

Khalaf, L. & Urga, G. (2014). Identification robust inference in cointegrating regressions. *Journal of Econometrics*, 182(2), pp. 385-396. doi: 10.1016/j.jeconom.2014.06.001



**CITY UNIVERSITY  
LONDON**

[City Research Online](http://www.city.ac.uk/researchonline/)

**Original citation:** Khalaf, L. & Urga, G. (2014). Identification robust inference in cointegrating regressions. *Journal of Econometrics*, 182(2), pp. 385-396. doi: 10.1016/j.jeconom.2014.06.001

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/14536/>

### **Copyright & reuse**

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

### **Versions of research**

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

### **Enquiries**

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).

# Identification robust inference in cointegrating regressions

Lynda Khalaf \*

Carleton University, Ottawa (Canada)

Giovanni Urga †

Cass Business School, City University London (UK)  
& Bergamo University (Italy)

May 30, 2014

---

\* Groupe de recherche en économie de l'énergie, de l'environnement et des ressources naturelles (GREEN) Université Laval, Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Economics Department, Carleton University, Canada. Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel (613) 520-2600-8697; FAX: (613)-520-3906. email: Lynda\_Khalaf@carleton.ca.

†Centre for Econometric Analysis, Faculty of Finance, Cass Business School, City University London, UK, and Dept. of Management, Economics and Quantitative Methods, University of Bergamo, Italy. Mailing address: 106 Bunhill Row, London EC1Y 8TZ U.K.. Tel. (44) 20 70408698 Fax. (44) 20 70408881. e-mail: g.urga@city.ac.uk

## Abstract

In cointegrating regressions, estimators and test statistics are nuisance parameter dependent. This paper addresses this problem from an identification-robust perspective. Confidence sets for the long-run coefficient (denoted  $\beta$ ) are proposed that invert LR-tests against an unrestricted or a cointegration-restricted alternative. For empirically relevant special cases, we provide analytical solutions to the inversion problem. A simulation study, imposing and relaxing strong exogeneity, analyzes our methods relative to standard Maximum Likelihood, Fully Modified and Dynamic OLS, and a stationarity-test based counterpart. In contrast with all the above, proposed methods have good size regardless of the identification status, and good power when  $\beta$  is identified.

**J.E.L. Classification Numbers:** C32, C12.

**Keywords:** Cointegration, Weak Identification, Bound Test, Simulation-Based Inference.

# 1 Introduction

Cointegration models - defined as stationary linear combinations between non-stationary variables - have wide applicability in econometrics. However, it is becoming increasingly clear from the literature that inference on cointegrating vectors is a challenging problem. In a recent survey, Johansen (2009) discusses, among others, two important reasons for the above. First, cointegrating equations have traditionally been interpreted as long-term relations, yet time series that can be modelled as such are short. Therefore, it becomes a natural part of the methodology to develop finite sample motivated methods. Second, finite sample methods have nevertheless been notably lacking. Available estimators and test statistics heavily rely on asymptotic theory, and more importantly, are nuisance parameter dependent which may cause severe finite sample distortions.

To set focus, consider the vector autoregressive framework of Johansen (1995) which, given a  $p$ -dimensional vector  $X_t$ , relies on the regression of  $\Delta X_t$  on  $X_{t-1}$ , and *e.g.* a constant and further lags of  $\Delta X_t$ . Let  $\Pi$  refer to the coefficient of  $X_{t-1}$  in the latter regression. The cointegrating relation and associated long-run coefficient, denoted as the  $(p \times r)$  matrix  $\beta$ , are defined in this context via a reduced rank restriction of the form  $\Pi = \gamma\beta'$ , where  $r$  refers to the cointegration rank. This paper focuses on estimating and testing long-run parameters without assuming that they are identified.

Identification failure typically occurs when the statistical objective function does not respond to some parameters, which is inherent to the above structure. This is because  $\beta$  cannot be recovered from the restriction  $\Pi = \gamma\beta'$  when  $\gamma$  is close to zero or is rank deficient, so within and close to this region, the likelihood function will inevitably be ill-behaved. Dufour (1997) is perhaps the first to formalize this issue via an illustrative bivariate process.

In traditional discussions of cointegration, related issues with  $\gamma$  are acknowledged although not widely recognized. Johansen (1988, 2000, 2002) show that standard likelihood ratio (LR) criteria are asymptotically  $\chi^2$  and Bartlett adjustable as long as  $\gamma \neq 0$  yet perform poorly otherwise.<sup>1</sup> Phillips (1994) argues that finite sample inference on  $\beta$  is

---

<sup>1</sup>In fact Johansen (2000, p. 741) defines the problematic parameter subspace as "the boundary where the order of integration or the number of cointegrating relations change".

possible in triangular systems setting  $\gamma = -(I_r, 0)'$  which amounts to imposing weak exogeneity and ruling out dynamics and feedback.<sup>2</sup> Johansen (1995, Chapter 8) formally links weak exogeneity to zero restrictions on components of  $\gamma$ . Further insights on less restrictive parametrizations of  $\gamma$  and their relevance and implications on inference may be traced back to the simulation design of Gonzalo (1994). One aim of the present paper is to provide an identification basis for understanding and solving such problems.

More generally, identification problems have previously been addressed in a variety of settings including the enduring weak-instruments case.<sup>3</sup> However, to our knowledge, cointegration has not been directly addressed. It may be worth remarking that Dufour (1997) raises yet does not solve the cointegration case. The contribution of the present paper is a formal solution for inference on  $\beta$  placing no prior restrictions on  $\gamma$ . In line with the above cited identification-robust literature, the main principles we follow and show can be summarized as follows. (1) Standard asymptotics provide poor approximations to the distributions of estimators and test statistics. (2) Wald-type confidence intervals of the form  $\{\text{estimate} \pm (\text{asymptotic standard error}) \times (\text{asymptotic critical point})\}$  will severely understate estimation uncertainty. (3) In contrast, likelihood-ratio type methods admit identification robust bounds which provides a first step towards a useful solution. (4) It is important to consider methods that allow for unbounded and possibly empty outcomes.

A few other papers have considered different although related problems in cointegrating regressions. In particular, Wright (2000) and Müller and Watson (2013) consider models in which regressors have roots local to unity while some linear combination of the regressand and regressors is stationary.<sup>4</sup> Tanaka (1993) and Jansson and Haldrup (2002) define setups in which regressors have unit roots yet some linear combination of the regressand and regressors is nearly stationary. Alternatively, Ioannidis and Chronis (2005) assume that nearly integrated series are nearly cointegrated when a linear combination exists

---

<sup>2</sup> $I_r$  refers to an  $r$ -dimensional identity matrix.

<sup>3</sup>See *e.g.* Dufour (2003), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Dufour and Jasiak (2001), Kleibergen (2002, 2005), Stock, Wright and Yogo (2002), Moreira (2003), Dufour and Taamouti (2005, 2007), Andrews, Moreira and Stock (2006), Guggenberger and Smith (2008), Antoine and Lavergne (2012), Guggenberger, Kleibergen, Mavroeidis and Chen (2012), Andrews and Cheng (2013).

<sup>4</sup>The near unit root issue may be traced back to Stock (1997) and Elliott (1998). See also Zivot (2000), Lanne (2000), Caner and Kilian (2001), Hjalmarrsson and Österholm (2010) and the references therein; on bootstraps with near-unit roots, see *e.g.* Andrews (2000) and Park (2006).

with a near integration order that is smaller than the order of near integration of the considered series. With the exception of Wright (2000) and more recently Müller and Watson (2013), this literature does not address inference. Wright (2000) tests a specified value of  $\beta$  by assessing the stationarity of resulting residuals for a single cointegrating vector. Müller and Watson (2013) relax the latter restriction yet work within a common trend definition of cointegration that introduces further complexities via high-dimensional nuisance parameters. Our approach in this papers remains within the tractable and by now well understood reduced rank regression likelihood framework.

Formally, we propose to invert LR-type statistics that test a specified value for  $\beta$  against (i) an unrestricted, or (ii) a cointegration-restricted alternative. Tests on  $\Pi$  in implicit form as also considered as in Phillips (1994). We underscore - as in Wright (2000) - the merits of a confidence set that can be empty, and characterize unbounded outcomes as well. Our results link unbounded and empty confidence sets to departures from the cointegration hypothesis, the consequences of which are of obvious concern. Formally, we show that unbounded confidence sets which suggest that available data is uninformative on  $\beta$  may result from overestimating the rank of  $\Pi$ . In contrast, empty sets may result from underestimating the rank of  $\Pi$  which also reflects departures from the exact unit root assumption on the components of  $X_t$ .

Allowing for possible weak identification, we propose three methods to adequately size the above defined statistics. The first method involves a bounds-based critical value; for general insights on the usefulness of bounds when nuisance parameters yield identification problems, see Dufour (1989, 1997), Dufour and Khalaf (2002) and Beaulieu, Dufour and Khalaf (2013a,b). The latter may be viewed as a *least favorable* (LF) critical value in the sense of Andrews and Cheng (2013). Second, we introduce a data-dependent critical value based on the "*Type 2 Robust*" approach from Andrews and Cheng (2013). The latter checks whether available data suggests weak identification and if so, adjusts the cut-off towards the bound via a smooth transition function. Said differently, the Type 2 robust procedure involves a data-based continuous transition from the standard to the bounds-based LF critical value that improves size-corrected power. Third, we examine a simulation-based method based on Dufour (2006) that may be interpreted, because of its

parametric basis, as an often unattainable full-information *first best* (FB).

For the special cases  $r = 1$  and  $r = p - 1$ , we provide analytical solutions to the inversion problem. These solutions use the mathematics of quadrics as in Dufour and Taamouti (2005). The proposed LF and Type 2 critical values do not vary with the tested value of  $\beta$  and thus preserve the quadrics form of the test inversion solution for these special cases.

Finally, we conduct a simulation study to assess the properties of our proposed inference methods. In addition, we also check whether and to what degree available competing methods, specifically the Maximum Likelihood of Johansen (1995), the Fully Modified OLS (FMOLS) of Phillips and Hansen (1990) and Phillips (1991, 1995), the Dynamic OLS (DOLS) of Stock and Watson (1993), and the stationarity-test based method from Wright (2000), suffer from identification problems. Our simulation design goes beyond triangular representations that facilitate finite sample analysis; see Gonzalo (1994) or Boswijk (1995) for early references in this regard. We thus follow Gonzalo's simulation design which allows us to control persistence as well as exogeneity. Results can be summarized as follows.

