Generalized C(α) tests with nonstandard convergence rates

Jean-Marie Dufour and Masaya Takano

Abstract We study hypothesis testing of linear and nonlinear restrictions on a finite-dimensional parameter vector, using estimating functions (or moment equations), when nuisance parameter estimators and the estimating functions converge at nonstandard rates. We focus on generalized $C(\alpha)$ tests [Dufour et al. (2016)], which allow one to use a wide class of root-*n* consistent restricted estimators, under weak assumptions on the asymptotic distribution of the estimators. However, root-*n* consistency remains notably restrictive, because it precludes estimators which converge at a slow rate, *e.g.* many estimators based on nonparametric regressions. We establish conditions under which generalized $C(\alpha)$-type statistics follow the usual chi-square distribution (under the null hypothesis) when the statistic is based on a restricted estimator which converges at a rate slower than the usual $n^{1/2}$ rate. We also allow for nonstandard convergence rates on the estimating functions and their derivatives. The conditions given depend on the relation between the different convergence rates. As a special case, when the estimating function converges to its limit at rate $n^{1/2}$, we show that the convergence rate of the restricted estimator need only be faster than $n^{1/4}$. We apply the proposed procedure to a testing problem on derivatives of the conditional expectation, involving multiple nonstandard rates.

1 Introduction

In this paper, we develop generalized $C(\alpha)$-type tests for linear and nonlinear hypotheses, in order to allow for nonstandard (possibly slow) convergence rates on

Jean-Marie Dufour
Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec, H3A 2T7, Canada. e-mail: jean-marie.dufour@mcgill.ca

Masaya Takano
Faculty of Business and Economics, Monash University, 29 Ancora Imparo Wy, Room 340, Clayton, VIC 3168 Australia. e-mail: masaya.takano@monash.edu

the parameter estimates used, in the context of models specified through estimating functions or moment equations [*e.g.*, using the generalized method of moments (GMM)].

In parametric models, likelihood ratio (LR) tests [Neyman and Pearson (1928)], Wald tests [Wald (1943)], and score tests [introduced by Rao (1948)], along with various extensions, constitute the basis of statistical hypothesis testing. In likelihood models, LR tests require one to estimate the model under both the null hypothesis and without restrictions, Wald tests only require unrestricted estimators, while score tests only use restricted estimators. Under standard regularity conditions, the three tests have local asymptotic efficiency. These general testing procedures can be extended to more general setups where moments (or estimating functions) play the role of a score-type function; see, for example, Dufour et al. (2017), Gouriéroux and Monfort (1995), Newey and West (1987).

Optimization of a likelihood function under constraints can be computationally expensive. The $C(\alpha)$ test procedure proposed by Neyman (1954, 1959) extends Rao's score test by allowing one to replace the maximum-likelihood estimator by any root-*n* consistent restricted estimator. As in the original score test, the $C(\alpha)$ test enjoys optimality properties and has been extended to more general setups. This provides great flexibility in the choice of estimator used, because the asymptotic distribution of the restricted estimator need not be known (or Gaussian).

The literature on $C(\alpha)$ tests and related procedures is extensive; see Bartoo and Puri (1967), Basawa (1985), Bera and Bilias (2001), Bera and Yoon (1993), Berger and Wallenstein (1989), Bernshtein (1976, 1978, 1980a,b, 1981), Bhat and Nagnur (1965), Bontemps (2019), Bontemps and Meddahi (2012), Bühler and Puri (1966), Chant (1974), Chaudhuri and Zivot (2011), Chibisov (1973), Dagenais and Dufour (1991), Davidson and MacKinnon (1991, 1993), Dufour and Valéry (2009), Dufour and Dagenais (1992), Dufour et al. (2016, 2017), Foutz (1976), Hall and Mathiason (1990), Jaggia and Trivedi (1994), Kocherlakota and Kocherlakota (1991), Le Cam and Traxler (1978), Le Cam (1956), Moran (1970, 1973), Neyman (1979), Pal (2003), Paul and Barnwal (1990), Rao (1996), Ray (1974), Ronchetti (1987), Singh and Zhurbenko (1975), Smith (1987b,a), Tarone (1979, 1985), Tarone and Gart (1980), Vorob'ev and Zhurbenko (1979), Wang (1981, 1982), Wooldridge (1990).

The initial idea of the $C(\alpha)$ test is to orthogonalize the scores associated with the parameters of interest (which are restricted by the null hypothesis) with respect to the scores of the nuisance parameters (at least, under the null hypothesis). This reduces the sensitivity of the test statistic to the distribution of the nuisance parameter estimate, and indeed evacuates it from its asymptotic distribution (under appropriate regularity conditions). The $C(\alpha)$ test can be further generalized to relax orthogonality conditions as well as the assumption of differentiability of the log-likelihood function (or score-type function); see Dufour et al. (2016). Previous work typically assumes the existence of a $n^{1/2}$-consistent restricted estimator, so that these results are not applicable in cases where the restricted estimator has a slower convergence rate (under the null hypothesis).

In this paper, we *first* extend the generalized $C(\alpha)$ test proposed in Dufour et al. (2016) for testing general parameter restrictions using a vector of estimating functions. We allow for both the restricted estimator and the estimating functions to converge at rates slower than $n^{1/2}$. When the estimating functions converge at the standard rate $n^{1/2}$, our conditions entail that the convergence rate of the estimator be faster than $n^{1/4}$. *In some cases, it could even be slower.* Naturally, since the results presented extend those of Dufour et al. (2016), they *do not require orthogonality* of the moment equations with respect to the score of the log-likelihood function, nor even the existence of a score. On the other hand, the specific form of the restriction (and its derivative) plays a central role in creating the required asymptotic invariance.

There are many examples where slow convergence can occur. In the context of discrete choice models, the maximum score estimator of Manski (1975) is known to converge at the cube rate $n^{1/3}$, and an improved estimator of Horowitz (1992) utilizing kernel smoothing enjoys a faster rate of convergence, but still does not reach the parametric rate $n^{1/2}$. The problem involving cubic-root asymptotics was first pointed out by Chernoff (1964), and has been studied by Kim and Pollard (1990) and Seo and Otsu (2018), among others. Caner (2006) shows that the convergence rate of an M-estimator (with weakly dependent data) depends on the decay rate of the mixing coefficients and the smoothness of the objective function, and may be slower than $n^{1/2}$. In general, $n^{1/2}$-consistent estimation in the presence of an infinite-dimensional parameter may not be feasible [*e.g.*, Firpo et al. (2009)], and this applies quite generally to estimators based on nonparametric regressions.

Second, we let the estimating functions converge to a non-degenerate limit at a more general rate than $n^{1/2}$. To show invariance of the asymptotic distribution of the $C(\alpha)$-type statistic to the distribution of the restricted estimator, we only require that the restricted estimator converge not too slowly relative to the estimating function. This allows one to use an estimating function and a restricted estimator based on different samples – whose sizes can be quite different – which can lead to different rates of convergence. This feature is easily accommodated by the asymptotic invariance of $C(\alpha)$-type statistics with respect to the distribution of the estimator used.

Antoine and Renault (2012) consider GMM inference when the estimating equations converge at multiple nonstandard rates, and they study the convergence rates of their estimators of structural parameters. In addition, they show that hypothesis testing can be conducted by a Wald-type test. However, their results are derived under the assumption that the estimating functions and their derivatives (with respect to model parameters) converge at the same rates, which holds automatically in the linear case, but may be violated in more general setups. For instance, the derivatives of the score function associated with local polynomial regression of order at least two converge at rates faster than the corresponding estimating functions. Our test procedures allow for such a structure, so that they are applicable to a wider class of testing problems. One can also remain agnostic on the asymptotic distribution of the restricted estimator used; in particular, the latter need not be obtained by minimizing a criterion function based on the estimating function considered. This yields computational advantages. When the estimating function consists of the derivatives of a local maximum likelihood function [Fan et al. (1998)], computing the restricted local maximum-likelihood estimator can be costly. In contrast, $C(\alpha)$-type procedures allow one to use a wider array of estimation methods, such as local polynomial fitting.

The paper is organized as follows. In Section 2, we describe the setup and extend the generalized $C(\alpha)$ test of Dufour et al. (2016) by relaxing assumptions on the convergence rates of the restricted estimator and estimating functions. Section 3 discusses an application of the test procedure to a testing problem on the derivatives of a nonparametric regression. We conclude in Section 4. The proof of the main result and details on the application are provided in the Supplementary material.

Notation $X_n = o_p(R_n)$ means: $X_n = Y_n R_n$ and $Y_n \xrightarrow[n\to\infty]{\mathrm{p}} 0$. $X_n = O_p(R_n)$ means: $X_n = R_n Y_n$ and $Y_n = O_p(1)$; if $R_n = 1$, $X_n = R_n Y_n$ simply means that X_n is asymptotically stochastically bounded. The symbol := means "equal by definition".

