Generalised Wald Type Test of Nonlinear Restrictions

Zaka Ratsimalahelo

CRESE, EA 3190, Universite Bourgogne Franche-Comté, Besançon, France
Email: zaka.ratsimalahelo@univ-fcomte.fr

Abstract

This paper proposes a generalised Wald type tests to test the hypothesis of the nonlinear restrictions. We circumvent the problem of singularity of the covariance matrix associated with the usual Wald test by proposing a generalised inverse procedure, and an alternative simple procedure which can be approximated by a suitable chi-square distribution. New threshold value is derived to estimate the rank of the covariance matrix.

Subject Areas

Statistics and Econometrics

Keywords

Nonlinear Restrictions, Deficient Rank, Singular Covariance Matrix, Generalised Wald Test

1. Introduction

Analyses of economic data often entail the testing of hypotheses that imply complex nonlinear restrictions on subsets of parameters, and it is thus desirable to employ econometric methods that are flexible both in terms of their applications and implementation for typical data sets. In this paper, we are interested in testing the null hypothesis H_0 in the following form

$$H_0 : g(\theta) = 0 \tag{1}$$

against the alternative hypothesis

$$H_1 : g(\theta) \neq 0 \tag{2}$$

where the vector parameter $\theta \in \Theta$; the parameter space $\theta \in \Theta$ is a k-dimensional compact subspace of $\mathbb{R}^k$ and $g(\theta)$ is a mapping from $\mathbb{R}^k$ to

$\mathbb{R}^{l}$ continuously differentiable functions of θ.

In econometrics literature, the Wald tests are commonly used to test this null hypothesis since the generality of its formulation affords the testing of several interesting economic hypotheses which might present formidable difficulties for other procedures (see, e.g. [1] [2] [3]). Moreover, it is well known that in a linear regression model with normally distributed errors, the Wald statistic for a set of linear restrictions is a monotonic transformation of the likelihood ratio (LR) test statistics.

Under regularity conditions, the null asymptotic distribution is chi-squared distribution, and this is the distribution that one usually uses to carry out hypothesis tests. On the other hand, under the nonregularity conditions, Wald tests fail to have limiting chi-squared distribution in general [4] as in the cases of multiple-hypothesis testing within a single model. The functional restrictions used to represent the hypotheses failed to satisfy regularity conditions. In particular, their Jacobian matrices failed to have full rank in certain circumstances [1]. Also, in vector autoregressive (VAR) processes, if the process is stationary, the multivariate least squares (LS) estimator of the coefficients has a non-singular asymptotic distribution whereas the distribution becomes singular if some variables are integrated or cointegrated. So the Wald test has a nonstandard asymptotic distribution.

Another example is the Granger non-causality test. As shown in [5] when there is a cointegrating relationship, in general the Wald statistic of the Granger non-causality test has a non-standard limiting distribution, depending on nuisance parameters.

This paper develops an asymptotic theory for Wald test in the case where the asymptotic covariance matrix is degenerate. That case occurs when the functional restriction representing the hypothesis of interest fails to satisfy the regular conditions that its Jacobian matrix has full rank at the true parameter values and/or the asymptotic covariance matrix of the estimator parameters is singular.

[6] used a generalized inverse of the asymptotic covariance matrix and they showed that the Wald test still has an asymptotically standard distribution. However, the serious problem is how to detect the degeneracy or the rank of asymptotic covariance matrix.

It is well known that the rank of matrix is equal to the number of nonzero eigenvalues. [6] used an ad-hoc threshold to determine if the eigenvalues are zero or not. Unfortunately, their choice of a threshold is more approximate. The values of the threshold do not appear to be based on any explicit analytical expressions. An obvious way of circumventing this difficulty would be to determine the distribution of the eigenvalues.

In this paper, we investigate the asymptotic distribution of the eigenvalues and new threshold values are proposed to test if the eigenvalues are significantly different from zero or not. This permits to determine the rank of the covariance

matrix. One important property of the eigenvalues is that they are sufficient statistics invariant with respect to the multiplication of the matrix from left and right by any nonsingular matrices.

The paper is organised as follows. Section 2 develops the general expression for the tests statistics. Section 3 proposes new threshold value for the eigenvalues and conclusions are given in Section 4.

2. Wald Tests

In this section, we will develop general expressions for Wald test statistics for nonlinear restrictions. Let $\hat{\theta}$ be an estimator of θ based on a sample of size n. We make the following assumption regarding the estimator $\hat{\theta}$.