Although high persistence causes size distortions for the considered LR statistics, these are easily corrected as proposed above, imposing and relaxing weak exogeneity. The size of DOLS and FMOLS based t-tests exceeds 90% at the boundary. Furthermore, failure of weak exogeneity causes very severe distortions for DOLS (size  $\simeq 88\%$  even with  $T = 300$ ) as well as for FMOLS, albeit to a lesser extent (size nevertheless remains around 37% with  $T = 300$ ), even when  $\beta$  is identified. The test from Wright (2000) is also oversized at the boundary. In contrast, even when weak exogeneity fails, our proposed methods have good size regardless of the identification status, and good power when  $\beta$  is identified. With regards to power, our proposed Type 2 robust method is as powerful as the FB bootstrap. This is noteworthy since the Type 2 method does not require full information, while the FB (here by construction) utilizes the often unavailable information on the dependence structure of residuals in the cointegrating equation.

The remainder of the paper is organized as follows. In Section 2, we set-up the framework and introduce the hypotheses associated with the test we propose to invert. The statistics underlying these tests are defined and analyzed in Section 3, and robust cut-off

points are introduced in Section 4. In Section 5, we present the test inversion strategy for the general case. Section 6 discusses the  $r = 1$  and  $r = p - 1$  special cases. The simulation study is discussed in Section 7, while Section 8 concludes the paper. The technical Appendix A.1 summarizes the general projection methods applied, while Appendix A.2 reports the proofs of Theorems and Lemmas.

## 2 Framework and methodological overview

Consider (see Johansen, 1995) the  $p$ -dimensional process  $X_t$  defined by

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + e_t, \quad t = 1, \dots, T, \quad (1)$$

$$\Pi = \gamma \beta' \quad (2)$$

where  $e_t$  are i.i.d.  $N_p(0, \Omega)$ , initial values  $X_{-k}, \dots, X_0$  are fixed and  $\gamma$  and  $\beta$  are unknown  $p \times r$  matrices with rank  $r$ ,  $r$  is the cointegration rank,  $D_t$  is the  $m$ -dimensional deterministic term. We use this framework as a basis to derive a confidence set for  $\beta$  that is robust to the identification problem arising from its definition via the non-linear restriction (2).

To introduce the considered test statistics, we adopt the notation from Johansen (1995) that will facilitate our presentation. Define  $Z_{0t} = \Delta X_t$ ,  $Z_{1t} = X_{t-1}$  and stack the variables  $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$  and  $D_t$  into the  $p(k-1) + m$  vector  $Z_{2t}$  and introduce the conformable  $p \times p(k-1) + m$  coefficient matrix  $\Psi$  comprising  $\Gamma_1, \dots, \Gamma_{k-1}, \Phi$ , leading to

$$Z_{0t} = \Pi Z_{1t} + \Psi Z_{2t} + \epsilon_t, \quad t = 1, \dots, T.$$

To concentrate the latter model into the standard reduced rank regression framework "in residuals" (Johansen, 1995, Chapter 6), we further define

$$R_{0t} = \Pi R_{1t} + \epsilon_t, \quad t = 1, \dots, T \quad (3)$$

$$R_0 = R_1 \Pi' + \epsilon, \quad (4)$$

where  $R_0$  and  $R_1$  are the  $T \times p$  matrices with rows  $R'_{0t}$  and  $R'_{1t}$ ,  $\epsilon$  is the  $T \times p$  matrix with



rows  $\epsilon'_t$ ,

$$R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t}, \quad R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t} \quad (5)$$

$$M_{ij} = M'_{ji} = T^{-1} \sum_{t=1}^T Z_{it}Z'_{jt}, \quad i, \quad j = 0, 1, 2. \quad (6)$$

For further reference and in line with our notation in the Introduction, let  $I_s$  refer to an  $s$ -dimensional identity matrix.

Our set estimates rely on the standard normalizations

$$\beta = (I_r, \mathbf{b}')' \quad (7)$$

where  $\mathbf{b}$  is the  $(p-r) \times r$  unknown matrix of interest. This normalization is convenient for various purposes as discussed in Johansen (1995, Chapter 13, Section 13.2). It implies that the  $p$  observables need to be classified in two groups, of dimension  $r$  and  $p-r$ , respectively, so that a set of  $p-r$  variables do not cointegrate. In contrast to traditional methods, our set estimates allow one to validate the considered classification. An empty outcome suggests the choice in question is incompatible with data and can be safely refuted. An unbounded outcome suggests the chosen observables do not lack fit yet contain sparse information on associated long-run parameters. Such checks are built into set estimates, so pre-tests are not needed.

In this context, when  $\gamma = 0$ , that is, when the components of  $X_t$  are not cointegrated,  $\beta$  is not identified. Identification problems would also occur if  $\gamma$  is rank deficient, which reflects misspecifying the number of cointegrating relations. Our objective is to provide a confidence set for  $\beta$  which is valid whether the rank condition on  $\gamma$  holds or not. We proceed by inverting test statistics for the hypothesis that fixes  $\beta$  to a known value

$$H_0(\beta_0) : \Pi = \gamma\beta'_0, \quad \beta_0 \quad \text{known}, \quad (8)$$

which subject to (7) gives

$$H_0(\mathbf{b}_0) : \Pi = \gamma\beta'_0, \quad \beta_0 = (I_r, \mathbf{b}'_0)', \quad \mathbf{b}_0 \quad \text{known},$$

Inverting a test of  $H_0(\mathbf{b}_0)$  at a given level  $\alpha$  consists in collecting, numerically or analytically, the  $\mathbf{b}_0$  values that are not rejected using the considered test at the considered

level. For example, given a right-tailed test statistic  $\mathcal{T}(\mathbf{b}_0)$  with  $\alpha$ -level cut-off point  $\mathcal{T}_c$ , test inversion involves solving, over  $\mathbf{b}_0$ , the inequality  $\mathcal{T}(\mathbf{b}_0) < \mathcal{T}_c$ . The solution of this inequality is a parameter space subset, denoted  $\text{CS}(\mathbf{b}; \alpha)$ , that satisfies

$$\mathbf{P}[\mathbf{b} \in \text{CS}(\mathbf{b}; \alpha)] \geq 1 - \alpha. \quad (9)$$

Identification robustness requires a convenient choice of  $\mathcal{T}_c$  so that (9) regardless of the rank of  $\gamma$ .

A complete description of our methodology thus requires: (i) defining the test statistics we propose to invert, (ii) obtaining identification-robust cut-off points for these statistics, and (iii) characterizing the inversion solution. Steps (i) - (iii) are discussed, in turn, in the following sections.

### 3 Test statistics

We consider LR type statistics that are well known in multivariate regression; see Dufour and Khalaf (2002) and the references therein. Specifically, we use the following three statistics. The first is the standard LR statistic associated with the null hypothesis (8):

$$LR(\mathbf{b}_0) = T \ln \left[ \frac{|S_0|}{|S|} \right], \quad (10)$$

$$S_0 = R'_0 \left( I - R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \beta'_0 R'_1 \right) R_0, \quad (11)$$

$$S = R'_0 \left( I - R_1 (R'_1 R_1)^{-1} R'_1 \right) R_0 \quad (12)$$

with  $\beta_0 = (I_r, \mathbf{b}'_0)'$ . In this context, if  $\beta = \beta_0$  then the multivariate regression

$$R_0 = R_1 \beta_0 \gamma' + \epsilon \quad (13)$$

is linear in parameters, so assessing  $\beta = \beta_0$  against the unrestricted counterpart involves a linear restriction. The numerator in (10) is the sum-of-squared residuals of (13) while its denominator is the fully unrestricted sum-of-squared residuals from (3).

The second statistic assesses (8) against a restricted alternative, by replacing the unrestricted estimate in the denominator of (10) with a cointegration restricted one, as follows.

Given

$$S = \left( R_0 - R_1 \hat{\Pi}' \right)' \left( R_0 - R_1 \hat{\Pi}' \right), \quad \hat{\Pi}' = (R'_1 R_1)^{-1} R'_1 R_0 \quad (14)$$

replace  $\hat{\Pi}$  by  $\tilde{\Pi} = \tilde{\gamma}\tilde{\beta}'$  where  $\tilde{\beta}$  is the MLE estimator of  $\beta$  and  $\tilde{\gamma}$  is obtained by inserting  $\tilde{\beta}$  into (13) and then applying OLS. Note that  $\tilde{\gamma} = \left(\tilde{\beta}'R_1'R_1\tilde{\beta}\right)^{-1}\tilde{\beta}'R_1'R_0$  so

$$R_0 - R_1\tilde{\Pi}' = \left(I - R_1\tilde{\beta}\left(\tilde{\beta}'R_1'R_1\tilde{\beta}\right)^{-1}\tilde{\beta}'R_1'\right)R_0$$

and the cointegration-restricted statistic obtains as

$$LRC(\mathbf{b}_0) = T \ln \left[ \frac{|S_0|}{|\tilde{S}|} \right], \quad \tilde{S} = R_0' \left( I - R_1\tilde{\beta}\left(\tilde{\beta}'R_1'R_1\tilde{\beta}\right)^{-1}\tilde{\beta}'R_1' \right) R_0. \quad (15)$$

Finally, we consider the counterpart of  $LR(\mathbf{b}_0)$  which assesses  $H_0(\mathbf{b}_0)$  in the implicit form

$$H_{0\perp}(\mathbf{b}_0) : \Pi\beta_{\perp 0} = \mathbf{0}, \quad \beta_{\perp 0} = (-\mathbf{b}'_0, I_{p-r})'. \quad (16)$$

Observe that the  $p \times (p-r)$  matrix  $\beta_{\perp 0}$  satisfies the orthogonality condition  $\beta'_{\perp 0}\beta_0 = 0$ .

The statistic is the LR criterion

$$LRP(\mathbf{b}_0) = T \ln \left[ \frac{|S_{\perp 0}|}{|S|} \right], \quad (17)$$

$$S_{\perp 0} = S + \hat{\Pi}\beta_{\perp 0}[\beta'_{\perp 0}(R_1'R_1)^{-1}\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\hat{\Pi}'.$$

This statistic and the underlying approach consisting in direct tests on  $\Pi$  relates to the Wald tests considered by Phillips (1994).

A well known result on determinants<sup>5</sup> leading to

$$\begin{aligned} |S_{\perp 0}|/|S| &= \left| I_p + S^{-1}\hat{\Pi}\beta_{\perp 0}[\beta'_{\perp 0}(R_1'R_1)^{-1}\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\hat{\Pi}' \right| \\ &= \left| I_{p-r} + [\beta'_{\perp 0}(R_1'R_1)^{-1}\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\hat{\Pi}'S^{-1}\hat{\Pi}\beta_{\perp 0} \right| \end{aligned}$$

serves to write  $LRP(\mathbf{b}_0)$  in the following form which will be used below

$$LRP(\mathbf{b}_0) = T \ln \left( \left| I_{p-r} + [\beta'_{\perp 0}(R_1'R_1)^{-1}\beta_{\perp 0}]^{-1}\beta'_{\perp 0}\hat{\Pi}'S^{-1}\hat{\Pi}\beta_{\perp 0} \right| \right). \quad (18)$$

---

<sup>5</sup>For any  $n \times m$  matrix  $S$  and any  $m \times n$  matrix  $U$ ,  $|I_n + SU| = |I_m + US|$ ; see e.g. Harville(1997, Section 18.1, p. 416).

The same result also yields the following useful decomposition for  $LR(\mathbf{b}_0)$ .

$$\begin{aligned}
LR(\mathbf{b}_0) &= T \ln \left[ \frac{|\frac{1}{T}R'_0R_0|}{|\frac{1}{T}R'_0(I_T - R_1(R'_1R_1)^{-1}R'_1)R_0|} \right] \\
&+ T \ln \left[ \frac{|\frac{1}{T}R'_0(I_T - R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1}\beta'_0R'_1)R_0|}{|\frac{1}{T}R'_0R_0|} \right] \\
&= -T \ln \left( \left| I_p - (R'_0R_0)^{-1}R'_0R_1(R'_1R_1)^{-1}R'_1R_0 \right| \right) \\
&+ T \ln \left( \left| I_p - (R'_0R_0)^{-1}R'_0R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1}\beta'_0R'_1R_0 \right| \right).
\end{aligned}$$

So rewriting the determinant of the second term gives

$$\begin{aligned}
LR(\mathbf{b}_0) &= -T \ln \left( \left| I_p - (R'_0R_0)^{-1}R'_0R_1(R'_1R_1)^{-1}R'_1R_0 \right| \right) \\
&+ T \ln \left( \left| I_r - \beta'_0R'_1R_0(R'_0R_0)^{-1}R'_0R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1} \right| \right).
\end{aligned} \tag{19}$$

Applying similar decompositions to  $LRC(\mathbf{b}_0)$  we get

$$\begin{aligned}
LRC(\mathbf{b}_0) &= T \ln \left[ \frac{|\frac{1}{T}R'_0R_0|}{\left| \frac{1}{T}R'_0 \left( I - R_1\tilde{\beta} \left( \tilde{\beta}'R'_1R_1\tilde{\beta} \right)^{-1} \tilde{\beta}'R'_1 \right) R_0 \right|} \right] \\
&+ T \ln \left[ \frac{|\frac{1}{T}R'_0(I_T - R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1}\beta'_0R'_1)R_0|}{|\frac{1}{T}R'_0R_0|} \right] \\
&= -T \ln \left( \left| I_p - (R'_0R_0)^{-1}R'_0R_1\tilde{\beta} \left( \tilde{\beta}'R'_1R_1\tilde{\beta} \right)^{-1} \tilde{\beta}'R'_1R_0 \right| \right) \\
&+ T \ln \left( \left| I_p - (R'_0R_0)^{-1}R'_0R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1}\beta'_0R'_1R_0 \right| \right)
\end{aligned}$$

which again yields

$$\begin{aligned}
LRC(\mathbf{b}_0) &= -T \ln \left( \left| I_r - \tilde{\beta}'R'_1R_0(R'_0R_0)^{-1}R'_0R_1\tilde{\beta} \left( \tilde{\beta}'R'_1R_1\tilde{\beta} \right)^{-1} \right| \right) \\
&+ T \ln \left( \left| I_r - \beta'_0R'_1R_0(R'_0R_0)^{-1}R'_0R_1\beta_0(\beta'_0R'_1R_1\beta_0)^{-1} \right| \right).
\end{aligned} \tag{20}$$

## 4 Identification robust critical points

The asymptotic null distribution of  $LRC(\mathbf{b}_0)$ , for any  $\beta_0$  value under test provided  $\gamma$  is not rank deficient (see e.g. Chapter 7 from Johansen, 1995) is  $\chi^2(l_c)$  with

$$l_c = [p^2 - (p - r)^2] - pr = r(p - r), \tag{21}$$

$$p^2 - (p - r)^2 = \text{number of parameters imposing cointegration}, \tag{22}$$

$$pr = \text{number of parameters in } \gamma. \tag{23}$$

The same asymptotics, provided  $\gamma$  is not rank deficient, suggest that the null distribution of  $LR(\mathbf{b}_0)$  can be approximated as  $\chi^2(l_u)$  with

$$l_u = p^2 - pr = p(p - r) \quad (24)$$

$$p^2 = \text{number of parameters in } \Pi \quad (25)$$

$$pr = \text{number of parameters in } \gamma. \quad (26)$$

Given the duality between the implicit and explicit tests, the null distribution of  $LRP(\mathbf{b}_0)$  can also be approximated as  $\chi^2(l_u)$ . These approximations will perform poorly when  $\gamma$  may be rank deficient and also possibly when weak exogeneity fails. We thus propose alternative cut-off points that will control size for both statistics.

Using an argument similar to the one in Dufour (1989), for a univariate regression, and Dufour and Khalaf (2002), for multivariate regression, we show that the above defined LR statistics have null distributions which admit an identification-robust bound that can be described as follows. We introduce a hypothesis (denoted  $H_0^*$ ) that fixes both  $\gamma$  and  $\beta$ , which is a special case of the restrictions to be tested. Then we argue that the LR criterion (denoted  $LR_*$ ) associated with  $H_0^*$  provides the desired bound. The result follows from two considerations. First, by construction, it is evident that  $LR_*$  is larger than the LR test statistics of interest, and thus its null distribution yields an upper bound (and conservative critical points) applicable to both  $LR(\mathbf{b}_0)$  and  $LRC(\mathbf{b}_0)$ . Second, the null distribution of  $LR_*$  can be approximated using a standard  $\chi^2$  cut-off point regardless of the specific values of  $\gamma$  and  $\beta$  in  $H_0^*$ . It is worth noting that the bound implicit in Dufour (1997, Theorem 5.1) may be obtained using the same rationale as the bounds presented here.

**Theorem 1** *In the context of (1)-(2) with  $1 < r < p - 1$  and the null hypothesis (8), consider the LR statistic, denoted  $LR_*$ , for testing, against an unrestricted alternative,*

$$H_0^* : \Pi = \gamma_0 \beta_0' \quad (27)$$

*such that  $\beta_0$  satisfies (8) and  $\gamma_0 \neq 0$  is known. Then*

$$P[LR(\mathbf{b}_0) \geq \lambda_*(\alpha)] \leq \alpha \quad (28)$$

*for all  $0 \leq \alpha \leq 1$ , where  $\lambda_*(\alpha)$  is determined such that  $P[LR_* \geq \lambda_*(\alpha)] = \alpha$ .*

The asymptotic null distribution of  $LR_*$ , with reference *e.g.* to Chapter 7 from Johansen (1995), is  $\chi^2(p^2)$ . The finite-sample dominance result in (28) suggests the  $\chi^2(p^2)$  as a LF bound for the null distribution of both  $LR(\mathbf{b}_0)$  and  $LRC(\mathbf{b}_0)$ . Indeed, it is also easy to see that  $LRC(\mathbf{b}_0) \leq LR(\mathbf{b}_0) \leq LR_*$  which suggests that although valid for both statistics, our proposed bound is tighter in the case of  $LR(\mathbf{b}_0)$ .

If  $\beta$  is identified, the above defined bound may prove to be conservative. We thus introduce an alternative critical value adapted from the Type 2 approach of Andrews and Cheng (2013). The idea is to define an  $\alpha$ -level critical value that provides a continuous transition from a weak-identification to a strong-identification cut-off point, using a data depend function which would assess the extent of weak-identification. For this purpose, we use the smooth transition function

$$\mathbf{s}(x) = \exp(-x/2)$$

recommended by Andrews and Cheng (2013) applied to the discrepancy between Johansen's cointegration test statistic and its  $\alpha$ -level cut-off point. Based on the magnitude of this discrepancy, we define a cut-off point denoted  $\hat{c}$  which transitions between the LF  $\chi^2(p^2)$  critical point and its strong-identification counterparts ( $\chi^2$  with degrees-of-freedom as in (24) or (24)). Formally, for an  $\alpha$ -level test, we propose:

$$\hat{c} = \left\{ \begin{array}{ll} c_B & \text{if } A_n \leq \kappa \\ c_s + [c_B - c_s]\mathbf{s}(A_n - \kappa) & \text{if } A_n > \kappa \end{array} \right\} \quad (29)$$

$$A_n = -T \ln \left( \sum_{i=1}^p (1 - \hat{\lambda}_i) \right) \quad (30)$$

where  $c_B$  is the  $\alpha$ -level  $\chi^2(p^2)$  cut-off point,  $c_s$  is the  $\alpha$ -level  $\chi^2(p^2 - pr)$  for  $LR(\mathbf{b}_0)$  and or the  $\alpha$ -level  $\chi^2(r(p - r))$  for  $LRC(\mathbf{b}_0)$ ,  $\hat{\lambda}_i$  are the eigen values of  $\hat{\Pi}$  so  $A_n$  is Johansen's statistic associated with  $rank(\Pi) = 0$  and  $\kappa$  is its  $\alpha$ -level tabulated cut-off point as reported *e.g.* in Chapter 15 in Johansen (1995). We use  $(A_n - \kappa)$  in the sense of Andrews and Cheng (2013) as a transition metric to gauge rather than pre-test the strength of identification.

To conclude, observe that  $LRP(\mathbf{b}_0)$  admits the same LF bound and Type 2 cut-off points, given the equivalence of the underlying testing procedures; see, for example, Gouriéroux, Monfort and Renault (1995).

## 5 Test inversion, the general case

Inverting the above defined test statistics involves solving the inequations

$$LR(\mathbf{b}_0) < \bar{c}, \quad (31)$$

$$LRC(\mathbf{b}_0) < \tilde{c}, \quad (32)$$

$$LRP(\mathbf{b}_0) < \bar{c}, \quad (33)$$

where  $\bar{c}$  refers to the cut-off point associated with both  $LR(\mathbf{b}_0)$  and  $LRP(\mathbf{b}_0)$  at a desired level (say 5%), and  $\tilde{c}$  to the critical point associated with  $LRC(\mathbf{b}_0)$  and the same test level. These critical points may be sets to either  $c_B$ , or  $\hat{c}$  as defined in (29).<sup>6</sup> When  $r$  is either 1 or  $p - 1$ , we can find an analytical solution to the inversion problem, which will be discussed in section 6.

When  $1 < r < p - 1$ , a numerical solution is required in which case inverting  $LR(\mathbf{b}_0)$  is equivalent to inverting  $LRP(\mathbf{b}_0)$ . This involves collecting, for example, by grid search, the  $\mathbf{b}_0$  values that are not rejected using the considered test at the considered level. The output of such a search is a joint confidence region, which we denoted above as  $CS(\mathbf{b}; \alpha)$ . Yet the object of interest may consist in deriving confidence intervals for e.g. the individual components of  $\mathbf{b}$ , or more generally, for a given scalar function  $g(\mathbf{b})$ . To do this, we proceed by projecting  $CS(\mathbf{b}; \alpha)$ , that is, by minimizing and maximizing  $g(\mathbf{b})$  over the  $\mathbf{b}$  values in  $CS(\mathbf{b}; \alpha)$ . Confidence intervals so obtained are simultaneous, in the following sense: for any set of  $m$  continuous real valued functions of  $\mathbf{b}$ ,  $g_i(\mathbf{b}) \in R$ ,  $i = 1, \dots, m$ , let  $g_i(CS(\mathbf{b}; \alpha))$  denote the image of  $CS(\mathbf{b}; \alpha)$  by the function  $g_i$ . Then

$$\mathbb{P}[g_i(\mathbf{b}) \in g_i(CS(\mathbf{b}; \alpha)), \quad i = 1, \dots, m] \geq 1 - \alpha. \quad (34)$$

If  $\mathcal{T}_c$  is defined so that (9) holds regardless of the rank of  $\gamma$ , then (34) would also hold whether the rank of  $\gamma$  is full or not.

Our MC experiments (reported in Section 7) show that  $c_B$  or  $\hat{c}$  provide robust approximations for  $\mathcal{T}_c$ . Alternatively, a simulation-based approach can be applied at every step of the above described inversion method. This would correspond to collecting the  $\mathbf{b}_0$  values

---

<sup>6</sup>The mechanics of test inversion will also work with  $c_s$  in the sense that solutions to the considered inequalities can be found using  $\bar{c} = c_s$  or  $\tilde{c} = c_s$ , yet adequate coverage will not be warranted which beats the purpose of inverting the considered tests. This is further illustrated in section 7.

such that a bootstrap-type p-value imposing  $\mathbf{b}_0$  exceeds the considered level. Bootstrap methods have long been available for cointegrating regressions (see e.g. Li 1994, Li and Maddala 1997, Psaradakis 2001, Chang, Park and Song 2006, Palm, Smeekes and Urbain 2010). However, as argued by Palm, Smeekes and Urbain (2010), the properties of these bootstraps are not fully understood. Whether available bootstraps work well in nearly cointegrated systems is so far an open question.<sup>7</sup> As a matter of fact, many bootstraps are known to fail because of identification or boundary issues (Dufour 1997, Andrews 2000, 2001). We thus next propose a Monte Carlo (MC) method to approximate assess (8), treating  $\gamma$  as a nuisance parameter as in Dufour (2006), which allows us to control the level exactly.

## 5.1 Simulation-based procedure

A baseline algorithm is provided drawing from the VECM (3), which assumes: (i) parameters other than  $\gamma$ ,  $\mathbf{b}$  and the variance covariance of errors are partialled-out, (ii) conditioning on initial values of the process, and (iii) imposing normality. Most importantly, the null hypothesis (8) is maintained throughout, so  $\mathbf{b}$  is set to  $\mathbf{b}_0$ . For clarity, the algorithm is presented for the  $LR$  statistic, yet it can be applied in exactly the same way to  $LRC$ ; again, we emphasize that  $LR$  and  $LRP$  will produce numerically identical results. Let  $\Omega$  refer to the Cholesky factor of the error covariance matrix and  $LR^{(0)}$  refer to the observed value of the test statistic.

- A1 For given values of  $\gamma$  and  $\Omega$  and setting  $\mathbf{b}$  to  $\mathbf{b}_0$ , draw  $N$  realizations of  $R_0$  from the Gaussian model (3) each of size  $T$ . Calculating the LR statistic (10) from each draw yields  $LR(\mathbf{b}_0; \gamma, \Omega)^{(j)}$ ,  $j = 1, \dots, N$  which we summarize as the vector

$$\mathcal{L}_N(\gamma, \Omega) = \left( LR(\mathbf{b}_0; \gamma, \Omega)^{(1)}, \dots, LR(\mathbf{b}_0; \gamma, \Omega)^{(N)} \right)'. \quad (35)$$

Our notation emphasizes dependence on  $(\gamma, \Omega)$ , which will become clear as we proceed.

---

<sup>7</sup>We note that available simulation studies that assess the properties of available bootstraps often impose weak exogeneity, an assumption that may be too restrictive and may undercut their reliability.



A2 Define the MC  $p$ -value function

$$p_N[LR^{(0)}|\mathcal{L}_N(\gamma, \Omega)] = \frac{NG_N [LR^{(0)}; \mathcal{L}_N(\gamma, \Omega)] + 1}{N + 1}, \quad (36)$$

$$G_N [LR^{(0)}; \mathcal{L}_N(\gamma, \Omega)] = \frac{1}{N} \sum_{j=1}^N I_{[0, \infty)} [\mathcal{L}_N(\gamma, \Omega) - LR^{(0)}], \quad (37)$$

where  $I_A[x] = 1$ , if  $x \in A$ , and  $I_A[x] = 0$ , if  $x \notin A$ . If  $\gamma$  and  $\Omega$  are given then a decision rule based on comparing

$$\hat{p}_N(LR; \gamma, \Omega) = p_N[LR^{(0)}|\mathcal{L}_N(\gamma, \Omega)] \quad (38)$$

to an  $\alpha$  cut-off where  $\alpha(N + 1)$  is an integer yields a test with size  $\alpha$ .

A3 Typically,  $\gamma$  and  $\Omega$  are not set by the null hypothesis. In this case, maximize  $p_N[LR^{(0)}|\mathcal{L}_N(\gamma, \Omega)]$  over all the  $(\gamma, \Omega)$  values compatible with the null hypothesis maintaining the rank restriction on  $\gamma$ , leading to

$$\hat{p}_N^*(LR) = \sup_{\gamma, \Omega} \{\hat{p}_N(LR; \gamma, \Omega)\}$$

and reject the null hypothesis latter if  $\hat{p}_N^*(LR)$  is less than or equal to  $\alpha$ . Then the probability of rejection under the null hypothesis is itself not larger than  $\alpha$ ; see Dufour (2006).

It is possible to partial  $(\gamma, \Omega)$ , provided the null hypothesis is imposed. The simulation study we report below supports this suggestion. The following modification of A1-A3 would achieve this purpose.

**A1\*** With  $\mathbf{b}$  set to  $\mathbf{b}_0$  estimate  $\gamma$  and  $\Omega$  from the observed data by running the regression of  $R_0$  on  $R_1\beta_0$ , and denote these estimates  $\hat{\gamma}(\mathbf{b}_0)$  and  $\hat{\Omega}(\mathbf{b}_0)$ . Proceed as in A1, replacing  $\gamma$  and  $\Omega$  by  $\hat{\gamma}(\beta_0)$  and  $\hat{\Omega}(\beta_0)$ .

**A2\*** Apply A2, substituting  $\gamma$  and  $\Omega$  by  $\hat{\gamma}(\mathbf{b}_0)$  and  $\hat{\Omega}(\mathbf{b}_0)$  leading to the empirical  $p$ -value

$$\hat{p}_N(LR; \hat{\gamma}(\mathbf{b}_0), \hat{\Omega}(\mathbf{b}_0)) = p_N[LR^{(0)}|\mathcal{L}_N(\hat{\gamma}(\mathbf{b}_0), \hat{\Omega}(\mathbf{b}_0))]. \quad (39)$$

**A3\*** Reject the null hypothesis latter if  $\hat{p}_N(LR; \hat{\gamma}(\mathbf{b}_0), \hat{\Omega}(\mathbf{b}_0))$  is less than or equal to  $\alpha$ .

Because  $\gamma$  affects identification, this does not guarantee that the associated limiting (for  $T \rightarrow \infty$  and finite  $N$ ) null rejection probability will not exceed  $\alpha$ , so the maximization, that is step A3 above, is thus recommended for exactness. Nevertheless, our simulation study suggests that maximization does not seem necessary. We recommend to initiate step A3 using  $\hat{\gamma}(\mathbf{b}_0)$  and  $\hat{\Omega}(\mathbf{b}_0)$ , and to stop maximization as soon as a p-value that exceeds  $\alpha$  is obtained, which speeds the process up substantially.

Test inversion requires *e.g.* running this MC test over a set of economically relevant value of  $\mathbf{b}_0$ . It is also worth noting that the p-value will vary with each  $\mathbf{b}_0$ ; said differently, in contrast with  $c_B$  or  $\hat{c}$ , the corresponding simulated cut-off point will vary with  $\mathbf{b}_0$ . Projection confidence sets at level  $\alpha$  for any linear function of  $\mathbf{b}_0$  require minimizing and maximizing this function imposing that the p-value associated with each  $\mathbf{b}_0$  is greater than  $\alpha$ . Recall that components of  $\mathbf{b}_0$  can be seen as linear function of this matrix of the selection form (zeros and ones). Such restricted optimization problem are not prohibitive, yet in view of our simulation results, using  $c_B$  or  $\hat{c}$  are worthy less expensive alternatives. This discussion also reinforces the usefulness of the analytical special cases we discuss below.

## 5.2 Empty and unbounded confidence sets, discussion

The resulting confidence sets can take several forms: (a) a closed interval; (b) unbounded intervals; (c) the entire real line; (d) an empty set. Case (a) corresponds to a situation where  $\beta$  is well identified, while (b) and (c) correspond to unbounded confidence sets and indicate (partial or complete) non-identification. The possibility of getting an empty confidence set may appear surprising. But, on hindsight, this is quite natural: it may suggest that no value of  $\beta_0$  does allow  $\Pi = \gamma\beta'_0$  to be acceptable for any  $\gamma$ .

In this section, we discuss empty and unbounded sets more formally. Specifically, we relate such outcomes to inference on two commonly assessed hypotheses:

$$H_r : \text{rank}(\Pi) = r, \quad H_n : \Pi = 0.$$

The statistic

$$LR_{\min} = \min_{\beta_0} LR(\beta_0)$$

coincides (see also Gouriéroux, Monfort and Renault 1995) with the LR criterion associated within (3) against the assumption of full rank. Furthermore, the statistic

$$LR_n = -T\bar{L} \quad (40)$$

where

$$\bar{L} = \ln \left( \left| I_p - (R'_0 R_0)^{-1} R'_0 R_1 (R'_1 R_1)^{-1} R'_1 R_0 \right| \right) \quad (41)$$

provides the LR criterion associated with  $H_n$  which corresponds to the no-cointegration null hypothesis (see for example equation (3.13) in Gouriéroux, Monfort and Renault 1995). Denote the  $\alpha$ -level cut-off point of  $LR_{\min}$ , which is associated with  $H_r$ , as  $c_r$ .

The following necessary, although not sufficient) conditions relate empty and unbounded confidence sets to these commonly assessed hypotheses.

**Lemma 1** *If  $LR_{\min} > c_B$  where  $c_B$  is the bound cut-off point defined in (29), which implies that the confidence set based on inverting  $LR(\mathbf{b}_0)$  at the  $\alpha$ -level is empty, then  $H_r$  is rejected at this level using the traditional LR test.*

Note that Lemma 1 provides a necessary but not sufficient condition: the (standard) reduced rank test may be significant yet we cannot be sure that the confidence set is empty, unless the bound cut-off point is used to assess  $H_r$ . Furthermore, Lemma 1 holds if  $c_B$  is replaced by  $c_s$  because  $c_r \leq c_s \leq c_B$  since  $H_0^* \subseteq H_0 \subseteq H_r$ . By construction  $c_s \leq \hat{c} \leq c_B$ , so Lemma 1 also holds for the Type 2 cut-off defined above.

**Lemma 2** *If  $LR_n \leq c_B$  where  $c_B$  is the bound cut-off point defined in (29), which implies that the confidence set based on inverting  $LR(\mathbf{b}_0)$  at the  $\alpha$ -level is the real line, then  $H_n$  is not rejected at this level using the traditional LR test for no-cointegration.*

Note that again Lemma 2 provides a necessary but not sufficient condition. Thus we cannot rule out the case where some information may be still available in the data on  $\beta_0$  even when Johansen's test fails to reject the no-cointegration null. Here again, Lemma 2 is also verified if  $c_B$  is replaced by  $c_s$  because  $c_s \leq c_B$  which also implies that Lemma 2 is verified for the Type 2 cut-off  $\hat{c}$ . Conditions similar to both Lemma 1 and 2 can also be derived for the  $LRC(\beta_0)$  statistic.

To sum up, Lemmas 1 and 2 confirm that our proposed confidence sets provide relevant information on whether cointegration is supported by the data, a property not shared by standard confidence intervals. Our confidence sets may turn out to be empty, which occurs when all possible values of  $\beta$  are rejected suggesting that its definition over-estimates the rank of  $\Pi$  in Johansen's framework. If the latter is underestimated, and in particular because the unknown  $\gamma$  may be close to zero or rank-deficient, then our confidence sets will be unbounded. We next proceed to discussing the special cases  $r = 1$  and  $r = p - 1$ , for which test inversion admits a useful and tractable analytical solution.

## 6 Analytical solutions for special cases

When  $r$  is either 1 or  $p - 1$ , we can find an analytical solution for (31)-(33). In both cases, we rewrite the inequalities in the *quadratic* form

$$\pi' A_{22} \pi + 2A_{12} \pi + A_{11} \leq 0 \quad (42)$$

where  $\pi$  is the  $a \times 1$  vector of unknown parameters in  $\beta_0$  and  $A_{22}$ ,  $A_{12}$  and  $A_{11}$  depend on the data and the considered cut-off point. To do this, we proceed as follows.

For  $r = 1$ , we first cast the inequalities under consideration in the

$$(1, \pi') Q (1, \pi')' \leq 0 \quad (43)$$

and  $Q$  is an  $(a + 1) \times (a + 1)$  data dependent matrix. We next partition  $Q$  as follows (see also Bolduc, Khalaf and Yelou 2010)

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \quad (44)$$

where  $Q_{11}$  is a scalar,  $Q_{22}$  is  $a \times a$  and  $Q_{12} = Q_{21}'$  is  $1 \times a$  so that (43) may be re-expressed as (42) where

$$A_{22} = Q_{22}, \quad A_{12} = Q_{12}, \quad A_{11} = Q_{11}.$$

For  $r = p - 1$ , we reduce the inequality to the form

$$(-\pi', 1) J (-\pi', 1)' \leq 0 \quad (45)$$

then partition  $J$  as follows

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \quad (46)$$

where  $J_{22}$  is a scalar,  $J_{11}$  is  $a \times a$  and  $J_{21} = J'_{12}$  is  $1 \times a$  so that (45) may be re-expressed as (42) with

$$A_{22} = J_{11}, \quad A_{12} = -J_{21}, \quad A_{11} = J_{22}.$$

A general solution to inequations of the form (42) was introduced by Dufour and Taamouti (2005, 2007). We summarize this solution in the Appendix and describe the resulting projection confidence sets pertaining to each component of  $\pi$  and any linear transformation of latter of the form  $\omega'\pi$  where  $\omega$  is a non-zero  $a \times 1$  vector. These sets can take several forms depending on the eigenvalues of  $A_{22}$ : (a) a closed interval; (b) the union of two unbounded intervals; (c) the entire real line; (d) an empty set. We thus proceed to show how (31) - (33) can be rewritten in the proposed quadric forms.

## 6.1 Inverting the LR criteria: the $r = 1$ case

Decomposition (19) when  $r = 1$  gives

$$\begin{aligned} LR(\mathbf{b}_0) &= -T \ln \left( \left| I_p - (R'_0 R_0)^{-1} R'_0 R_1 (R'_1 R_1)^{-1} R'_1 R_0 \right| \right) \\ &\quad + T \ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right) \end{aligned}$$

which allows us to write inequality (31) as a quadratic inequations in  $\mathbf{b}_0$ , which is summarized in the following Theorem.

**Theorem 2** *In the context of (1)-(2) with  $r = 1$ ,  $\beta = (1, \mathbf{b}')'$  and the null hypothesis (8) inverting the statistic  $LR(\mathbf{b}_0)$  defined in (10) at the  $\alpha$ -level corresponds to the following inequality, in  $\mathbf{b}_0$ ,*

$$(1, \mathbf{b}'_0) \left[ R'_1 \left( \bar{d} I_T - R_0 (R'_0 R_0)^{-1} R'_0 \right) R_1 \right] (1, \mathbf{b}'_0)' < 0 \quad (47)$$

where

$$\bar{d} = 1 - \exp \left[ \bar{c}/T + \ln \left( \left| I_p - (R'_0 R_0)^{-1} R'_0 R_1 (R'_1 R_1)^{-1} R'_1 R_0 \right| \right) \right] \quad (48)$$

and  $\bar{c}$  is the statistic's  $\alpha$ -level critical point.

Inequation (47) thus coincides with (43) for  $\pi = \mathbf{b}_0$  and  $Q = R'_1 (\bar{d}I_T - R_0 (R'_0 R_0)^{-1} R'_0) R_1$ . The same reasoning holds with  $LRC(\mathbf{b}_0)$  since decomposition (20) with  $r = 1$  gives

$$\begin{aligned} LRC(\mathbf{b}_0) &= -T \ln \left( 1 - \frac{\tilde{\beta}' R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \tilde{\beta}}{\tilde{\beta}' R'_1 R_1 \tilde{\beta}} \right) \\ &\quad + T \ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right). \end{aligned}$$

We are thus back to a quadratic inequation, as shown in the following Theorem.

**Theorem 3** *In the context of (1)-(2) with  $r = 1$ ,  $\beta = (1, b')'$  and the null hypothesis (8) inverting the statistic  $LRC(\mathbf{b}_0)$  defined in (15) at the  $\alpha$ -level corresponds to the following inequality, in  $\mathbf{b}_0$ ,*

$$(1, \mathbf{b}'_0) \left[ R'_1 \left( \tilde{d}I_T - R_0 (R'_0 R_0)^{-1} R'_0 \right) R_1 \right] (1, \mathbf{b}'_0)' < 0 \quad (49)$$

where

$$\tilde{d} = 1 - \exp \left[ \tilde{c}/T + \ln \left( 1 - \frac{\tilde{\beta}' R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \tilde{\beta}}{\tilde{\beta}' R'_1 R_1 \tilde{\beta}} \right) \right] \quad (50)$$

and  $\tilde{c}$  is the statistic's  $\alpha$ -level critical point.

## 6.2 Inverting the implicit form statistic: the $r = p - 1$ case

When  $r = p - 1$ , (18) gives

$$LRP(\mathbf{b}_0) = T \ln \left( 1 + \frac{\beta'_{\perp 0} \hat{\Pi}' S^{-1} \hat{\Pi} \beta_{\perp 0}}{\beta'_{\perp 0} (R'_1 R_1)^{-1} \beta_{\perp 0}} \right), \quad \beta_{\perp 0} = (-\mathbf{b}'_0, 1)'$$

**Theorem 4** *In the context of (1)-(2) with  $r = p - 1$  and  $\beta_{\perp 0} = (-\mathbf{b}'_0, 1)'$  and the null hypothesis (16) inverting the statistic  $LRP(\mathbf{b}_0)$  defined in (17) at the  $\alpha$ -level corresponds to the following inequality, in  $\mathbf{b}_0$ ,*

$$(-\mathbf{b}'_0, 1) \left[ \hat{\Pi}' S^{-1} \hat{\Pi} - (R'_1 R_1)^{-1} (\exp(\bar{c}/T) - 1) \right] (-\mathbf{b}'_0, 1)' \leq 0. \quad (51)$$

where  $\bar{c}$  is the statistic's  $\alpha$ -level critical point.

Inequation (51) thus coincides with (45), for  $\pi = \mathbf{b}_0$  and  $J = \hat{\Pi}' S^{-1} \hat{\Pi} - (R'_1 R_1)^{-1} (\exp(\bar{c}/T) - 1)$ .

## 7 Simulation study

The above considered example from Dufour (1997) may serve as a base case since  $\mathbf{b}$  is a scalar and perhaps easier to interpret. One should check the size of the inverted test, and then obtain a confidence interval for  $\mathbf{b}$  using one or more available usual methods. Checking whether this set covers the hypothesized value provides the size of the Wald-type test associated with the confidence interval. It matters to assess size given various choices for  $\gamma$ , as  $\gamma$  approaches the non-identification boundary, here 0.

### 7.1 Monte Carlo design

A simulation design which fits the objectives of this paper must provide a basis for understanding the parametrization of  $\gamma$ . We thus consider the model by Gonzalo (1994) which allows  $\gamma$  to embed near unit roots and persistence as well as departures from weak exogeneity. The model in structural form is the following:

$$\begin{aligned} y_t - \mathbf{b}x_t &= z_t, \quad z_t = \rho z_{t-1} + e_{z_t} \\ a_1 y_t - a_2 x_t &= w_t, \quad w_t = w_{t-1} + e_{w_t} \end{aligned}$$

with

$$\begin{pmatrix} e_{z_t} \\ e_{w_t} \end{pmatrix} \equiv iid \text{ N } \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \zeta\sigma \\ \zeta\sigma & \sigma^2 \end{pmatrix} \right].$$

Gonzalo derives its reduced rank regression representation which yields

$$\gamma = - \left( (\rho - 1) \frac{a_2}{a_1 \mathbf{b} - a_2}, (\rho - 1) \frac{a_1}{a_1 \mathbf{b} - a_2} \right)'$$

We thus see that  $\gamma$  may approach zero if  $\rho$  approaches 1, while weak exogeneity can be imposed by setting  $a_1 = 0$ .

An alternative expression of this model is useful to shed further light on the  $a_1 \neq 0$  case. To write this model in a triangular form, substitute  $z_t + \mathbf{b}x_t$  for  $y_t$  in the second equation then solve for  $x_t$  which yields

$$\begin{aligned} a_1 (z_t + \mathbf{b}x_t) - a_2 x_t &= w_t \\ (a_1 \mathbf{b} - a_2) x_t &= w_t - a_1 z_t \\ (a_1 \mathbf{b} - a_2) [x_t - x_{t-1}] &= w_t - w_{t-1} - a_1 (z_t - z_{t-1}). \end{aligned}$$

and from there on to

$$y_t - \mathbf{b}x_t = z_t, \quad (52)$$

$$z_t = \rho z_{t-1} + e_{z_t} \Leftrightarrow z_t - z_{t-1} = (\rho - 1) z_{t-1} + e_{z_t}, \quad (53)$$

$$x_t - x_{t-1} = \delta (z_t - z_{t-1}) + e_{v_t} \quad (54)$$

$$e_{v_t} = \phi e_{w_t} \quad (55)$$

$$\phi = \frac{1}{a_1 \mathbf{b} - a_2}, \quad \delta = \frac{-a_1}{a_1 \mathbf{b} - a_2} \quad (56)$$

$$\begin{pmatrix} e_{z_t} \\ e_{v_t} \end{pmatrix} \equiv iid \text{ N } \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \zeta \phi \sigma \\ \zeta \phi \sigma & \sigma^2 \phi^2 \end{pmatrix} \right].$$

When  $a_1 \neq 0$ , then as argued by Gonzalo (1994), "the error correction term ( $z_{t-1}$ ) can be present in both equations of the ECM. In this case  $x_t$ , is no longer weakly exogenous. Our expression (54) illustrates that indeed, feedback caused by a non-zero coefficient on ( $z_t - z_{t-1}$ ), violates weak exogeneity in this context. This model is empirically relevant and its formulation that sets the feedback coefficient [here  $\delta$ ] forth helps disentangle two different although related sources of identification concerns: the unit root boundary, and departure from weak exogeneity.

Under the null hypothesis,  $\mathbf{b}_0 = 1$ . We set, as in Gonzalo (1994),  $a_2 = -1$ ,  $\zeta = -0.5$ ,  $\sigma = .25$ , for both size and power study. We compare two choices for  $a_1$ ,  $a_1 = 0$  and  $a_1 = 1$ , to assess deviations from weak-exogeneity, and two choices for  $\rho$ ,  $\rho = 0.8$  and  $\rho = 0.99$  to check weak-identification, with two sample sizes  $T = 100, 300$ .<sup>8</sup>

## 7.2 Monte Carlo p-values

The nuisance parameters associated with testing  $\mathbf{b} = \mathbf{b}_0$  can be narrowed down, in this context, to the following:  $\rho$  and  $\delta$ , both of which control identification and the variance/covariance matrix of  $e_{z_t}$  and  $e_{v_t}$ , so all in all, five free parameters.

If  $\mathbf{b} = \mathbf{b}_0$  with  $\mathbf{b}_0$  known, then  $\rho$  is estimable consistently by e.g. the OLS regression of  $(y_t - \mathbf{b}_0 x_t)$  on its first lag [using (53)]. Similarly,  $\delta$  is estimable consistently by e.g. the OLS regression of  $(x_t - x_{t-1})$  on  $(y_t - y_{t-1} - \mathbf{b}_0 (x_t - x_{t-1}))$  [using (54)]. These two regressions also yield consistent estimates of the variance/covariance matrix of  $e_{z_t}$  and

---

<sup>8</sup>Results for  $\rho = .96$  [available upon request] are qualitatively similar to the weak-identification case.



$e_{v_t}$ . So whether  $\delta$  is imposed to be zero or not [that is whether  $a_1$  is imposed to zero or not], the variance/covariance parameters can be partialled-out so that a maximization over  $\delta$  and  $\rho$  may be sufficient in practice to control the size of the test. Simulation results available so far support this suggestion. For presentation clarity, we denote the MC test implemented in this way as the RPLMC test, which stands for *Restricted Partialled-out local MC* test. *Restricted* formally means that  $\mathbf{b}$  is fixed to  $\mathbf{b}_0$ . *Partialled-out* implies that nuisance parameters are narrowed down to  $\delta$  and  $\rho$ , while the remaining parameters that define the DGP are inferred fixing  $\delta$  and  $\rho$ . *Local* implies that an estimate of nuisance parameters [here  $\delta$  and  $\rho$ ] that is consistent under the null hypothesis [here  $\mathbf{b} = \mathbf{b}_0$ ] is used; refer to (39) in the above defined algorithm.

### 7.3 Results

We analyze size and power of  $LR(\beta_0)$  and  $LRC(\beta_0)$ , relative to available procedures. These include DOLS, FMOLS, Wright's (2000) test and the Bartlett corrected  $LRC(\beta_0)$  from Johansen (2002), derived with the true parameters to illustrate the best case scenario even if infeasible. 1000 replications are applied in all trials and the MC methods are implemented with 99 simulated samples. We report the results in Tables 1-3. The results for DOLS, FMOLS and the Wright (2000) test use the Newey-West HAC with the Andrews (1991) automatic bandwidth. Rejection probabilities reported under the alternative are not size corrected, yet we study power for the procedures with empirical size not exceeding 7%. The nominal size is 5% for all procedures. Results can be summarized as follows.

In the considered bivariate system, although high persistence causes size distortions for the considered LR statistics, these are easily corrected via our proposed simulation method, imposing and relaxing weak exogeneity. The Bartlett correction does not work when it is mostly needed. This is not surprising in view of the discussion in Johansen (2000, 2002), yet is worth documenting. Recall that we have implemented the infeasible correction here, with known parameter values. In contrast, the bound and Type 2 corrections work well even with  $T = 100$  and  $a_1 \neq 0$ . One useful result further emerges from our experiments: in this design, the RPLMC method suffices to control the size of  $LR(\beta_0)$ . There seems to be no need for the maximized MC procedure in this design, so results reported below

under the MC heading use the RPLMC method as described above.

**[Insert somewhere here Tables 1 and 2]**

The size of DOLS and FMOLS based t-tests exceeds 90% at the boundary. Furthermore, failure of weak-exogeneity causes very severe distortions for DOLS (size  $\simeq 88\%$  even with  $T = 300$ ) as well as for FMOLS (size remains around 37% with  $T = 300$ ), even when  $\beta$  is identified. The test from Wright (2000) is also oversized at the boundary, with distortions worsening as  $T$  increases. This underscores the fact that identification problems are not just small sample concerns. All tests behave much worse than the uncorrected LR-based statistics. These results are noteworthy, particularly because the LR framework is not popular when weak exogeneity is in doubt.

Power of our LR tests is good even when relying on the LF bound. Recall that the MC method treats the dependence structure as known: nuisance parameters are estimated but the AR(1) structure as in the true model is imposed. Similarly, exogeneity, in the form of  $\delta = 0$  is also imposed when  $a_1 = 0$ . As implemented, this method may be viewed as a often unattainable first best bootstrap. The fact that the Type 2 correction meets and in some cases beats this first best is noteworthy. Indeed, because the Type 2 critical value does not vary with the tested value of  $\beta$  and thus preserves the quadrics form of the test inversion solution, this correction emerges as a very promising and very useful practical solution. Power results for  $\rho = .99$  reflect the extent of weak identification given available data and the considered alternative space. Despite the serious power problems we still find with  $T = 300$  in our design, results are not meant to suggest the test has no power. Since  $\beta$  remains identified, information on  $b$  would eventually mount up and power would pick up, possibly mildly, as  $T$  grows and  $b$  departs further from the null.

**[Insert somewhere here Table 3]**

## 8 Conclusion

This paper was concerned with identification problems in the context of cointegrating regression. We proposed confidence sets for long-run parameters that do not require identification by inverting simulation or bound-based LR tests.

In contrast to standard Wald-type intervals, our proposed confidence sets provide, in addition to correct coverage, built-in specification checks. In particular, unbounded confidence sets may occur when, in the underlying reduced rank regression, the rank is over-estimated. Unbounded sets result from underestimating the rank in question, which signals - among other issues - slow adjustment to the long-run equilibrium.

We showed, via a Monte Carlo study, that even within a small-scale bivariate system, commonly used procedures for inference on long-run coefficients can be severely oversized at the model boundary, that is when the parameter that controls cointegration approaches the unit-root boundary. Of the four methods we compared (DOLS, FMOLS, the method of Wright (2000) and LR) to address the nuisance parameter dependency problem arising from weak identification, only the LR achieved size control via the simulation or bound-based corrections we introduced, even when weak exogeneity fails. Our results suggest that further research should proceed in this direction, in line with the general weak-identification literature, with emphasis on larger dimensions or on models with potential breaks.

## Acknowledgments

We wish to thank participants in the New York Camp Econometrics VII (The Otesaga Resort Hotel, Cooperstown, NY, 13-15 April 2012) in particular James MacKinnon, in the 12th Oxmetrics User Conference (London UK, 3-4 September 2012) in particular Bent Nielsen, in the 29th Canadian Econometric Study Group Annual Meeting (Kingston, Ontario, Canada, 27-28 October 2012) in particular Bertille Antoine, and in the 23rd EC2 Conference (Maastricht, The Netherlands, 14-15 December 2012), for useful comments and suggestions. We are greatly indebted to Soren Johansen, who read previous versions of this paper and provided very insightful comments. We wish to thank the Editor, Yacine Ait-Sahalia, and two anonymous referees for providing us with very insightful comments and suggestions that greatly helped to improve the paper. The usual disclaimer applied. We are grateful to Michele Bergamelli for excellent research assistance. This work was supported by the Social Sciences and Humanities Research Council of Canada, the Fonds FQRSC (Government of Québec), and the Centre for Econometric Analysis of Cass Business School.

# Appendix

## A.1 Eigenvalue Based Inequations

From Dufour and Taamouti (2005, 2007) so projections based confidence sets (CSs) based on (42), for any linear transformation of  $\pi$  of the form  $\omega'\pi$  can be obtained as follows. Let  $\tilde{A} = -A_{22}^{-1}A'_{12}$ ,  $\tilde{D} = A_{12}A_{22}^{-1}A_{12} - A_{11}$ . If all the eigenvalues of  $A_{22}$  [as defined in (44)] are positive so  $A_{22}$  is positive definite then:

$$\text{CS}_\alpha(\omega'\pi) = \left[ \omega'\tilde{A} - \sqrt{\tilde{D}(\omega'A_{22}^{-1}\omega)}, \omega'\tilde{A} + \sqrt{\tilde{D}(\omega'A_{22}^{-1}\omega)} \right], \quad \text{if } \tilde{D} \geq 0 \quad (57)$$

$$\text{CS}_\alpha(\omega'\pi) = \emptyset, \quad \text{if } \tilde{D} < 0. \quad (58)$$

If  $A_{22}$  is non-singular and has one negative eigenvalue then: (i) if  $\omega'A_{22}^{-1}\omega < 0$  and  $\tilde{D} < 0$ :

$$\text{CS}_\alpha(\omega'\pi) = \left] -\infty, \omega'\tilde{A} - \sqrt{\tilde{D}(\omega'A_{22}^{-1}\omega)} \right] \cup \left[ \omega'\tilde{A} + \sqrt{\tilde{D}(\omega'A_{22}^{-1}\omega)}, +\infty \right[; \quad (59)$$

(ii) if  $\omega'A_{22}^{-1}\omega > 0$  or if  $\omega'A_{22}^{-1}\omega \leq 0$  and  $\tilde{D} \geq 0$  then:

$$\text{CS}_\alpha(\omega'\pi) = \mathbb{R}; \quad (60)$$

(iii) if  $\omega'A_{22}^{-1}\omega = 0$  and  $\tilde{D} < 0$  then:

$$\text{CS}_\alpha(\omega'\pi) = \mathbb{R} \setminus \left\{ \omega'\tilde{A} \right\}. \quad (61)$$

The projection is given by (60) if  $A_{22}$  is non-singular and has at least two negative eigenvalues.

## A.2 Proof of Theorems and Lemmas

### Proof of Theorem 1

In the context of (3), consider testing the null hypothesis

$$H_0^* : \Pi = \gamma_0\beta'_0$$

where  $\beta_0$  conforms with  $H_0$  in (8) and  $\gamma_0$  is known and has full rank. Let  $LR_*$  denote the LR statistic associated with testing  $H_0^*$  against an unrestricted alternative. By construction,

$H_0^* \subseteq H_0$ , or in other words,  $H_0^*$  is more restricted than  $H_0$  and consequently

$$LR(\beta_0) \leq LR_*.$$

It follows that  $P[LR(\beta_0) \geq x] \leq P[LR_* \geq x]$ ,  $\forall x$ , with yields (28). ■

### Proof of Lemma 1

First recall that  $H_0^* \subseteq H_0 \subseteq H_r$ . To see this, recall that  $H_r$  corresponds to  $\Pi = \gamma\beta'$  for some  $\gamma$  and  $\beta$  both of rank  $r$ .  $H_0$  is more restricted than  $H_r$  since it sets  $\beta$  to the  $\beta_0$  value which is known. In turn,  $H_0^*$  sets both  $\gamma$  and  $\beta$  to known values. It follows that  $c_r \leq c_s \leq c_B$ . So

$$LR_{\min} > c_B \Rightarrow LR_{\min} > c_r$$

which proves the lemma. ■

### Proof of Lemma 2

The statistic  $LR(\mathbf{b}_0)$  may be expressed as

$$LR(\mathbf{b}_0) = LR_n + T \ln \left( \left| I_r - \beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \right| \right),$$

so the inequality under consideration is

$$T \ln \left( \left| I_r - \beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \right| \right) < (c_B - LR_n). \quad (62)$$

Observe that

$$T \ln \left( \left| I_r - \beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \right| \right) < 0$$

since it is equal to

$$\begin{aligned} & T \ln \left[ \frac{\left| \frac{1}{T} R'_0 (I_T - R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \beta'_0 R'_1) R_0 \right|}{\left| \frac{1}{T} R'_0 R_0 \right|} \right] \\ &= -T \ln \left[ \frac{\left| \frac{1}{T} R'_0 R_0 \right|}{\left| \frac{1}{T} R'_0 (I_T - R_1 \beta_0 (\beta'_0 R'_1 R_1 \beta_0)^{-1} \beta'_0 R'_1) R_0 \right|} \right] \end{aligned}$$

which corresponds to the likelihood ratio associated with testing  $\Pi = 0$  against  $\Pi = \delta\beta'_0$  (the former is more restrictive than the latter). When  $LR_n < c_B$  then  $c_B - LR_n > 0$  so

inequality (62) will be satisfied for any  $\beta_0$  since its right-hand-side which is negative for any  $\beta_0$  will always be less than a positive  $(c_B - LR_n)$ . It follows that confidence set will be the real line. Lemma 2 thus follows on recalling that the cut-off points associated with no-cointegration null [here  $\Pi = 0$ ] are larger than their their  $\chi^2$  counterpart. ■

### Proof of Theorem 2

The problem consists in solving the inequality

$$-T\bar{L} + T \ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right) < \bar{c}$$

or alternatively

$$\ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right) < [\bar{c}/T + \bar{L}]. \quad (63)$$

Taking exponential on both sides leads to

$$-\frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} < -\bar{d}$$

with as  $\bar{d}$  as in (48). Assuming  $\beta = (1, \mathbf{b}')'$  leads to

$$-(1, \mathbf{b}'_0) R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 (1, \mathbf{b}'_0)' + \bar{d} (1, \mathbf{b}'_0) R'_1 R_1 (1, \mathbf{b}'_0)' < 0$$

which gives (47). ■

### Proof of Theorem 3

The problem consists in solving the inequality

$$-T\bar{L}C + T \ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right) < \tilde{c}$$

where

$$\bar{L}C = \ln \left( 1 - \frac{\tilde{\beta}' R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \tilde{\beta}}{\tilde{\beta}' R'_1 R_1 \tilde{\beta}} \right)$$

or alternatively

$$\ln \left( 1 - \frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} \right) < [\tilde{c}/T + \bar{L}C].$$

Taking exponential on both sides leads to

$$-\frac{\beta'_0 R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 \beta_0}{\beta'_0 R'_1 R_1 \beta_0} < -\tilde{d}$$

with  $\tilde{d}$  as in (50). Assuming  $\beta = (1, \mathbf{b}')'$  leads to

$$-(1, \mathbf{b}'_0) R'_1 R_0 (R'_0 R_0)^{-1} R'_0 R_1 (1, \mathbf{b}'_0)' + \tilde{d} (1, \mathbf{b}'_0) R'_1 R_1 (1, \mathbf{b}'_0)' < 0$$

which gives (49). ■

### Proof of Theorem 4

The problem consists in solving the inequality

$$\ln \left( 1 + \frac{\beta'_{\perp 0} \hat{\Pi}' S^{-1} \hat{\Pi} \beta_{\perp 0}}{\beta'_{\perp 0} (R'_1 R_1)^{-1} \beta_{\perp 0}} \right) \leq \frac{\bar{c}}{T}.$$

Taking exponential on both sides leads to

$$\frac{\beta'_{\perp 0} \hat{\Pi}' S^{-1} \hat{\Pi} \beta_{\perp 0}}{\beta'_{\perp 0} (R'_1 R_1)^{-1} \beta_{\perp 0}} < \left( \exp \left( \frac{\bar{c}}{T} \right) - 1 \right)$$

which gives (51). ■

Table 1. Size of non-LR based tests for $b = b_0$					
Test	$\rho$	$T = 100$		$T = 300$	
		$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$
Wright (2002)	.80	.222	.224	.154	.154
	.99	.722	.788	.843	.864
DOLS	.80	.282	.878	.194	.883
	.99	.748	.992	.764	.995
FMOLS	.80	.160	.689	.096	.375
	.99	.512	.962	.419	.925

Notes: Numbers reported are empirical rejections. The underlying model is described by (52)-(54), with  $b = b_0 = 1$ . The  $a_1 \neq 0$  case corresponds to the design for which weak exogeneity fails;  $\rho = .99$  suggests that  $b$  is weakly identified.



		$LR(\beta_0)$				$LRC(\beta_0)$			
Test	$\rho$	$T = 100$		$T = 300$		$T = 100$		$T = 300$	
		$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$
$\chi^2$ standard	.80	.073	.073	.063	.063	.078	.078	.061	.061
	.99	.190	.184	.138	.139	.302	.302	.188	.188
$\chi^2$ , Bartlett	.80	-	-	-	-	.057	.057	.050	.050
	.99	-	-	-	-	.150	.150	.087	.087
MC:	.80	.052	.055	.048	.047	.072	.062	.060	.056
	.99	.056	.060	.064	.060	.076	.068	.051	.050
$\chi^2$ Bound	.80	.017	.017	.010	.010	.002	.002	.002	.002
	.99	.050	.052	.030	.030	.043	.043	.018	.018
Type 2	.80	.068	.068	.071	.071	.062	.062	.061	.061
	.99	.061	.061	.045	.045	.055	.055	.032	.032

Notes: The underlying model is described by (52)-(54), with  $b = b_0$ . Numbers reported are empirical rejections under the null hypothesis. The  $a_1 \neq 0$  case corresponds to the design for which weak exogeneity fails;  $\rho = .99$  suggests that  $b$  is weakly identified. MC refers to the above defined parametric bootstrap-type Restricted Partialled-out local Monte Carlo test. Restricted formally means that  $b$  is fixed to  $b_0$ . Partialled-out implies that nuisance parameters are narrowed down to  $\delta$  and  $\rho$ , while the remaining parameters that define the DGP are inferred given  $\delta$  and  $\rho$ . Local implies that an estimate of  $\delta$  and  $\rho$  that is consistent under the null hypothesis [here  $b = b_0$ ] is used.  $\delta$  is the feedback coefficient which allows to impose and relax exogeneity in the bootstrap samples. Reported results impose  $\delta = 0$  for the  $a_1 = 0$  case. The Bartlett correction from Johansen (2002) is derived with the true parameter values.

<b>Table 3.</b> Power of tests based on $LR(\beta_0)$									
$\rho$	$b$	MC		Bound		$LR(\beta_0)$ Type 2		$LRC(\beta_0)$ Type2	
		$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$	$a_1 = 0$	$a_1 \neq 0$
$T = 100$									
.8	1.10	.092	.070	.023	.018	.081	.070	.080	.067
	1.20	.128	.082	.045	.023	.119	.081	.130	.080
	1.50	.346	.144	.229	.059	.357	.150	.375	.169
	1.60	.414	.184	.304	.085	.436	.185	.440	.214
	2.00	.666	.338	.550	.230	.675	.357	.648	.375
.99	1.10	.063	.060	.051	.052	.059	.061	.053	.053
	1.20	.064	.058	.049	.052	.059	.059	.049	.053
	1.50	.068	.066	.050	.049	.057	.058	.049	.049
	1.60	.072	.066	.050	.049	.059	.057	.047	.049
	2.00	.072	.062	.049	.051	.056	.057	.046	.049
$T = 300$									
.8	1.10	.152	.058	.067	.018	.190	.091	.255	.090
	1.20	.412	.148	.293	.067	.497	.190	.570	.255
	1.50	.920	.542	.818	.427	.932	.629	.920	.679
	1.60	.966	.658	.906	.530	.967	.719	.954	.751
	2.00	1.00	.914	.996	.818	1.00	.932	.998	.920
.99	1.10	.062	.056	.029	.030	.043	.044	.030	.030
	1.20	.060	.054	.030	.029	.045	.043	.031	.030
	1.50	.054	.050	.029	.028	.041	.043	.031	.030
	1.60	.058	.048	.031	.028	.044	.043	.031	.030
	2.00	.066	.046	.038	.029	.053	.041	.040	.031

Notes: Numbers reported are empirical rejections. For the definition of the model and test methods, see notes to Table 2. Under the null hypothesis  $b = b_0 = 1$ .

## References

- Andrews D. W. K. , 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.
- Andrews, D. W. K., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399-405.
- Andrews, D. W. K., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683-734.
- Andrews, D. W. K., Cheng, X., 2013. Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics* 173, 36-56.
- Andrews, D. W. K., Moreira, M. J., Stock, J. H., 2006. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715-752.
- Antoine B., Lavergne, P., 2012. Conditional moment models under weak identification. Working paper, Simon Fraser University.

Beaulieu M.-C., Dufour J.-M., Khalaf, L., 2013a. Testing portfolio efficiency with an unobservable zero-beta rate and non-Gaussian distributions: an exact identification-robust approach. *Review of Economic Studies* 80, 892-924.

Beaulieu M.-C., Dufour J.-M., Khalaf, L., 2013b. Inference in structural multivariate factor models with rank restrictions: factor proliferation and weak identification. Working Paper, McGill University, Canada.

Bolduc, D., Khalaf, L., Yelou, C., 2010. Identification robust confidence sets methods for inference on parameter ratios with application to discrete choice models. *Journal of Econometrics* 157, 317-327.

Boswijk, H. P., 1995. Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* 69,133-158.

Caner, M., Kilian, L., 2001. Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. *Journal of International Money and Finance* 20, 639-657.

Chang, Y., Park, J. Y., Song, K., 2006. Bootstrapping cointegrating regressions. *Journal of Econometrics* 133, 703-739.

Dufour, J.-M., 1989. Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica* 57, 335-355.

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

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767-808.

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., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review* 42, 815-843.

Dufour, J.-M., Khalaf, L., 2002. Simulation-based finite and large sample tests in multivariate regressions. *Journal of Econometrics* 111, 303-332.

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

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

Elliott G., 1998. On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* 66, 149-158.

Gonzalo, J., 1994. Five alternative methods of estimating long run equilibrium relationships. *Journal of Econometrics* 60, 203-233.

Gouriéroux, C., Monfort, A., Renault, E., 1995. Inference in factor models, in G. S. Maddala, P. C. B. Phillips and T. N. Srinivasan, eds, *Advances in Econometrics and Quantitative Economics*, Blackwell, Oxford, U.K., Chapter 13, 311-353.

Guggenberger P., Kleibergen F., Mavroeidis S., Chen, L., 2012. On the asymptotic sizes of subset Anderson-Rubin and Lagrange Multiplier tests in linear instrumental variables regression. *Econometrica* 80, 2649-2666.

- Guggenberger P., Smith, R., 2008. Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics* 142, 134-161.
- Harville, D. A., 1997. *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York.
- Hjalmarsson, E., Österholm, P., 2010. Testing for cointegration using the Johansen methodology when variables are near-integrated: size distortions and partial remedies. *Empirical Economics* 39, 51-76.
- Ioannidis, E. E., Chronis, G. A., 2005. Extreme spectra of VAR models and orders of near-cointegration. *Journal of Time Series Analysis* 26, 399-421.
- Jansson, M., Haldrup, N., 2002. Regression theory for nearly cointegrated time series. *Econometric Theory*, 18, 1309-1335.
- Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254.
- Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Model*. 2nd Edition, Oxford University Press, New York.
- Johansen, S., 2000. A Bartlett correction factor for tests on the cointegration relations. *Econometric Theory* 16, 740-778.
- Johansen, S., 2002. A small sample correction of the test for cointegration rank in the vector autoregressive model. *Econometrica* 70, 1929-1961.
- Johansen, S., 2009. Cointegration: overview and development, in (eds) Andersen, T.G., Davis, R.A., Kreiß, J.P., Mikosch, T. (2009), *Handbook of Financial Time Series*, Springer-Verlag, 671-693.
- Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781-1803.
- Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103-1123.
- Lanne, M., 2000. Near unit roots, cointegration, and the term structure of interest rates. *Journal of Applied Econometrics* 15, 513-529.
- Li, Y., 1994. Bootstrapping cointegrated regression. *Economics Letters* 44, 229-233.
- Li, H., Maddala, G. S., 1997. Bootstrapping cointegrated regressions. *Journal of Econometrics* 80, 297-318.
- Moreira, M. J., 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027-1048.
- Muller, U. K., Watson, M. W., 2013. Low-frequency robust cointegration testing. *Journal of Econometrics* 174, 66-81.
- Palm, F. C., Smeeks, S., Urbain, J.-P., 2010. A sieve bootstrap test for cointegration in a conditional error correction model. *Econometric Theory* 26, 647-681
- Park, J. Y., 2006. A bootstrap theory for weakly integrated processes. *Journal of Econometrics* 133, 639-672.
- Phillips, P. C. B., 1991. Optimal inference in cointegrating system. *Econometrica* 59, 283-306.
- Phillips, P. C. B., 1994. Some exact distribution theory for maximum likelihood estimators of cointegration coefficients in error correction models. *Econometrica* 62, 73-94.
- Phillips, P. C. B., 1995. Fully modified least-squares and vector autoregression. *Econometrica* 63, 1023-1078.

- Phillips, P. C. B., Hansen, B. E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99-125.
- Psaradakis, Z., 2001. On bootstrap inference in cointegrating regressions. *Economics Letters* 72, 1-10.
- Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.
- Stock, J. H., 1997. Cointegration, long-run comovements, and long-horizon forecasting. in (eds) Kreps D.M. and K. Wallis, *Advances in Economics and Econometrics: Theory and Applications*, Volume 3, Cambridge: Cambridge University Press, Chapter 2, 34-70
- Stock, J. H., Watson, M. W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783-820.
- Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20, 518-529.
- Tanaka, K., 1993. An alternative approach to the asymptotic theory of spurious regression, cointegration, and near cointegration. *Econometric Theory* 9, 36-61.
- Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variable regression with weak instruments. *Econometrica* 66, 1389-1404.
- Wright, J. H., 2000. Confidence set for cointegrating coefficients based on stationarity tests. *Journal of Business and Economic Statistics* 18, 211-222.
- Zivot, E., 2000. The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* 16, 407-439.
- Zivot, E., Startz, R., Nelson, C. R., 1998. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 39, 1119-1144.