2 Asymptotic distribution of generalized $C(\alpha)$ statistics

Dufour et al. (2016) consider the problem of testing a general (possibly nonlinear) restriction on a finite-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^p$ of the form

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

where $\psi : \Theta \mapsto \mathbb{R}^{p_1}$. The test procedure relies on the existence of an $m \times 1$ estimating function $D_n(\theta)$ such that

$$\sqrt{n} D_n(\theta_0) \xrightarrow[n\to\infty]{L} \mathrm{N}[0,\, I_0] \tag{2}$$

where $p_1 \leq p \leq m$ and I_0 is a $m \times m$ nonsingular matrix, along with a restricted estimator $\tilde{\theta}_n^0$ of θ which converges in probability to the "true value" θ_0 under H_0, *i.e.*

$$\tilde{\theta}_n^0 - \theta_0 = O_p(n^{-1/2}) \quad \text{under } H_0\,. \tag{3}$$

Under appropriate regularity conditions, the proposed generalized $C(\alpha)$ statistic is asymptotically distributed as $\chi^2(p_1)$ under H_0.

In this section, we study the asymptotic distribution of generalized $C(\alpha)$ statistics when $\tilde{\theta}_n^0$ converges to θ_0 at a rate slower than $n^{1/2}$ under the null hypothesis.[1] This relaxation is achieved by exploiting further expansions of the score $D_n(\theta_0)$ and the restriction function $\psi(\theta)$ around $\theta = \theta_0$. We also allow the score $D_n(\theta_0)$ to converge at a general rate n^{r_D} such that $r_D > 0$. The regularity conditions used [adapted from Dufour et al. (2016)] are stated below.

Let

[1] For similar results on related problems, see Bontemps (2019) and Chernozhukov et al. (2018).

$$D_n(\theta) = [D_{1n}(\theta), \ldots, D_{mn}(\theta)]', \; n = 1, 2, \ldots \tag{4}$$

be a sequence of $m \times 1$ random vectors which are functions of a $p \times 1$ parameter vector θ, where $\theta \in \Theta \subseteq \mathbb{R}^p$ $(p \leq m)$ and Θ is a non-empty open subset of $\mathbb{R}^p$. There is a unique vector $\theta_0 \in \Theta$ called the "true parameter value". r_D, r_θ and r_M represent positive constants.

We first assume that the score $D_n(\theta)$ evaluated at $\theta = \theta_0$ converges to a Gaussian limit at the rate r_D.

Assumption 1 (Score asymptotic normality)

$$n^{r_D} D_n(\theta_0) \xrightarrow[n\to\infty]{L} \mathrm{N}[0, I(\theta_0)] \tag{5}$$

where $I(\theta_0)$ is a nonsingular $m \times m$ matrix.

Normality of the limiting distribution of $n^{r_D} D_n(\theta_0)$ is used to obtain an asymptotically chi-square test statistic. This condition is not necessary for the asymptotic negligibility of the estimation error of $\tilde{\theta}_n$ in the asymptotic distribution of the test statistic under the null hypothesis.

Assumption 2 (Convergence rate of the restricted estimator)
$\tilde{\theta}_n^0, \; n \geq 1$ *is a random sequence on* Θ *such that*

$$\|\tilde{\theta}_n^0 - \theta_0\| = O_p(n^{-r_\theta}) \tag{6}$$

under H_0 [*in* (1)].

Assumption 2 means that the scaled estimation error $n^{r^\theta}(\tilde{\theta}_n^0 - \theta_0)$ is stochastically bounded. The asymptotic distribution of $\tilde{\theta}_n^0$ need not be known nor even exist. The following two assumptions play a key role for allowing $r_\theta < r_D$.

Assumption 3 (Score expansion)

For any θ in some non-empty open neighborhood $\mathcal{V}_D(\theta_0)$ of θ_0, and for some $p \times 1$ nonnegative vector $\beta = (\beta_1, \ldots, \beta_p)'$ with nonnegative components ($\beta_i \geq 0, \; i = 1, \ldots, p$), we have:

$$D_n(\theta) = D_n(\theta_0; \omega) + J(\theta_0)\mathbf{n}^{-\beta}(\theta - \theta_0) + B_n(\theta, \theta_0) \quad a.s. \tag{7}$$

where $J(\theta)$ is an $m \times p$ (nonrandom) function of θ, $\mathbf{n}^{-\beta}$ is a $p \times p$ diagonal matrix with i-th entry $n^{-\beta_i}$, $i = 1, \ldots, p$, and the remainder $B_n(\theta, \theta_0)$ satisfies

$$r_n(\delta, \theta_0) = o_p(\delta) \tag{8}$$

where $\delta > 0$ and

$$r_n(\delta, \theta_0) := \sup\{\|B_n(\theta, \theta_0)\| : \theta \in \mathcal{V}_D(\theta_0) \text{ and } 0 < \|\theta - \theta_0\| \leq \delta\}. \tag{9}$$

In contrast with Assumption 3.4 in Dufour et al. (2016), Assumption 4 below imposes the existence of a second-order expansion for $D_n(\theta)$ around $\theta = \theta_0$. This allows the linear expansion term to be scaled by some orders of n. In standard cases, β is a zero vector, so that $\mathbf{n}^{-\beta}$ is an identity matrix. However, there are problems where taking $\beta = 0$ leads to rank deficiency of $J(\theta_0)$, for instance, when the model is defined by local moment conditions [Carroll et al. (1998), Gagliardini et al. (2011), Lewbel (2007), Xu (2020)]. The scaling matrix $\mathbf{n}^{-\beta}$ captures the rate at which the sensitivity of $D_n(\theta)$ with respect to $\theta - \theta_0$ decays near $\theta = \theta_0$.

Assumption 4 (Score second-order expansion)

Assumption 3 *holds with condition* (8) *replaced by*

$$r_n(\delta, \theta_0) = O_p(\delta^2) \tag{10}$$

where $\delta > 0$.

Assumption 4 strengthens Assumption 3.4 in Dufour et al. (2016), which only entails $B_n(\theta, \theta_0; \omega) = O_p(\|\theta - \theta_0\|)$. It is satisfied when $D_n(\theta)$ is twice differentiable with probability one, but it also covers a class of nonsmooth functions [see Bontemps (2019)]. We now consider regularity assumptions on the restriction function $\psi(\theta)$.

Assumption 5 (Restriction function: continuous differentiability)

The function

$$\psi(\theta) = [\psi_1(\theta), \ldots, \psi_{p_1}(\theta)]' \tag{11}$$

is a $p_1 \times 1$ *twice differentiable vector function of* θ *with bounded first derivative*

$$P(\theta) = [P_1(\theta)', \ldots, P_{p_1}(\theta)']' \tag{12}$$

where

$$P_l(\theta) = \frac{\partial \psi_l(\theta)}{\partial \theta'}, \quad l = 1, \ldots, p_1. \tag{13}$$

Assumption 6 (Restriction function: bound on second derivatives)

Under Assumption 5*, there exists some* $\epsilon > 0$ *such that an open ball* $\mathcal{N}_\epsilon(\theta_0)$ *of* θ_0 *with radius* ϵ *satisfies the following condition: for any* $\theta^* \in \mathcal{N}_\epsilon(\theta_0)$*, the derivative* $\mathbb{H}_l(\theta^*)$ *of* $P_l(\theta^*)$ *exists and is bounded for all* $l = 1, \ldots, p_l$*,* i.e.

$$\sup_{1 \le l \le p_1} \sup_{\theta^* \in \mathcal{N}_\epsilon(\theta_0)} \|\mathbb{H}_l(\theta^*)\| < C_P(\theta_0) \tag{14}$$

for some positive constant $C_P(\theta_0)$.

While Assumption 6 imposes additional smoothness on ψ [relative to Assumption 5 originally made in Dufour et al. (2016)], it accommodates empirically relevant equality constraints, such as linear and polynomial restrictions.

Assumption 7 (Lipschitz condition on $J(\theta)$)

There exists a nonempty open neighborhood $\mathcal{V}_{J,1}(\theta_0)$ *of* θ_0 *such that, for any* $\theta \in \mathcal{V}_{J,1}(\theta_0)$,

$$\|J(\theta) - J(\theta_0)\| \le C_J(\theta_0)\|\theta - \theta_0\| \quad \text{for some constant } C_J(\theta_0) > 0. \tag{15}$$

Assumption 8 (Consistency rates of the estimators of $(I(\theta_0), J(\theta))$ and W_n)
The following conditions hold:

1. $\{\tilde{I}_n, n \geq 1\}$ *is a sequence of* $p \times p$ *random matrices such that*

$$\|\tilde{I}_n - I(\theta_0)\| = o_p(1); \tag{16}$$

2. $\{\tilde{J}_n(\theta) : n \geq 1\}$ *is a sequence of* $m \times p$ *random matrices such that*

$$\sup_{\theta \in \mathcal{V}_{J,2}(\theta_0)} \|\tilde{J}_n(\theta) - J(\theta)\| = O_p(n^{-r_M}) \tag{17}$$

where $\mathcal{V}_{J,2}(\theta_0)$ *is a nonempty open neighborhood of* θ_0 *and* $r_M > 0$;

3. $\{W_n,\ n \geq 1\}$ *is a sequence of* $m \times m$ *symmetric nonsingular* (*random*) *matrices such that*

$$\|W_n - W_0\| = O_p(n^{-r_M}) \tag{18}$$

where W_0 *is a nonsingular nonrandom matrix.*

Assumption 7 characterizes the smoothness of $J(\theta)$ in the neighborhood of θ_0 through a Lipschitz condition, while Assumption 8 give rates of convergence for $\tilde{I}_n$, $\tilde{J}_n(\theta)$ and W_n. The estimator $\tilde{I}_n$ may involve corrections for heteroskedasticity and autocorrelation (HAC); for further discussion, see Dufour et al. (2016).