Assumption (i) $\hat{\theta} = \theta + O_p\left(n^{-1/2}\right)$,

(ii) $\sqrt{n}\left(\hat{\theta} - \theta\right) \to^d N_k\left(0, \Omega\right)$, *where* Ω *is finite non zero but possibly singular*,

(iii) *a consistent estimator* $\hat{\Omega}$ *of* Ω *is available*.

The asymptotic normality of $\hat{\theta}$ implies that under the null hypothesis, $\sqrt{n}g\left(\hat{\theta}\right)$ is asymptotically normally distributed with a finite asymptotic variance matrix Σ.

$$\sqrt{n}g\left(\hat{\theta}\right) \to^d N_l\left(0, \Sigma\right) \tag{3}$$

where $\Sigma = G(\theta)\Omega G(\theta)'$ is of dimension $(l \times l)$ and $G(\theta) = \dfrac{\partial g(\theta)}{\partial \theta'}$ is the $l \times k$ Jacobian matrix of $g = \left(g_1, \cdots, g_l\right)'$.

To establish the limiting distribution of the standard Wald test, we assume the following two regularity conditions.

C1. The Jacobian matrix $G(\theta)$ is of full rank for all θ in the parameter space.

C2. The asymptotic covariance matrix Ω is non singular, it is symmetric and positive definite.

Let $\hat{\Sigma}$ be a consistent estimator of Σ, obtained by replacing $G(\theta)$ and Ω by their consistent estimator $G\left(\hat{\theta}\right)$ and $\hat{\Omega}$ respectively. Hence under the two regularity conditions C1 and C2, the Wald statistic for testing (1) is given by

$$W = ng\left(\hat{\theta}\right)' \hat{\Sigma}^{-1} g\left(\hat{\theta}\right) \tag{4}$$

and W is asymptotically distributed as chi-square with l degrees of freedom on H_0.

Now, when the covariance matrix Σ is singular, *i.e.* if at least one of these regularity conditions (C1 and C2) is not satisfied, then the Wald statistic may not have an asymptotic chi-square distribution. The singularity of Σ comes from three possible situations. First the matrix of first-order partial derivatives of the restrictions, $G(\theta)$ has reduced rank. Secondly the asymptotic covariance Ω is a matrix degenerate and finally the combination of the two:

$rank\left(G(\theta)\right) < l$ and Ω is singular.

If the function $g(\theta)$ involves products of the elements of θ then the matrix of first-order partial derivatives $G(\theta)$ is likely to have a reduced rank over part of the parameter space. Such functions are relatively common in time series analysis. For instance, impulse responses and related quantities of interest in a vector autoregressive (VAR) analysis involve products of the VAR coefficients. Similarly, such functions come up in analyzing multi-step causality in VAR models [6]; or Granger causality in VARMA models [7] and in testing of restrictions on the levels of parameters of a cointegrated VAR process.

The singularity of the asymptotic covariance matrix of the estimator parameters occurs for instance in VAR processes. If the process is nonstationary, *i.e.* some variables are integrated or cointegrated, the multivariate least squares (LS) estimator of the coefficients has a singular asymptotic distribution. This singularity problem has also been noted or discussed in the context of the long-run impact matrix (see [8]).

In that case, it is a usual practice to resort a generalized inverse procedure when we have inverted a singular matrix. That is, we have under the null hypothesis,

$$W_+ = n g\left(\hat{\theta}\right)' \hat{\Sigma}^+ g\left(\hat{\theta}\right) \tag{5}$$

where $\hat{\Sigma}^+$ is the Moore-Penrose generalized inverse of a matrix $\hat{\Sigma}$. Under additional condition, $\operatorname{rank}\left(\hat{\Sigma}\right) = \operatorname{rank}(\Sigma) = r < l$, the Wald statistic W_+ still have an asymptotic chi-square distribution with r degrees of freedom. (see [4]). Unfortunately, this rank condition is not always easy to verify in practice.

