial M_{n,R}}{\partial\theta_j}(\hat{\theta}_{n,R})}_{0} &= \frac{\partial M_{n,R}}{\partial\theta_j}(\theta_0) + \left[(\hat{\theta}_{n,R} - \theta_0)'\nabla\right]^{(1)} \frac{\partial M_{n,R}}{\partial\theta_j}(\bar{\theta}_{n,0}^{(j)}) \\
&= \frac{\partial M_{n,R}}{\partial\theta_j}(\theta_0) + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\bar{\theta}_{n,0}^{(j)}) \\
&= \frac{\partial M_{n,R}}{\partial\theta_j}(\theta_0) + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta_0) \\
&\quad + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \left[\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\bar{\theta}_{n,0}^{(j)}) - \frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta_0)\right]
\end{aligned}$$

$$= \quad \frac{\partial M_{n,R}}{\partial \theta_j}(\theta_0) + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta_0) + e_{jn} \tag{B.31}$$

where $\bar{\theta}_{n,0}^{(j)} = \xi_{n,j}\theta_0 + (1-\xi_{n,j})\hat{\theta}_{n,R}$, $0 < \xi_{n,j} < 1$, with $\bar{\theta}_{n,0}^{(j)} \in \mathbb{R}^P$, for $j = 1, \dots, p$. Let

$$\sqrt{n} e_{jn} = \sum_{i=1}^{p} \sqrt{n}(\hat{\theta}_{n,R} - \theta_0)_i \left[\frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\bar{\theta}_{n,0}^{(j)}) - \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta_0) \right] \tag{B.32}$$

By Assumption **B.5**,

$$\left[\frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\bar{\theta}_{n,0}^{(j)}) - \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta_0) \right] < \sup_{\theta} \left| \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta) - \frac{\partial^2 M_{\infty,R}}{\partial \theta_j \partial \theta_i}(\theta) \right| \overset{a.s.}{\to} 0 \ .$$

Hence,

$$\sqrt{n} e_{jn} = o_p(1) \quad \Leftrightarrow \quad \sqrt{n}(\hat{\theta}_{n,R} - \theta_0)_i = O_p(1) \ \text{ for } i = 1, \dots, p \ .$$

As this holds for any arbitrary j, $\sqrt{n} e_{jn} = o_p(1)$ for all $j = 1, \dots, p$.

Rearranging equation (B.31) yields in an equivalent way:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{n,R} - \theta_0) &= -\left[\frac{\partial^2 M_{n,R}}{\partial \theta \partial \theta'}(\theta_0) \right]^{-1} \sqrt{n} \frac{\partial M_{n,R}}{\partial \theta}(\theta_0) + \sqrt{n} e_{1n}^* \\ &= -V_0^{-1} \sqrt{n} X_{n,R} + \sqrt{n} e_{1n}^* \end{aligned} \tag{B.33}$$

where $X_{n,R} = \frac{\partial M_{n,R}}{\partial \theta}(\theta_0)$ and e_{1n} and e_{1n}^* are both $o_p(1/\sqrt{n})$. Similarly we have for the restricted estimator based on the two other FOC under Assumption **2.5**:

$$\frac{\partial M_{n,R}}{\partial \theta}(\hat{\theta}_{n,R}^0) - \frac{\partial \psi'}{\partial \theta}(\hat{\theta}_{n,R}^0) \hat{\lambda}_n^{\mathcal{L}} = 0 \tag{B.34}$$

$$\psi(\hat{\theta}_{n,R}^0) = 0 \ . \tag{B.35}$$

So we have for $j = 1, \dots, p$:

$$\frac{\partial M_{n,R}}{\partial \theta_j}(\theta_0) - \frac{\partial \psi'}{\partial \theta_j}(\theta_0) \hat{\lambda}_{j,n}^{\mathcal{L}} + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\bar{\theta}_{n,0}^{(j)}) = 0$$

or equivalently

$$\frac{\partial M_{n,R}}{\partial \theta_j}(\theta_0) - \frac{\partial \psi'}{\partial \theta_j}(\theta_0) \hat{\lambda}_{j,n}^{\mathcal{L}} + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta_0) + \sum_{i=1}^{p}(\hat{\theta}_{n,R} - \theta_0)_i \left[\frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\bar{\theta}_{n,0}^{(j)}) - \frac{\partial^2 M_{n,R}}{\partial \theta_j \partial \theta_i}(\theta_0) \right] = 0 \ .$$

Let this time e_{jn} be

$$\sqrt{n}e_{jn} \quad = \quad \sum_{i=1}^{p}\sqrt{n}(\hat{\theta}_{n,R}-\theta_0)_i\left[\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\bar{\theta}_{n,0}^{(j)})-\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta_0)\right] \tag{B.36}$$

and $\bar{\theta}_{n,0}^{(i)}=\xi_{n,i}\theta_0+(1-\xi_{n,i})\hat{\theta}_{n,R}^0$, $0<\xi_{n,i}<1$. Again by Assumption **B.5**,

$$\left[\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\bar{\theta}_{n,0}^{(j)})-\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta_0)\right] < \sup_{\theta}\left|\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta)-\frac{\partial^2 M_{\infty,R}}{\partial\theta_j\partial\theta_i}(\theta)\right| \overset{a.s.}{\to} 0\ .$$

Hence,

$$\sqrt{n}e_{jn}=o_p(1) \ \ \Leftrightarrow\ \ \sqrt{n}(\hat{\theta}_{n,R}-\theta_0)_i=O_p(1) \ \text{ for } i=1,\ldots,p\ .$$

Recall $P(\theta)=\frac{\partial\psi}{\partial\theta'}(\theta)$. Under Assumptions **B.1**-**B.8** and **2.5**, both estimators $\hat{\theta}_{n,R}$ and $\hat{\theta}_{n,R}^0$ are strongly consistent and asymptotically normal.

As for the third FOC, we have for $j=1,\ldots,q$:

$$\underbrace{\psi_j(\theta_0)}_{0}+\sum_{i=1}^{p}(\hat{\theta}_{n,R}^0-\theta_0)_i\frac{\partial\psi_j}{\partial\theta_i}(\theta_0)+e_{jn}=0$$

with

$$e_{jn}=\sum_{i=1}^{p}(\hat{\theta}_{n,R}^0-\theta_0)_i\left(\frac{\partial\psi_j}{\partial\theta_i}(\bar{\theta}_{n,0}^{(j)})-\frac{\partial\psi_j}{\partial\theta_i}(\theta_0)\right)\ .$$

By Assumption **2.5**, $\psi(\theta)$ is continuously differentiable, hence

$$\left(\frac{\partial\psi_j}{\partial\theta_i}(\bar{\theta}_{n,0}^{(j)})-\frac{\partial\psi_j}{\partial\theta_i}(\theta_0)\right)\overset{\mathrm{P}}{\to}\ 0\ .$$

Hence,

$$\sqrt{n}e_{jn}=o_p(1) \ \ \Leftrightarrow\ \ \sqrt{n}(\hat{\theta}_{n,R}-\theta_0)_i=O_p(1) \ \text{ for } i=1,\ldots,p\ .$$

As j is arbitrary, it holds for any $j=1,\ldots,q$. The three FOC gathered together yield

$$(\hat{\theta}_{n,R}-\theta_0)=-V_0^{-1}X_{n,R}+\epsilon_{1n} \tag{B.37}$$

$$(\hat{\theta}_{n,R}^0-\theta_0)=V_0^{-1}\left[P_0'\hat{\lambda}_n^{\mathcal{L}}-X_{n,R}\right]+\epsilon_{2n} \tag{B.38}$$

$$P_0(\hat{\theta}_{n,R}^0-\theta_0)+e_{3n}=0\ . \tag{B.39}$$

Multiplying (B.38) by P_0 and solving for $\hat{\lambda}_n^{\mathcal{L}}$ yields

$$\hat{\lambda}_n^{\mathcal{L}}=\left(P_0V_0^{-1}P_0'\right)^{-1}P_0V_0^{-1}X_{n,R} \tag{B.40}$$

Moreover, we have:

$$\begin{aligned}(\hat{\theta}^0_{n,R}-\theta_0) &= V_0^{-1}\Big[P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}X_{n,R}\Big]-V_0^{-1}X_{n,R}+\epsilon_{2n}\\ &= -AX_{n,R}+\epsilon_{2n} \qquad \text{(B.41)}\end{aligned}$$

with $A=V_0^{-1}-V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}$.

The LR test statistic is given by

$$LR=n\xi_n=n\underbrace{[M_{n,R}(\hat{\theta}^0_{n,R})-M_{n,R}(\hat{\theta}_{n,R})]}_{\xi_n}$$

and developing ξ_n up to order 2 yields

$$\begin{aligned}\xi_n=&\frac{\partial M_{n,R}}{\partial\theta'}(\theta_0)(\hat{\theta}^0_{n,R}-\theta_0)-\frac{\partial M_{n,R}}{\partial\theta'}(\theta_0)(\hat{\theta}_{n,R}-\theta_0)\\ &+\frac{1}{2}\Big\{(\hat{\theta}^0_{n,R}-\theta_0)'\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0)(\hat{\theta}^0_{n,R}-\theta_0)-(\hat{\theta}_{n,R}-\theta_0)'\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0)(\hat{\theta}_{n,R}-\theta_0)\Big\}+e_n\end{aligned} \qquad \text{(B.42)}$$

with

$$\begin{aligned}ne_n &= \sqrt{n}(\hat{\theta}^0_{n,R}-\theta_0)'\Big[\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\bar{\theta}_{n,0})-\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0)\Big]\sqrt{n}(\hat{\theta}^0_{n,R}-\theta_0)\\ &\quad -\sqrt{n}(\hat{\theta}_{n,R}-\theta_0)'\Big[\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\bar{\theta}_{n,0})-\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0)\Big]\sqrt{n}(\hat{\theta}_{n,R}-\theta_0)\\ &= o_p(1) \;\;\Leftrightarrow\;\; \sqrt{n}(\hat{\theta}^0_{n,R}-\theta_0)=O_p(1)\;, \qquad \text{(B.43)}\end{aligned}$$

since by Assumption **B.5**, we have

$$\Big[\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\bar{\theta}^{(j)}_{n,0})-\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta_0)\Big] < \sup_\theta\Big|\frac{\partial^2 M_{n,R}}{\partial\theta_j\partial\theta_i}(\theta)-\frac{\partial^2 M_{\infty,R}}{\partial\theta_j\partial\theta_i}(\theta)\Big|\overset{a.s.}{\to}0\;.$$

Using (B.37), (B.41) and the fact that

$$V_0^{-1}V_0=Id,\;\;(V_0=2J_0'I^R(\theta_0)J_0)\;\text{ and }\;AV_0A=A$$

we have:

$$\begin{aligned}\xi_n &= \frac{1}{2}X_{n,R}'(V_0^{-1}-A)X_{n,R}+e_n^* \qquad \text{(B.44)}\\ &= \frac{1}{2}X_{n,R}'V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}X_{n,R}+e_n^*,\;\;e_n^*=o_p(1/n) \qquad \text{(B.45)}\end{aligned}$$

Given that

$$\sqrt{n}X_{n,R} \xrightarrow{\mathcal{L}} N[0,U_0] \text{ with } U_0 \leq 2V_0 = 4J_0' I^R(\theta_0)J_0 \ ,$$

we deduce

$$\begin{aligned} LR(\psi)^R = n\xi_n \xrightarrow{\mathcal{L}} LR(\psi)^R_\infty \leq \frac{1}{2}u'\sqrt{2}V_0^{1/2}V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}\sqrt{2}V_0^{1/2}u \\ LR(\psi)^R_\infty \leq u'V_0^{-1/2}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1/2}u \\ LR(\psi)^R_\infty \leq u'Qu \sim \chi^2(q) \end{aligned}$$

with $Q' = Q$ and $Q^2 = Q$ and $rank(Q) = q$.

□

PROOF of Proposition 6.14

We now exploit the decomposition of the *regularized* LR statistic into three components:

$$\begin{aligned} LR(\psi)^R &= n\xi_n = n[M_{n,R}(\hat{\theta}^0_{n,R}) - M_{n,R}(\hat{\theta}_{n,R})] \\ &= n[M_{1n,R}(\hat{\theta}^0_{n,R}) - M_{1n,R}(\hat{\theta}_{n,R})] + n[M_{2n,R}(\hat{\theta}^0_{n,R}) - M_{2n,R}(\hat{\theta}_{n,R})] + n[M_{3n,R}(\hat{\theta}^0_{n,R}) - M_{3n,R}(\hat{\theta}_{n,R})] \\ &= n\xi_{1n} + n\xi_{2n} + n\xi_{3n} \ , \end{aligned}$$

where $\xi_{in} = M_{in,R}(\hat{\theta}^0_{n,R}) - M_{in,R}(\hat{\theta}_{n,R})$, for $i = 1, 2, 3$. It has been shown that $\sqrt{n}\frac{\partial M_{n,R}}{\partial \theta}(\theta_0) \xrightarrow{\mathcal{L}} N[0, U_0]$ and as $\sqrt{n}D_n(\theta_0) \xrightarrow{\mathcal{L}} N[0, I(\theta_0)]$, we have:

$$\begin{aligned} \sqrt{n}\frac{\partial M_{n,R}}{\partial \theta}(\theta_0) &= \sqrt{n}\frac{\partial M_{1n,R}}{\partial \theta}(\theta_0) + \sqrt{n}\frac{\partial M_{2n,R}}{\partial \theta}(\theta_0) + \sqrt{n}\frac{\partial M_{3n,R}}{\partial \theta}(\theta_0) \\ &= \sum_{i=1}^{3} 2\frac{\partial D_n'}{\partial \theta}(\theta_0)\bar{I}^R_{ii}\sqrt{n}D_n(\theta_0) = \sum_{i=1}^{3}\sqrt{n}\frac{\partial M_{in,R}}{\partial \theta}(\theta_0) \end{aligned} \quad \text{(B.46)}$$

where

$$\sqrt{n}\frac{\partial M_{in,R}}{\partial \theta}(\theta_0) = 2\frac{\partial D_n'}{\partial \theta}(\theta_0)\bar{I}^R_{ii}\sqrt{n}D_n((\theta_0)) \xrightarrow{\mathcal{L}} N[0, U_{i0}]$$

with

$$U_{i0} = 4J_0' I^R_{ii}(\theta_0)I(\theta_0)I^R_{ii}(\theta_0)J_0 \ \text{ for } i = 1, 2, 3.$$

Using results stated in Property 1, we have:

$$U_{10} = 4J_0' I^R_{11}(\theta_0)I(\theta_0)I^R_{11}(\theta_0)J_0 \ = 4J_0' I^R_{11}(\theta_0)J_0 = 2V_{10} \quad \text{(B.47)}$$

$$U_{20} = 4J_0' I^R_{22}(\theta_0)I(\theta_0)I^R_{22}(\theta_0)J_0 \ \leq 4J_0' I^R_{22}(\theta_0)J_0 = 2V_{20} \quad \text{(B.48)}$$

$$U_{30} = 4J_0' I^R_{33}(\theta_0)I(\theta_0)I^R_{33}(\theta_0)J_0 \ \leq 4J_0' I^R_{33}(\theta_0)J_0 = 2V_{30} \quad \text{(B.49)}$$

and

$$U_0 = U_{10} + U_{20} + U_{30} \ . \quad \text{(B.50)}$$

Regarding the second derivatives $\frac{\partial^2 M_{in,R}}{\partial\theta\partial\theta'}(\theta_0)$, we have:

$$\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0) = \frac{\partial^2}{\partial\theta\partial\theta'}\big[M_{1n,R}+M_{2n,R}+M_{3n,R}\big] = \sum_{i=1}^{3}\frac{\partial^2 M_{in,R}}{\partial\theta\partial\theta'}(\theta_0) \tag{B.51}$$

$$= \sum_{i=1}^{3}\left\{2\frac{\partial D_n'}{\partial\theta}(\theta_0)\bar{I}^R_{ii,n}\frac{\partial D_n}{\partial\theta'}(\theta_0)+2\varSigma_{ii,n}\right\} \tag{B.52}$$

where $\varSigma_{ii,n}$ is a $p\times p$ matrix with typical elements given by:

$$\underbrace{\frac{\partial^2 D_n'}{\partial\theta_i\partial\theta_j}(\theta_0)\underbrace{\bar{I}^R_{ii,n}}_{O_p(1)}\underbrace{D_n(\theta_0)}_{o_p(1)}}_{o_p(1)},$$

and $\varSigma_{ii,n}=o_p(1)$. Hence,

$$\frac{\partial^2 M_{in,R}}{\partial\theta\partial\theta'}(\theta_0) = 2\frac{\partial D_n'}{\partial\theta}(\theta_0)\bar{I}^R_{ii,n}\frac{\partial D_n}{\partial\theta'}(\theta_0) \xrightarrow{\text{P}} 2J_0'I^R_{ii}(\theta_0)J_0 \equiv V_{i0}\ . \tag{B.53}$$

Recall that

$$\frac{\partial^2 M_{n,R}}{\partial\theta\partial\theta'}(\theta_0) \xrightarrow{\text{P}} 2J_0'I^R(\theta_0)J_0 = V_0\ .$$

Developing ξ_n up to order 2 component by component yields:

$$\begin{aligned}
\xi_n &= \sum_{i=1}^{3}\Bigg\{\frac{\partial M_{in,R}}{\partial\theta'}(\theta_0)(\hat\theta^0_{n,R}-\theta_0)-\frac{\partial M_{in,R}}{\partial\theta'}(\theta_0)(\hat\theta_{n,R}-\theta_0)\\
&\quad+\frac{1}{2}\Bigg[(\hat\theta^0_{n,R}-\theta_0)'\frac{\partial^2 M_{in,R}}{\partial\theta\partial\theta'}(\theta_0)(\hat\theta^0_{n,R}-\theta_0)-(\hat\theta_{n,R}-\theta_0)'\frac{\partial^2 M_{in,R}}{\partial\theta\partial\theta'}(\theta_0)(\hat\theta_{n,R}-\theta_0)\Bigg]\Bigg\}+e_n\\
&= \sum_{i=1}^{3}\Big\{-X_{in,R}'AX_{n,R}+X_{in,R}'V_0^{-1}X_{n,R}+\frac{1}{2}[X_{n,R}'AV_{i0}AX_{n,R}-X_{n,R}'V_0^{-1}V_{i0}V_0^{-1}AX_{n,R}]+e_n\\
&= \sum_{i=1}^{3}\Big\{X_{in,R}'\big(-A+V_0^{-1}\big)X_{n,R}+\frac{1}{2}X_{n,R}'\big(AV_{i0}A-V_0^{-1}V_{i0}V_0^{-1}\big)X_{n,R}\Big\}+e_n
\end{aligned}$$

where $X_{in,R}=\frac{\partial M_{in,R}}{\partial\theta}(\theta_0)$. It can be shown that

$$AV_{i0}A-V_0^{-1}V_{i0}V_0^{-1} = -V_0^{-1}V_{i0}V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}\ ,$$

or in other words

$$AV_{i0}A = V_0^{-1}V_{i0}V_0^{-1} - V_0^{-1}V_{i0}V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1} = V_0^{-1}V_{i0}A\ .$$

We know that

$$V_0^{-1} - A = V_0^{-1}P_0'\big(P_0V_0^{-1}P_0'\big)^{-1}P_0V_0^{-1}\ .$$

□

References

[0]

Amemya, Y. (1986), Limiting distribution of certain characteristic roots under general conditions and their applications, Technical report 17, economic workshop, Stanford University.

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.

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.

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

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

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

Dufour, J.-M. and Trognon, A. (2001), Invariant tests based on M-estimators, estimating functions, and the generalized method of moments. CIREQ working paper, University of Montreal.

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.

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

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

Khuri, A. I. (2003), *Advanced Calculus with applications in statistics*, 2nd edn, Wiley.

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

Kolmogorov, a. N. and Fomin, S. V. (1970), *Introductory real analysis*, Dover Publications, inc.

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

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

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

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

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 West, K. D. (1987), 'Hypothesis testing with efficient method of moments estimators', *International Economic Review* **28**, 777–787.

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.

Staiger, D. and Stock, J. (1997), 'Instrumental variable regressions with weak instruments', *Econometrica* **65**(3), 557–586.

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.