Assumption 9 (Relations between convergence rates)

$$r_\theta > r_D/2 \quad \textit{and} \quad r_\theta + r_M > r_D. \tag{19}$$

Assumption 9 imposes restrictions on the constants (r_D, r_θ, r_M). It can be rewritten in the form

$$r_\theta > \max\{r_D/2,\ r_D - r_M\}. \tag{20}$$

When $r_D = 1/2$ (the usual convergence rate of the empirical estimating function), it means that $r_\theta > \max\{0.25,\ 0.5 - r_M\}$. This assumption departs from Dufour et al. (2016), who consider the case where $r_D = r_\theta = 1/2$ [with $\|W_n - W_0\| = o_p(1)$]. The implications of Assumption 9 are further discussed after the main result of this paper (Theorem).

Assumption 10 (Full-rank matrices $I(\theta_0)$ and $J(\theta_0)$ and their estimators)
For any $n \geq 1$,

$$\text{rank}\,[I(\theta_0)] = \text{rank}\left[\tilde{I}_n\right] = m$$

and

$$\text{rank}\,[J(\theta_0)] = \text{rank}\left[\tilde{J}_n(\tilde{\theta}_n^0)\right] = p \tag{21}$$

for some nonnegative value of β *satisfying Assumption* 3 *with probability one.*

Assumption 11 (Non-degeneracy of restriction Jacobian)
For any $\theta \in \Theta$ *such that* $\psi(\theta) = 0$, *we have*

$$\text{rank}\,[P(\theta)] = p_1. \tag{22}$$

Assumption 10 requires that the matrices $I(\theta_0)$ and $J(\theta_0)$ as well as their estimators have full rank. Assumption 11 means that $P(\theta)$ has full rank for any θ which satisfies the null hypothesis.

Assumption 12 (Restricted estimator)
$\psi(\theta_0) = \psi(\tilde{\theta}_n^0) = 0$ and $\psi(\check{\theta}_n) = O_p(n^{-2r_\theta})$, where

$$\check{\theta}_n := \theta_0 + \boldsymbol{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0) \tag{23}$$

and β satisfies Assumption 10.

In the standard case where β is a zero vector, the above assumption holds trivially since $\tilde{\theta}_n^0 = \check{\theta}_n$. If β is not a zero vector, $\check{\theta}_n$ is as an (infeasible) estimator of θ_0 which strictly improves $\tilde{\theta}_n^0$ in terms of bias, *i.e.* $\|\check{\theta}_n - \theta_0\| < \|\tilde{\theta}_n^0 - \theta_0\|$ everywhere, while it may not satisfy the restriction $\psi(\theta) = 0$ exactly. Assumption 12 requires that the distance $\psi(\check{\theta}_n) - \psi(\theta_0)$ is of order n^{-2r_θ}. This condition can be checked, for example, when the order of bias of $\tilde{\theta}_n^0$ is known. In addition, for a restriction fixing a subvector of θ, $\psi(\tilde{\theta}_n^0) = 0$ implies $\psi(\check{\theta}_n) = 0$.

Set

$$\tilde{Q}_n := \tilde{Q}[W_n] = \tilde{P}_n[\tilde{J}_n W_n \tilde{J}_n]^{-1}\tilde{J}_n W_n \tag{24}$$

where $\tilde{J}_n := \tilde{J}_n(\tilde{\theta}_n^0)$ and $\tilde{P}_n := P(\tilde{\theta}_n^0)$. Consider now the $p_1 \times 1$ function

$$s_n(\theta) := \tilde{Q}_n D_n(\theta) \tag{25}$$

and the generalized $C(\alpha)$ statistic

$$PC(\tilde{\theta}_n^0; \psi) = n^{2r_D}\, s_n(\tilde{\theta}_n^0)' \left[\tilde{Q}_n\, \tilde{I}_n\, \tilde{Q}_n'\right]^{-1} s_n(\tilde{\theta}_n^0). \tag{26}$$

Under the above assumptions, we can establish the asymptotic distributions of $s_n(\tilde{\theta}_n^0)$ and $PC(\tilde{\theta}_n^0; \psi)$.

Theorem (Asymptotic distribution of generalized $C(\alpha)$ statistic)
Suppose that the assumptions 1 *to* 12 *are satisfied. Then, under H_0,*

$$n^{r_D}\, s_n(\tilde{\theta}_n^0) \xrightarrow[n\to\infty]{L} \mathrm{N}\big[0,\, Q(\theta_0)I(\theta_0)Q(\theta_0)'\big] \tag{27}$$

where $s_n(\theta)$ is defined in (25) *and*

$$Q(\theta_0) = P(\theta_0)\big[J(\theta_0)'W_0 J(\theta_0)\big]^{-1} J(\theta_0)'W_0. \tag{28}$$

Furthermore, the generalized $C(\alpha)$ statistic $PC(\tilde{\theta}_n^0; \psi)$ in (26) *is asymptotically distributed as $\chi^2(p_1)$. When $r_\theta \geq r_D$, the above assertions hold without Assumptions* 4 *and* 6.

This theorem generalizes Proposition 3.1 of Dufour et al. (2016), which establishes the asymptotic distribution of $PC(\tilde{\theta}_n^0; \psi)$ when $r_D = r_\theta = 1/2$. For $r_D = 1/2$, the result holds if $r_\theta > 1/4$, *i.e.* when $\tilde{\theta}_n^0$ converges at a rate faster than $n^{1/4}$ with $\tilde{Q}_n$ [in (24)] converging to $Q(\theta_0)$ at an appropriately fast rate. This is ensured by the condition

$$r_\theta + r_M > r_D \tag{29}$$

from Assumption 9. In the worse case where r_θ is arbitrarily close to $1/4$, (29) requires that r_M be also larger than $1/4$. On the other hand, when $r_\theta = 1/2$, (29) always holds when $r_M > 0$ so that the convergence rate of $\tilde{Q}_n$ can be arbitrarily slow.

Given the asymptotic normality of $s_n(\tilde{\theta}_n^0)$ in (27), an alternative $C(\alpha)$-type statistic can be constructed by considering a linear transformation of $s_n(\tilde{\theta}_n^0)$: by some $p^* \times p_1$ matrix R_S where $p^* \in \{1, \ldots, p_1\}$:

$$s_n^{(R_s)}(\tilde{\theta}_n^0) = R_s s_n(\tilde{\theta}_n^0). \tag{30}$$

Such a transformation puts on different weights on the p_1 restrictions specified by each element of ψ. A test statistic based on the estimating function $s_n^{(R_s)}(\tilde{\theta}_n^0)$ may yield power improvements against certain alternatives, while it can have lower power in other directions.

3 Testing derivatives of a conditional expectation function

In this section, we apply the generalized $C(\alpha)$ testing procedure to a hypothesis testing problem on derivatives of the conditional expectation function. For a pair of integrable random variables (Y, X), consider the nonparametric regression of Y of X:

$$Y = g(X) + \epsilon, \qquad \mathbb{E}[\epsilon \mid X] = 0, \tag{31}$$

where the conditional expectation function $g(x) = \mathbb{E}[Y \mid X = x]$ is infinitely continuously differentiable. For a positive integer m, denote by $g^{(m)}(x_0)$ the m-th derivative of $g(x)$ evaluated at $x = x_0$. We are interested in testing whether

$$g^{(m)}(x_0) = 0, \qquad \forall m \geq m_0, \tag{32}$$

i.e. all the derivatives of order higher than or equal to the positive integer m_0 are all zero. In particular, when $m_0 = 1$, the marginal effect of x on $g(x)$ at $x = x_0$ is zero under the null hypothesis. The case where $m_0 = 2$ corresponds to testing the local linearity of m at $x = x_0$. Furthermore, whether higher order derivatives are zero is often of interest, because it determines the order of the bias of kernel-based estimators [Calonico et al. (2018), Hall (1992)].

We formulate this testing problem in the local estimating equation framework [Carroll et al. (1998), Gagliardini et al. (2011), Lewbel (2007), Xu (2020)] and then apply the generalized $C(\alpha)$ test procedure. Further details on the local estimating equation are given in the Supplementary material. As shown below, the estimating equations and the restricted estimator of the model parameter converge at nonstandard possibly distinct rates. We only require a restricted estimator which converges at a faster rate than an unrestricted estimator. Such an estimator is often easier to compute. Further, our procedure allows for rank deficiency of the Jacobian matrix due to a faster convergence rate than the one for the estimating equations. This problem is common in the local moment equation setup. The Wald-type test considered by Antoine and Renault (2012) assumes that the moment equations and their derivatives converge at the same rate, so that it is not applicable in such a case. We circumvent this issue by appropriately scaling the Jacobian matrix by the diagonal matrix $\mathbf{n}^{-\beta}$ to derive a valid test statistic.

In what follows, we assume that $\sigma^2(x) = \mathrm{Var}(\epsilon \mid X = x)$ is continuous, and the density $f(x)$ of X is continuously differentiable, both bounded away from zero in the neighborhood of x_0. Let $\{h_n\}, \{h_n^*\}, \{h_n^{(f)}\}, \{h_n^{(\sigma)}\}$ be sequences of positive real numbers all converging to zero as $n \to \infty$. We assume that K is a symmetric probability density function with bounded support on $\mathbb{R}$, and set

$$\mu_l = \int z^l K(z)dz,\ \upsilon_l = \int z^l K^2(z)dz, \qquad l = 0, 1, \dots . \tag{33}$$

Note that $\mu_0 = 1$, and $\mu_l = \nu_l = 0$ for l odd.

For a positive integer M such that $M \geq m_0$, a Taylor expansion of $g(x)$ of order M in the neighborhood of $x = x_0$ yields:

$$g(x) \approx \sum_{s=0}^{M} \theta_{0,s}(x - x_0)^s \tag{34}$$

where

$$\begin{aligned}\theta_0 &= (\theta_{0,0}, \dots , \theta_{0,s}, \dots , \theta_{0,M})' \\ &= (g(x_0), \dots , \frac{1}{s!} g^{(s)}(x_0), \dots , \frac{1}{M!} g^{(M)}(x_0))'.\end{aligned} \tag{35}$$

Then, we have the local moment function:

$$m(\theta;\ Z) = Y - \sum_{s=0}^{M} \theta_s (X - x_0)^s \tag{36}$$

where $Z := (Y, X)'$; see the Supplementary material. We wish to test the hypothesis:[2]

$$H_0(m_0)\ :\ \theta_m = 0, \qquad \forall m\ \ (m_0 \leq m \leq M). \tag{37}$$

[2] Note that (37) is not equivalent to (32) since the former does not restrict the derivatives of order higher than M. As (37) implies (32), any valid test for the former also has correct size for the latter. On the other hand, the rejection probability against cases where (32) holds but $g^{(m^*)} \neq 0$ for some $m^* > M + 1$ [so that (32) does not hold] is also no greater than the size of a test for (37). However, the choice of M is typically implied by a particular problem of interest at hand, and these cases are often not practically important. For example, if one wishes to test whether the marginal effect of $m(x_0)$ is zero, it is sufficient to consider $m_0 = 1,\ M = 2$.

For the function $\psi : \mathbb{R}^M \to \mathbb{R}^{M-m_0+1}$ defined by

$$\psi(\theta) = (\theta_{m_0}, \theta_{m_0+1}, \ldots, \theta_M)', \tag{38}$$

the restriction (37) is equivalent to

$$\psi(\theta) = 0. \tag{39}$$

Given a set of i.i.d. observations $\{x_i, y_i\}_{i=1}^n$ and the local moment equations in (36), we can then consider the following $(M+1) \times 1$ vector of estimating functions:

$$D_n(\theta) = \begin{bmatrix} \frac{1}{nh_n} \sum_{i=1}^n \left[y_i - \sum_{s=0}^M \theta_s (x_i - x_0)^s\right] K\left(\frac{x_i - x_0}{h_n}\right) \\ \frac{1}{nh_n^2} \sum_{i=1}^n \left[y_i - \sum_{s=0}^M \theta_s (x_i - x_0)^s\right] (x_i - x_0) K\left(\frac{x_i - x_0}{h_n}\right) \\ \vdots \\ \frac{1}{nh_n^{M+1}} \sum_{i=1}^n \left[y_i - \sum_{s=0}^M \theta_s (x_i - x_0)^s\right] (x_i - x_0)^M K\left(\frac{x_i - x_0}{h_n}\right) \end{bmatrix}. \tag{40}$$

$D_n(\theta)$ can be interpreted as the vector of scaled first-order conditions of the objective function of the local polynomial regression of order M defined as

$$S_n^{(M)}(\theta) = \sum_{i=1}^n \left[y_i - \sum_{j=0}^M \theta_s (x_i - x_0)^s\right]^2 K\left(\frac{x_i - x_0}{h_n}\right) \tag{41}$$

with respect to $\theta = (\theta_0, \ldots, \theta_M)'$.

While an unrestricted estimator $\hat{\theta}_n$ of θ may be obtained by minimizing (41), a restricted estimator $\tilde{\theta}_n^0$ with a faster rate of convergence can be achieved by considering local polynomial regression of order $m_0 - 1$: this leads one to minimize

$$S_n^0(\delta) = \sum_{i=1}^n \left[y_i - \sum_{s=0}^{m_0-1} \delta_s (x_i - x_0)^s\right]^2 K\left(\frac{x_i - x_0}{h_n^*}\right) \tag{42}$$

with respect to $\delta := \left(\delta_0, \ldots, \delta_{m_0-1}\right)'$, yielding the estimate

$$\hat{\delta}_n = \left(\hat{\delta}_{n,0}, \ldots, \hat{\delta}_{n,m_0-1}\right)'. \tag{43}$$

The constrained estimator of θ is then $\tilde{\theta}_n^0 = (\tilde{\theta}_{n,0}^0, \ldots, \tilde{\theta}_{n,M}^0)'$ where

$$\tilde{\theta}_{n,s}^0 = \begin{cases} \hat{\delta}_{n,s}, & s = 0, \ldots, m_0 - 1, \\ 0, & s = m_0, \ldots, M. \end{cases} \tag{44}$$

Other choices of the bandwidth parameter h_n^* are possible. By Theorem 3.1 of Fan and Gijbels (1996), we have: if $h_n^* = O_p(n^{-1/(2m_0+3)})$, then

$$\|\tilde{\theta}_n^0 - \theta_0\| = O\big(n^{-2/(2m_0+3)}\big). \tag{45}$$

On the other hand, the unrestricted estimator $\tilde{\theta}_n^0$ converges at the slower rate $n^{-2/(2M_0+5)}$. This is one advantage of using a restricted estimator in this context.

On setting

$$h_n = \begin{cases} o_p\big(n^{-1/2(M+2)}\big), & \text{if } M \text{ is odd,} \\ o_p\big(n^{-1/(2M+3)}\big), & \text{if } M \text{ is even,} \end{cases} \tag{46}$$

we then get convergence to normality of the estimating function:

$$\sqrt{nh_n} D_n(\theta_0) \xrightarrow[n\to\infty]{L} \mathrm{N}\left[0,\ I(\theta_0)\right] \tag{47}$$

where $I(\theta_0)$ is a positive-definite matrix of size $M+1$ defined as

$$I(\theta_0) = \sigma^2(x_0) f(x_0) [v_{i+j-2}]_{1\le i,j\le M+1}. \tag{48}$$

Now, $D_n(\theta)$ can be expressed as

$$D_n(\theta) = D_n(\theta_0) + J_{\theta,n}(\theta_0)\mathbf{n}^{-\beta}(\theta - \theta_0) \tag{49}$$

where $J_{\theta,n}(\theta)$ is a symmetric matrix of size $M+1$ defined as

$$J_{\theta,n}(\theta) = \left[\frac{1}{nh_n^{k+j-1}} \sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h_n}\right)(x_i - x_0)^{k+j-2}\right]_{1\le k,j\le M+1} \tag{50}$$

and β is a $(M+1)$ dimensional vector satisfying[3]

$$\mathbf{n}^{-\beta} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & h_n & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & h_n^M \end{bmatrix}. \tag{51}$$

Then,

$$\|J_{\theta,n}(\theta_0) - J_\theta(\theta_0)\| = o_p\big((nh_n)^{-1/2}\big) \tag{52}$$

where

$$J_\theta(\theta_0) = f(x_0) \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_M \\ \mu_1 & \mu_2 & \cdots & \mu_{M+1} \\ \vdots & & \ddots & \vdots \\ \mu_M & \cdots & \mu_{2M-1} & \mu_{2M} \end{bmatrix}. \tag{53}$$

Hence, we can write

[3] While we assume here that h_n is of the form $n^{-\gamma}$ for $\gamma > 0$, a more general diagonal matrix of size $M+1$ which depends on n may replace $\mathbf{n}^{-\beta}$ as long as all its elements converge to zero.

$$D_n(\theta) = D_n(\theta_0) + J_\theta(\theta_0)\mathbf{n}^{-\beta}(\theta - \theta_0) + o_p\big((nh_n)^{-1/2}\big). \tag{54}$$

Consider estimators of $(\tilde{J}_{\theta,n}, \tilde{I}_{\theta,n})$ of $(J_\theta(\theta_0), I(\theta_0))$ given by:

$$\tilde{J}_{\theta,n} = \hat{f}(x_0)\begin{bmatrix} 1 & \mu_1 & \cdots & \mu_M \\ \mu_1 & \mu_2 & \cdots & \mu_{M+1} \\ \vdots & & \ddots & \vdots \\ \mu_M & \cdots & \mu_{2M-1} & \mu_{2M} \end{bmatrix}, \tag{55}$$

$$\tilde{I}_n = \hat{\sigma}^2(x_0)\hat{f}(x_0)[v_{i+j-2}]_{1\le i,j\le M+1}, \tag{56}$$

where $\big(\hat{f}(x_0), \hat{\sigma}^2(x_0)\big)$ is a consistent estimator of $\big(f(x_0), \sigma^2(x_0)\big)$. For example, $\hat{f}(x_0)$ may be a usual kernel density estimator

$$\hat{f}(x_0) = \frac{1}{nh_n^{(f)}}\sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h_n^{(f)}}\right). \tag{57}$$

Then, by choosing $h_n^{(f)}$ so that $h_n^{(f)} = O(n^{-1/5})$, we have

$$\hat{f}(x_0) - f(x_0) = O_p(n^{-2/5}). \tag{58}$$

Similarly, on taking

$$\hat{\sigma}^2(x_0) = \sum_{i=1}^{n} \frac{K\left(\frac{x_i - x_0}{h_n^{(\sigma)}}\right)e_i^2}{K\left(\frac{x_i - x_0}{h_n^{(\sigma)}}\right)} \tag{59}$$

where $e_i = y_i - \tilde{\theta}_0^0, i = 1, \ldots, n$, with $h_n^{(\sigma)} = O(n^{-1/5})$, we have:

$$\hat{\sigma}^2(x_0) - \sigma^2(x_0) = O_p(n^{-2/5}). \tag{60}$$

Under the above conditions,

$$\|\tilde{J}_{\theta,n} - J_\theta(\theta_0)\| = O_p(n^{-2/5}),\ \|\tilde{I}_n - I(\theta_0)\| = O_p(n^{-2/5}). \tag{61}$$

Finally, the derivative $\tilde{P}_n$ of $\psi(\theta)$ evaluated at $\tilde{\theta}_n$ is given by

$$\tilde{P}_n = \big[\mathbf{0}_{(M-m_0+1)\times m_0} \vdots \mathbf{1}_{(M-m_0+1)\times(M-m_0+1)}\big] \tag{62}$$

where $\mathbf{0}_{(M-m_0+1)\times m_0}$is a $(M - m_0 + 1) \times m_0$ zero matrix, and $\mathbf{1}_{(M-m_0+1)\times(M-m_0+1)}$ is an identity matrix of size $(M - m_0 + 1)$. On setting $W_n = \tilde{I}_n^{-1}$, the generalized $C(\alpha)$ statistic given in (26) can be expressed as

$$PC(\tilde{\theta}_n^0; \psi) = \frac{nh_n D_n(\tilde{\theta}_n^0)' \mathcal{U}^{-1} \tilde{P}_n' \big[\tilde{P}_n \mathcal{U}^{-1} \mathcal{V} \mathcal{U}^{-1} \tilde{P}_n\big]' \tilde{P}_n \mathcal{U}^{-1} D_n(\tilde{\theta}_n^0)}{\hat{f}(x_0)\hat{\sigma}^2(x_0)} \tag{63}$$

where $\mathcal{U} = [\mu_{i+j-2}]_{1\le i,j\le M+1}$ and $\mathcal{V} = [\nu_{i+j-2}]_{1\le i,j\le M+1}$. In particular, if $M = m_0 = 1$, we have

$$PC(\tilde{\theta}_n^0; \psi) = (nh_n)\frac{\nu_2}{\hat{f}(x_0)\hat{\sigma}^2(x_0)} \left(\frac{1}{nh_n^2}\sum_{i=1}^{n}\Big[y_i - \sum_{s=0}^{1}\tilde{\theta}_{n,s}^0 (x_i - x_0)^s\Big](x_i - x_0) K\left(\frac{x_i - x_0}{h_n}\right)\right)^2. \tag{64}$$

See the supplementary material for derivations. Then, by our Theorem, we have

$$PC(\tilde{\theta}_n^0; \psi) \xrightarrow[n\to\infty]{L} \chi^2(M - m_0 + 1). \tag{65}$$

This test procedure only requires estimation of the derivatives of order up to $m_0 - 1$. On the other hand, other methods, such as the Wald-type test, requires an unrestricted estimator so that all M derivatives need be estimated. Such an estimator suffers from a slower convergence than the restricted estimator in (44). In addition, we do not require that the bandwidth h_n^* for the restricted estimator and h_n for the test statistic be the same. This flexibility can be beneficial in practice. Finally, we allow for rank deficiency of the Jacobian matrix of $D_n(\theta)$ by appropriately scaling with the diagonal matrix $\mathbf{n}^{-\beta}$, which captures differences in convergence rates of $D_n(\theta)$ and its derivatives.

It is important to point out the limitation that the convergence in (65) is only pointwise, not uniform. Unless further restrictions are imposed, the testing problem considered in this section satisfies conditions of Theorem 1 of Romano (2004) [p. 570], so that there is no uniformly asymptotically valid test. See Kamat (2018) for a similar problem in the context of the regression discontinuity design, as well as Bahadur and Savage (1956), Dufour (1997, 2003), and Bertanha and Moreira (2020) for general results on "near non-identification" and (non-)testability. Modifications to allow uniform size validity (in finite samples or asymptotically) go beyond the scope of the present paper.

4 Conclusion

In this paper, we have studied generalized $C(\alpha)$ tests for restrictions on a finite-dimensional parameter when estimating equations and nuisance-parameter estimators converge at nonstandard rates. Under general conditions, we have shown that the null asymptotic distributions of the proposed test statistics are chi-square. In particular, the restricted estimator may converge at a rate slower than the one of the estimating equations. Such a case arises when an estimating function and the restricted estimator are based on different samples of vastly different sizes or, in the context of models defined by local moment conditions, different bandwidths may be employed to evaluate an estimating function and to construct the restricted estimator.

As discussed in Dufour et al. (2016), the generalized $C(\alpha)$ statistic nests existing $C(\alpha)$-type statistics as special cases. The results presented here broaden the applicability of such statistics to problems involving nonstandard rates. Further extensions as well as a large array of applications to specific statistical and econometric problems are discussed in Takano (2022) [Chapter 3].

As pointed out in Section 3, the distributional results presented in this paper constitute pointwise convergence properties, not finite-sample or uniform convergence results. It would be of interest to develop finite-sample procedures or give conditions for asymptotic size validity. Such extensions are left to later work.

Acknowledgements The authors thank Éric Renault, Victoria Zinde-Walsh, two anonymous referees, and the Editor Davy Paindaveine, for several useful comments. This work was supported by the William Dow Chair of Political Economy (McGill University), Monash University (Faculty of Business and Economics), the Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada.

References

B. Antoine and E. Renault. Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics*, 170(2):350–367, 2012.

R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 27(4):1115–1122, 1956.

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

I. V. Basawa. Neyman-LeCam tests based on estimating functions. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 811–825. Wadsworth, 1985.

A. Bera and Y. Bilias. 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, 2001.

A. K. Bera and M. J. Yoon. Specification testing with locally misspecified alternatives. *Econometric theory*, 9(4):649–658, 1993.

A. Berger and S. Wallenstein. On the theory of $C(\alpha)$-tests. *Statistics & Probability Letters*, 7(5):419–424, 1989.

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

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

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

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

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

M. Bertanha and M. J. Moreira. Impossible inference in econometrics: Theory and applications. *Journal of Econometrics*, 218(2):247–270, 2020.

B. R. Bhat and B. N. Nagnur. Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. *Biometrika*, 52(3-4):459–468, 1965. doi: 10.1093/biomet/52.3-4.459.

C. Bontemps. Moment-based tests under parameter uncertainty. *Review of Economics and Statistics*, 101(1):146–159, 2019.

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

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

Z. Cai, Y. Ren, and L. Sun. Pricing kernel estimation: A local estimating equation approach. *Econometric Theory*, 31(3):560–580, 2015.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

S. Calonico, M. D. Cattaneo, and M. H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

M. Caner. M-estimators with non-standard rates of convergence and weakly dependent data. *Journal of Statistical Planning and Inference*, 136(4):1207–1219, 2006.

R. J. Carroll, D. Ruppert, and A. H. Welsh. Local estimating equations. *Journal of the American Statistical Association*, 93(441):214–227, 1998.

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

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

H. Chernoff. Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16(1):31–41, 1964.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, 2018.

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

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

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

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

J.-M. Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387, 1997.

J.-M. Dufour. Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*, 36 (4):767–808, 2003.

J.-M. Dufour and M. G. Dagenais. Nonlinear models, rescaling and test invariance. *Journal of Statistical Planning and Inference*, 32(1):111–135, 1992.

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

J.-M. Dufour, A. Trognon, and P. Tuvaandorj. Generalized $C(\alpha)$ tests for estimating functions with serial dependence. In *Advances in Time Series Methods and Applications*, pages 151–178. Springer, 2016.

J.-M. Dufour, A. Trognon, and P. Tuvaandorj. Invariant tests based on m-estimators, estimating functions, and the generalized method of moments. *Econometric Reviews*, 36(1-3):182–204, 2017.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Routledge, 1996.

J. Fan, M. Farmen, and I. Gijbels. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):591–608, 1998.

S. Firpo, N. M. Fortin, and T. Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

R. V. Foutz. On the consistency of locally asymptotically most stringent tests. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 4(2): 211–219, 1976. ISSN 03195724.

P. Gagliardini, C. Gouriéroux, and E. Renault. Efficient derivative pricing by the extended method of moments. *Econometrica*, 79(4):1181–1232, 2011.

C. Gouriéroux and A. Monfort. *Statistics and econometric models*, volume 1. Cambridge University Press, 1995.

P. Hall. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, 20(2):675–694, 1992.

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

J. L. Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531, 1992.

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

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

V. Kamat. On nonparametric inference in the regression discontinuity design. *Econometric Theory*, 34(3):694–703, 2018.

J. Kim and D. Pollard. Cube root asymptotics. *Annals of Statistics*, 18(1):191–219, 1990.

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

L. Le Cam. On the asymptotic theory of estimation and testing hypotheses. In *Contribution to the Theory of Statistics*. University of California Press, 1956.

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

D. S. Lee. Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697, 2008.

D. S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

A. Lewbel. A local generalized method of moments estimator. *Economics Letters*, 94(1):124–128, 2007.

C. F. Manski. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228, 1975.

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

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

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

J. Neyman. Sur une famille de tests asymptotiques des hypothéses statistiques composées. *Trabajos de Estadistica*, 5(2):161–168, 1954. ISSN 0490-219X.

J. Neyman. Optimal asymptotic tests of composite hypotheses. *Probability and Statsitics*, pages 213–234, 1959.

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

J. Neyman and E. S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A(1/2):263–294, 1928.

C. Pal. Higher order $c(\alpha)$ tests with applications to mixture models. *Journal of Statistical Planning and Inference*, 113(1):179–187, 2003. ISSN 0378-3758.

S. R. Paul and R. K. Barnwal. 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):19–24, 1990. ISSN 00390526.

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

C. R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the*

Cambridge Philosophical Society, volume 44, pages 50–57. Cambridge University Press, 1948.

R. M. Ray. Maxmin $C(\alpha)$ tests against two-sided alternatives. *Annals of Statistics*, 2(6):1175–1188, 1974. ISSN 00905364.

J. P. Romano. On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scandinavian Journal of Statistics*, 31(4):567–584, 2004.

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

M. H. Seo and T. Otsu. Local m-estimation with discontinuous criterion for dependent and limited observations. *Annals of Statistics*, 46(1):344–369, 2018.

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

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

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

M. Takano. *Essays on Nonparametric and High-Dimensional Econometrics*. Ph.D. Thesis, Department of Economics, McGill University, Montréal, Canada, 2022.

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

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

R. E. Tarone and J. J. Gart. On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association*, 75(369):110–116, 1980. ISSN 01621459.

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

A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482, 1943.

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

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

J. M. Wooldridge. A unified approach to robust, regression-based specification tests. *Econometric Theory*, 6(1):17–43, 1990.

K.-L. Xu. Inference of local regression in the presence of nuisance parameters. *Journal of Econometrics*, 218(2):532–560, 2020.

Generalized C(α) tests with nonstandard convergence rates: Supplement

Jean-Marie Dufour and Masaya Takano

In this supplementary material, we provide a proof to Theorem in Section 2 and a brief review of probems defined by local estimating equations and moment conditions as well as detailed derivations of the test statistic in Section 3.

Proof (Theorem) The proof involves five major steps.
(a) There exists some $\epsilon > 0$ such that such that Assumption 6 holds and

$$\mathcal{N}_\epsilon(\theta_0) \subset \mathcal{V}_D(\theta_0) \cap \mathcal{V}_{J,1}(\theta_0) \cap \mathcal{V}_{J,2}(\theta_0). \tag{S–1}$$

By the assumptions 3, 6, 7 and 8.2, θ_0 belongs to the open convex set

$$\mathcal{V}(\theta_0) := \mathcal{V}_D(\theta_0) \cap \mathcal{N}_\epsilon(\theta_0) \cap \mathcal{V}_{J,1}(\theta_0) \cap \mathcal{V}_{J,2}(\theta_0). \tag{S–2}$$

On applying the convergence property given by Assumption 2, for any positive constant $\delta \in (0, 1)$, we can find a positive integer $n_0 = n_0(\delta)$ – which depends on δ – such that the event $\tilde{\theta}_n^0 \in \mathcal{V}(\theta_0)$ has probability at least $1 - \delta$ for $n > n_0(\delta)$, *i.e.*

$$\mathbb{P}[\tilde{\theta}_n^0 \in \mathcal{V}(\theta_0)] \geq 1-\delta \text{ for } n > n_0(\delta). \tag{S–3}$$

For δ given, define the event

$$A_\delta := \{\tilde{\theta}_n^0 \in \mathcal{V}(\theta_0)\ \ \forall n > n_0(\delta)\}. \tag{S–4}$$

Throughout the rest of the proof, we condition on A_δ and $n > n_0(\delta)$.

Jean-Marie Dufour
McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7 Canada, e-mail: jean-marie.dufour@mcgill.ca

Masaya Takano
Monash University, 29 Ancora Imparo Wy, Room 340, Clayton VIC 3168 Australia
e-mail: masaya.takano@monash.edu

By Assumption 7, we have:

$$\|J(\tilde{\theta}_n^0) - J(\theta_0)\| \leq C_J(\theta_0)\|\tilde{\theta}_n^0 - \theta_0\| \tag{S–5}$$

where $C_J(\theta_0)$ is a positive constants. Using the triangle inequality along with the assumptions 2 and 8.2, it follows that

$$\|\tilde{J}_n(\tilde{\theta}_n^0) - J\,(\theta_0)\,\| = O_p(n^{-\min(r_M,\, r_\theta)}). \tag{S–6}$$

Set $P_l(\theta) = [P_{l1}(\theta), \ldots, P_{lp}(\theta)]$. By Assumptions 5 and 6, and applying a Taylor expansion to each function $P_{lk}(\theta)$ in a neighborhood of θ_0, we can write:

$$P_{lk}(\tilde{\theta}_n^0) = P_{lk}\,(\theta_0) + \mathbb{H}_l(\theta^*_{n,lk})(\tilde{\theta}_n^0 - \theta_0), \quad k = 1, \ldots, p,\ l = 1, \ldots, p_1, \tag{S–7}$$

where $\theta^*_{n,lk}$ is some point between $\tilde{\theta}_n^0$ and θ_0 [in $\mathbb{R}^p$] – and thus in $\mathcal{N}_P(\theta_0)$ – hence

$$\begin{aligned}\|P(\tilde{\theta}_n^0) - P(\theta_0)\| &\leq p_1 \sup_{1\leq l\leq p_l}\ \sup_{\theta\in\mathcal{N}_P(\theta_0)} \|\mathbb{H}_l\,(\theta)\|\,\left\|\tilde{\theta}_n^0 - \theta_0\right\| \\ &\leq p_1\, C_P(\theta_0)\left\|\tilde{\theta}_n^0 - \theta_0\right\|, \ \text{ for } n > n_0(\delta),\end{aligned} \tag{S–8}$$

and, using Assumption 2,

$$\|P(\tilde{\theta}_n^0) - P(\theta_0)\| = O_p(n^{-r_\theta}). \tag{S–9}$$

From (S–6), (S–9) and Assumption 8.3, we get:

$$\|\tilde{Q}_n - Q(\theta_0)\| = O_p(n^{-\min(r_\theta,\, r_M)}). \tag{S–10}$$

By the assumptions 8.3, 10 and 11, $\tilde{Q}_n$ and $\tilde{I}_n$ have ranks p_1 and p respectively, so that

$$\text{rank}[\tilde{Q}_n\, \tilde{I}_n\, \tilde{Q}_n'] = p_1 \tag{S–11}$$

and $\tilde{Q}_n\, \tilde{I}_n\, \tilde{Q}_n'$ is invertible. Then, on using the assumptions 10 and 11, we have:

$$\|(\tilde{Q}_n\, \tilde{I}_n^{-1}\, \tilde{Q}_n')^{-1} - (Q_0\, I_0^{-1}\, Q_0')^{-1}\| \ = o_p\,(1)\,. \tag{S–12}$$

(b) We next show that

$$n^{r_\theta}\, P(\theta_0)\mathbf{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0) \xrightarrow[n\to\infty]{\mathsf{p}} 0. \tag{S–13}$$

By the definition of $\check{\theta}_n$, we have

$$\mathbf{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0) = \check{\theta}_n - \theta_0. \tag{S–14}$$

On using Assumption 6, $\psi(\check{\theta}_n)$ has the following expansion:

$$\psi(\check{\theta}_n) = \psi(\theta_0) + P(\theta_0)(\check{\theta}_n - \theta_0) + \mathbf{H}(\check{\theta}_n - \theta_0;\, \theta_n^*) \tag{S–15}$$

where $\mathbf{H}(\check{\theta}_n - \theta_0;\, \theta_n^*)$ is the $p_1 \times 1$ vector defined by

$$\mathbf{H}(\check{\theta}_n - \theta_0;\, \theta_n^*) := [\mathbf{H}_1(\check{\theta}_n - \theta_0;\, \theta_{1,n}^*),\, \ldots\, ,\, \mathbf{H}_{p_1}(\check{\theta}_n - \theta_0;\, \theta_{p_1,n}^*)], \tag{S–16}$$

$$\mathbf{H}_l(\check{\theta}_n - \theta_0;\, \theta_{l,n}^*) := (\check{\theta}_n - \theta_0)' \mathbb{H}_l\big(\theta_{l,n}^*\big)(\check{\theta}_n - \theta_0), \quad l = 1,\, \ldots\, ,\, p_1, \tag{S–17}$$

and each $\theta_{l,n}^* \in \Theta$ belongs to the line between $\check{\theta}_n$ and θ_0. By Assumption 12, we have $\psi(\theta_0) = 0$, so that

$$n^{r_D} P(\theta_0)(\check{\theta}_n - \theta_0) = n^{r_D}\psi(\check{\theta}_n) - n^{r_D}\mathbf{H}(\check{\theta}_n - \theta_0;\, \theta_n^*), \tag{S–18}$$

while Assumptions 2 and 6 imply that

$$\|\mathbf{H}(\check{\theta}_n - \theta_0;\, \theta_n^*)\| \le p_1 C_{P,\,\theta^*}\|\tilde{\theta}_n^0 - \theta_0\|^2 \tag{S–19}$$

with probability approaching one [by an argument similar to the one for (S–8)]. Since $\psi(\check{\theta}_n) = O_p(n^{-2r_\theta})$ [by Assumption 12], it follows from Assumption 9 that

$$\begin{aligned}\|n^{r_D} P(\theta_0)(\check{\theta}_n - \theta_0)\| &\le n^{r_D}\psi(\check{\theta}_n) + p_1 C_{P,\,\theta^*} n^{r_D}\|\tilde{\theta}_n^0 - \theta_0\|^2 \\ &= O_p(n^{r_D - 2r_\theta}) = o_p(1),\end{aligned} \tag{S–20}$$

which entails (S–13). Further, by the assumptions 3 and 4,

$$\begin{aligned} n^{r_D}\, \tilde{Q}_n D_n(\tilde{\theta}_n^0) = n^{r_D}\tilde{Q}_n D_n\,(\theta_0) + n^{r_D}\tilde{Q}_n J\,(\theta_0)\ \mathbf{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0) \\ + n^{r_D}\tilde{Q}_n B_n(\tilde{\theta}_n^0,\, \theta_0). \end{aligned} \tag{S–21}$$

(c) We will now show that

$$n^{r_D}\,[\tilde{Q}_n D_n(\tilde{\theta}_n^0) - Q(\theta_0)D_n(\theta_0)] \xrightarrow[n\to\infty]{\mathsf{p}} 0 \tag{S–22}$$

which, by Assumption 1 along with Slutsky's theorem, implies

$$n^{r_D}\tilde{Q}_n D_n(\tilde{\theta}_n^0) \xrightarrow[n\to\infty]{L} \mathrm{N}\big[0,\, Q(\theta_0)I(\theta_0)Q\,(\theta_0)'\,\big]. \tag{S–23}$$

We first note that Assumption 9 implies

$$r_D - \min(r_M,\, r_\theta) - r_\theta = r_D - [\min(r_M,\, r_\theta) + r_\theta] < 0. \tag{S–24}$$

Using (S–10), we can then write

$$\tilde{Q}_n J\,(\theta_0) = P(\theta_0) + O_p(n^{-\min(r_M,\, r_\theta)}) \tag{S–25}$$

hence, by (S–13),

$$\begin{aligned}\|n^{r_D}\tilde{Q}_n J\,(\theta_0)\ \mathbf{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0)\| \le \|n^{r_D} P(\theta_0)\mathbf{n}^{-\beta}(\tilde{\theta}_n^0 - \theta_0)\| \\ + \|O_p(n^{r_D - \min(r_M,\, r_\theta) - r_\theta})\| = o_p(1).\end{aligned} \tag{S–26}$$

(d) We have shown that

$$n^{r_D}\tilde{Q}_n D_n(\tilde{\theta}_n^0) = n^{r_D} Q(\theta_0) D_n(\theta_0) + o_p(1). \tag{S–27}$$

Hence, by (S–12),

$$PC(\tilde{\theta}_n^0;\, \psi) = n^{2r_\theta}\, D_n(\theta_0)' Q(\theta_0)' \left[Q(\theta_0) I(\theta_0) Q(\theta_0)'\right]^{-1} Q(\theta_0) D_n(\theta_0) + o_p(1), \tag{S–28}$$

where

$$\operatorname{rank}[Q(\theta_0) I(\theta_0) Q(\theta_0)'] = p_1. \tag{S–29}$$

(e) To prove the final claim, we observe that, under Assumption 3,

$$\|n^{r_D}\tilde{Q}_n B_n(\tilde{\theta}_n^0,\, \theta_0)\| \le n^{r_D}\|\tilde{Q}_n\|\|B_n(\tilde{\theta}_n^0,\, \theta_0)\| = O_p(n^{r_D - r_\theta}). \tag{S–30}$$

Then, by Assumption 5,

$$\psi(\tilde{\theta}_n^0) = \psi(\theta_0) + P(\theta_0)(\tilde{\theta}_n^0 - \theta_0) + B_p(\tilde{\theta}_n^0,\, \theta_0) \tag{S–31}$$

where

$$B_p(\tilde{\theta}_n^0,\, \theta_0) = o_p(\|\tilde{\theta}_n - \theta_0\|), \tag{S–32}$$

hence

$$\|n^{r_D} P(\theta_0)(\tilde{\theta}_n^0 - \theta_0)\| = o_p(n^{r_D - r_\theta}). \tag{S–33}$$

□

Local estimating equations and moment conditions

The finite-dimensional parameter of interest θ is often defined by local moment conditions of the form

$$\mathbb{E}[m(\theta_0;\ Z) \mid X = x_0] = 0 \tag{S–34}$$

where m is a finite-dimensional vector of moment equations which hold for the true value θ_0 of θ locally at a fixed value x_0 of the conditioning variable but not necessarily uniformly. Such an instance arises, for example, when θ is the value of a functional evaluated at x_0. More generally, local estimating equations are characterized by equations defined locally at a point. The framework of local estimating equations was first introduced by Carroll et al. (1998), and then extended by Xu (2020) to allow for non-smooth criterion functions and the presence of nuisance parameters . Lewbel (2007) considers inference based on local moment conditions in the generalized method of moments framework. Gagliardini et al. (2011) propose the extended method of moments, which accommodates both global and local moment restrictions.

When X has no mass at x_0, inference on θ based on local estimating equations is typically carried out by approximating (S–34) by kernel smoothing. For example, given a set of observations $\{z_i\}_{i=1}^n := \{(y_i, x_i)\}_{i=1}^n$, (S–34) may imply an estimating equation

$$D_n(\theta) = \frac{1}{nh_n}\sum_{i=1}^n h(\theta;\, z_i) K\left(\frac{x_i - x_0}{h_n}\right) \tag{S–35}$$

where K is a kernel function, and $h_n \to 0$ as $n \to \infty$. The convergence rate of the estimating equation (S–35) depends on the bandwidth parameter h_n and is slower than $n^{1/2}$. It is known that the convergence rate of an estimator of θ_0 based on the estimating equation (S–35) is slower than $n^{1/2}$, and each element of the estimator may converge to a non-degenerate distribution at a different rate as demonstrated in Section 3. Furthermore, the coefficient $J(\theta_0)$ of the linear expansion of $D_n(\theta)$ around $\theta = \theta_0$ may not have full rank without appropriately scaling by the diagonal matrix $\mathbf{n}^{-\beta}$ as shown in succeeding examples. For hypothesis testing in the local estimating function framework, Antoine and Renault (2012) consider a Wald-type test, and the empirical likelihood Lagrange multiplier test is proposed by Xu (2020).

The generalized $C(\alpha)$ test has the following advantages over these methods:

(1) Only a restricted estimator needs be estimated in contrast with the Wald test, which requires an unrestricted estimator. A restricted estimator is often easier to estimate as the applications illustrate.

(2) The bandwidth used in the test statistic may be different from the one(s) employed for the restricted estimator. The Wald test by Antoine and Renault (2012) implicitly requires undersmoothing to establish asymptotic normality of the GMM estimator. On the other hand, the restricted estimator may be estimated at an optimal rate.

(3) The Lagrange multiplier test only considers a class of null hypotheses where the parameter of interest takes some fixed hypothetical value.

Section 3 applies the proposed test procedure to a testing problem on the derivatives of the conditional expectation function. also provides applications to hypothesis testing in the regression discontinuity design [Calonico et al. (2014), Imbens and Lemieux (2008), Lee (2008), Lee and Lemieux (2010)] and the stochastic discount factor model [Cai et al. (2015)].

Derivation of the test statistic in Section 3

We have:

$$\tilde{J}_{\theta,n} = \hat{f}(x_0)\mathcal{U},\ \tilde{I}_{\theta,n} = \hat{\sigma}^2(x_0)\hat{f}(x_0)\mathcal{V}, \tag{S–36}$$

$$\tilde{P}_n = \left[\mathbf{0}_{(M-m_0+1)\times m_0} \vdots \mathbf{1}_{(M-m_0+1)\times(M-m_0+1)}\right],\ \tilde{W}_n = \tilde{I}_n^{-1}, \tag{S–37}$$

where $\mathcal{U} = [\mu_{i+j-2}]_{1\le i,j\le M+1}$ and $\mathcal{V} = [\nu_{i+j-2}]_{1\le i,j\le M+1}$. Then,

$$\tilde{Q}_n\,\tilde{I}_n\,\tilde{Q}_n' = \tilde{P}_n\left[\tilde{J}_{\theta,n}'\tilde{I}_{\theta,n}^{-1}\tilde{J}_{\theta,n}\right]^{-1}\tilde{P}_n' = \frac{\hat{\sigma}^2(x_0)}{\hat{f}(x_0)}\tilde{P}_n\mathcal{U}^{-1}\mathcal{V}\mathcal{U}^{-1}\tilde{P}_n' \tag{S–38}$$

hence

$$\left[\tilde{Q}_n \, \tilde{I}_n \, \tilde{Q}_n'\right]^{-1} = \frac{\hat{f}(x_0)}{\hat{\sigma}^2(x_0)} \left[\tilde{P}_n \mathcal{U}^{-1} \mathcal{V} \mathcal{U}^{-1} \tilde{P}_n'\right]^{-1}. \tag{S–39}$$

Note also that

$$\tilde{Q}_n D_n(\tilde{\theta}_n) = \tilde{P}_n \left[\tilde{J}_{\theta,n}' \tilde{I}_n^{-1} \tilde{J}_{\theta,n}\right]^{-1} \tilde{J}_{\theta,n}' \tilde{I}_n^{-1} D_n(\tilde{\theta}_n) = \frac{1}{\hat{f}(x_0)} \tilde{P}_n \mathcal{U}^{-1} D_n(\tilde{\theta}_n). \tag{S–40}$$

Since $D_n(\theta_0) = O_p\left((nh_n)^{-1/2}\right)$, we then have:

$$PC(\tilde{\theta}_n; \psi) = \frac{(nh_n) D_n(\tilde{\theta}_n^0)' \mathcal{U}^{-1} \tilde{P}_n' \left[\tilde{P}_n \mathcal{U}^{-1} \mathcal{V} \mathcal{U}^{-1} \tilde{P}_n'\right]^{-1} \tilde{P}_n \mathcal{U}^{-1} D_n(\tilde{\theta}_n^0)}{\hat{f}(x_0) \hat{\sigma}^2(x_0)}. \tag{S–41}$$

Suppose now that $M = m_0 = 1$. In this case,

$$\mathcal{U} = \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix}, \; \mathcal{V} = \begin{bmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{bmatrix}, \; \tilde{P}_n = \begin{bmatrix} 0 & 1 \end{bmatrix}, \tag{S–42}$$

$$\begin{aligned} D_n(\theta) &:= \begin{bmatrix} D_{1,n}(\theta) \\ D_{2,n}(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{nh_n} \sum_{i=1}^n \left[y_i - \sum_{s=0}^1 \theta_s (x_i - x_0)^s\right] K\left(\frac{x_i - x_0}{h_n}\right) \\ \frac{1}{nh_n^2} \sum_{i=1}^n \left[y_i - \sum_{s=0}^1 \theta_s (x_i - x_0)^s\right] (x_i - x_0) K\left(\frac{x_i - x_0}{h_n}\right) \end{bmatrix}, \end{aligned} \tag{S–43}$$

$$\tilde{P}_n \mathcal{U}^{-1} D_n(\tilde{\theta}_n) = \frac{\mu_2^{-1}}{nh_n^2} \sum_{i=1}^n \left[y_i - \sum_{s=0}^1 \tilde{\theta}_{n,s}^0 (x_i - x_0)^s\right] (x_i - x_0) K\left(\frac{x_i - x_0}{h_n}\right) \tag{S–44}$$

and

$$\tilde{P}_n \mathcal{U}^{-1} \mathcal{V} \mathcal{U}^{-1} \tilde{P}_n = \begin{bmatrix} 0 & \mu_2^{-1} \end{bmatrix} \begin{bmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{bmatrix} \begin{bmatrix} 0 \\ \mu_2^{-1} \end{bmatrix} = \mu_2^{-2} \nu_2. \tag{S–45}$$

On substituting the latter expressions into (S–41), we get:

$$PC(\tilde{\theta}_n^0; \psi) = (nh_n) \frac{\nu_2}{\hat{f}(x_0) \hat{\sigma}^2(x_0)} \left(\frac{1}{nh_n^2} \sum_{i=1}^n \left[y_i - \sum_{s=0}^1 \tilde{\theta}_{n,s}^0 (x_i - x_0)^s\right] (x_i - x_0) K\left(\frac{x_i - x_0}{h_n}\right)\right)^2. \tag{S–46}$$

References

B. Antoine and E. Renault. Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics*, 170(2):350–367, 2012.

R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 27(4):1115–1122, 1956.

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

I. V. Basawa. Neyman-LeCam tests based on estimating functions. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 811–825. Wadsworth, 1985.

A. Bera and Y. Bilias. 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, 2001.

A. K. Bera and M. J. Yoon. Specification testing with locally misspecified alternatives. *Econometric theory*, 9(4):649–658, 1993.

A. Berger and S. Wallenstein. On the theory of $C(\alpha)$-tests. *Statistics & Probability Letters*, 7(5):419–424, 1989.

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

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

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

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

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

M. Bertanha and M. J. Moreira. Impossible inference in econometrics: Theory and applications. *Journal of Econometrics*, 218(2):247–270, 2020.

B. R. Bhat and B. N. Nagnur. Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. *Biometrika*, 52(3-4):459–468, 1965. doi: 10.1093/biomet/52.3-4.459.

C. Bontemps. Moment-based tests under parameter uncertainty. *Review of Economics and Statistics*, 101(1):146–159, 2019.

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

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

Z. Cai, Y. Ren, and L. Sun. Pricing kernel estimation: A local estimating equation approach. *Econometric Theory*, 31(3):560–580, 2015.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

S. Calonico, M. D. Cattaneo, and M. H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

M. Caner. M-estimators with non-standard rates of convergence and weakly dependent data. *Journal of Statistical Planning and Inference*, 136(4):1207–1219, 2006.

R. J. Carroll, D. Ruppert, and A. H. Welsh. Local estimating equations. *Journal of the American Statistical Association*, 93(441):214–227, 1998.

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

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

H. Chernoff. Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16(1):31–41, 1964.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, 2018.

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

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

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

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

J.-M. Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387, 1997.

J.-M. Dufour. Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*, 36 (4):767–808, 2003.

J.-M. Dufour and M. G. Dagenais. Nonlinear models, rescaling and test invariance. *Journal of Statistical Planning and Inference*, 32(1):111–135, 1992.

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

J.-M. Dufour, A. Trognon, and P. Tuvaandorj. Generalized $C(\alpha)$ tests for estimating functions with serial dependence. In *Advances in Time Series Methods and Applications*, pages 151–178. Springer, 2016.

J.-M. Dufour, A. Trognon, and P. Tuvaandorj. Invariant tests based on m-estimators, estimating functions, and the generalized method of moments. *Econometric Reviews*, 36(1-3):182–204, 2017.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Routledge, 1996.

J. Fan, M. Farmen, and I. Gijbels. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):591–608, 1998.

S. Firpo, N. M. Fortin, and T. Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

R. V. Foutz. On the consistency of locally asymptotically most stringent tests. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 4(2): 211–219, 1976. ISSN 03195724.

P. Gagliardini, C. