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


* 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 Universidad 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 (0) 41 41 75 131; FAX: 33 (0) 41 75 13 80. 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  

2. Generalized $C(\alpha)$ statistic  

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

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

5. Testing a subvector  

6. Two-stage procedures  
   6.1. Tests based on general two-step estimation  
   6.2. Tests based on a two-step GMM estimation  

7. Conclusion  

A. Proofs
### List of propositions and theorems

<table>
<thead>
<tr>
<th>Assumption 3.1</th>
<th>Existence of score-type functions</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption 3.2</td>
<td>Score asymptotic normality</td>
<td>5</td>
</tr>
<tr>
<td>Assumption 3.3</td>
<td>Non-singularity of the score variance</td>
<td>5</td>
</tr>
<tr>
<td>Assumption 3.4</td>
<td>Score expansion</td>
<td>5</td>
</tr>
<tr>
<td>Assumption 3.5</td>
<td>Consistent estimator of $J(\theta_0)$</td>
<td>6</td>
</tr>
<tr>
<td>Assumption 3.6</td>
<td>Asymptotic score non-degeneracy</td>
<td>6</td>
</tr>
<tr>
<td>Assumption 3.7</td>
<td>Restriction differentiability</td>
<td>6</td>
</tr>
<tr>
<td>Assumption 3.8</td>
<td>Restriction rank</td>
<td>6</td>
</tr>
<tr>
<td>Assumption 3.9</td>
<td>Estimator $\sqrt{n}$ convergence</td>
<td>6</td>
</tr>
<tr>
<td>Assumption 3.10</td>
<td>Restricted estimator</td>
<td>7</td>
</tr>
<tr>
<td>Assumption 3.11</td>
<td>Consistent estimator of score covariance matrix</td>
<td>7</td>
</tr>
<tr>
<td>Assumption 3.12</td>
<td>Weight matrix convergence</td>
<td>7</td>
</tr>
<tr>
<td>Proposition 3.1</td>
<td>Asymptotic distribution of generalized $C(\alpha)$ statistic</td>
<td>7</td>
</tr>
<tr>
<td>Assumption 3.13</td>
<td>Score differentiability</td>
<td>7</td>
</tr>
<tr>
<td>Assumption 3.14</td>
<td>Score derivative uniform convergence</td>
<td>8</td>
</tr>
<tr>
<td>Proposition 3.2</td>
<td>Sufficiency of score Jacobian continuity and uniform convergence</td>
<td>8</td>
</tr>
<tr>
<td>Proof of Proposition 3.1</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Proof of Proposition 3.2</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>
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.


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 (1987a) 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. \textit{Generalized} \( C(\alpha) \) \textit{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

\[
D_n(\theta; Z_n) \xrightarrow{p} D_\infty(\theta; \theta_0) \quad (2.1)
\]

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

\[
M_n(\theta, W_n) = D_n(\theta; Z_n)' W_n D_n(\theta; Z_n) \quad (2.2)
\]


A common assumption in such contexts consists in assuming that

\[
E_{\theta_0}[D_n(\theta_0; Z_n)] = 0 \quad (2.3)
\]
where $E_\theta[\cdot]$ represents the expected value under any data distribution such that $\theta$ 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:

$$\sqrt{n}D_n(\theta_0; Z_n) \xrightarrow{n \to \infty} N\left[0, I(\theta_0)\right],$$  

(2.4)

$$J_n(\theta_0; Z_n) = \frac{\partial D_n(\theta_0; Z_n)}{\partial \theta'} \xrightarrow{p \to \infty} J(\theta_0),$$  

(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

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

(2.6)

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

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

(2.7)

has full row-rank $p_1$ (at least in an open neighborhood of $\theta_0$).

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

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

(2.8)

we may use the standard formula

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

(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

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

(2.10)

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

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

(2.11)
where $\hat{h}(\theta) = \frac{1}{n} \sum_{t=1}^{n} h(\theta; z_t)$, 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 $\{h(\theta; z_t)\}_{t=1}^{\infty}$:

$$\hat{I}_n(\theta) = \frac{1}{n} \sum_{j=-n+1}^{n-1} \tilde{k}(j/B_n) \hat{I}_n(j, \theta)$$

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

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


We now consider the problem of formulating a test statistic for $H_0$ using a general restricted estimator of $\theta_0$. 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:

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

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

$$\bar{Q}[W_n] := \tilde{P}_n(\tilde{I}_n W_n \tilde{I}_n)\tilde{I}_n W_n,$$

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

$$\plim_{n \to \infty} W_n = W_0, \quad \det(W_0) \neq 0.$$
3. DISTRIBUTION OF THE GENERALIZED C(α) STATISTIC

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

\[ s(\tilde{\theta}^0_n; W_n) \xrightarrow{L} \mathcal{N} \left[ 0, Q(\theta_0) I(\theta_0) Q(\theta_0)' \right] \]

where

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

and \( \text{rank} [Q(\theta_0)] = p_1 \). This suggests the following generalized C(α) criterion:

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

(2.14)

where \( \tilde{D}_n = D_n(\theta^0_n; Z_n) \) and \( \tilde{I}_n = I_n(\theta^0_n) \). We show in Section 3 that the asymptotic distribution of \( PC(\tilde{\theta}^0_n; \psi, W_n) \) 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(α) statistic

In this section, we derive the asymptotic distribution of the generalized C(α) 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.

\[ D_n(\theta, \omega) = (D_{1n}(\theta, \omega), \ldots, D_{mn}(\theta, \omega))' \in \mathcal{S}, n = 1, 2, \ldots \]

is a sequence of \( m \times 1 \) random vectors, defined on a common probability space \( (\mathcal{S}, \mathcal{G}, \mathbb{P}) \), which are functions of a \( p \times 1 \) parameter vector \( \theta \), where \( \theta \in \Theta \subseteq \mathbb{R}^p \) and \( \Theta \) 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 \( \omega \), so the symbol \( \omega \) 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.

\[ \sqrt{n} D_n(\theta_0) \xrightarrow{p} D(\theta_0) \text{ where } D(\theta_0) \sim \mathcal{N}[0, I(\theta_0)] \]

**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 \( \theta \) in a non-empty open neighborhood \( N_0 \) of \( \theta_0 \), \( D_n(\theta) \) 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 $\omega \in \mathcal{D}$, where $\mathcal{D}$ is an event with probability one (i.e., $\mathbb{P}[\omega \in \mathcal{D}] = 1$) and $J(\theta)$ is an $m \times p$ (nonrandom) matrix function of $\theta$ and the remainder $R_n(\theta, \theta_0, \omega)$ satisfies the following condition: for any $\varepsilon > 0$ and $\delta > 0$, we have

$$\limsup_{n \to \infty} \mathbb{P} \left[ \{ \omega : r_n(\delta, \theta_0, \omega) > \varepsilon \} \right] \leq U_D(\delta, \varepsilon, \theta_0)$$

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

$U_D(\delta, \varepsilon, \theta_0) \geq 0$ and $\lim\limits_{\delta \downarrow 0} U_D(\delta, \varepsilon, \theta_0) = 0$.

**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 $\theta_0$ such that, for all $\varepsilon > 0$ and $\delta > 0$,

$$\limsup_{n \to \infty} \mathbb{P} \left[ \left\{ \omega : \Delta_n(\theta_0, \delta, \omega) > \varepsilon \right\} \right] \leq U_J(\delta, \varepsilon, \theta_0)$$

where

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

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

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

**Assumption 3.6 Asymptotic Score Non-Degeneracy**. $\text{rank}[J(\theta)] = p$ 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 $\theta$ with derivative $P(\theta) := \frac{\partial \psi}{\partial \theta}$.

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

**Assumption 3.9 Estimator $\sqrt{n}$ Convergence**. $\hat{\theta}_n^0 := \tilde{\theta}_n^0(\omega)$ is a consistent estimator of $\theta_0$, i.e.,

$$\text{plim}_{n \to \infty} (\hat{\theta}_n^0 - \theta_0) = 0,$$

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

$$\limsup_{n \to \infty} \mathbb{P} \left[ \left\{ \omega : \sqrt{n}\| \hat{\theta}_n^0 - \theta_0 \| \geq y \right\} \right] \leq U(y; \theta_0), \forall y > 0,$$

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 $\sqrt{n}(\tilde{\theta}_n^0 - \theta_0) = O_p(1)$ 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. \( \tilde{J}_n, n \geq 1, \) is a sequence of \( m \times m \) symmetric nonsingular (random) matrices such that \( \text{plim} \tilde{J}_n = J(\theta_0) \).

**Assumption 3.12** Weight Matrix Convergence. \( W_n, n \geq 1, \) is a sequence of \( m \times m \) symmetric nonsingular (random) matrices such that \( \text{plim} W_n = W_0 \) where \( W_0 \) is nonsingular.

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

**Proposition 3.1** Asymptotic Distribution of Generalized \( C(\alpha) \) Statistic. Let \( \tilde{Q}_n := \tilde{Q}[W_n] = \tilde{P}[\tilde{J}_n W_n \tilde{J}_n]^{-1} \tilde{J}_n W_n \) 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 \( H_0 \),

\[
\sqrt{n} \tilde{Q}_n D_n(\tilde{\theta}_n^0; Z_n) \xrightarrow{L} n \mathbb{N}[0, Q(\theta_0)I(\theta_0)Q(\theta_0)^\top] \quad (3.1)
\]

where \( Q(\theta_0) = P(\theta_0)[J(\theta_0)W_0J(\theta_0)]^{-1}J(\theta_0)^\top W_0 \), and

\[
\text{PC}(\tilde{\theta}_n^0; \psi, W_n) = n D_n(\tilde{\theta}_n^0; Z_n)^\top \tilde{Q}_n' \tilde{Q}_n [\tilde{Q}_n \tilde{J}_n \tilde{Q}_n']^{-1} \tilde{Q}_n D_n(\tilde{\theta}_n^0; Z_n) \xrightarrow{L} \chi^2(p_1). \quad (3.2)
\]

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 \( \theta \). This is allowed by making a direct assumption on the existence of a linear expansion of \( D_n(\theta, \omega) \) around \( \theta_0 \) [Assumption 3.4]. For the same reason, \( J_n(\theta, \omega) \) does not have to be continuous with respect to \( \theta \).

Since the differentiability of \( D_n(\theta, \omega) \) with respect to \( \theta \) 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 \( \theta_0 \). More precisely, consider the following assumptions.

**Assumption 3.13** Score Differentiability. \( D_n(\theta, \omega) \) is almost surely (a.s.) differentiable with respect to \( \theta \), for all \( n \), in a non-empty open neighborhood \( N_1 \) of \( \theta_0 \). The derivative matrix of \( D_n(\theta, \omega) \) is denoted

\[
J_n(\theta, \omega) = \frac{\partial D_n(\theta, \omega)}{\partial \theta} \quad (3.3)
\]

where the sequence of matrices \( J_n(\theta, \omega), n \geq 1, \) is well-defined for \( \omega \in \mathcal{D}_j \) and \( \mathcal{D}_j \) is an event with probability one (i.e., \( P[\omega \in \mathcal{D}_j] = 1 \)).
Proposition 3.2

Assumption 3.14 (ing conditions:

(a) \( J_n(\theta, \omega) \) is continuous with respect to \( \theta \) for all \( \theta \in N_2, \omega \in D_f \) and \( n \geq 1 \);

(b) \( \sup_{\theta \in N_2} ||J_n(\theta, \omega) - J(\theta)|| \overset{p}{\rightarrow} 0 \).

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

\[
PC(\tilde{\theta}_n^0; \psi) = n D_n(\tilde{\theta}_n^0; Z_n)^\top W_n D_n(\tilde{\theta}_n^0; Z_n)
\]

(4.1)

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

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

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

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 \( W_n \):

\[
PC(\tilde{\theta}_n^0; \psi, W_n) = PC(\tilde{\theta}_n^0; \psi)
\]

\[
= n D_n(\tilde{\theta}_n^0; Z_n)^\top (\tilde{J}_n^{-1})^\top \tilde{P}_n^\top [\tilde{P}_n(\tilde{J}_n \tilde{I}_n^{-1} \tilde{J}_n)^{-1} \tilde{P}_n^\top]^{-1} \tilde{P}_n \tilde{J}_n^{-1} D_n(\tilde{\theta}_n^0; Z_n).
\]

(4.2)

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

5. Testing a subvector

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

$$H_0 : \theta_1 = \tilde{\theta}_{10}$$

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

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

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 $\theta$. We can then write $\tilde{\theta}_n^0 = (\tilde{\theta}_{10}^0, \tilde{\theta}_{2n}^0)'$ where $\tilde{\theta}_{2n}^0$ is a $\sqrt{n}$-consistent estimator of $\theta_2$. 

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 $PC(\hat{\theta}_n^0; \psi, W_n)$ is identical to the score-type statistic suggested by Newey and West (1987a):

$$S(\psi) = n D_n(\tilde{\theta}_n^0; Z_n)' \tilde{I}_n^{-1} \tilde{f}_n(\tilde{\theta}_n^0) \tilde{I}_n^{-1} \tilde{f}_n' \tilde{I}_n^{-1} D_n(\tilde{\theta}_n^0; Z_n)$$

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

$$LM(\psi) = n \hat{\lambda}_n \hat{\lambda}_n = \hat{\lambda}_n$$

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

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

we see easily that

$$S(\psi) = LM(\psi).$$

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

$$PC(\hat{\theta}_n^0; \psi) = n D_n(\tilde{\theta}_n^0; Z_n)' \tilde{I}_n^{-1} \tilde{P}_n(\tilde{\theta}_n^0) \tilde{I}_n^{-1} \tilde{P}_n' \tilde{I}_n^{-1} D_n(\tilde{\theta}_n^0; Z_n)$$

where $D_n(\tilde{\theta}_n^0; Z_n)$ 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 (1987b) in a likelihood setting; see Dagenais and Dufour (1991) for further discussion of this test statistic.
5. TESTING A SUBVECTOR

Let us partition \( J(\theta) \) and \( J_n = J_n(\tilde{\theta}^0_n; Z_n) \) conformably with \( \theta = (\theta'_1, \theta'_2)' \):

\[
J(\theta) = [J_1(\theta), J_2(\theta)], \quad J_n = [J_{n,1}, J_{n,2}] = [\tilde{J}_{n,1}(\tilde{\theta}^0_n; Z_n), \tilde{J}_{n,2}(\tilde{\theta}^0_n; Z_n)],
\]

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

\[
\tilde{J}_n = W_n^{1/2} J_n = [\tilde{J}_{n,1}, \tilde{J}_{n,2}], \quad \tilde{J}_{n,i} = W_n^{1/2} J_{n,i} \quad i = 1, 2,
\]

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

\[
\begin{pmatrix}
(\tilde{J}_n W_n \tilde{J}_n)_11 & (\tilde{J}_n W_n \tilde{J}_n)_12 \\
(\tilde{J}_n W_n \tilde{J}_n)_21 & (\tilde{J}_n W_n \tilde{J}_n)_22
\end{pmatrix}
=
\begin{pmatrix}
\tilde{J}_{n,1} W_n \tilde{J}_{n,1} & \tilde{J}_{n,1} W_n \tilde{J}_{n,2} \\
\tilde{J}_{n,2} W_n \tilde{J}_{n,1} & \tilde{J}_{n,2} W_n \tilde{J}_{n,2}
\end{pmatrix}
(\tilde{J}_n W_n \tilde{J}_n)_{21}
\begin{pmatrix}
(\tilde{J}_n W_n \tilde{J}_n)_{21} & (\tilde{J}_n W_n \tilde{J}_n)_{22}
\end{pmatrix}
,
\]

(5.3)

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,

\[
[(\tilde{J}_n W_n \tilde{J}_n)_{11}]^{-1} (\tilde{J}_n W_n \tilde{J}_n)_{12} = -(\tilde{J}_n W_n \tilde{J}_n)_{12} [(\tilde{J}_n W_n \tilde{J}_n)_{22}]^{-1} = -(\tilde{J}_{n,1} W_n \tilde{J}_{n,2})(\tilde{J}_{n,2} W_n \tilde{J}_{n,2})^{-1},
\]

(5.7)

\[
(\tilde{J}_n W_n \tilde{J}_n)_{11} = [(\tilde{J}_{n,1} W_n \tilde{J}_{n,1}) - \tilde{J}_{n,1} W_n \tilde{J}_{n,2}(\tilde{J}_{n,2} W_n \tilde{J}_{n,2})^{-1} \tilde{J}_{n,2} W_n \tilde{J}_{n,1}]^{-1};
\]

(5.8)

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

\[
\tilde{Q}[W_n] = \tilde{F}_n^{-1} (\tilde{J}_n W_n \tilde{J}_n)_{11}^{-1} (\tilde{J}_n W_n \tilde{J}_n)_{12} W_n
\]

\[
= [I_{p_1}, 0_{p_1 \times p_2}] \begin{pmatrix}
(\tilde{J}_n W_n \tilde{J}_n)_{11} & (\tilde{J}_n W_n \tilde{J}_n)_{12} \\
(\tilde{J}_n W_n \tilde{J}_n)_{21} & (\tilde{J}_n W_n \tilde{J}_n)_{22}
\end{pmatrix}
\begin{pmatrix}
\tilde{J}_n^{-1} W_n \\
\tilde{J}_n^{-1}
\end{pmatrix}
,
\]

(5.9)

where

\[
J_{n,1|2} = J_{n,1} - J_{n,2}(J_{n,2} W_n \tilde{J}_{n,2})^{-1} J_{n,2} W_n \tilde{J}_{n,1} = W_n^{-1/2} M[\tilde{J}_{n,2}] J_{n,1},
\]

(5.10)

\[
\tilde{V}_{n,1|2} = (\tilde{J}_n^{-1} W_n \tilde{J}_n)_{11} - \tilde{J}_n^{-1} W_n \tilde{J}_{n,2}(J_{n,2} W_n \tilde{J}_{n,2})^{-1} J_{n,2} W_n \tilde{J}_{n,1}
\]

\[
= J_{n,1} M[\tilde{J}_{n,2}] J_{n,1},
\]

(5.11)
5. TESTING A SUBVECTOR

Using (5.9), we get:

\[ \tilde{Q}[W_n] \tilde{D}_n = \tilde{V}_{n-1/2} J_n' W_n \tilde{D}_n \]
\[ = \tilde{V}_{n-1/2} \left[ J_{n-1}' W_n \tilde{D}_n - (J_{n-1}' W_n J_{n-2}) (J_{n-2}' W_n J_{n-2})^{-1} J_{n-2}' W_n \tilde{D}_n \right], \tag{5.12} \]
\[ \tilde{J}_{n-1/2}' W_n \tilde{D}_n = J_{n-1}' W_n \tilde{D}_n - (J_{n-1}' W_n J_{n-2}) (J_{n-2}' W_n J_{n-2})^{-1} J_{n-2}' W_n \tilde{D}_n, \tag{5.13} \]
\[ \tilde{Q}[W_n] \tilde{I}_n \tilde{Q}[W_n]' = \tilde{V}_{n-1/2} J_n' W_n \tilde{I}_n W_n J_{n-1/2} \tilde{V}_{n-1/2} \]
\[ = \tilde{V}_{n-1/2} J_n' M_{[\tilde{J}_{n-2}]} W_n^{1/2} \tilde{I}_n W_n^{1/2} M_{[\tilde{J}_{n-2}]} J_{n-1} \tilde{V}_{n-1/2}, \tag{5.14} \]

where \( \tilde{D}_n = D_n(\theta_0^*; Z_n) \). The generalized \( C(\alpha) \) statistic then takes the form:

\[ PC_1(\tilde{\theta}_n^0, \bar{\theta}_{10}, W_n) = PC(\tilde{\theta}_n^0, \psi, W_n) \]
\[ = n \tilde{D}_n' W_n \tilde{I}_{n-1/2} (J_{n-1/2}' W_n \tilde{I}_n W_n J_{n-1/2})^{-1} J_{n-1/2}' W_n \tilde{D}_n \]
\[ = n \tilde{D}_n' W_n^{1/2} M_{[\tilde{J}_{n-2}]} J_{n-1} \tilde{\Sigma}(W_n)^{-1} \tilde{J}_n' M_{[\tilde{J}_{n-2}]} W_n^{1/2} \tilde{D}_n \tag{5.15} \]

where

\[ \tilde{\Sigma}(W_n) = \tilde{J}_n' M_{[\tilde{J}_{n-2}]} W_n^{1/2} I_n W_n^{1/2} M_{[\tilde{J}_{n-2}]} J_{n-1} \]

and the matrix \( \tilde{V}_{n-1/2} \) cancels out.

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

\[ \sqrt{n} \left[ \begin{array}{c} \tilde{S}_{n-1/2} \\ \tilde{S}_{n-2} \end{array} \right] = \sqrt{n} \tilde{R}_n \tilde{D}_n \xrightarrow{n \to \infty} N \left[ 0, \tilde{R}(\theta_0) I(\theta_0) \tilde{R}(\theta_0)' \right] \tag{5.16} \]

where

\[ \begin{bmatrix} J_{n-1/2}' W_n \\ J_{n-2}' \tilde{I}_n^{-1} \end{bmatrix} \xrightarrow{n \to \infty} R(\theta_0) = \begin{bmatrix} J_{1/2}(\theta_0)' W_0 \\ J_{2}(\theta_0)' I(\theta_0)^{-1} \end{bmatrix}, \tag{5.17} \]

\[ J_{1/2}(\theta_0) = J_1(\theta_0) - J_2(\theta_0) [J_2(\theta_0)' W_0 J_2(\theta_0)]^{-1} J_2(\theta_0)' W_0 J_1(\theta_0), \tag{5.18} \]

and the asymptotic covariance matrix between \( \sqrt{n} \tilde{S}_{n-2} \) and \( \sqrt{n} \tilde{S}_{n-1/2} \) is

\[ [J_2(\theta_0)' I(\theta_0)^{-1}] I(\theta_0) [W_0 J_{1/2}(\theta_0)] = J_2(\theta_0)' [W_0 J_{1/2}(\theta_0)] = 0. \tag{5.19} \]

Indeed, the above orthogonality can be viewed as the source of the evacuation of the distribution of \( \sqrt{n}(\tilde{\theta}_n^0 - \theta_0) \) 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 \( H_0 \),

\[ J_{1/2}(\theta_0)' W_0 \sqrt{n} \left[ D_n(\tilde{\theta}_n^0 - D_n(\theta_0)) \right] = J_{1/2}(\theta_0)' W_0 J(\theta_0) \sqrt{n}(\tilde{\theta}_n^0 - \theta_0) + o_p(1) \]
5. TESTING A SUBVECTOR

\[ J_{1/2}(\theta_0)W_0J_{2}(\theta_0)\sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) + o_p(1) = o_p(1). \]  

(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 \(\hat{\theta}_{n}^{0}\), and similarly for the generalized C(\(\alpha\)) statistic.

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

\[
PC_1(\hat{\theta}_{n}^{0}; \theta_{10}) = n\bar{D}_n^{0}\bar{T}_n^{-1}\bar{J}_n^{-1/2}M[\bar{J}_n^{-1}\bar{T}_n^{-1}\bar{J}_n^{-1/2}]^{-1}\bar{J}_n^{-1/2}\bar{T}_n^{-1}\bar{D}_n
\]

(5.21)

where

\[
\bar{J}_n^{-1/2} = [I_n - J_{n-2}J_{n-2}^{-1}J_{n-1}]^{-1}J_{n-1} = I_n^{-1/2}M[\bar{T}_n^{-1/2}J_{n-2}]I_n^{-1/2}J_{n-1},
\]

(5.22)

\[
J_{n-1}^{-1} = I_n^{-1/2}J_{n-1}, \quad J_{n-2}^{-1} = I_n^{-1/2}J_{n-2}.
\]

(5.23)

Upon using (5.16)- (5.19), we see that \(\bar{J}_n^{-1/2}\bar{T}_n^{-1}\bar{D}_n\) and \(\bar{J}_{n-2}^{-1}\bar{T}_n^{-1}\bar{D}_n\) are asymptotically uncorrelated, and

\[
\bar{J}_n^{-1/2}\bar{T}_n^{-1}\bar{D}_n = \bar{J}_n^{-1/2}M[\bar{T}_n^{-1/2}J_{n-2}]I_n^{-1/2}\bar{D}_n
\]

(5.24)

where \(M[\bar{T}_n^{-1/2}J_{n-2}]I_n^{-1/2}\bar{D}_n\) is the residual from the projection of \(\bar{T}_n^{-1/2}\bar{D}_n\) on \(\bar{T}_n^{-1/2}J_{n-2}\). Further, on applying the Frisch-Waugh-Lovell theorem, we see that

\[
P[\bar{J}_n^{-1/2}] = P[\bar{J}_{n-2}] + P[M[\bar{J}_{n-2}]\bar{J}_n^{-1/2}],
\]

(5.25)

hence

\[
PC_1(\hat{\theta}_{n}^{0}; \theta_{10}) = n\bar{D}_n^{0}\bar{T}_n^{-1/2}\{P[\bar{J}_n] - P[\bar{J}_{n-2}]\}I_n^{-1/2}\bar{D}_n
\]

(5.26)

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

\[
\bar{D}_n = \begin{bmatrix}
\bar{D}_{n1} \\
\bar{D}_{n2}
\end{bmatrix}, \quad \bar{I}_n = \begin{bmatrix}
\bar{I}_{n11} & \bar{I}_{n12} \\
\bar{I}_{n21} & \bar{I}_{n22}
\end{bmatrix},
\]

(5.27)

where \(\bar{D}_{ni} = D_n(\hat{\theta}_{ni}^{0}; Z_n)\) is a \(p_i \times 1\) vector and \(\bar{I}_{nij}\) 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 \(\bar{J}_n = -\bar{I}_n\), the formulas in (5.21) and (5.26) reduce to a
6. **TWO-STAGE PROCEDURES**

simple difference between two statistics:

\[
PC_1(\hat{\theta}_n^0; \bar{\theta}_{10}) = n \left( \tilde{D}_{n1} - \tilde{I}_{n12} \tilde{I}^{-1}_{n22} \tilde{D}_{n2} \right) \left( \tilde{I}_{n11} - \tilde{I}_{n12} \tilde{I}^{-1}_{n22} \tilde{I}_{n21} \right)^{-1} \left( \tilde{D}_{n1} - \tilde{I}_{n12} \tilde{I}^{-1}_{n22} \tilde{D}_{n2} \right) \\
= n \left[ \tilde{D}_n^r \tilde{I}^{-1}_n \tilde{D}_n - \tilde{D}^r_n \tilde{I}^{-1}_{n22} \tilde{D}_{n2} \right].
\]  

(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 \( \theta_1 \) is the parameter vector of interest (on which inference focuses), and \( \theta_2 \) denotes a vector of nuisance parameters which is consistently estimated by an auxiliary estimate \( \tilde{\theta}_2^n \), 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 \( \theta_2 \) is estimated in a first stage, and then treated as known for the purpose of testing the value of another parameter vector \( \theta_1 \). 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

\[
D_{n1}(\theta; X_n) = D_{n1}(\theta_1, \theta_2; X_n).
\]  

(6.1)

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

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

(6.2)

A typical setup is the one where

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

(6.3)

\[
E_{\theta} [h_1(\theta_1, \theta_2; x_t)] = 0, \quad t = 1, \ldots, n,
\]  

(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.
6. TWO-STAGE PROCEDURES

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 \( H_0 : \theta_1 = \theta_{10} \) 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 \( \tilde{\theta}_{2n}^0 \) be an estimator of the nuisance parameter vector \( \theta_2 \) obtained from the data \( Y_n = [y_1, \ldots, y_n] \) which may be different from \( X_n \). For example, \( \tilde{\theta}_{2n}^0 \) may be based on an “auxiliary” estimating function \( D_{n2}(\theta; X_n) \), but this is not required. Consider now the restricted estimator \( \tilde{\theta}_0 = (\tilde{\theta}_{10}, \tilde{\theta}_{2n}^0)' \), and denote \( \tilde{D}_{n1} = D_{n1}(\tilde{\theta}_0; X_n) \), \( \tilde{I}_{n1}, \tilde{J}_{n1} := \tilde{J}_{n1}(\tilde{\theta}_0^0) \) and \( W_{n11}, \tilde{W}_{n11} \), the matrices corresponding to \( D_{n1} = D_{n1}(\theta^0; Z_n) \), \( I_n, J_{n;i} \) and \( W_n \) respectively for the system based on the estimating function \( D_{n1}(\theta; X_n) \); \( D_{n1} \) has dimension \( m_1 \times 1 \), \( J_{n;i} \) 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), \( I_{n1}, J_{n;i} \) and \( W_{n11} \) are the corresponding submatrices of \( I_n, J_{n;i} \) and \( W_n \) respectively, where

\[
W_n = \begin{bmatrix} W_{n11} & W_{n12} \\ W_{n21} & W_{n22} \end{bmatrix}
\]

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 \( H_0 : \theta_1 = \theta_{10} \):

\[
PC_1(\tilde{\theta}_0^0; \tilde{\theta}_{10}, W_{n11}) = n \tilde{D}_{n1}^t W_{n11} \tilde{J}_{n1|2} \tilde{\Sigma}_{n1|2}^{-1} \tilde{J}_{n1|2}^t W_{n11} \tilde{D}_{n1}
\]

(6.6)

where \( \tilde{\Sigma}_{n1|2} = \tilde{J}_{n1|2}^t W_{n11} \tilde{I}_{n1} W_{n11} \tilde{J}_{n1|2} \), and

\[
\tilde{J}_{n1|2} = \tilde{J}_{n1} - \tilde{J}_{n12} (\tilde{J}_{n12} W_{n11} \tilde{J}_{n12})^{-1} \tilde{J}_{n12} W_{n11} \tilde{J}_{n11}
\]

\[
= W_{n11}^{-1/2} M [W_{n11}^{1/2} \tilde{J}_{n12}] W_{n11}^{1/2} \tilde{J}_{n11},
\]

(6.7)

\[
\tilde{\Sigma}_{n1|2} = \tilde{J}_{n1|2}^t W_{n11} \tilde{I}_{n1} W_{n11} \tilde{J}_{n1|2}
\]

\[
= \tilde{J}_{n11} W_{n11}^{1/2} M [W_{n11}^{1/2} \tilde{J}_{n12}] W_{n11}^{1/2} \tilde{J}_{n11} W_{n11}^{1/2} M [W_{n11}^{1/2} \tilde{J}_{n12}] W_{n11}^{1/2} \tilde{J}_{n11}.
\]

(6.8)

By Proposition 3.1, \( PC_1(\tilde{\theta}_0^0; \tilde{\theta}_{10}, W_{n11}) \) has a \( \chi^2(p_1) \) asymptotic distribution under \( H_0 \). On taking \( W_{n11} = \tilde{I}_{n11}^{-1} \), \( PC_1 \) takes the following simplified form:

\[
PC_1(\tilde{\theta}_0^0; \tilde{\theta}_{10}, \tilde{I}_{n11}^{-1}) = n \tilde{D}_{n1}^t \tilde{I}_{n11}^{-1/2} \tilde{M}_{12} \tilde{I}_{n11}^{-1/2} \tilde{J}_{n11} \tilde{\Sigma}_{n1|2}^{-1} \tilde{J}_{n1|2}^t \tilde{M}_{12} \tilde{I}_{n11}^{-1/2} \tilde{D}_{n1}
\]

(6.9)

where \( \tilde{M}_{12} = M [\tilde{J}_{n11}]^{1/2} \tilde{J}_{n12} \) and \( \tilde{\Sigma}_{n1|2} = \tilde{J}_{n1|2}^t \tilde{I}_{n11}^{-1} \tilde{J}_{n1|2} = \tilde{J}_{n11} \tilde{I}_{n11}^{-1/2} \tilde{M}_{12} \tilde{I}_{n11}^{-1/2} \tilde{J}_{n11} \).

\(^2\)The number of observations in the dataset \( Y \) could be different from \( n \), say equal to \( n_2 \), \( n_2 \neq n \). If the auxiliary estimate \( \tilde{\theta}_{2n2}^0 \), obtained from the second dataset satisfies \( \sqrt{n_2}(\tilde{\theta}_{2n2}^0 - \theta_{20}) = O_p(1) \), then \( \sqrt{n_2}(\tilde{\theta}_{2n2}^0 - \theta_{20}) = \sqrt{n_2}\sqrt{n_2}(\tilde{\theta}_{2n2}^0 - \theta_{20}) = O_p(1) \) provided \( n/n_2 = O(1) \), and the arguments that follow remain valid. When a set of estimating functions \( D_{n2}(\theta_2) \) for the second dataset is considered, the argument presented here remains valid provided \( \sqrt{n_2}D_{n2}(\theta_{20}) \) obeys a central limit theorem in addition to the previous conditions on the auxiliary estimate and the sample sizes.
6. TWO-STAGE PROCEDURES

When calculating the standard error of the estimator of $\theta_1$, one needs to take into account the sampling error associated with the first-stage estimator of the parameter $\theta_2$; 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 asymptotic distribution of $PC_1(\tilde{\theta}_n; \tilde{\theta}_{10}, W_{n11})$ with respect to the distribution of $\tilde{\theta}_n^0$ is entailed by the orthogonality relation

$$J_{12}(\theta_0)'[W_{011}J_{1112}(\theta_0)] = J_{12}(\theta_0)'W_{011}W_{011}^{-1/2}M[W_{011}^{1/2}J_{12}(\theta_0)]W_{011}^{1/2}J_{11}(\theta_0)$$

$$= [W_{011}^{1/2}J_{12}(\theta_0)]'M[W_{011}^{1/2}J_{12}(\theta_0)]W_{011}^{1/2}J_{11}(\theta_0) = 0,$$

(6.10)

where $\text{plim} W_{n11} = W_{011}$. This in turn implies that $\sqrt{n} J_{1112} W_{n11} \tilde{D}_{n1}$ is asymptotically uncorrelated with $\sqrt{n} J_{1112} W_{n11} \tilde{D}_{n1}$; 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 $\theta_2$ and account for the uncertainty of $\theta_2$ estimates, where $m_2 \geq p_2$. Further, we suppose here that $D_{n2}(\theta; X_n)$ only depends on $\theta_2$, 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

$$E_{\theta} \left[ \begin{array}{c} h_1(\theta_1, \theta_2; x_t) \\ h_2(\theta_2; y_t) \end{array} \right] = 0, \ t = 1, \ldots, n,$$

(6.11)

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

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

(6.12)

In this context, the sample estimating function is

$$\tilde{D}_n = \begin{bmatrix} D_{n1} \\ D_{n2} \end{bmatrix} = \begin{bmatrix} D_{n1}(\theta_1^0; X_n) \\ D_{n2}(\theta_2; Y_n) \end{bmatrix}$$

(6.13)

with

$$J(\theta) = [J_1(\theta), J_2(\theta)] = \begin{bmatrix} J_{11}(\theta) & J_{12}(\theta) \\ 0_{m_2 \times p_1} & J_{22}(\theta) \end{bmatrix}.$$  

(6.14)

The partitioned Jacobian estimator is then given by

$$\tilde{J}_n = [\tilde{J}_{n1}, \tilde{J}_{n2}] = \begin{bmatrix} \tilde{j}_{n11} & \tilde{j}_{n12} \\ 0_{m_2 \times p_1} & \tilde{j}_{n22} \end{bmatrix}.$$  

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

$$PC_1(\tilde{\theta}_n^0, \tilde{\theta}_{10}, W_n) = n \tilde{D}_n' W_n \tilde{J}_{n|2} (\tilde{J}_{n|2}' W_n \tilde{J}_{n|2})^{-1} \tilde{J}_{n|2}' W_n \tilde{D}_n$$

(6.16)

with

$$\tilde{J}_{n|2}' W_n \tilde{D}_n = \tilde{J}_{n|2}' W_{n11} \tilde{D}_{n1} + \left[ \tilde{J}_{n|2}' W_{n12} \tilde{D}_{n2} - \tilde{J}_{n|2}' W_n \tilde{J}_{n|2} (\tilde{J}_{n|2}' W_n \tilde{J}_{n|2})^{-1} \tilde{J}_{n|2}' W_n \tilde{D}_n \right].$$

(6.17)

In this case, the correction for the estimation of $\theta_2$ is accounted by the two last terms in the above expression for $\tilde{J}_{n|2}' W_n \tilde{D}_n$.

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.

$$W_{n12} = W_{n21} = 0.$$  

(6.18)

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

$$\tilde{J}_{n|2}' W_n \tilde{D}_n = \tilde{J}_{n|2}' W_{n11} \tilde{D}_{n1} - \tilde{J}_{n11} W_{n11} \tilde{J}_{n12} (\tilde{J}_{n2}' W_{n2} \tilde{J}_{n2})^{-1} \left[ \tilde{J}_{n12} W_{n11} \tilde{D}_{n1} + \tilde{J}_{n22}' W_{n22} \tilde{D}_{n2} \right],$$

(6.19)

and

$$\tilde{J}_{n|2}' W_n \tilde{J}_{n|2} = \tilde{J}_{n12}' W_{n11} \tilde{J}_{n12} + \tilde{J}_{n22}' W_{n22} \tilde{J}_{n22}.$$  

(6.20)

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

7. Conclusion

In this paper, we have introduced a comprehensive $C(\alpha)$ statistic based on estimating functions (or GMM setups). As in Smith (1987a), 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

A. PROOFS

PROOF OF PROPOSITION 3.1 To simplify notation, we shall assume throughout that \( \omega \in \mathcal{D}_J \) (an event with probability 1) and drop the symbol \( \omega \) 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 \( P(\hat{\theta}_n^0) \) and \( J_n(\hat{\theta}_n^0) \) converge in probability to \( P(\theta_0) \) and \( J(\theta_0) \) respectively. The consistency of \( P(\hat{\theta}_n^0) \), i.e.

\[
\lim_{n \to \infty} [P(\hat{\theta}_n^0) - P(\theta_0)] = 0, \quad (A.1)
\]

follows simply from the consistency of \( \hat{\theta}_n^0 \) [Assumption 3.9] and the continuity of \( P(\theta) \) at \( \theta_0 \) [Assumption 3.7]. Further, by Assumption 3.8, since \( P(\theta) \) is continuous in open neighborhood of \( \theta_0 \), we also have

\[
\text{rank } [\hat{P}_n] = \text{rank } [P(\theta_0)] = p_1. \quad (A.2)
\]

Consider now \( J_n(\hat{\theta}_n^0) \). 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

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

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

\[
P[\|J_n(\hat{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon] \geq P[\|\hat{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \|J_n(\hat{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon]
\geq P[\|\hat{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \Delta_n(\theta_0, \delta_1) \leq \varepsilon]
\geq 1 - P[\|\hat{\theta}_n^0 - \theta_0\| > \delta_1] - P[\Delta_n(\theta_0, \delta_1) > \varepsilon]
\geq 1 - (\varepsilon_1/2) - (\varepsilon_1/2) = 1 - \varepsilon_1.
\]

Thus,

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

hence

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

or, equivalently,

\[
\lim_{n \to \infty} [J_n(\hat{\theta}_n^0) - J(\theta_0)] = 0. \quad (A.4)
\]

By Assumption 3.4, we can write [setting \( 0/0 = 0 \)]:

\[
\|\sqrt{n}[D_n(\hat{\theta}_n^0) - D_n(\theta_0)] - J(\theta_0)\sqrt{n}(\hat{\theta}_n^0 - \theta_0)\| = \sqrt{n}[R_n(\hat{\theta}_n^0, \theta_0)]
\]
A. PROOFS

\[ \frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} - \sqrt{n}\|\hat{\theta}_n^0 - \theta_0\| \]

where

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

and

\[ \limsup_{n \to \infty} P\left[r_n(\delta, \theta_0) > \varepsilon\right] < U_D(\delta, \varepsilon, \theta_0). \]

Thus, for any \( \varepsilon > 0 \) and \( \delta > 0 \), we have:

\[ P\left[\frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} \leq \varepsilon\right] \geq P\left[r_n(\delta, \theta_0) \leq \varepsilon, \hat{\theta}_n^0 \in N_0 \text{ and } \|\hat{\theta}_n^0 - \theta_0\| \leq \delta\right] \]

\[ \geq 1 - P\left[r_n(\delta, \theta_0) > \varepsilon\right] - P\left[\hat{\theta}_n^0 \notin N_0 \text{ or } \|\hat{\theta}_n^0 - \theta_0\| > \delta\right] \]

hence, using the consistency of \( \hat{\theta}_n^0 \),

\[ \liminf_{n \to \infty} P\left[\frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} \leq \varepsilon\right] \geq 1 - \limsup_{n \to \infty} P\left[r_n(\delta, \theta_0) > \varepsilon\right] \]

\[ - \limsup_{n \to \infty} P\left[\hat{\theta}_n^0 \notin N_0 \text{ or } \|\hat{\theta}_n^0 - \theta_0\| > \delta\right] \]

\[ \geq 1 - U_D(\delta, \varepsilon, \theta_0). \]

Since \( \lim_{\delta \to 0} U_D(\delta, \varepsilon, \theta_0) = 0 \), it follows that \( \lim_{n \to \infty} P\left[\frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} \leq \varepsilon\right] = 1 \) for any \( \varepsilon > 0 \), or equivalently,

\[ \frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} \xrightarrow{p_{n \to \infty}} 0. \]

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

\[ \sqrt{n}\|R_n(\hat{\theta}_n^0, \theta_0)\| = \frac{\|R_n(\hat{\theta}_n^0, \theta_0)\|}{\|\hat{\theta}_n^0 - \theta_0\|} \sqrt{n}\|\hat{\theta}_n^0 - \theta_0\| \xrightarrow{p_{n \to \infty}} 0 \]  \hspace{1cm} (A.5)

and

\[ \|\sqrt{n}[D_n(\hat{\theta}_n^0) - D_n(\theta_0)] - J(\theta_0)\sqrt{n}(\hat{\theta}_n^0 - \theta_0)\| \xrightarrow{p_{n \to \infty}} 0. \]  \hspace{1cm} (A.6)

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

\[ \psi(\theta) = \psi(\theta_0) + P(\theta_0)(\theta - \theta_0) + R_2(\theta, \theta_0), \]  \hspace{1cm} (A.7)

for \( \theta \in N \subseteq N_0 \cap V_0 \), where \( N \) is a non-empty open neighborhood of \( \theta_0 \) and

\[ \lim_{\theta \to \theta_0} \frac{\|R_2(\theta, \theta_0)\|}{\|\theta - \theta_0\|} = 0. \]
\[ i.e., R_2(\theta, \theta_0) = o(||\theta - \theta_0||), \text{ so that, using Assumption 3.10,} \]
\[
\sqrt{n} P(\theta_0)(\tilde{\theta}_n^0 - \theta_0) = \sqrt{n}[\psi(\tilde{\theta}_n^0) - \psi(\theta_0)] - \sqrt{n} R_2(\tilde{\theta}_n^0, \theta_0)
\]
\[
= -\sqrt{n} R_2(\tilde{\theta}_n^0, \theta_0) \tag{A.8}
\]
for \( \tilde{\theta}_n^0 \in N \), and
\[
||\sqrt{n} P(\theta_0)(\tilde{\theta}_n^0 - \theta_0)|| = \frac{||R_2(\tilde{\theta}_n^0, \theta_0)||}{||\tilde{\theta}_n^0 - \theta_0||} \rightarrow 0 \tag{A.9}
\]
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:
\[
\text{rank } [\bar{P}_n] = p_1, \quad \text{rank } [\bar{J}_n] = p, \quad \text{rank } [\bar{I}_n] = m, \quad \text{rank } [W_n] = m, \tag{A.10}
\]
so the matrices \( \bar{J}_n, \bar{I}_n \), and \( W_n \) all have full column rank. Since \( \text{plim } \bar{P}_n = P(\theta_0) \) and \( \text{plim } \bar{J}_n = J(\theta_0) \), we can then write:
\[
\text{plim } \bar{J}_n W_n \bar{J}_n^{-1} = [J(\theta_0)'W_0 J(\theta_0)]^{-1}, \quad \text{plim } \bar{Q}_n = Q(\theta_0),
\]
\[
\text{plim } \bar{Q}_n \bar{J}_n = \text{plim } \bar{Q}_n J(\theta_0) = Q(\theta_0) J(\theta_0) = P(\theta_0),
\]
where \( \bar{Q}_n := \bar{Q}[W_n] = \bar{P}_n[J_n' W_n J_n]^{-1} J_n' W_n \). Then, using (A.6) and (A.9), it follows that:
\[
\text{plim } \frac{\sqrt{n} \bar{Q}_n D_n(\tilde{\theta}_n^0) - \sqrt{n} Q(\theta_0) D_n(\theta_0)}{\sqrt{n} \bar{Q}_n D_n(\tilde{\theta}_n^0) - \sqrt{n} Q(\theta_0) D_n(\theta_0)}
\]
\[
= \text{plim } \left\{ \sqrt{n} \bar{Q}_n D_n(\tilde{\theta}_n^0) - Q(\theta_0) \sqrt{n} D_n(\theta_0) \right\} - \text{plim } \left\{ P(\theta_0) \sqrt{n} (\tilde{\theta}_n^0 - \theta_0) \right\}
\]
\[
= \text{plim } \left\{ \bar{Q}_n \left[ D_n(\tilde{\theta}_n^0) - D_n(\theta_0) \right] - J(\theta_0) \sqrt{n} (\tilde{\theta}_n^0 - \theta_0) \right\}
\]
\[
+ \text{plim } \left\{ \left[ \bar{Q}_n - Q(\theta_0) \right] \sqrt{n} D_n(\theta_0) + \left[ \bar{Q}_n J(\theta_0) - P(\theta_0) \right] \sqrt{n} (\tilde{\theta}_n^0 - \theta_0) \right\}
\]
\[
= \text{plim } \left\{ \bar{Q}_n \left[ D_n(\tilde{\theta}_n^0) - D_n(\theta_0) \right] - J(\theta_0) \sqrt{n} (\tilde{\theta}_n^0 - \theta_0) \right\} = 0.
\]
We conclude that the asymptotic distribution of \( \sqrt{n} \bar{Q}_n D_n(\tilde{\theta}_n^0) \) is the same as the one of \( Q(\theta_0) \sqrt{n} D_n(\theta_0) \), namely (by Assumption 3.2) a N \([0, V_\psi(\theta_0)]\) 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
\[
\hat{V}_\psi(\theta_n^0) = \bar{Q}_n \bar{I}_n \bar{Q}_n'
\]
\[
\tag{A.11}
\]
A. **PROOFS**

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

\[
\text{rank} \left[ V_\psi(\hat{\theta}_n^0) \right] = p_1. \tag{A.12}
\]

Thus the test criterion

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

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

**PROOF OF PROPOSITION 3.2**  Consider the (non-empty) open neighborhood \( N = N_1 \cap N_2 \) of \( \theta_0 \). For any \( \theta \in N \) and \( \omega \in \mathcal{F} \), we can write

\[
\|J(\theta) - J(\theta_0)\| \leq \|J_n(\theta, \omega) - J(\theta)\| + \|J_n(\theta_0, \omega) - J(\theta_0)\|
\]

By Assumption 3.14(b), we have

\[
\text{plim} \left( \sup_{\theta \in N} \|J_n(\theta, \omega) - J(\theta)\| \right) \leq \text{plim} \left( \sup_{\theta_n \in N_2} \|J_n(\theta, \omega) - J(\theta)\| \right) = 0
\]

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

\[
\sup_{\theta \in N} \{\|J_n(\theta, \omega) - J(\theta)\|\} \underset{t \to \infty}{\longrightarrow} 0 \ a.s.
\]

Let

\[
CS = \{ \omega \in \mathcal{F} : \lim_{t \to \infty} \left( \sup_{\theta \in N} \|J_n(\theta, \omega) - J(\theta)\| \right) = 0 \}
\]

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

\[
t \geq t_0(\varepsilon, \omega) \Rightarrow 2 \sup_{\theta \in N} \{\|J_n(\theta, \omega) - J(\theta)\|\} < \varepsilon/2.
\]

Further, since \( J_n(\theta, \omega) \) is continuous in \( \theta \) at \( \theta_0 \), 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 \( n = n_0 \), we find that \( \|\theta - \theta_0\| < \delta(n_0, \omega) \) implies

\[
\|J(\theta) - J(\theta_0)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]
A. PROOFS

In other words, for any \( \varepsilon > 0 \), we can choose \( \delta = \delta(n_0, \varepsilon) > 0 \) such that

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

and the function \( J(\theta) \) must be continuous at \( \theta_0 \). Part (a) of the Proposition is established.

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

\[
\Delta_n(\theta_0, \delta, \omega) := \sup \{ \| J_n(\theta, \omega) - J(\theta_0) \| : \theta \in N_2 \text{ and } 0 \leq \| \theta - \theta_0 \| \leq \delta \}
\]

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

\[
\limsup_{n \to \infty} P \{ \omega : \Delta_n(\theta_0, \delta, \omega) > \varepsilon \} \leq \limsup_{n \to \infty} P \{ \omega : \overline{A}_n(N_2, \omega) > \varepsilon \} \leq U_f(\delta, \varepsilon, \theta_0)
\]

for any function \( U_f(\delta, \varepsilon, \theta_0) \) that satisfies the conditions of Assumption 3.5. The latter thus holds with \( V_0 \) any non-empty open neighborhood of \( \theta_0 \) 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 \( \theta_0 \) for all \( \omega \in \mathcal{D}_f \), 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 \( \theta \) in an open neighborhood \( U \) of \( \theta_0 \) (with \( U \subseteq N_2 \)), we can write

\[
D_n(\theta, \omega) = D_m(\theta_0, \omega) + J_n(\theta_0, \omega)(\theta - \theta_0)
\]

where \( J_n(\theta, \omega)_i \) and \( J(\theta)_i \) are the \( i \)-th rows of \( J_n(\theta, \omega) \) and \( J(\theta) \) respectively,

\[
R_n(\theta, \theta_0, \omega) = [J_n(\theta_0, \omega)_i - J(\theta_0)_i](\theta - \theta_0)
\]

and \( \tilde{\theta}_n(\omega) \) belongs to the line joining \( \theta \) and \( \theta_0 \). Further, for \( \theta \in U \),

\[
|R_n(\tilde{\theta}_n(\omega), \theta_0, \omega)| \leq \| J_n(\tilde{\theta}_n(\omega), \omega)_i - J(\theta_0)_i \| \| \theta - \theta_0 \|
\]

hence, on defining \( N_0 = U \),

\[
R_n(\theta, \theta_0, \omega) = [R_{1n}(\tilde{\theta}_n(\omega), \theta_0, \omega), \ldots, R_{mn}(\tilde{\theta}_n(\omega), \theta_0, \omega)]',
\]

we see that

\[
|R_n(\theta, \theta_0, \omega)| \leq \sum_{i=1}^{m} |R_{in}(\tilde{\theta}_n(\omega), \theta_0, \omega)|
\]
A. PROOFS

\[ \leq m \| \theta - \theta_0 \| \sup_{\theta \in N_0} \{ \| J_n(\theta, \omega) - J(\theta) \| \} \]

and

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

\[ \leq m \sup \{ \| J_n(\theta, \omega) - J(\theta) \| : \theta \in N_2 \} \]

Thus \( r_n(\delta, \theta_0, \omega) \overset{p}{\to} 0 \) and

\[ \limsup_{n \to \infty} P \left[ \{ \omega : r_n(\delta, \theta_0, \omega) > \varepsilon \} \right] \leq U_D(\delta, \varepsilon, \theta_0) \quad (A.13) \]

must hold for any function that satisfies the conditions of Assumption 3.4. This completes the proof. \( \square \)
References


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.


REFERENCES


