Generalized $C(\alpha)$ tests for estimating functions with serial dependence*

Jean-Marie Dufour † McGill University

Alain Trognon ‡ CREST-ENSAE, Paris

Purevdorj Tuvaandorj § CREST-ENSAI, Rennes

First version: February 2013
Revised: April 2015, August 2015
This version: May 2016
Compiled: January 1, 2017, 9:35

This paper has been published in: *Advances in Time Series Methods and Applications: The A. Ian McLeod Festschrift*, edited by Wai Keung Li, David Stanford and Hao Yu, Springer, Berlin and New York, 2016, 151-178.

* The authors thank Marine Carrasco, Jean-Pierre Cotton, Russell Davidson, Abdeljelil Farhat, V. P. Godambe, Christian Genest, Christian Gouriéroux, Stéphane Grégoir, Tyaniu He, Frank Kleibergen, Sophocles Mavroeidis, Hervé Mignon, Julien Neves, Denis Pelletier, Mohamed Taamouti, Masaya Takano, and Pascale Valéry for several useful comments. Earlier versions of this paper were presented at the Canadian Statistical Society 1997 annual meeting and at INSEE (CREST, Paris). This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

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

‡ Member, CREST-ENSAE (Centre de recherche en économie et statistique) and Associate Professor, University Paris 1. Mailing address: CREST PARIS, Timbre J310, 15 Boulevard Gabriel Péri, 92254 MALAKOFF CEDEX, France. TEL.: 33 1 41175131; FAX: 33 1 41176480. e-mail: trognon@ensae.fr . Web page: www.crest.fr/component/

§ CREST-ENSAI - Campus de Ker-Lann, Rue Blaise Pascal - BP 37203, 35712 BRUZ cedex, France. TEL.: 33 (0)2 99 05 32 63; FAX: 33 (0)2 99 05 32 05. e-mail: purevdorj.tuvaandorj@ensai.fr .

ABSTRACT

We propose generalized $C(\alpha)$ tests for testing linear and nonlinear parameter restrictions in models specified by estimating functions. The proposed procedures allow for general forms of serial dependence and heteroskedasticity, and can be implemented using any root-n consistent restricted estimator. The asymptotic distribution of the proposed statistic is established under weak regularity conditions. We show that earlier $C(\alpha)$-type statistics are included as special cases. The problem of testing hypotheses fixing a subvector of the complete parameter vector is discussed in detail as another special case. We also show that such tests provide a simple general solution to the problem of accounting for estimated parameters in the context of two-step procedures where a subvector of model parameters is estimated in a first step and then treated as fixed.

Key words: Testing; $C(\alpha)$ test; Estimating function; Generalized method of moment (GMM); Serial dependence; Pseudo-likelihood; M-estimator; Nonlinear model; Score test; Lagrange multiplier test; Heteroskedasticity.

Contents

List of propositions and theorems

1. Introduction

The $C(\alpha)$ statistic introduced by Neyman (1959) embodies a general mechanism for dealing with nuisance parameters in tests of composite hypotheses. The basic idea of the method can be conveniently explained by using parameter subvector testing as an example. One first considers a score-type function for the tested parameter. The score function is then orthogonalized with respect to directions associated with the nuisance parameters under the null hypothesis. This removes the impact of the estimation error on the nuisance parameter: the residual vector from the projection – the *effective score function* – evaluated at the auxiliary estimator of the nuisance parameter is asymptotically equivalent to the effective score function evaluated at the true parameter. It is easy to see that the latter is asymptotically normally distributed, and consequently its normalized form – the $C(\alpha)$ statistic – has an asymptotic chi-square distribution under the null hypothesis.

The $C(\alpha)$ test enjoys a local optimality property while being computationally attractive (a few artificial regressions would be enough in many circumstances) and uses only $\sqrt{n}$-consistent estimator for the nuisance parameters which may not be asymptotically normal or even may not have an asymptotic distribution. When the restricted maximum likelihood (ML) estimator is used, the statistic reduces to Rao's score statistic. It is also useful to stress that the objects projected on the space spanned by the nuisance parameter scores can be more general functions [called Cramér functions by Neyman (1959)], not necessarily the score function associated with the parameters of interest. For further discussions of $C(\alpha)$ tests and references, see Le Cam (1956), Bhat and Nagnur (1965), Bühler and Puri (1966), Bartoo and Puri (1967), Moran (1970, 1973), Chibisov (1973), Chant (1974), Ray (1974), Singh and Zhurbenko (1975), Foutz (1976), Vorob'ev and Zhurbenko (1979), Bernshtein (1976, 1978, 1980*a*, 1980*b*, 1981), Le Cam and Traxler (1978), Neyman (1979), Tarone (1979, 1985), Tarone and Gart (1980), Wang (1981, 1982), Basawa (1985), Ronchetti (1987), Smith (1987*a*, 1987*b*), Berger and Wallenstein (1989), Hall and Mathiason (1990), Paul and Barnwal (1990), Wooldridge (1990), Dagenais and Dufour (1991), Davidson and MacKinnon (1991, 1993), Kocherlakota and Kocherlakota (1991), Dufour and Dagenais (1992), Bera and Yoon (1993), Jaggia and Trivedi (1994), Rao (1996), Bera and Bilias (2001), Pal (2003), Dufour and Valéry (2009), Chaudhuri and Zivot (2011), Bontemps and Meddahi (2012).

In spite of numerous generalizations and modifications in parametric models, extensions of the $C(\alpha)$ test to other types of estimation criteria, *e.g.* estimating equations [Durbin (1960), Godambe (1960, 1991), Small and McLeish (1994), Basawa, Godambe and Taylor (1997), Heyde (1997)], minimum distance, or the generalized method of moments [GMM, Hansen (1982), Hall (2004)], appear to be scarce. In particular, work on such tests has focused on linear hypotheses (especially, hypothesis setting the value of a parameter subvector) and/or independent observations; see Basawa (1985).

In this paper, we propose and study a general $C(\alpha)$-type statistic in estimating-function and GMM setups, with weakly specified temporal dependence and heteroskedasticity. The proposed generalized statistic is quite comprehensive and includes earlier $C(\alpha)$-type statistics as special cases, as well as a wide spectrum of new ones. The null hypothesis takes the form of a general constraint (linear or nonlinear) on model parameters. This extends the $C(\alpha)$ test proposed by Smith (1987*a*) for nonlinear restrictions in parametric likelihood models. The asymptotic distribution of the test statistic is derived under a set of weak regularity conditions, allowing for general forms of serial dependence and heteroskedasticity.

A number of important special cases of the extended test statistic are discussed in detail. These include testing whether a parameter subvector has a given value – for which we give a number of alternative forms and special cases – and accounting for parameter uncertainty in two-stage procedures. The latter problem has considerable practical importance. Due to the fact that nonlinear estimating functions are often difficult to estimate, it is convenient to estimate some parameters by an alternative simpler method, and then use these estimates as if they were known. Such procedures can however modify the distributions of test statistics and induce distortions in test levels; see Gong and Samaniego (1981), Pagan (1984, 1986), Murphy and Topel (1985), and Newey and McFadden (1994). So it is important to make corrections for such effects. We underscore that generalized $C(\alpha)$ tests can provide relatively simple solutions to such difficulties in the context of estimating functions and GMM estimation, again in presence of general forms of serial dependence and heteroskedasticity. We first discuss tests based on a general first-stage estimator, as well as tests based on a two-stage GMM estimation.

The paper is organized as follows. Section 2 lays out the general framework considered in the paper and introduces the $C(\alpha)$ statistic. The regularity conditions are stated and the asymptotic properties of the generalized $C(\alpha)$ statistic are studied in Section 3. We discuss the forms that the $C(\alpha)$ statistic takes in some special cases in Section 4. Section 5 considers the problem of testing the value of parameter subvector. We formulate the $C(\alpha)$ statistic for models estimated by two-step procedures in Section 6. We briefly conclude in Section 7.

2. Generalized $C(\alpha)$ statistic

We consider an $m \times 1$ vector estimating (or score-type) function $D_n(\theta; Z_n)$ which depends on an $n \times k$ data matrix $Z_n = [z_1, z_2, \ldots, z_n]'$ and a parameter vector $\theta \in \Theta \subseteq \mathbb{R}^p$ such that

$$D_n(\theta; Z_n) \xrightarrow[n\to\infty]{\mathsf{p}} D_\infty(\theta; \theta_0) \tag{2.1}$$

where $D_n(\theta; Z_n)$ is typically the sample mean of an estimating function, such as $D_n(\theta; Z_n) = \frac{1}{n}\sum_{t=1}^{n} h(\theta; z_t)$, $D_\infty(\cdot; \theta_0)$ is a mapping from Θ to $\mathbb{R}^m$, and θ_0 denotes the "true" parameter vector. The parameter θ is estimated by minimizing a criterion function of the form

$$M_n(\theta, W_n) = D_n(\theta; Z_n)' W_n D_n(\theta; Z_n) \tag{2.2}$$

where W_n is a symmetric positive definite matrix. This setup comprises as special cases the method of estimating functions [Durbin (1960), Godambe (1960, 1991), Small and McLeish (1994), Basawa et al. (1997), Heyde (1997)], the generalized method of moments [Hansen (1982), Hall (2004)], maximum likelihood, pseudo-maximum likelihood, M-estimation and instrumental-variable methods.

A common assumption in such contexts consists in assuming that

$$\mathsf{E}_{\theta_0}[D_n(\theta_0; Z_n)] = 0 \tag{2.3}$$

where $\mathsf{E}_{\theta}[\cdot]$ represents the expected value under any data distribution such that θ can be interpreted as the true parameter vector, along with a number of additional regularity assumptions which allow the application of central limit theorems and laws of large numbers, such as:

$$\sqrt{n} D_n(\theta_0; Z_n) \xrightarrow[n\to\infty]{L} \mathrm{N}[0, I(\theta_0)] , \tag{2.4}$$

$$J_n(\theta_0; Z_n) = \frac{\partial D_n(\theta_0; Z_n)}{\partial \theta'} \xrightarrow[n\to\infty]{\mathsf{p}} J(\theta_0) , \tag{2.5}$$

where $I(\theta_0)$ and $J(\theta_0)$ are $m \times m$ and $m \times p$ full-column rank matrices. In Section 3, we relax the assumptions (2.3) and (2.5).

The hypothesis we wish to test has the form

$$H_0 : \psi(\theta) = 0 \tag{2.6}$$

where $\psi(\theta)$ is a $p_1 \times 1$ continuously differentiable function of θ with $1 \leq p_1 \leq p$, and the $p_1 \times p$ matrix

$$P(\theta) = \frac{\partial \psi}{\partial \theta'} \tag{2.7}$$

has full row-rank p_1 (at least in an open neighborhood of θ_0).

Let $\hat{\theta}_n$ be the unrestricted estimator of θ obtained by minimizing $M_n(\theta, W_n)$, $\hat{\theta}_n^0$ the corresponding constrained estimator under H_0, and $\tilde{\theta}_n^0$ any other restricted estimator of θ under H_0. Let us also denote estimators of $I(\theta)$ and $J(\theta)$ by $\hat{I}_n(\theta)$ and $\hat{J}_n(\theta)$ respectively, where θ may be replaced by unrestricted and restricted estimators of θ to obtain estimators of $I(\theta_0)$ and $J(\theta_0)$. If

$$D_n(\theta; Z_n) = \frac{1}{n} \sum_{t=1}^{n} h(\theta; z_t) , \tag{2.8}$$

we may use the standard formula

$$\hat{J}_n(\theta) = \frac{\partial D_n(\theta; Z_n)}{\partial \theta'} = J_n(\theta; Z_n) . \tag{2.9}$$

Depending on the problem at hand, different forms of $\hat{I}_n(\theta)$ may be considered. The standard estimator appropriate for random sampling models is

$$\hat{I}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} h(\theta; z_t) h(\theta; z_t)' . \tag{2.10}$$

Some authors also argue that the centered version of (2.10) given by

$$\hat{I}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left[h(\theta; z_t) - \overline{h}(\theta)\right] \left[h(\theta; z_t) - \overline{h}(\theta)\right]' \tag{2.11}$$

where $\overline{h}(\theta) = \frac{1}{n}\sum_{t=1}^{n} h(\theta; z_t)$, can yield power improvements; see Hall (2000).

In this paper, we stress applications to time series data where serial dependence is present. In view of this, we focus on "heteroskedasticity-autocorrelation consistent" (HAC) covariance matrix estimators which account for the potential serial correlation and heteroskedasticity in the sequence $\{h(\theta; z_t)\}_{t=1}^{\infty}$:

$$\hat{I}_n(\theta) = \sum_{j=-n+1}^{n-1} \bar{k}(j/B_n)\, \hat{\Gamma}_n(j, \theta) \tag{2.12}$$

where $\bar{k}(\cdot)$ is a kernel function, B_n is a bandwidth parameter (which depends on the sample size and, possibly, on the data), and

$$\hat{\Gamma}_n(j, \theta) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} h(\theta; z_t)\, h(\theta; z_{t-j})', & \text{if } j \geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n} h(\theta; z_{t+j})\, h(\theta; z_t)', & \text{if } j < 0. \end{cases} \tag{2.13}$$

The reader is referred to Newey and West (1987*b*), Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Cushing and McGarvey (1999), Kiefer, Vogelsang and Bunzel (2000), and Kiefer and Vogelsang (2002, 2005) for further properties of covariance estimators of the form (2.12).

We now consider the problem of formulating a test statistic for H_0 using a general restricted estimator of θ_0. This means that we wish to use statistics based on estimators which may not be obtained by minimizing the objective function M_n in (2.2). This is motivated by the fact that minimizing M_n often constitutes a difficult numerical problem plagued by instabilities. Similarly, while some local efficiency arguments suggest taking $W_n = \hat{I}_n^{-1}$ [see Hansen (1982, Theorem 3.2), Davidson and MacKinnon (1993, Section 17.3), Gouriéroux and Monfort (1995, Section 9.5.2), Hall (2004, Section 3.6)], ill-conditioning can make this choice infeasible or harmful. So we allow here for a general weighting matrix W_n.

In order to obtain a unified test criterion which includes several other score-type statistics, we consider the following general "score-type" function:

$$s(\tilde{\theta}_n^0; W_n) = \sqrt{n}\, \tilde{Q}[W_n]\, D_n(\tilde{\theta}_n^0; Z_n)$$

where $\tilde{\theta}_n^0$ is a consistent restricted estimate of θ_0 such that $\psi(\tilde{\theta}_n^0) = 0$ and $\sqrt{n}(\tilde{\theta}_n^0 - \theta_0)$ is asymptotically bounded in probability,

$$\tilde{Q}[W_n] := \tilde{P}_n(\tilde{J}_n' W_n \tilde{J}_n)^{-1} \tilde{J}_n' W_n\,,$$

$\tilde{P}_n = P(\tilde{\theta}_n^0)$, $\tilde{J}_n = \hat{J}_n(\tilde{\theta}_n^0)$, and W_n is a symmetric positive definite (possibly random) $m \times m$ matrix such that

$$\operatorname*{plim}_{n\to\infty} W_n = W_0\,, \quad \det(W_0) \neq 0.$$

Under general regularity conditions (see Section 3), we have:

$$s\big(\tilde{\theta}_n^0; W_n\big) \xrightarrow[n\to\infty]{L} \mathrm{N}\left[0, Q(\theta_0)\, I(\theta_0)\, Q(\theta_0)'\right]$$

where

$$Q(\theta_0) = \operatorname*{plim}_{n\to\infty} \tilde{Q}[W_n] = P(\theta_0)\left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1} J(\theta_0)' W_0$$

and rank $[Q(\theta_0)] = p_1$. This suggests the following generalized $C(\alpha)$ criterion:

$$PC\big(\tilde{\theta}_n^0; \psi, W_n\big) = n\tilde{D}_n' \tilde{Q}[W_n]' \left\{\tilde{Q}[W_n]\,\tilde{I}_n\,\tilde{Q}[W_n]'\right\}^{-1} \tilde{Q}[W_n]\,\tilde{D}_n \tag{2.14}$$

where $\tilde{D}_n = D_n\big(\tilde{\theta}_n^0; Z_n\big)$ and $\tilde{I}_n = \hat{I}_n(\tilde{\theta}_n^0)$. We show in Section 3 that the asymptotic distribution of $PC\big(\tilde{\theta}_n^0; \psi, W_n\big)$ is $\chi^2(p_1)$ under H_0. The proposed test statistic includes as a special case several statistics proposed in the statistical and econometric literatures. We discuss these as well as other special cases in sections 4, 5 and 6.

3. Distribution of the generalized $C(\alpha)$ statistic

In this section, we derive the asymptotic distribution of the generalized $C(\alpha)$ statistic defined in (2.14) under the following set of assumptions. $\|\cdot\|$ refers to the Euclidean distance, applied to either vectors or matrices.

Assumption 3.1 EXISTENCE OF SCORE-TYPE FUNCTIONS.

$$D_n(\theta, \omega) = \big(D_{1n}(\theta, \omega), \ldots, D_{mn}(\theta, \omega)\big)', \ \omega \in \mathscr{X},\ n = 1, 2, \ldots$$

is a sequence of $m\times 1$ random vectors, defined on a common probability space $(\mathscr{X}, \mathscr{A}_{\mathscr{X}}, \mathsf{P})$, which are functions of a $p\times 1$ parameter vector θ, where $\theta \in \Theta \subseteq \mathbb{R}^p$ and Θ is a non-empty open subset of $\mathbb{R}^p$. All the random variables considered here as well in the following assumptions are functions of ω, so the symbol ω may be dropped to simplify notations [e.g., $D_n(\theta) := D_n(\theta, \omega)$]. *There is a unique vector $\theta_0 \in \Theta$ called the "true parameter value".*

Assumption 3.2 SCORE ASYMPTOTIC NORMALITY.

$$\sqrt{n}D_n(\theta_0) \xrightarrow[n\to\infty]{\mathsf{p}} \bar{D}_\infty(\theta_0) \text{ where } \bar{D}_\infty(\theta_0) \sim \mathrm{N}[0, I(\theta_0)]\,.$$

Assumption 3.3 NON-SINGULARITY OF THE SCORE VARIANCE. *$I(\theta)$ is nonsingular for any $\theta \in \Theta$ which satisfies the restriction $\psi(\theta) = 0$.*

Assumption 3.4 SCORE EXPANSION. *For θ in a non-empty open neighborhood N_0 of θ_0, $D_n(\theta)$ admits an expansion of the form*

$$D_n(\theta, \omega) = D_n(\theta_0, \omega) + J(\theta_0)(\theta - \theta_0) + R_n(\theta, \theta_0, \omega)$$

for $\omega \in \mathscr{D}_J$, *where* $\mathscr{D}_J$ *is an event with probability one* (i.e., $\mathsf{P}[\omega \in \mathscr{D}_J] = 1$) *and* $J(\theta)$ *is an* $m \times p$ (*nonrandom*) *matrix function of* θ *and the remainder* $R_n(\theta,\, \theta_0,\, \omega)$ *satisfies the following condition: for any* $\varepsilon > 0$ *and* $\delta > 0$, *we have*

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : r_n(\delta,\, \theta_0,\, \omega) > \varepsilon\}\big] \le U_D(\delta,\, \varepsilon,\, \theta_0)$$

$$r_n(\delta,\, \theta_0,\, \omega) = \sup\left\{ \frac{\|R_n(\theta,\, \theta_0,\, \omega)\|}{\|\theta - \theta_0\|} : \theta \in N_0 \text{ and } 0 < \|\theta - \theta_0\| \le \delta \right\},$$

$U_D(\delta,\, \varepsilon,\, \theta_0) \ge 0$ *and* $\lim_{\delta \downarrow 0} U_D(\delta,\, \varepsilon,\, \theta_0) = 0$.

Assumption 3.5 CONSISTENT ESTIMATOR OF $J(\theta_0)$. *There is a sequence of* $m \times p$ *random matrices* $J_n(\theta,\, \omega)$ *and a non-empty open neighborhood* V_0 *of* θ_0 *such that, for all* $\varepsilon > 0$ *and* $\delta > 0$,

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : \Delta_n(\theta_0,\, \delta,\, \omega) > \varepsilon\}\big] \le U_J(\delta,\, \varepsilon,\, \theta_0)$$

where

$$\Delta_n(\theta_0,\, \delta,\, \omega) := \sup\{\|J_n(\theta,\, \omega) - J(\theta_0)\| : \theta \in V_0 \text{ and } 0 \le \|\theta - \theta_0\| \le \delta\}$$

and $U_J(\delta,\, \varepsilon,\, \theta_0)$ *is a non-random function such that*

$$U_J(\delta,\, \varepsilon,\, \theta_0) \ge 0 \quad \text{and} \quad \lim_{\delta \downarrow 0} U_J(\delta,\, \varepsilon,\, \theta_0) = 0.$$

Assumption 3.6 ASYMPTOTIC SCORE NON-DEGENERACY. $\operatorname{rank}[J(\theta)] = p$ *for any* $\theta \in \Theta$ *which satisfies the restriction* $\psi(\theta) = 0$.

Assumption 3.7 RESTRICTION DIFFERENTIABILITY. $\psi(\theta)$ *is a* $p_1 \times 1$ *continuously differentiable vector function of* θ *with derivative* $P(\theta) := \dfrac{\partial \psi}{\partial \theta'}$.

Assumption 3.8 RESTRICTION RANK. $\operatorname{rank}[P(\theta)] = p_1$ *for any* $\theta \in \Theta$ *which satisfies the restriction* $\psi(\theta) = 0$.

Assumption 3.9 ESTIMATOR $\sqrt{n}$ CONVERGENCE. $\tilde{\theta}_n^0 := \tilde{\theta}_n^0(\omega)$ *is a consistent estimator of* θ_0, i.e.,

$$\operatorname*{plim}_{n\to\infty} \big(\tilde{\theta}_n^0 - \theta_0\big) = 0,$$

such that $\sqrt{n}\big(\tilde{\theta}_n^0 - \theta_0\big)$ *is asymptotically bounded in probability,* i.e.,

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : \sqrt{n}\,\big\|\tilde{\theta}_n^0 - \theta_0\big\| \ge y\}\big] \le U(y;\, \theta_0),\ \forall y > 0,$$

where $U(y;\, \theta_0)$ *is a function such that* $\lim_{y\to\infty} U(y;\, \theta_0) = 0$.

The latter assumption requires that the auxiliary estimator $\tilde{\theta}_n^0$ be $\sqrt{n}$-consistent only under the null hypothesis H_0, and corresponds to Neyman's (1959) local $\sqrt{n}$-consistency assumption. It may also be written $\sqrt{n}\big(\tilde{\theta}_n^0 - \theta_0\big) = O_p(1)$ under H_0.

Assumption 3.10 RESTRICTED ESTIMATOR. $\psi(\tilde{\theta}_n^0) = \psi(\theta_0) = 0$ *with probability* 1.

Assumption 3.11 CONSISTENT ESTIMATOR OF SCORE COVARIANCE MATRIX. $\tilde{I}_n$, $n \geq 1$, *is a sequence of* $m \times m$ *symmetric nonsingular* (*random*) *matrices such that* $\operatorname*{plim}_{n\to\infty} \tilde{I}_n = I(\theta_0)$.

Assumption 3.12 WEIGHT MATRIX CONVERGENCE. W_n, $n \geq 1$, *is a sequence of* $m \times m$ *symmetric nonsingular* (*random*) *matrices such that* $\operatorname*{plim}_{n\to\infty} W_n = W_0$ *where* W_0 *is nonsingular.*

The following proposition establishes the asymptotic distribution of the generalized $C(\alpha)$ statistic $PC\big(\tilde{\theta}_n^0;\, \psi, W_n\big)$ in (2.14).

Proposition 3.1 ASYMPTOTIC DISTRIBUTION OF GENERALIZED $C(\alpha)$ STATISTIC. *Let* $\tilde{Q}_n := \tilde{Q}[W_n] = \tilde{P}_n\,[\tilde{J}_n'\, W_n\, \tilde{J}_n]^{-1} \tilde{J}_n'\, W_n$ *where* $\tilde{J}_n = J_n(\tilde{\theta}_n^0;\, Z_n)$, $\tilde{P}_n = P(\tilde{\theta}_n^0)$. *If the assumptions 3.1 to 3.12 are satisfied, then, under* H_0,

$$\sqrt{n}\,\tilde{Q}_n D_n(\tilde{\theta}_n^0;\, Z_n) \xrightarrow[n\to\infty]{L} \mathrm{N}\big[0,\, Q(\theta_0)I(\theta_0)Q(\theta_0)'\big] \tag{3.1}$$

where $Q(\theta_0) = P(\theta_0)\big[J(\theta_0)'W_0 J(\theta_0)\big]^{-1} J(\theta_0)'W_0$, *and*

$$PC(\tilde{\theta}_n^0;\, \psi, W_n) = n D_n\big(\tilde{\theta}_n^0;\, Z_n\big)'\tilde{Q}_n'\,\big[\tilde{Q}_n \tilde{I}_n \tilde{Q}_n'\big]^{-1}\, \tilde{Q}_n D_n\big(\tilde{\theta}_n^0;\, Z_n\big) \xrightarrow[n\to\infty]{L} \chi^2(p_1)\,. \tag{3.2}$$

It is of interest to note here that the assumptions 3.4 and 3.5 do not require that $D_n(\theta,\, \omega)$ be differentiable with respect to θ. This is allowed by making a direct assumption on the existence of a linear expansion of $D_n(\theta,\, \omega)$ around θ_0 [Assumption 3.4]. For the same reason, $J_n(\theta,\, \omega)$ does not have to be continuous with respect to θ.

Since the differentiability of $D_n(\theta,\, \omega)$ with respect to θ is a common assumption, we will now show that the high-level assumptions 3.4 and 3.5 hold in the standard case where $D_n(\theta,\, \omega)$ is differentiable, with probability limit $J(\theta)$, and both $J_n(\theta,\, \omega)$ and $J(\theta)$ are continuous at least at every point in a neighborhood of θ_0. More precisely, consider the following assumptions.

Assumption 3.13 SCORE DIFFERENTIABILITY. $D_n(\theta,\omega)$ *is almost surely* (*a.s.*) *differentiable with respect to* θ, *for all* n, *in a non-empty open neighborhood* N_1 *of* θ_0. *The derivative matrix of* $D_n(\theta,\omega)$ *is denoted*

$$J_n(\theta,\, \omega) = \frac{\partial D_n(\theta,\omega)}{\partial \theta'} \tag{3.3}$$

where the sequence of matrices $J_n(\theta,\, \omega)$, $n \geq 1$, *is well-defined for* $\omega \in \mathscr{D}_J$ *and* $\mathscr{D}_J$ *is an event with probability one* (i.e., $\mathsf{P}[\omega \in \mathscr{D}_J] = 1$).

Assumption 3.14 SCORE DERIVATIVE UNIFORM CONVERGENCE. *$D_n(\theta, \omega)$ satisfies the following conditions:*

(a) *$J_n(\theta, \omega)$ is continuous with respect to θ for all $\theta \in N_2$, $\omega \in \mathscr{D}_J$ and $n \geq 1$;*

(b) $\sup_{\theta \in N_2} \|J_n(\theta, \omega) - J(\theta)\| \xrightarrow[n\to\infty]{\mathsf{p}} 0$.

We then have the following implication, which shows that Proposition **3.1** still holds if the assumptions 3.4 and 3.5 are replaced by the (stronger) assumptions 3.13 and 3.14. Another implication is that $J(\theta)$ is continuous at $\theta = \theta_0$ in this special case.

Proposition 3.2 SUFFICIENCY OF SCORE JACOBIAN CONTINUITY AND UNIFORM CONVERGENCE. *Suppose the assumptions 3.1 to 3.3 hold. Then the assumptions 3.13 and 3.14 entail that:*

(a) *$J(\theta)$ is continuous at $\theta = \theta_0$;*

(b) *both the assumptions 3.4 and 3.5 also hold.*

4. Alternative $C(\alpha)$-type statistics

It will be of interest to examine a number of special forms of the general statistic proposed in Section 2. In particular, the statistic $PC(\tilde{\theta}_n^0; \psi, W_n)$ nests several $C(\alpha)$-type and score-based statistics proposed in the statistical and econometric literatures, as well as new ones.[1] It will be of interest to spell out some of these.

On taking $W_n = \tilde{I}_n^{-1}$, as suggested by efficiency arguments, $PC(\tilde{\theta}_n^0; \psi, W_n)$ reduces to

$$PC(\tilde{\theta}_n^0; \psi) = n D_n\big(\tilde{\theta}_n^0; Z_n\big)' \tilde{W}_n D_n\big(\tilde{\theta}_n^0; Z_n\big) \tag{4.1}$$

where $\tilde{\theta}_n^0$ is any root-n consistent estimator of θ which satisfies $\psi(\tilde{\theta}_n^0) = 0$, and

$$\tilde{W}_n = \tilde{I}_n^{-1} \tilde{J}_n' \big(\tilde{J}_n' \tilde{I}_n^{-1} \tilde{J}_n\big)^{-1} \tilde{P}_n' \big[\tilde{P}_n \big(\tilde{J}_n' \tilde{I}_n^{-1} \tilde{J}_n\big)^{-1} \tilde{P}_n'\big]^{-1} \tilde{P}_n \big(\tilde{J}_n' \tilde{I}_n^{-1} \tilde{J}_n\big)^{-1} \tilde{J}_n' \tilde{I}_n^{-1}$$

with $\tilde{P}_n = P(\tilde{\theta}_n^0)$, $\tilde{I}_n = \hat{I}_n(\tilde{\theta}_n^0)$ and $\tilde{J}_n = \hat{J}_n(\tilde{\theta}_n^0)$.

When the number of equations equals the number of parameters $(m = p)$, we have $\tilde{Q}[W_n] = \tilde{P}_n \tilde{J}_n^{-1}$ and $PC(\tilde{\theta}_n^0; \psi, W_n)$ does not depend on the choice of W_n:

$$\begin{aligned} PC(\tilde{\theta}_n^0; \psi, W_n) &= PC\big(\tilde{\theta}_n^0; \psi\big) \\ &= n D_n\big(\tilde{\theta}_n^0; Z_n\big)' (\tilde{J}_n^{-1})' \tilde{P}_n' \big[\tilde{P}_n \big(\tilde{J}_n' \tilde{I}_n^{-1} \tilde{J}_n\big)^{-1} \tilde{P}_n'\big]^{-1} \tilde{P}_n \tilde{J}_n^{-1} D_n\big(\tilde{\theta}_n^0; Z_n\big) . \end{aligned} \tag{4.2}$$

In particular, this will be the case if $D_n(\theta; Z_n)$ is the derivative of a (pseudo) log-likelihood function.

[1] For further discussion of $C(\alpha)$ tests, the reader may consult Basawa (1985), Ronchetti (1987), Smith (1987*a*), Berger and Wallenstein (1989), Dagenais and Dufour (1991), and Kocherlakota and Kocherlakota (1991).

For $m \geq p$, when $\tilde{\theta}_n^0$ is obtained by minimizing $M_n(\theta) = D_n(\theta; Z_n)' \tilde{I}_n^{-1} D_n(\theta; Z_n)$ subject to $\psi(\theta) = 0$, where $\tilde{I}_n$ is an estimator of $I(\theta_0)$, we can write $\tilde{\theta}_n^0 = \hat{\theta}_n^0$ and $PC(\tilde{\theta}_n^0; \psi, W_n)$ is identical to the s*core-type statistic* suggested by Newey and West (1987*a*):

$$S(\psi) = n D_n(\hat{\theta}_n^0; Z_n)' \hat{I}_n^{-1} \hat{J}_n \big(\hat{J}_n' \hat{I}_n^{-1} \hat{J}_n\big)^{-1} \hat{J}_n' \hat{I}_n^{-1} D_n(\hat{\theta}_n^0; Z_n) \tag{4.3}$$

where $\hat{I}_n = \hat{I}_n(\hat{\theta}_n^0)$ and $\hat{J}_n = \hat{J}_n(\hat{\theta}_n^0)$. This statistic is closely related with the *Lagrange-multiplier-type* (LM-type) statistic

$$LM(\psi) = n \hat{\lambda}_n' \hat{P}_n \big(\hat{J}_n' \hat{I}_n^{-1} \hat{J}_n\big)^{-1} \hat{P}_n' \hat{\lambda}_n \tag{4.4}$$

where $\hat{P}_n = P(\hat{\theta}_n^0)$ and $\hat{\lambda}_n$ is the Lagrange multiplier in the corresponding constrained optimization problem. Indeed, upon using the first-order condition

$$J_n\big(\hat{\theta}_n^0; Z_n\big)' \tilde{I}_n^{-1} D_n\big(\hat{\theta}_n^0; Z_n\big) = P\big(\hat{\theta}_n^0\big)' \hat{\lambda}_n\,, \tag{4.5}$$

we see easily that

$$S(\psi) = LM(\psi)\,. \tag{4.6}$$

In (correctly specified) parametric models, we have $I(\theta) = -J(\theta)$ and the $C(\alpha)$ statistic in (4.2) reduces to

$$PC\big(\tilde{\theta}_n^0; \psi\big) = n D_n\big(\tilde{\theta}_n^0; Z_n\big)' \tilde{I}_n^{-1} \tilde{P}_n' \big[\tilde{P}_n \tilde{I}_n^{-1} \tilde{P}_n'\big]^{-1} \tilde{P}_n \tilde{I}_n^{-1} D_n\big(\tilde{\theta}_n^0; Z_n\big) \tag{4.7}$$

where $D_n\big(\tilde{\theta}_n^0; Z_n\big)$ is the score of the log-likelihood function and $\tilde{I}_n$ is the Fisher information matrix or a consistent estimate, each evaluated at the auxiliary estimator $\tilde{\theta}_n^0$. The extension of $C(\alpha)$ statistics to a general parameter constraint given in (4.7) was first proposed by Smith (1987*b*) in a likelihood setting; see Dagenais and Dufour (1991) for further discussion of this test statistic.

5. Testing a subvector

A common problem in statistics consists in testing an hypothesis of the form

$$H_0 : \theta_1 = \bar{\theta}_{10} \tag{5.1}$$

where θ_1 is a subvector of θ, and $\bar{\theta}_{10}$ is a given possible value of θ_1, *i.e.* we consider $\psi(\theta) = \theta_1 - \bar{\theta}_{10}$. Without loss of generality, we can assume that $\theta = (\theta_1', \theta_2')'$ where θ_1 is a $p_1 \times 1$ vector and θ_2 is a $p_2 \times 1$ vector, and denote $\theta_0 = (\theta_{10}', \theta_{20}')'$ the "true value" of θ. In this case,

$$P(\theta) = [I_{p_1}, 0_{p_1 \times p_2}] \tag{5.2}$$

where I_{p_1} is the identity matrix of order p_1 and $0_{p_1 \times p_2}$ is the $p_1 \times p_2$ zero matrix. Let $\tilde{\theta}_n^0$ be a restricted $\sqrt{n}$-consistent estimator of θ. We can then write $\tilde{\theta}_n^0 = (\bar{\theta}_{10}', \tilde{\theta}_{2n}^{0\prime})'$ where $\tilde{\theta}_{2n}^0$ is a $\sqrt{n}$-consistent estimator of θ_2.

Let us partition $J(\theta)$ and $\tilde{J}_n = J_n(\tilde{\theta}_n^0; Z_n)$ conformably with $\theta = (\theta_1', \theta_2')'$:

$$J(\theta) = [J_{\cdot 1}(\theta), J_{\cdot 2}(\theta)], \; \tilde{J}_n = [\tilde{J}_{n\cdot 1}, \tilde{J}_{n\cdot 2}] = [\tilde{J}_{n\cdot 1}(\tilde{\theta}_n^0; Z_n), \tilde{J}_{n\cdot 2}(\tilde{\theta}_n^0; Z_n)], \tag{5.3}$$

where $J_{\cdot i}(\theta)$ and $\tilde{J}_{n\cdot i} = \tilde{J}_{n\cdot i}(\tilde{\theta}_n^0; Z_n)$ are $m \times p_i$ matrices, $i = 1, 2$. Let also

$$\tilde{J}_n^* = W_n^{1/2} \tilde{J}_n = [\tilde{J}_{n\cdot 1}^*, \tilde{J}_{n\cdot 2}^*], \; \tilde{J}_{n\cdot i}^* = W_n^{1/2} \tilde{J}_{n\cdot i} \;\; i = 1, 2, \tag{5.4}$$

and conformably partition the matrix $\tilde{J}_n' W_n \tilde{J}_n$ and its inverse $(\tilde{J}_n' W_n \tilde{J}_n)^{-1}$:

$$\tilde{J}_n' W_n \tilde{J}_n = \begin{bmatrix} (\tilde{J}_n' W_n \tilde{J}_n)_{11} & (\tilde{J}_n' W_n \tilde{J}_n)_{12} \\ (\tilde{J}_n' W_n \tilde{J}_n)_{21} & (\tilde{J}_n' W_n \tilde{J}_n)_{22} \end{bmatrix} = \begin{bmatrix} \tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 1} & \tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 2} \\ \tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 1} & \tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2} \end{bmatrix}, \tag{5.5}$$

$$(\tilde{J}_n' W_n \tilde{J}_n)^{-1} = \begin{bmatrix} (\tilde{J}_n' W_n \tilde{J}_n)^{11} & (\tilde{J}_n' W_n \tilde{J}_n)^{12} \\ (\tilde{J}_n' W_n \tilde{J}_n)^{21} & (\tilde{J}_n' W_n \tilde{J}_n)^{22} \end{bmatrix}, \tag{5.6}$$

where $(\tilde{J}_n' W_n \tilde{J}_n)_{ij}$ and $(\tilde{J}_n' W_n \tilde{J}_n)^{ij}$ are $p_i \times p_j$ matrices, $i, j = 1, 2$. We denote $P[Z] = Z(Z'Z)^{-1}Z'$ the projection matrix on the space spanned by the columns of a full-column rank matrix Z, and $M[Z] = I - Z(Z'Z)^{-1}Z'$.

Let us now assume that the matrix $(\tilde{J}_n' W_n \tilde{J}_n)_{22}$ is invertible. This entails that $(\tilde{J}_n' W_n \tilde{J}_n)^{11}$ is invertible and, on using standard rules for multiplying partitioned matrices,

$$\begin{aligned} \left[(\tilde{J}_n' W_n \tilde{J}_n)^{11}\right]^{-1} (\tilde{J}_n' W_n \tilde{J}_n)^{12} &= -(\tilde{J}_n' W_n \tilde{J}_n)_{12} [(\tilde{J}_n' W_n \tilde{J}_n)_{22}]^{-1} \\ &= -(\tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 2})(\tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2})^{-1}, \end{aligned} \tag{5.7}$$

$$(\tilde{J}_n' W_n \tilde{J}_n)^{11} = \left[(\tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 1}) - \tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 2} (\tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2})^{-1} \tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 1}\right]^{-1}; \tag{5.8}$$

see Harville (1997, Theorem 8.5.11). We can then rewrite $\tilde{Q}[W_n]$ as

$$\begin{aligned} \tilde{Q}[W_n] &= \tilde{P}_n (\tilde{J}_n' W_n \tilde{J}_n)^{-1} \tilde{J}_n' W_n \\ &= [I_{p_1}, 0_{p_1 \times p_2}] \begin{bmatrix} (\tilde{J}_n' W_n \tilde{J}_n)^{11} & (\tilde{J}_n' W_n \tilde{J}_n)^{12} \\ (\tilde{J}_n' W_n \tilde{J}_n)^{21} & (\tilde{J}_n' W_n \tilde{J}_n)^{22} \end{bmatrix} \begin{bmatrix} \tilde{J}_{n\cdot 1}' \\ \tilde{J}_{n\cdot 2}' \end{bmatrix} W_n \\ &= \left[(\tilde{J}_n' W_n \tilde{J}_n)^{11} \tilde{J}_{n\cdot 1}' + (\tilde{J}_n' W_n \tilde{J}_n)^{12} \tilde{J}_{n\cdot 2}'\right] W_n \\ &= (\tilde{J}_n' W_n \tilde{J}_n)^{11} \left[\tilde{J}_{n\cdot 1}' + \left((\tilde{J}_n' W_n \tilde{J}_n)^{11}\right)^{-1} (\tilde{J}_n' W_n \tilde{J}_n)^{12} \tilde{J}_{n\cdot 2}'\right] W_n \\ &= (\tilde{J}_n' W_n \tilde{J}_n)^{11} \left[\tilde{J}_{n\cdot 1}' - (\tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 2})(\tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2})^{-1} \tilde{J}_{n\cdot 2}'\right] W_n \\ &= \tilde{V}_{n\cdot 1|2}^{-1} \tilde{J}_{n\cdot 1|2}' W_n \end{aligned} \tag{5.9}$$

where

$$\tilde{J}_{n\cdot 1|2} = \tilde{J}_{n\cdot 1} - \tilde{J}_{n\cdot 2} (\tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2})^{-1} \tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 1} = W_n^{-1/2} M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^*, \tag{5.10}$$

$$\begin{aligned} \tilde{V}_{n\cdot 1|2} &= (\tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 1}) - \tilde{J}_{n\cdot 1}' W_n \tilde{J}_{n\cdot 2} \left(\tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 2}\right)^{-1} \tilde{J}_{n\cdot 2}' W_n \tilde{J}_{n\cdot 1} \\ &= \tilde{J}_{n\cdot 1}^{*\prime} M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^*. \end{aligned} \tag{5.11}$$

Using (5.9), we get:

$$\begin{aligned}\tilde{Q}[W_n]\,\tilde{D}_n &= \tilde{V}_{n\cdot 1|2}^{-1}\,\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n \\ &= \tilde{V}_{n\cdot 1|2}^{-1}\left[\tilde{J}_{n\cdot 1}'\,W_n\,\tilde{D}_n-(\tilde{J}_{n\cdot 1}'\,W_n\,\tilde{J}_{n\cdot 2})(\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{J}_{n\cdot 2})^{-1}\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{D}_n\right], \end{aligned} \tag{5.12}$$

$$\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n=\tilde{J}_{n\cdot 1}'\,W_n\,\tilde{D}_n-(\tilde{J}_{n\cdot 1}'\,W_n\,\tilde{J}_{n\cdot 2})(\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{J}_{n\cdot 2})^{-1}\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{D}_n\,, \tag{5.13}$$

$$\begin{aligned}\tilde{Q}[W_n]\,\tilde{I}_n\,\tilde{Q}[W_n]' &= \tilde{V}_{n\cdot 1|2}^{-1}\,\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{I}_n\,W_n\,\tilde{J}_{n\cdot 1|2}\,\tilde{V}_{n\cdot 1|2}^{-1} \\ &= \tilde{V}_{n\cdot 1|2}^{-1}\,\tilde{J}_{n\cdot 1}^{*\prime}M[\tilde{J}_{n\cdot 2}^{*}]W_n^{1/2}\,\tilde{I}_n\,W_n^{1/2}M[\tilde{J}_{n\cdot 2}^{*}]\tilde{J}_{n\cdot 1}^{*}\,\tilde{V}_{n\cdot 1|2}^{-1}\,, \end{aligned} \tag{5.14}$$

where $\tilde{D}_n=D_n\big(\tilde{\theta}_n^0;Z_n\big)$. The generalized $C(\alpha)$ statistic then takes the form:

$$\begin{aligned}PC_1\big(\tilde{\theta}_n^0;\bar{\theta}_{10},W_n\big) &= PC\big(\tilde{\theta}_n^0;\psi,W_n\big) \\ &= n\,\tilde{D}_n'\,W_n\,\tilde{J}_{n\cdot 1|2}\big(\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{I}_n\,W_n\,\tilde{J}_{n\cdot 1|2}\big)^{-1}\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n \\ &= n\,\tilde{D}_n'\,W_n^{1/2}M[\tilde{J}_{n\cdot 2}^{*}]\tilde{J}_{n\cdot 1}^{*}\,\tilde{\Sigma}_n(W_n)^{-1}\,\tilde{J}_{n\cdot 1}^{*\prime}M[\tilde{J}_{n\cdot 2}^{*}]W_n^{1/2}\tilde{D}_n \end{aligned} \tag{5.15}$$

where

$$\tilde{\Sigma}_n(W_n)=\tilde{J}_{n\cdot 1}^{*\prime}M[\tilde{J}_{n\cdot 2}^{*}](W_n^{1/2}\tilde{I}_n\,W_n^{1/2})M[\tilde{J}_{n\cdot 2}^{*}]\tilde{J}_{n\cdot 1}^{*}$$

and the matrix $\tilde{V}_{n\cdot 1|2}^{-1}$ cancels out.

It is also of interest to note that the transformed score $\tilde{S}_{n\cdot 1|2}=\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n$ in $PC_1\big(\tilde{\theta}_n^0;\bar{\theta}_{10},W_n\big)$ is by construction uncorrelated with $\tilde{S}_{n\cdot 2}=\tilde{J}_{n\cdot 2}'\tilde{I}_n^{-1}\,\tilde{D}_n$ asymptotically. This follows on observing that:

$$\sqrt{n}\left[\begin{array}{c}\tilde{S}_{n\cdot 1|2}\\ \tilde{S}_{n\cdot 2}\end{array}\right]=\sqrt{n}\,\tilde{R}_n\,\tilde{D}_n \xrightarrow[n\to\infty]{L} \mathrm{N}\big[0,\bar{R}(\theta_0)I(\theta_0)\bar{R}(\theta_0)'\big] \tag{5.16}$$

where

$$\tilde{R}_n=\left[\begin{array}{c}\tilde{J}_{n\cdot 1|2}'W_n\\ \tilde{J}_{n\cdot 2}'\,\tilde{I}_n^{-1}\end{array}\right]\xrightarrow[n\to\infty]{p} R(\theta_0)=\left[\begin{array}{c}J_{\cdot 1|2}(\theta_0)'W_0\\ J_{\cdot 2}(\theta_0)'I(\theta_0)^{-1}\end{array}\right], \tag{5.17}$$

$$J_{\cdot 1|2}(\theta_0)=J_{\cdot 1}(\theta_0)-J_{\cdot 2}(\theta_0)[J_{\cdot 2}(\theta_0)'W_0\,J_{\cdot 2}(\theta_0)]^{-1}J_{\cdot 2}(\theta_0)'W_0\,J_{\cdot 1}(\theta_0)\,, \tag{5.18}$$

and the asymptotic covariance matrix between $\sqrt{n}\,\tilde{S}_{n\cdot 2}$ and $\sqrt{n}\,\tilde{S}_{n\cdot 1|2}$ is

$$\big[J_{\cdot 2}(\theta_0)'I(\theta_0)^{-1}\big]\,I(\theta_0)\,\big[W_0\,J_{\cdot 1|2}(\theta_0)\big]=J_{\cdot 2}(\theta_0)'\,\big[W_0\,J_{\cdot 1|2}(\theta_0)\big]=0\,. \tag{5.19}$$

Indeed, the above orthogonality can be viewed as the source of the evacuation of the distribution of $\sqrt{n}\big(\,\tilde{\theta}_n^0-\theta_0\big)$ from the asymptotic distribution of the generalized $C(\alpha)$ statistic: using the assumptions 3.4 and 3.9 [see (A.5)], we see easily that, under H_0,

$$J_{\cdot 1|2}(\theta_0)'W_0\sqrt{n}\,\big[D_n(\tilde{\theta}_n^0)-D_n(\theta_0)\big]=J_{\cdot 1|2}(\theta_0)'W_0\,J(\theta_0)\sqrt{n}\big(\,\tilde{\theta}_n^0-\theta_0\big)+o_p(1)$$

$$= J_{\cdot 1|2}(\theta_0)' W_0 J_{\cdot 2}(\theta_0) \sqrt{n}\big(\tilde{\theta}_{2n}^{0} - \theta_{20}\big) + o_p(1) = o_p(1)\,. \tag{5.20}$$

Thus the asymptotic null distribution of the modified score used by the generalized $C(\alpha)$ statistic does not depend on the limit distribution of the nuisance parameter estimator $\tilde{\theta}_n^0$, and similarly for the generalized $C(\alpha)$ statistic.

When $W_n = \tilde{I}_n^{-1}$, the formula in (5.15) simplifies to:

$$\begin{aligned} PC_1(\tilde{\theta}_n^0; \bar{\theta}_{10}) &= n\tilde{D}_n' \tilde{I}_n^{-1} \tilde{J}_{n\cdot 1|2} \left[\tilde{J}_{n\cdot 1|2}' \tilde{I}_n^{-1} \tilde{J}_{n\cdot 1|2}\right]^{-1} \tilde{J}_{n\cdot 1|2}' \tilde{I}_n^{-1} \tilde{D}_n \\ &= n\tilde{D}_n' \tilde{I}_n^{-1/2} M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^* \left[\tilde{J}_{n\cdot 1}^{*\prime} M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^*\right]^{-1} \tilde{J}_{n\cdot 1}^{*\prime} M[\tilde{J}_{n\cdot 2}^*] \tilde{I}_n^{-1/2} \tilde{D}_n \\ &= n\tilde{D}_n' \tilde{I}_n^{-1/2} P\big[M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^*\big] \tilde{I}_n^{-1/2} \tilde{D}_n \end{aligned} \tag{5.21}$$

where

$$\tilde{J}_{n\cdot 1|2} = \big[I_m - \tilde{J}_{n\cdot 2}\left(\tilde{J}_{n\cdot 2}' \tilde{I}_n^{-1} \tilde{J}_{n\cdot 2}\right)^{-1} \tilde{J}_{n\cdot 2}' \tilde{I}_n^{-1}\big] \tilde{J}_{n\cdot 1} = \tilde{I}_n^{1/2} M[\tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}] \tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 1}\,, \tag{5.22}$$

$$\tilde{J}_{n\cdot 1}^* = \tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 1}\,,\; \tilde{J}_{n\cdot 2}^* = \tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}\,. \tag{5.23}$$

Upon using (5.16) - (5.19), we see that $\tilde{J}_{n\cdot 1|2}' \tilde{I}_n^{-1} \tilde{D}_n$ and $\tilde{J}_{n\cdot 2}' \tilde{I}_n^{-1} \tilde{D}_n$ are asymptotically uncorrelated, and

$$\begin{aligned} \tilde{J}_{n\cdot 1|2}' \tilde{I}_n^{-1} \tilde{D}_n &= \tilde{J}_{n\cdot 1}' \tilde{I}_n^{-1/2} M[\tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}] \tilde{I}_n^{-1/2} \tilde{D}_n \\ &= \tilde{J}_{n\cdot 1}' \tilde{I}_n^{-1/2} \{I_m - P[\tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}]\} \tilde{I}_n^{-1/2} \tilde{D}_n \end{aligned} \tag{5.24}$$

where $M[\tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}] \tilde{I}_n^{-1/2} \tilde{D}_n$ is the residual from the projection of $\tilde{I}_n^{-1/2} \tilde{D}_n$ on $\tilde{I}_n^{-1/2} \tilde{J}_{n\cdot 2}$. Further, on applying the Frisch-Waugh-Lovell theorem, we see that

$$P\left[\tilde{J}_n^*\right] = P\left[\tilde{J}_{n\cdot 2}^*\right] + P\big[M[\tilde{J}_{n\cdot 2}^*] \tilde{J}_{n\cdot 1}^*\big]\,, \tag{5.25}$$

hence

$$\begin{aligned} PC_1(\tilde{\theta}_n^0; \bar{\theta}_{10}) &= n\tilde{D}_n' \tilde{I}_n^{-1/2} \big\{P\left[\tilde{J}_n^*\right] - P\left[\tilde{J}_{n\cdot 2}^*\right]\big\} \tilde{I}_n^{-1/2} \tilde{D}_n \\ &= n\left[\tilde{D}_n' \tilde{I}_n^{-1} \tilde{J}_n \big(\tilde{J}_n' \tilde{I}_n^{-1} \tilde{J}_n\big)^{-1} \tilde{J}_n' \tilde{I}_n^{-1} \tilde{D}_n - \tilde{D}_n' \tilde{I}_n^{-1} \tilde{J}_{n\cdot 2} \big(\tilde{J}_{n\cdot 2}' \tilde{I}_n^{-1} \tilde{J}_{n\cdot 2}\big)^{-1} \tilde{J}_{n\cdot 2}' \tilde{I}_n^{-1} \tilde{D}_n\right]. \end{aligned} \tag{5.26}$$

Finally, let us consider parametric models where $m = p$ and $D_{ni}(\theta; Z_n)$ denotes the $p_i \times 1$ score function (the derivative of the log-likelihood function) corresponding to θ_i, $i = 1, 2$, along with the corresponding partition of $\tilde{D}_n$ and $\tilde{I}_n$:

$$\tilde{D}_n = \begin{bmatrix} \tilde{D}_{n1} \\ \tilde{D}_{n2} \end{bmatrix} = \begin{bmatrix} D_{n1}(\tilde{\theta}_n^0; Z_n) \\ D_{n2}(\tilde{\theta}_n^0; Z_n) \end{bmatrix},\; \tilde{I}_n = \begin{bmatrix} \tilde{I}_{n11} & \tilde{I}_{n12} \\ \tilde{I}_{n21} & \tilde{I}_{n22} \end{bmatrix}, \tag{5.27}$$

where $\tilde{D}_{ni} = D_{ni}(\tilde{\theta}_n^0; Z_n)$ is a $p_i \times 1$ vector and $\tilde{I}_{nij}$ is $p_i \times p_j$ matrix, $i, j = 1, 2$. In such cases, we have $J(\theta_0) = -I(\theta_0)$, and upon setting $\tilde{J}_n = -\tilde{I}_n$, the formulas in (5.21) and (5.26) reduce to a simple difference between two statistics:

$$\begin{aligned} PC_1(\tilde{\theta}_n^0; \bar{\theta}_{10}) &= n\big(\tilde{D}_{n1} - \tilde{I}_{n12}\tilde{I}_{n22}^{-1}\tilde{D}_{n2}\big)'\big(\tilde{I}_{n11} - \tilde{I}_{n12}\tilde{I}_{n22}^{-1}\tilde{I}_{n21}\big)^{-1}\big(\tilde{D}_{n1} - \tilde{I}_{n12}\tilde{I}_{n22}^{-1}\tilde{D}_{n2}\big) \\ &= n\big[\tilde{D}_n'\tilde{I}_n^{-1}\tilde{D}_n - \tilde{D}_{n2}'\tilde{I}_{n22}^{-1}\tilde{D}_{n2}\big]. \end{aligned} \tag{5.28}$$

6. Two-stage procedures

In this section, we formulate the $C(\alpha)$ statistic for estimating functions (or GMM-type) models estimated by two-step procedures. The $C(\alpha)$ test procedure applies in a natural way to moment condition models estimated by a two-step procedure, because a correction for the first-stage estimation error is readily built into the statistic. Models of this kind typically involve a parameter vector $\theta = (\theta_1', \theta_2')'$ where θ_1 is the parameter vector of interest (on which inference focuses), and θ_2 denotes a vector of nuisance parameters which is consistently estimated by an auxiliary estimate $\tilde{\theta}_{2n}^0$ obtained from the first-stage estimation. Gong and Samaniego (1981), Pagan (1984, 1986), and Murphy and Topel (1985) among others study the properties of two-step estimation and testing procedures in a likelihood framework. Newey and McFadden (1994) deal with the problem in a GMM framework, but do not consider the $C(\alpha)$ statistic.

In this section, we describe how generalized $C(\alpha)$ tests can provide relatively simple solutions to such problems in the context of estimating functions and GMM estimation, with serial dependence. We first consider the generic case where the nuisance vector θ_2 is estimated in a first stage, and then treated as known for the purpose of testing the value of another parameter vector θ_1. Second, we examine the special case of a two-step GMM estimation, where the estimation of the nuisance parameter is based on a separate set of estimating functions (or moment conditions).

6.1. Tests based on general two-step estimation

Suppose we are interested in testing the restriction $H_0 : \theta_1 = \bar{\theta}_{10}$ based on data $X_n = [x_1, \ldots, x_n]$ and an $m_1 \times 1$ vector of estimating functions

$$D_{n1}(\theta; X_n) = D_{n1}(\theta_1, \theta_2; X_n). \tag{6.1}$$

In particular, we may assume $D_{n1}(\theta; X_n)$ is a subvector of a larger vector

$$D_n(\theta; X_n) = [D_{n1}(\theta; X_n)', D_{n2}(\theta; X_n)']'. \tag{6.2}$$

A typical setup is the one where

$$D_{n1}(\theta; X_n) = \frac{1}{n}\sum_{t=1}^{n} h_1(\theta_1, \theta_2; x_t), \tag{6.3}$$

$$\mathsf{E}_\theta\,[h_1(\theta_1, \theta_2; x_t)] = 0, \quad t = 1, \ldots, n, \tag{6.4}$$

and $h_1(\theta; x_t) = h_1(\theta_1, \theta_2; x_t)$ is a subvector of a higher-dimensional vector $h(\theta; x_t) = [h_1(\theta; x_t)', h_2(\theta; x_t)']'$ of estimating functions.

If the dimension of $D_{n1}(\theta_1, \theta_2; X_n)$ is large enough ($m_1 \geq p$) and the regularity conditions of Proposition **3.1** are satisfied when $D_n(\theta; X_n)$ is replaced by $D_{n1}(\theta; X_n)$, we can build general $C(\alpha)$-type tests of $H_0 : \theta_1 = \bar{\theta}_{10}$ based on $D_{n1}(\theta_1, \theta_2; X_n)$. No information on the (eventual) left-out estimating functions $D_{n2}(\theta; X_n)$ is required. These features underscore the remarkable versatility of estimating functions in conjunction with the generalized $C(\alpha)$ procedure described in this paper.

Let $\tilde{\theta}^0_{2n}$ be an estimator of the nuisance parameter vector θ_2 obtained from the data $Y_n = [y_1, \ldots, y_n]$ which may be different from X_n.[2] For example, $\tilde{\theta}^0_{2n}$ may be based on an "auxiliary" estimating function $D_{n2}(\theta; X_n)$, but this is not required. Consider now the restricted estimator $\tilde{\theta}^0_n = \big(\bar{\theta}'_{10}, \tilde{\theta}^{0\prime}_{2n}\big)'$, and denote $\tilde{D}_{n1} = D_{n1}\big(\tilde{\theta}^0_n; X_n\big)$, $\tilde{I}_{n11}$, $\tilde{J}_{n1i} := \hat{J}_{n1i}\big(\tilde{\theta}^0_n\big)$ and W_{n11}, the matrices corresponding to $\tilde{D}_n = D_n(\tilde{\theta}^0_n; Z_n)$, $\tilde{I}_n$, $\tilde{J}_{n\cdot i}$ and W_n respectively for the system based on the estimating function $D_{n1}(\theta; X_n)$; $\tilde{D}_{n1}$ has dimension $m_1 \times 1$, $\tilde{J}_{n1i}$ is $m_1 \times p_i$, and W_{n11} is $m_1 \times m_1$. In the case where $D_{n1}(\theta; X_n)$ is a subvector of $D_n(\theta; X_n)$ as in (6.2), $\tilde{I}_{n11}$, $\tilde{J}_{n1i}$ and W_{n11} are the corresponding submatrices of $\tilde{I}_n$, $\tilde{J}_{n\cdot i}$ and W_n respectively, where

$$W_n = \begin{bmatrix} W_{n11} & W_{n12} \\ W_{n21} & W_{n22} \end{bmatrix} \tag{6.5}$$

and W_{nij} is a $p_i \times p_j$ matrix, $i, j = 1, 2$.

Making the appropriate substitutions in (5.15), we then get the following $C(\alpha)$-type statistic for $H_0 : \theta_1 = \bar{\theta}_{10}$:

$$PC_1\big(\tilde{\theta}^0_n; \bar{\theta}_{10}, W_{n11}\big) = n\,\tilde{D}'_{n1}\, W_{n11}\, \tilde{J}_{n11|2}\, \tilde{\Sigma}^{-1}_{n11|2}\, \tilde{J}'_{n11|2}\, W_{n11}\, \tilde{D}_{n1} \tag{6.6}$$

where $\tilde{\Sigma}_{n11|2} = \tilde{J}'_{n11|2}\, W_{n11}\, \tilde{I}_{n11}\, W_{n11}\, \tilde{J}_{n11|2}$, and

$$\begin{aligned} \tilde{J}_{n11|2} &= \tilde{J}_{n11} - \tilde{J}_{n12}(\tilde{J}'_{n12}\, W_{n11}\, \tilde{J}_{n12})^{-1}\tilde{J}'_{n12}\, W_{n11}\, \tilde{J}_{n11} \\ &= W^{-1/2}_{n11}\, M[W^{1/2}_{n11}\, \tilde{J}_{n12}]W^{1/2}_{n11}\, \tilde{J}_{n11}\,, \end{aligned} \tag{6.7}$$

$$\begin{aligned} \tilde{\Sigma}_{n11|2} &= \tilde{J}'_{n11|2}\, W_{n11}\, \tilde{I}_{n11}\, W_{n11}\, \tilde{J}_{n11|2} \\ &= \tilde{J}'_{n11}\, W^{1/2}_{n11} M[W^{1/2}_{n11}\, \tilde{J}_{n12}]W^{1/2}_{n11}\, \tilde{I}_{n11} W^{1/2}_{n11} M[W^{1/2}_{n11}\, \tilde{J}_{n12}]W^{1/2}_{n11}\, \tilde{J}_{n11}\,. \end{aligned} \tag{6.8}$$

By Proposition **3.1**, $PC_1\big(\tilde{\theta}^0_n; \bar{\theta}_{10}, W_{n11}\big)$ has a $\chi^2(p_1)$ asymptotic distribution under H_0. On taking $W_{n11} = \tilde{I}^{-1}_{n11}$, PC_1 takes the following simplified form:

$$PC_1\big(\tilde{\theta}^0_n; \bar{\theta}_{10}, \tilde{I}^{-1}_{n11}\big) = n\,\tilde{D}'_{n1}\, \tilde{I}^{-1/2}_{n11} \tilde{M}_{12}\, \tilde{I}^{-1/2}_{n11}\, \tilde{J}_{n11}\, \tilde{\Sigma}^{-1}_{n11|2}\, \tilde{J}'_{n11} \tilde{I}^{-1/2}_{n11}\, \tilde{M}_{12}\, \tilde{I}^{-1/2}_{n11}\, \tilde{D}_{n1} \tag{6.9}$$

where $\tilde{M}_{12} = M[\tilde{I}^{-1/2}_{n11}\, \tilde{J}_{n12}]$ and $\tilde{\Sigma}_{n11|2} = \tilde{J}'_{n11|2}\, \tilde{I}^{-1}_{n11}\, \tilde{J}_{n11|2} = \tilde{J}'_{n11}\, \tilde{I}^{-1/2}_{n11}\, \tilde{M}_{12}\, \tilde{I}^{-1/2}_{n11}\, \tilde{J}_{n11}$.

---

[2] The number of observations in the dataset Y could be different from n, say is equal to n_2, $n_2 \neq n$. If the auxiliary estimate $\tilde{\theta}^0_{2n_2}$ obtained from the second dataset satisfies $\sqrt{n_2}(\tilde{\theta}^0_{2n_2} - \theta_{20}) = O_p(1)$, then $\sqrt{n}(\tilde{\theta}^0_{2n_2} - \theta_{20}) = \sqrt{n/n_2}\sqrt{n_2}(\tilde{\theta}^0_{2n_2} - \theta_{20}) = O_p(1)$ provided $n/n_2 = O(1)$, and the arguments that follow remain valid. When a set of estimating functions $D_{n_2 2}(\theta_2)$ for the second dataset is considered, the argument presented here remains valid provided $\sqrt{n_2}D_{n_2 2}(\theta_{20})$ obeys a central limit theorem in addition to the previous conditions on the auxiliary estimate and the sample sizes.

When calculating the standard error of the estimator of θ_1, one needs to take into account the sampling error associated with the first-stage estimator of the parameter θ_2; see Newey and McFadden (1994). This is achieved transparently by the $C(\alpha)$ statistic, because its asymptotic distribution does not depend on the asymptotic distribution of the first-stage estimator. Here, the invariance of the the asymptotic distribution of $PC_1\big(\tilde{\theta}_n^0;\bar{\theta}_{10},W_{n11}\big)$ with respect to the distribution of $\tilde{\theta}_n^0$ is entailed by the orthogonality relation

$$\begin{aligned} J_{12}(\theta_0)'[W_{011}J_{11|2}(\theta_0)] &= J_{12}(\theta_0)'W_{011}\,W_{011}^{-1/2}M\big[W_{011}^{1/2}J_{12}(\theta_0)\big]W_{011}^{1/2}J_{11}(\theta_0) \\ &= \big[W_{011}^{1/2}J_{12}(\theta_0)\big]'M\big[W_{011}^{1/2}J_{12}(\theta_0)\big]W_{011}^{1/2}J_{11}(\theta_0) = 0\,, \end{aligned} \tag{6.10}$$

where $\operatorname*{plim}_{n\to\infty} W_{n11} = W_{011}$. This in turn implies that $\sqrt{n}\tilde{J}_{n11|2}'W_{n11}\,\tilde{D}_{n1}$ is asymptotically uncorrelated with $\sqrt{n}\tilde{J}_{n12}'\tilde{I}_{n11}^{-1}\tilde{D}_{n1}$; see (5.16) - (5.20) for a similar argument.

6.2. Tests based on a two-step GMM estimation

We now consider the case where the condition $m_1 \geq p$ may not hold – so rank conditions for applying a $C(\alpha)$-type test only based on h_1 cannot hold (without other restrictions) – but we have m_2 estimating functions $D_{n2}(\theta;X_n)$ as in (6.2) which can be used to draw inference on θ_2 and account for the uncertainty of θ_2 estimates, where $m_2 \geq p_2$. Further, we suppose here that $D_{n2}(\theta;X_n)$ only depends on θ_2, *i.e.* $D_{n2}(\theta;X_n) = D_{n2}(\theta_2;X_n)$, with $m_1 \geq p_1$ and $m_2 \geq p_2$.

In particular, these assumptions may be based on a system of moment equations

$$\mathsf{E}_\theta \begin{bmatrix} h_1(\theta_1,\theta_2;x_t) \\ h_2(\theta_2;y_t) \end{bmatrix} = 0\,,\ t = 1,\ldots,n, \tag{6.11}$$

where $h_2(\theta_2;y_t)$ is typically used to estimate the nuisance parameter θ_2 and

$$D_{n2}(\theta_2;Y_n) = \frac{1}{n}\sum_{t=1}^{n} h_2(\theta_2;y_t)\,. \tag{6.12}$$

In this context, the sample estimating function is

$$\tilde{D}_n = \begin{bmatrix} \tilde{D}_{n1} \\ \tilde{D}_{n2} \end{bmatrix} = \begin{bmatrix} D_{n1}(\tilde{\theta}_n^0;X_n) \\ D_{n2}(\tilde{\theta}_{2n}^0;Y_n) \end{bmatrix} \tag{6.13}$$

with

$$J(\theta) = [J_{\cdot 1}(\theta), J_{\cdot 2}(\theta)] = \begin{bmatrix} J_{11}(\theta) & J_{12}(\theta) \\ 0_{m_2\times p_1} & J_{22}(\theta) \end{bmatrix}. \tag{6.14}$$

The partitioned Jacobian estimator is then given by

$$\tilde{J}_n = [\tilde{J}_{n\cdot 1}, \tilde{J}_{n\cdot 2}] = \begin{bmatrix} \tilde{J}_{n11} & \tilde{J}_{n12} \\ 0_{m_2\times p_1} & \tilde{J}_{n22} \end{bmatrix}. \tag{6.15}$$

On assuming that the regularity conditions of Proposition **3.1** are satisfied, we can use here the statistic $PC_1\big(\tilde{\theta}_n^0;\,\bar{\theta}_{10},W_n\big)$ defined in (5.15). Further, the form (6.15) yields useful restrictions on the test statistic. We then have

$$PC_1\big(\tilde{\theta}_n^0;\,\bar{\theta}_{10},W_n\big) = n\tilde{D}_n'\,W_n\,\tilde{J}_{n\cdot 1|2}\big(\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{I}_n\,W_n\,\tilde{J}_{n\cdot 1|2}\big)^{-1}\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n \tag{6.16}$$

with

$$\begin{aligned}\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n &= \tilde{J}_{n11}'\,W_{n11}\,\tilde{D}_{n1} \\ &\quad + \big[\tilde{J}_{n11}'\,W_{n12}\,\tilde{D}_{n2} - \tilde{J}_{n\cdot 1}'\,W_n\,\tilde{J}_{n\cdot 2}(\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{J}_{n\cdot 2})^{-1}\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{D}_n\big]\,. \end{aligned} \tag{6.17}$$

In this case, the correction for the estimation of θ_2 is accounted by the two last terms in the above expression for $\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n$.

For moment equations of the form (6.11), it is natural to consider separate weightings for $\tilde{D}_{n1}$ and $\tilde{D}_{n2}$, *i.e.*

$$W_{n12} = W_{n21}' = 0\,. \tag{6.18}$$

On using both conditions (6.15) and (6.18), we see that

$$\begin{aligned}\tilde{J}_{n\cdot 1|2}'\,W_n\,\tilde{D}_n &= \tilde{J}_{n11}'\,W_{n11}\,\tilde{D}_{n1} \\ &\quad -\tilde{J}_{n11}'W_{n11}\tilde{J}_{n12}(\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{J}_{n\cdot 2})^{-1}\big[\tilde{J}_{n12}'\,W_{n11}\,\tilde{D}_{n1} + \tilde{J}_{n22}'\,W_{n22}\,\tilde{D}_{n2}\big]\,, \end{aligned} \tag{6.19}$$

$$\tilde{J}_{n\cdot 2}'\,W_n\,\tilde{J}_{n\cdot 2} = \tilde{J}_{n12}'\,W_{n11}\,\tilde{J}_{n12} + \tilde{J}_{n22}'\,W_{n22}\,\tilde{J}_{n22}\,. \tag{6.20}$$

Again the asymptotic distribution of the test statistic $PC_1\big(\tilde{\theta}_n^0;\,\bar{\theta}_{10},W_n\big)$ is $\chi^2(p_1)$ under the null hypothesis $H_0 : \theta_1 = \bar{\theta}_{10}$, irrespective of the asymptotic distribution of $\tilde{\theta}_{2n}^0$.

7. Conclusion

In this paper, we have introduced a comprehensive $C(\alpha)$ statistic based on estimating functions (or GMM setups). As in Smith (1987*a*), the null hypothesis is specified in terms of a general possibly nonlinear constraint, rather than a restriction fixing a parameter subvector. The proposed procedure allows for general forms of serial dependence and heteroskedasticity, and can be implemented using any root-*n* consistent restricted estimator. A detailed derivation of the asymptotic null distribution of the statistic was provided under weak regularity conditions.

The proposed generalized $C(\alpha)$-type statistic includes earlier ones as special cases, as well as a wide spectrum of new ones. A number of important special cases of the extended test statistic were discussed in detail. These include testing whether a parameter subvector has a given value – for which we give a number of alternative forms and special cases – and the important problem of accounting for parameter uncertainty in two-stage procedures.

Appendix

A. Proofs

PROOF OF PROPOSITION **3.1** To simplify notation, we shall assume throughout that $\omega \in \mathscr{D}_J$ (an event with probability 1) and drop the symbol ω from the random variables considered. In order to obtain the asymptotic null distribution of the generalized $C(\alpha)$ statistic defined in (2.14), we first need to show that $P(\tilde{\theta}_n^0)$ and $J_n(\tilde{\theta}_n^0)$ converge in probability to $P(\theta_0)$ and $J(\theta_0)$ respectively. The consistency of $P(\tilde{\theta}_n^0)$, *i.e.*

$$\operatorname*{plim}_{n\to\infty} [P(\tilde{\theta}_n^0) - P(\theta_0)] = 0\,, \tag{A.1}$$

follows simply from the consistency of $\tilde{\theta}_n^0$ [Assumption 3.9] and the continuity of $P(\theta)$ at θ_0 [Assumption 3.7]. Further, by Assumption 3.8, since $P(\theta)$ is continuous in open neighborhood of θ_0, we also have

$$\text{rank}\,[\tilde{P}_n] = \text{rank}\,[P(\theta_0)] = p_1\,. \tag{A.2}$$

Consider now $J_n(\tilde{\theta}_n^0)$. By Assumption 3.5, for any $\varepsilon > 0$ and $\varepsilon_1 > 0$, we can choose $\delta_1 := \delta(\varepsilon_1, \varepsilon) > 0$ and a positive integer $n_1(\varepsilon, \delta_1)$ such that: (i) $U_J(\delta_1, \varepsilon, \theta_0) \leq \varepsilon_1/2$, and (ii) $n > n_1(\varepsilon, \delta_1)$ entails

$$\mathsf{P}\,[\Delta_n(\theta_0, \delta) > \varepsilon] = \mathsf{P}\big[\{\omega : \Delta_n(\theta_0, \delta, \omega) > \varepsilon\}\big] \leq U_J(\delta_1, \varepsilon, \theta_0) \leq \varepsilon_1/2\,.$$

Further, by the consistency of $\tilde{\theta}_n^0$ [Assumption 3.9], we can choose $n_2(\varepsilon_1, \delta_1)$ such that $n > n_2(\varepsilon_1, \delta_1)$ entails $\mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1\big] \geq 1 - (\varepsilon_1/2)$. Then, using the Boole-Bonferroni inequality, we have for $n > \max\{n_1(\varepsilon, \delta_1), n_2(\varepsilon_1, \delta_1)\}$:

$$\begin{aligned}
\mathsf{P}\big[\|J_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] &\geq \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \|J_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] \\
&\geq \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \Delta_n(\theta_0, \delta_1) \leq \varepsilon\big] \\
&\geq 1 - \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| > \delta_1\big] - \mathsf{P}\big[\Delta_n(\theta_0, \delta_1) > \varepsilon\big] \\
&\geq 1 - (\varepsilon_1/2) - (\varepsilon_1/2) = 1 - \varepsilon_1\,.
\end{aligned}$$

Thus,

$$\liminf_{n\to\infty} \mathsf{P}\big[\|J_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] \geq 1 - \varepsilon_1\,,\ \text{for all } \varepsilon > 0,\ \varepsilon_1 > 0\,,$$

hence

$$\lim_{n\to\infty} \mathsf{P}\big[\|J_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] = 1\,,\ \text{for all } \varepsilon > 0\,, \tag{A.3}$$

or, equivalently,

$$\operatorname*{plim}_{n\to\infty} \big[J_n(\tilde{\theta}_n^0) - J(\theta_0)\big] = 0\,. \tag{A.4}$$

By Assumption 3.4, we can write [setting $0/0 = 0$] :

$$\|\sqrt{n}\,\big[D_n(\tilde{\theta}_n^0) - D_n(\theta_0)\big] - J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\| = \sqrt{n}\,\|R_n(\tilde{\theta}_n^0, \theta_0)\|$$

$$= \frac{\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0 - \theta_0\|$$

where

$$\frac{\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|} \leq r_n(\delta,\, \theta_0) \text{ when } \tilde{\theta}_n^0 \in N_0 \text{ and } \|\tilde{\theta}_n^0 - \theta_0\| \leq \delta$$

and $\limsup_{n\to\infty} \mathsf{P}\,[r_n(\delta,\, \theta_0) > \varepsilon] < U_D(\delta,\, \varepsilon,\, \theta_0)$. Thus, for any $\varepsilon > 0$ and $\delta > 0$, we have:

$$\begin{aligned} \mathsf{P}\left[\frac{\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|} \leq \varepsilon\right] &\geq \mathsf{P}\big[r_n(\delta,\, \theta_0) \leq \varepsilon\,, \tilde{\theta}_n^0 \in N_0 \text{ and } \|\tilde{\theta}_n^0 - \theta_0\| \leq \delta\big] \\ &\geq 1 - \mathsf{P}\big[r_n(\delta,\, \theta_0) > \varepsilon\big] - \mathsf{P}\big[\,\tilde{\theta}_n^0 \notin N_0 \text{ or } \|\tilde{\theta}_n^0 - \theta_0\| > \delta\big] \end{aligned}$$

hence, using the consistency of $\tilde{\theta}_n^0$,

$$\begin{aligned} \liminf_{n\to\infty} \mathsf{P}\left[\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|/\|\tilde{\theta}_n^0 - \theta_0\| \leq \varepsilon\right] &\geq 1 - \limsup_{n\to\infty} \mathsf{P}\,[r_n(\delta,\, \theta_0) > \varepsilon\,] \\ &\quad - \limsup_{n\to\infty} \mathsf{P}\big[\,\tilde{\theta}_n^0 \notin N_0 \text{ or } \|\tilde{\theta}_n^0 - \theta_0\| > \delta\big] \\ &\geq 1 - U_D(\delta,\, \varepsilon,\, \theta_0)\,. \end{aligned}$$

Since $\lim_{\delta\downarrow 0} U_D(\delta,\, \varepsilon,\, \theta_0) = 0$, it follows that $\lim_{n\to\infty} \mathsf{P}\big[\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|/\|\tilde{\theta}_n^0 - \theta_0\| \leq \varepsilon\big] = 1$ for any $\varepsilon > 0$, or equivalently,

$$\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|/\|\tilde{\theta}_n^0 - \theta_0\| \xrightarrow[n\to\infty]{\mathsf{p}} 0\,.$$

Since $\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)$ is asymptotically bounded in probability (by Assumption 3.9), this entails:

$$\sqrt{n}\,\|R_n(\tilde{\theta}_n^0,\, \theta_0)\| = \frac{\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0 - \theta_0\| \xrightarrow[n\to\infty]{\mathsf{p}} 0 \tag{A.5}$$

and

$$\|\sqrt{n}\,[D_n(\tilde{\theta}_n^0) - D_n(\theta_0)] - J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\| \xrightarrow[n\to\infty]{\mathsf{p}} 0\,. \tag{A.6}$$

By Taylor's theorem and assumptions 3.7 - 3.8, we also have the expansion:

$$\psi(\theta) = \psi(\theta_0) + P(\theta_0)(\theta - \theta_0) + R_2(\theta,\, \theta_0)\,, \tag{A.7}$$

for $\theta \in N \subseteq N_0 \cap V_0$, where N is a non-empty open neighborhood of θ_0 and

$$\lim_{\theta\to\theta_0} \|R_2(\theta,\, \theta_0)\,\|/\|\,\theta - \theta_0\| = 0\,,$$

i.e., $R_2(\theta,\,\theta_0)=o(\|\theta-\,\theta_0\|)$, so that, using Assumption 3.10,

$$\begin{aligned}\sqrt{n}P(\theta_0)(\tilde{\theta}_n^0-\theta_0) &= \sqrt{n}[\psi(\tilde{\theta}_n^0)-\psi(\theta_0)]-\sqrt{n}R_2(\tilde{\theta}_n^0,\,\theta_0)\\ &= -\sqrt{n}R_2(\tilde{\theta}_n^0,\,\theta_0)\end{aligned} \tag{A.8}$$

for $\tilde{\theta}_n^0\in N$, and

$$\|\sqrt{n}P(\theta_0)(\tilde{\theta}_n^0-\theta_0)\|=\frac{\|R_2(\tilde{\theta}_n^0,\,\theta_0)\|}{\|\tilde{\theta}_n^0-\theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0-\theta_0\|\xrightarrow[n\to\infty]{\mathsf{p}}0\,. \tag{A.9}$$

By (A.2) and (A.4) jointly with the assumptions 3.3, 3.6, 3.7, 3.8, 3.11 and 3.12, we have:

$$\text{rank}\,[\tilde{P}_n]=p_1\,,\quad \text{rank}\,[\tilde{J}_n]=p\,,\quad \text{rank}\,[\tilde{I}_n]=m,\quad \text{rank}\,[W_n]=m\,, \tag{A.10}$$

so the matrices $\tilde{J}_n$, $\tilde{I}_n$, and W_n all have full column rank. Since $\underset{n\to\infty}{\text{plim}}\,\tilde{P}_n=P(\theta_0)$ and $\underset{n\to\infty}{\text{plim}}\,\tilde{J}_n=J(\theta_0)$, we can then write:

$$\begin{aligned}\underset{n\to\infty}{\text{plim}}\left[\tilde{J}_n'W_n\tilde{J}_n\right]^{-1} &= \left[J(\theta_0)'W_0J(\theta_0)\right]^{-1},\ \underset{n\to\infty}{\text{plim}}\,\tilde{Q}_n=Q(\theta_0)\,,\\ \underset{n\to\infty}{\text{plim}}\,\tilde{Q}_n\tilde{J}_n &= \underset{n\to\infty}{\text{plim}}\,\tilde{Q}_nJ(\theta_0)=Q(\theta_0)J(\theta_0)=P(\theta_0)\,,\end{aligned}$$

where $\tilde{Q}_n:=\tilde{Q}[W_n]=\tilde{P}_n[\tilde{J}_n'W_n\tilde{J}_n]^{-1}\tilde{J}_n'W_n$. Then, using (A.6) and (A.9), it follows that:

$$\begin{aligned}\underset{n\to\infty}{\text{plim}} & \left\{\sqrt{n}\tilde{Q}_nD_n(\tilde{\theta}_n^0)-\sqrt{n}Q(\theta_0)D_n(\theta_0)\right\}\\ &= \underset{n\to\infty}{\text{plim}}\left\{\sqrt{n}\tilde{Q}_nD_n(\tilde{\theta}_n^0)-Q(\theta_0)\sqrt{n}D_n(\theta_0)\right\}-\underset{n\to\infty}{\text{plim}}\left\{P(\theta_0)\sqrt{n}(\tilde{\theta}_n^0-\theta_0)\right\}\\ &= \underset{n\to\infty}{\text{plim}}\left\{\tilde{Q}_n\left[\sqrt{n}[D_n(\tilde{\theta}_n^0)-D_n(\theta_0)]-J(\theta_0)\sqrt{n}(\tilde{\theta}_n^0-\theta_0)\right]\right\}\\ &\quad +\underset{n\to\infty}{\text{plim}}\left\{\left[\tilde{Q}_n-Q(\theta_0)\right]\sqrt{n}D_n(\theta_0)+\left[\tilde{Q}_nJ(\theta_0)-P(\theta_0)\right]\sqrt{n}(\tilde{\theta}_n^0-\theta_0)\right\}\\ &= \underset{n\to\infty}{\text{plim}}\left\{\tilde{Q}_n\left[\sqrt{n}[D_n(\tilde{\theta}_n^0)-D_n(\theta_0)]-J(\theta_0)\sqrt{n}(\tilde{\theta}_n^0-\theta_0)\right]\right\}=0\,.\end{aligned}$$

We conclude that the asymptotic distribution of $\sqrt{n}\tilde{Q}_nD_n(\tilde{\theta}_n^0)$ is the same as the one of $Q(\theta_0)\sqrt{n}D_n(\theta_0)$, namely (by Assumption 3.2) a $\mathrm{N}\left[0,V_\psi(\theta_0)\right]$ distribution where

$$V_\psi(\theta)=Q(\theta)I(\theta)Q(\theta)'$$

and $V_\psi(\theta_0)$ has rank $p_1=\text{rank}\,[Q(\theta_0)]=\text{rank}[P(\theta_0)]$. Consequently, the estimator

$$\tilde{V}_\psi(\tilde{\theta}_n^0)=\tilde{Q}_n\tilde{I}_n\tilde{Q}_n' \tag{A.11}$$

converges to $V_{\psi}(\theta_0)$ in probability and, by (A.10),

$$\operatorname{rank}\left[\tilde{V}_{\psi}(\tilde{\theta}_n^0)\right] = p_1 . \tag{A.12}$$

Thus the test criterion

$$PC(\tilde{\theta}_n^0;\, \psi, W_n) = nD_n(\tilde{\theta}_n^0;\, Z_n)'\tilde{Q}[W_n]'\left\{\tilde{Q}[W_n]\,\tilde{I}_n\,\tilde{Q}[W_n]'\right\}^{-1}\tilde{Q}[W_n]\,D_n(\tilde{\theta}_n^0;\, Z_n)$$

has an asymptotic $\chi^2(p_1)$ distribution. □

PROOF OF PROPOSITION **3.2** Consider the (non-empty) open neighborhood $N = N_1 \cap N_2$ of θ_0. For any $\theta \in N$ and $\omega \in \mathscr{Z}$, we can write

$$\begin{aligned} \|J(\theta) - J(\theta_0)\| &\leq \|J_n(\theta,\, \omega) - J(\theta)\| + \|J_n(\theta_0,\, \omega) - J(\theta_0)\| \\ &\quad + \|J_n(\theta,\, \omega) - J_n(\theta_0,\, \omega)\| \\ &\leq 2\sup_{\theta \in N} \|J_n(\theta,\, \omega) - J(\theta)\| + \|J_n(\theta,\, \omega) - J_n(\theta_0,\, \omega)\| \end{aligned}$$

By Assumption 3.14(b), we have

$$\operatorname*{plim}_{n\to\infty}\Big(\sup_{\theta \in N}\|J_n(\theta,\, \omega) - J(\theta)\|\Big) \leq \operatorname*{plim}_{n\to\infty}\Big(\sup_{\theta \in N_2}\|J_n(\theta,\, \omega) - J(\theta)\|\Big) = 0$$

and we can find a subsequence $\{J_{n_t}(\theta, \omega) : t = 1, 2, \dots\}$ of $\{J_n(\theta, \omega) : n = 1, 2, \dots\}$ such that

$$\sup_{\theta \in N}\{\|J_{n_t}(\theta,\, \omega) - J(\theta)\|\} \underset{t\to\infty}{\longrightarrow} 0 \quad a.s.$$

Let

$$CS = \Big\{\omega \in \mathscr{Z} : \lim_{t\to\infty}\Big(\sup_{\theta \in N}\|J_{n_t}(\theta,\, \omega) - J(\theta)\|\Big) = 0\Big\}$$

and $\varepsilon > 0$. By definition, $\mathsf{P}[\omega \in CS] = 1$. For $\omega \in CS$, we can choose $t_0(\varepsilon,\, \omega)$ such that

$$t \geq t_0(\varepsilon, \omega) \Rightarrow 2\sup_{\theta \in N}\{\|J_{n_t}(\theta,\, \omega) - J(\theta)\|\} < \varepsilon/2\ .$$

Further, since $J_n(\theta,\, \omega)$ is continuous in θ at θ_0, we can find $\delta(n,\, \omega) > 0$ such that

$$\|\theta - \theta_0\| < \delta(n, \omega) \Rightarrow \|J_n(\theta,\, \omega) - J_n(\theta_0,\, \omega)\| < \varepsilon/2\ .$$

Thus, taking $t_0 = t_0(\varepsilon,\, \omega)$ and $n = n_{t_0}$, we find that $\|\theta - \theta_0\| < \delta(n_{t_0}, \omega)$ implies

$$\|J(\theta) - J(\theta_0)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\ .$$

In other words, for any $\varepsilon > 0$, we can choose $\delta = \delta(n_{t_0}, \varepsilon) > 0$ such that

$$\|\theta - \theta_0\| < \delta \Rightarrow \|J(\theta) - J(\theta_0)\| < \varepsilon \,,$$

and the function $J(\theta)$ must be continuous at θ_0. Part (a) of the Proposition is established.

Set $\overline{\Delta}_n(N_2, \omega) := \sup\{\|J_n(\theta, \omega) - J(\theta)\| : \theta \in N_2\}$. To get Assumption 3.5, we note that

$$\begin{aligned} \Delta_n(\theta_0, \delta, \omega) \quad &: \; = \sup\{\|J_n(\theta, \omega) - J(\theta_0)\| : \theta \in N_2 \text{ and } 0 \le \|\theta - \theta_0\| \le \delta\} \\ &\le \; \overline{\Delta}_n(N_2, \omega) \end{aligned}$$

for any $\delta > 0$, hence, by Assumption 3.14(b),

$$\begin{aligned} \limsup_{n\to\infty} \mathsf{P}\big[\{\omega : \Delta_n(\theta_0, \delta, \omega) > \varepsilon\}\big] \quad &\le \quad \limsup_{n\to\infty} \mathsf{P}\left[\left\{\omega : \overline{\Delta}_n(N_2, \omega) > \varepsilon\right\}\right] \\ &\le \quad U_J(\delta, \varepsilon, \theta_0) \end{aligned}$$

for any function $U_J(\delta, \varepsilon, \theta_0)$ that satisfies the conditions of Assumption 3.5. The latter thus holds with V_0 any non-empty open neighborhood of θ_0 such that $V_0 \subseteq N_2$.

To obtain 3.4, we note that 3.14 entails $D_n(\theta, \omega)$ is continuously differentiable in an open neighborhood of θ_0 for all $\omega \in \mathscr{D}_J$, so that we can apply Taylor's formula for a function of several variables [see Edwards (1973, Section II.7)] to each component of $D_n(\theta, \omega)$: for all θ in an open neighborhood U of θ_0 (with $U \subseteq N_2$), we can write

$$\begin{aligned} D_{in}(\theta, \omega) \quad &= \quad D_{in}(\theta_0, \omega) + J_n\big(\bar{\theta}_n^{\,i}(\omega), \omega\big)_{i\cdot}(\theta - \theta_0) \\ &= \quad D_{in}(\theta_0, \omega) + J(\theta_0)_{i\cdot}(\theta - \theta_0) + R_{in}\big(\bar{\theta}_n^{\,i}(\omega), \theta_0, \omega\big)\,,\; i = 1, \ldots, m, \end{aligned}$$

where $J_n(\theta, \omega)_{i\cdot}$ and $J(\theta)_{i\cdot}$ are the i-th rows of $J_n(\theta, \omega)$ and $J(\theta)$ respectively,

$$R_{in}\big(\bar{\theta}_n^{\,i}(\omega), \theta_0, \omega\big) = \big[J_n(\bar{\theta}_n^{\,i}(\omega), \omega)_{i\cdot} - J(\theta_0)_{i\cdot}\big](\theta - \theta_0)$$

and $\bar{\theta}_n^{\,i}(\omega)$ belongs to the line joining θ and θ_0. Further, for $\theta \in U$,

$$\begin{aligned} \big|R_{in}\big(\bar{\theta}_n^{\,i}(\omega), \theta_0, \omega\big)\big| \quad &\le \quad \big\|J_n(\bar{\theta}_n^{\,i}(\omega), \omega)_{i\cdot} - J(\theta_0)_{i\cdot}\big\| \, \|\theta - \theta_0\| \\ &\le \quad \big\|J_n(\bar{\theta}_n^{\,i}(\omega), \omega) - J(\theta_0)\big\| \, \|\theta - \theta_0\| \\ &\le \quad \|\theta - \theta_0\| \sup\{\|J_n(\theta, \omega) - J(\theta)\| : \theta \in N_2\}\,,\; i = 1, \ldots, m, \end{aligned}$$

hence, on defining $N_0 = U$,

$$R_n(\theta, \theta_0, \omega) = \big[R_{1n}\big(\bar{\theta}_n^{\,1}(\omega), \theta_0, \omega\big), \ldots, R_{mn}\big(\bar{\theta}_n^{\,m}(\omega), \theta_0, \omega\big)\big]',$$

we see that

$$\|R_n(\theta, \theta_0, \omega)\| \quad \le \quad \sum_{1=1}^{m} \big|R_{in}\big(\bar{\theta}_n^{\,i}(\omega), \theta_0, \omega\big)\big|$$

$$\leq \quad m\,\|\theta - \theta_0\| \sup_{\theta \in N_2} \{\|J_n(\theta, \omega) - J(\theta)\|\}$$

and

$$\begin{aligned} r_n(\delta, \theta_0, \omega) \quad &: \quad = \sup\left\{ \frac{\|R_n(\theta, \theta_0, \omega)\|}{\|\theta - \theta_0\|} : \theta \in N_0 \text{ and } 0 < \|\theta - \theta_0\| \leq \delta \right\} \\ &\leq \quad m \sup\{\|J_n(\theta, \omega) - J(\theta)\| : \theta \in N_2\} \end{aligned}$$

Thus $r_n(\delta, \theta_0, \omega) \xrightarrow[n\to\infty]{\mathsf{p}} 0$ and

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : r_n(\delta, \theta_0, \omega) > \varepsilon\}\big] \leq U_D(\delta, \varepsilon, \theta_0) \tag{A.13}$$

must hold for any function that satisfies the conditions of Assumption 3.4. This completes the proof. □

References

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Andrews, D. W. K. and Monahan, J. C. (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica* **60**, 953–966.

Bartoo, J. B. and Puri, P. S. (1967), 'On optimal asymptotic tests of composite statistical hypotheses', *The Annals of Mathematical Statistics* **38**(6), 1845–1852.

Basawa, I. V. (1985), Neyman-Le Cam tests based on estimating functions, *in* L. Le Cam and R. A. Olshen, eds, 'Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer', Wadsworth, Belmont, California, pp. 811–825.

Basawa, I. V., Godambe, V. P. and Taylor, R. L., eds (1997), *Selected Proceedings of the Symposium on Estimating Functions*, Vol. 32 of *IMS Lecture Notes Monograph Series*, Institute of Mathematical Statistics, Hayward, California.

Bera, A. and Bilias, Y. (2001), 'Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results', *Journal of Statistical Planning and Inference* **97**, 9–44.

Bera, A. K. and Yoon, M. J. (1993), 'Specification testing with locally misspecified alternatives', *Econometric theory* **9**(04), 649–658.

Berger, A. and Wallenstein, S. (1989), 'On the theory of C_α-tests', *Statistics and Probability Letters* **7**, 419–424.

Bernshtein, A. V. (1976), 'On optimal asymptotic tests for composite hypotheses under non-standard conditions', *Theory of Probability and its Applications* **21**, 34–47.

Bernshtein, A. V. (1978), 'On optimal asymptotic tests of homogeneity', *Theory of Probability and its Applications* **22**, 377–383.

Bernshtein, A. V. (1980*a*), 'On the construction of majorizing tests', *Theory of Probability and its Applications* **25**, 16–26.

Bernshtein, A. V. (1980*b*), 'On verifying composite hypotheses with nuisance parameters in the multivariate case', *Theory of Probability and its Applications* **25**, 287–298.

Bernshtein, A. V. (1981), 'Asymptotically similar criteria', *Journal of Soviet Mathematics* **17**(3), 1825–1857.

Bhat, B. R. and Nagnur, B. N. (1965), 'Locally asymptotically most stringent tests and lagrangian multiplier tests of linear hypotheses', *Biometrika* **52**(3-4), 459–468.

Bontemps, C. and Meddahi, N. (2012), 'Testing distributional assumptions: A GMM approach', *Journal of Applied Econometrics* **27**, 978–1012.

Bühler, W. J. and Puri, P. S. (1966), 'On optimal asymptotic tests of composite hypotheses with several constraints', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **5**(1), 71–88.

Chant, D. (1974), 'On asymptotic tests of composite hypotheses in nonstandard conditions', *Biometrika* **61**(2), 291–298.

Chaudhuri, S. and Zivot, E. (2011), 'A new method of projection-based inference in GMM with weakly identified nuisance parameters', *Journal of Econometrics* **164**(2), 239–251.

Chibisov, D. M. (1973), Asymptotic expansions for Neyman's $C(\alpha)$ tests, *in* 'Proceedings of the Second Japan-USSR Symposium on Probability Theory', Springer, Berlin, pp. 16–45.

Cushing, M. J. and McGarvey, M. G. (1999), Covariance matrix estimation, *in* L. Mátyás, ed., 'Generalized Method of Moments Estimation', Cambridge University Press, Cambridge, U.K., chapter 3, pp. 63–95.

Dagenais, M. G. and Dufour, J.-M. (1991), 'Invariance, nonlinear models and asymptotic tests', *Econometrica* **59**, 1601–1615.

Davidson, R. and MacKinnon, J. G. (1991), 'Artificial regressions and $C(\alpha)$ tests', *Economics Letters* **35**, 149–153.

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Dufour, J.-M. and Dagenais, M. G. (1992), 'Nonlinear models, rescaling and test invariance', *Journal of Statistical Planning and Inference* **32**, 111–135.

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

Durbin, J. (1960), 'Estimation of parameters in time series regression models', *Journal of the Royal Statistical Society, Series A* **22**, 139–153.

Edwards, C. H. (1973), *Advanced Calculus of Several Variables*, Dover Publications, New York.

Foutz, R. V. (1976), 'On the consistency of locally asymptotically most stringent tests', *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **4**(2), 211–219.

Godambe, V. P. (1960), 'An optimum property of regular maximum likelihood estimation', *The Annals of Mathematical Statistics* **31**, 1208–1212. Ackowledgement 32 (1960), 1343.

Godambe, V. P., ed. (1991), *Estimating Functions*, Clarendon Press, Oxford, U.K.

Gong, G. and Samaniego, F. J. (1981), 'Pseudo maximum likelihood estimation: Theory and applications', *The Annals of Statistics* **13**, 861–869.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K. Translated by Quang Vuong.

Hall, A. R. (2000), 'Covariance matrix estimation and the power of the overidentifying restrictions test', *Econometrica* **68**, 1517–1528.

Hall, A. R. (2004), *Generalized Method of Moments*, Advanced Texts in Econometrics, Oxford University Press, Oxford, U.K.

Hall, W. J. and Mathiason, D. J. (1990), 'On large-sample estimation and testing in parametric models', *International Statistical Review* **58**(1), 77–97.

Hansen, B. E. (1992), 'Consistent covariance matrix estimation for dependent heterogeneous processes', *Econometrica* **60**, 967–972.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Heyde, C. C. (1997), *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*, Springer Series in Statistics, Springer-Verlag, New York.

Jaggia, S. and Trivedi, P. K. (1994), 'Joint and separate score tests for state dependence and unobserved heterogeneity', *Journal of Econometrics* **60**(1), 273–291.

Kiefer, N. M., Vogelsang, T. and Bunzel, H. (2000), 'Simple robust testing of regression hypotheses', *Econometrica* **68**, 695–714.

Kiefer, N. M. and Vogelsang, T. J. (2002), 'Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation', *Econometrica* **70**, 2093–2095.

Kiefer, N. M. and Vogelsang, T. J. (2005), 'A new asymptotic theory for heteroskedasticity-autocorrelation robust tests', *Econometric Theory* **21**(6), 1130–1164.

Kocherlakota, S. and Kocherlakota, K. (1991), 'Neyman's $C(\alpha)$ test and Rao's efficient score test for composite hypotheses', *Statistics and Probability Letters* **11**, 491–493.

Le Cam, L. (1956), On the asymptotic theory of estimation and testing hypotheses, *in* J. Neyman, ed., 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability', The University of California Press, Berkeley, California.

Le Cam, L. and Traxler, R. (1978), 'On the asymptotic behavior of mixtures of Poisson distributions', *Probability Theory and Related Fields* **44**(1), 1–45.

Moran, P. A. P. (1970), 'On asymptotically optimal tests of composite hypotheses', *Biometrika* **57**(1), 47–55.

Moran, P. A. P. (1973), 'Asymptotic properties of homogeneity tests', *Biometrika* **60**(1), 79–85.

Murphy, K. M. and Topel, R. H. (1985), 'Estimation and inference in two-step econometric models', *Journal of Business and Economic Statistics* **3**, 370–379.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume 4', North-Holland, Amsterdam, chapter 36, pp. 2111–2245.

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

Newey, W. K. and West, K. D. (1987*b*), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Neyman, J. (1959), Optimal asymptotic tests of composite statistical hypotheses, *in* U. Grenander, ed., 'Probability and Statistics, the Harald Cramér Volume', Almqvist and Wiksell, Uppsala, Sweden, pp. 213–234.

Neyman, J. (1979), 'C(α) tests and their use', *Sankhyā: The Indian Journal of Statistics, Series A* **41**, 1–21.

Pagan, A. (1984), 'Econometric issues in the analysis of regressions with generated regressors', *International Economic Review* **25**, 221–247.

Pagan, A. (1986), 'Two stage and related estimators and their applications', *Review of Economic Studies* **53**, 517–538.

Pal, C. (2003), 'Higher order $C(\alpha)$ tests with applications to mixture models', *Journal of Statistical Planning and Inference* **113**(1), 179–187.

Paul, S. R. and Barnwal, R. K. (1990), 'Maximum likelihood estimation and a $C(\alpha)$ test for a common intraclass correlation', *Journal of the Royal Statistical Society. Series D (The Statistician)* **39**(1), pp. 19–24.

Rao, B. L. S. P. (1996), 'Optimal asymptotic tests of composite hypotheses for continuous time stochastic processes', *Sankhyā: The Indian Journal of Statistics, Series A* **58**, 8–24.

Ray, R. M. (1974), 'Maxmin $C(\alpha)$ tests against two-sided alternatives', *The Annals of Statistics* **2**(6), 1175–1188.

Ronchetti, E. (1987), 'Robust $C(\alpha)$-type tests for linear models', *Sankhyā Series A* **49**, 1–16.

Singh, A. C. and Zhurbenko, I. G. (1975), 'The power of the optimal asymptotic tests of composite statistical hypotheses', *Proceedings of the National Academy of Sciences* **72**(2), 577–580.

Small, C. G. and McLeish, D. L. (1994), *Hilbert Space Methods in Probability and Statistical Inference*, John Wiley and Sons, New York.

Smith, R. J. (1987*a*), 'Alternative asymptotically optimal tests and their application to dynamic specification', *Review of Economic Studies* **LIV**, 665–680.

Smith, R. J. (1987*b*), 'Testing the normality assumption in multivariate simultaneous limited dependent variable models', *Journal of Econometrics* **34**, 105–123.

Tarone, R. E. (1979), 'Testing the goodness of fit of the binomial distribution', *Biometrika* **66**(3), 585–590.

Tarone, R. E. (1985), 'On heterogeneity tests based on efficient scores', *Biometrika* **72**(1), 91–95.

Tarone, R. E. and Gart, J. J. (1980), 'On the robustness of combined tests for trends in proportions', *Journal of the American Statistical Association* **75**(369), 110–116.

Vorob'ev, L. S. and Zhurbenko, I. G. (1979), 'Bounds for $C(\alpha)$-tests and their applications', *Theory of Probability and its Applications* **24**, 253–268.

Wang, P. C. C. (1981), 'Robust asymptotic tests of statistical hypotheses involving nuisance parameters', *The Annals of Statistics* **9**(5), 1096–1106.

Wang, P. C. C. (1982), 'On the computation of a robust version of the optimal $C(\alpha)$ test', *Communications in Statistics-Simulation and Computation* **11**(3), 273–284.

Wooldridge, J. M. (1990), 'A unified approach to robust, regression-based specification tests', *Econometric Theory* **6**, 17–43.