To overcome this difficulty, the Moore-Penrose generalized inverse can be obtained in using the spectral decomposition of $\hat{\Sigma}^+$. Let λ_i be the *i-th* eigenvalue of Σ and u_i the corresponding *i-th* (orthonormal) eigenvector, satisfying $u_i' u_j = \delta_{ij}$, where δ_{ij} is the usual Kronecker delta $(i, j = 1, \cdots, l)$. The eigenvalue-eigenvector equation is $\Sigma u_i = \lambda_i u_i$, then $u_i' \Sigma u_i = \lambda_i, i = 1, \cdots, l$ and collecting together we obtain $U' \Sigma U = \Lambda$, or $\Sigma = U \Lambda U'$, where $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_l)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$, is the $(l \times l)$ diagonal matrix of ordered eigenvalues and $U = [u_1, u_2, \cdots, u_l]$ is the $(l \times l)$ matrix whose columns are the orthonormal eigenvectors of Σ. That is $U'U = I_l$ and $UU' = I_l$ where I_l is the unity matrix of dimension $l \times l$. If $\lambda_i = 0, i = r + 1, \cdots, l$, so that $\operatorname{rank}(\Sigma) = r$, we can write

$$\Sigma = U_1 \Lambda_1 U_1', (l \times l)$$

$$U = [U_1, U_2], (l \times r, l \times l - r)$$

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where $\Lambda_1 = \operatorname{diag}(\lambda_1, \cdots, \lambda_r)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$, is the $(r \times r)$ diagonal matrix of the ordered positive eigenvalues of Σ and the columns of U_1, $u_i (i = 1, \cdots, r)$ are the corresponding eigenvectors to the positive eigenvalues.

The Moore-Penrose generalized inverse of Σ is $\Sigma^+ = U_1 \Lambda_1^{-1} U_1'$, consistently estimated by $\hat{\Sigma}^+ = \hat{U}_1 \hat{\Lambda}_1^{-1} \hat{U}_1'$ which can be used to construct W_+. Hence the Wald statistic

$$W_+ = n g\left(\hat{\theta}\right)' \hat{U}_1 \hat{\Lambda}_1^{-1} \hat{U}_1' g\left(\hat{\theta}\right) \tag{6}$$

is asymptotically chi-squared with r degrees of freedom on H_0, and $r = \mathrm{rank}\left(\Sigma\right)$.

This test statistic has fewer degrees of freedom than the usual Wald test since $r = \mathrm{rank}\left(\Sigma\right) \leq l$. Hence it has improved power, in particular if superfluous restrictions are removed. [9] finds in a small sample Monte Carlo investigation that power advantages may in fact result from taking into account fewer restrictions.

3. New Threshold Values for the Eigenvalues

However the problem associated with this test statistic is the determination of the number of nonzero eigenvalues of the matrix Σ that is the rank of Σ, since in the construction of generalised Wald test statistics, the rank of the variance matrix estimate should be equal to the rank of the true variance matrix. In this section we propose a sequence pretest procedure. We develop the test procedures for

$$H_0' : \lambda_i = 0, \ i = r+1, \cdots, l \tag{7}$$

against the alternative hypothesis

$$H_1' : \lambda_i > 0$$

where the λ_i are the eigenvalues of the matrix Σ. The null hypothesis H_0' is equivalent to the hypothesis $\mathrm{rank}\left(\Sigma\right) = r$. To decide if the eigenvalues are nonzero or not, [6] used a threshold c which depends on different factors such as the sample size n and the value of the largest eigenvalue $\hat{\lambda}_1$ of $\hat{\Sigma}$. Then λ_i is nonzero if the eigenvalue $\hat{\lambda}_i$ of the estimator $\hat{\Sigma}$ of Σ is greater than c. In other words, for our purposes one reject the null hypothesis if $\hat{\lambda}_i > c$. Unfortunately their choice of a threshold c is more approximative, the values of the threshold do not appear to be based on any explicit analytical expressions but are selected on an ad hoc basis. Consequently it will not be certain to take all restrictions which correspond to nonzero eigenvalues of the covariance matrix Σ. Thus the rank estimated may be larger than the true rank. We shall propose an appropriate threshold level criteria. The approach consist to derive the statistical properties of $\hat{\lambda}_i$. To this end, we use some results of the matrix perturbation theory.

Let us consider the matrix perturbation theory

$$\hat{\Sigma} = \Sigma + \varepsilon B \tag{8}$$

where εB, with small ε, is the error matrix $\hat{\Sigma} - \Sigma$ or the matrix perturbation. These results of the matrix perturbation theory indicate how much the eigenvalues and eigenvectors of the matrix $\hat{\Sigma}$ can differ from those of Σ for small ε. Assume that the estimator $\hat{\Sigma}$ of Σ is *root-n* consistent and has a

limiting normal distribution, then the perturbation term εB is of order $O_p\left(n^{-1/2}\right)$ and can be seen as a zero mean Gaussian random matrix [10].

Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_l$ be the ordered eigenvalues of matrix perturbed $\hat{\Sigma}$ and $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_l$ the corresponding eigenvectors, the eigenvalue-eigenvector equation

$$\hat{\Sigma}\hat{u}_i = \hat{\lambda}_i\hat{u}_i, \ i = 1, \cdots, l$$

can be written as

$$(\Sigma + \varepsilon B)\hat{u}_i = \hat{\lambda}_i\hat{u}_i.$$

Following the spectral decomposition of Σ, the eigenvalue-eigenvector equation can be equivalently expressed as

$$(U\Lambda U' + \varepsilon B)\hat{u}_i = \hat{\lambda}_i\hat{u}_i.$$

Now, premultiplying by U', we obtain

$$(\Lambda + \varepsilon U'BU)U'\hat{u}_i = \hat{\lambda}_i\left(U'\hat{u}_i\right)$$

that is, $\varepsilon U'BU$ is the perturbation matrix of the diagonal matrix Λ and $\hat{\lambda}_i$ is an eigenvalue of $\Lambda + \varepsilon U'BU$ corresponding to the eigenvector $U'\hat{u}_i$. The first-order approximation of $\hat{\lambda}_i$ is given by

$$\hat{\lambda}_i = \lambda_i + \varepsilon u_i'Bu_i + O_p\left(\varepsilon^2\right).$$

A similar result is given by ([11], p. 65). For $\varepsilon = n^{-1/2}$, we have

$$\hat{\lambda}_i = \lambda_i + n^{-1/2}u_i'Bu_i + O_p\left(n^{-1}\right). \tag{9}$$

Hence for the zero eigenvalue of Σ that is $\lambda_i = 0$, we get

$$\hat{\lambda}_i = n^{-1/2}u_i'Bu_i + O_p\left(n^{-1}\right) \tag{10}$$

which corresponds to the *i-th* smallest eigenvalue of the estimated matrix $\hat{\Sigma}$.

The asymptotic distribution of $\hat{\lambda}_i$ is obtained from the following equation

$$\sqrt{n}\left(\hat{\lambda}_i - \lambda_i\right) = u_i'Bu_i + o_p(1). \tag{11}$$

Since the elements of the perturbation term B are asymptotically normally distributed with mean zero, therefore $u_i'Bu_i$ which is a linear combination of the elements B will also be asymptotically normally distributed with mean zero, $E\left(u_i'Bu_i\right) = 0 \ \forall i$ and finite variance $\sigma_i^2 = Var\left(u_i'Bu_i\right) = \left(u_i' \otimes u_i'\right)Cov(B)\left(u_i \otimes u_i\right)$, where $\otimes$ denotes the Kronecker product. Moreover $\sqrt{n}\left(\hat{\lambda}_i - \lambda_i\right)$ has the same asymptotic distribution as $u_i'Bu_i$, and it can immediately be established that $\sqrt{n}\left(\hat{\lambda}_i - \lambda_i\right)$ converge in distribution in normally variate with mean zero and finite variance σ_i^2.

$$\sqrt{n}\left(\hat{\lambda}_i - \lambda_i\right) \to^d N\left(0, \sigma_i^2\right). \tag{12}$$

In particular, under $H_0' : \lambda_i = 0$, the smallest eigenvalue of $\hat{\Sigma}$, $\hat{\lambda}_i$ is asymptotically normally distributed with mean zero and finite asymptotic

variance σ_i^2/n.

In [12] [13], the perturbation theory has been adapted to bounded perturbations that permits to use it, due to the iterated logarithm law, in consistent rank estimation of matrix.

Having fully defined the statistical properties of $\hat{\lambda}_i$ we shall propose a one-side test statistic to decide whether eigenvalue should be declared zero or not.

Let $\hat{\sigma}_i^2$ be a consistent estimator of σ_i^2. According to the asymptotic distribution of $\hat{\lambda}_i$ we reject the null hypothesis H_0' if $\hat{\lambda}_i > z_\alpha \hat{\sigma}_i/\sqrt{n}$ where z_α represents the percentile of the standard normal distribution.

We have defined a value of the threshold adequately. To assess the test statistic, approximate p-values are computed with reference to standard normal distribution.

The procedure of test is based on a sequential testing of the smallest eigenvalues of $\hat{\Sigma}$. If the null hypothesis H_0' is rejected for the smallest $l-r$ eigenvalues then the rank of Σ is found to be r. Then, in the construction of generalised Wald test statistics W_+, we keep only the largest r eigenvalues of $\hat{\Sigma}$.

More precisely, the rank of Σ is detected sequentially using the pretest procedure. Since the eigenvalues of the estimated $\hat{\Sigma}$ are ordered, and tests for individual eigenvalues are considered, it is more appropriate to use an upward testing procedure. Starting with the smallest eigenvalue $\hat{\lambda}$ of $\hat{\Sigma}$, we carry out tests with progressively larger $\hat{\lambda}$ until we find that the test reject the null hypothesis that the rank considered is correct.

In sum, the sequential testing procedure proposed in this paper consists of the following two steps:

Step 1: Determine the rank of Σ by testing $H_0' : \lambda_i = 0$ by above pre-testing procedure (upward testing procedure).

Step 2: If Σ is found to be full rank, test the null hypothesis $H_0 : g(\theta) = 0$ with W (Equation (4)).

Otherwise, test the null H_0 with W_+ in using the decomposition spectral of $\hat{\Sigma}^+$ (Equation (6)).

4. Conclusions

In this paper, we have proposed an asymptotic theory for Wald type test of nonlinear restrictions. The Wald statistic is known to be asymptotically chi-square distributed under the null hypothesis, provided that the Jacobian of the restriction function describing the null hypothesis has full rank and the asymptotic covariance matrix of the estimator parameters is non-singular. There are, however, examples of important models and hypotheses in economics which do not satisfy theses regular conditions and the Wald statistic fails to have limiting chi-squared distribution. We derive a generalised Wald test statistics which guarantees the asymptotic chi-square distribution under nonregular conditions.

The limiting distribution of the Wald statistic for testing nonlinear restrictions depends substantially on the singularity or not of the asymptotic covariance matrix, and the determination of its rank is of great importance in this context. It is well known that the rank of the matrix is equal to the number of the eigenvalues nonzero. Using the perturbation theory, we derive the asymptotic distribution of the eigenvalues. We propose two new threshold values for the eigenvalues which permit to decide if the eigenvalues are nonzero or not.

References

[1] Sargan, J.D. (1980) Some Tests of Dynamic Specification for a Single Equation. *Econometrica*, **48**, 879-898. https://doi.org/10.2307/1912938

[2] Gregory, A.W. and Veall, M.R. (1985) Formulating Wald Tests of Nonlinear Restrictions. *Econometrica*, **53**, 1465-1468. https://doi.org/10.2307/1913221

[3] Gregory, A.W. and Veall, M.R. (1986) Wald Tests of Common Factor Restrictions. *Economics Letters*, **22**, 203-208. https://doi.org/10.1016/0165-1765(86)90232-6

[4] Andrew, D.W.K. (1987) Asymptotic Results for Generalized Wald Tests. *Econometric Theory*, **3**, 348-358. https://doi.org/10.1017/S0266466600010434

[5] Toda, H.Y. and Phillips, P.C.B. (1993) Vector Autoregression and Causality. *Econometrica*, **59**, 229-255. https://doi.org/10.1016/0304-4076(93)90024-Y

[6] Lütkepohl, H. and Burda (1997) Modified Wald Tests under Nonregular Conditions. *Journal of Econometrics*, **78**, 315-332.
https://doi.org/10.1016/S0304-4076(97)80015-2

[7] Lütkepohl, H. (1993) Introduction to Multiple Time Series Analysis. 2nd Edition, Springer, Berlin. https://doi.org/10.1007/978-3-642-61695-2

[8] Johansen, S. (1995) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.
https://doi.org/10.1093/0198774508.001.0001

[9] Gallant, A.R. (1977) Testing a Nonlinear Regression Specification: A Nonregular case. *Journal of American Statistical Association*, **72**, 523-530.
https://doi.org/10.1080/01621459.1977.10480606

[10] Ratsimalahelo, Z. (2001) Specification of VARMA Echelon Form Models. In: Neck, R. Eds., *Proceedings IFAC in Symposium on Modelling and Control of Economic Systems*, Elsevier, Pergamon Press, 129-137.
https://doi.org/10.1016/S1474-6670(17)33052-5

[11] Wilkinson, W.H. (1965) Algebraic Eigenvalue Problems. Oxford University Press, Oxford.

[12] Ratsimalahelo, Z. (2003) Strongly Consistent Determination of the Rank of Matrix. Working Paper, University of Franche-Comté, U.F.R Sciences Economiques.

[13] Lardies, J. and Ratsimalahelo, Z. (2005) Model Order Estimation of Multivariable Stochastic Process. *Communications in Statistics-Simulation and Computation*, **34**, 863-877. https://doi.org/10.1080/03610910500308123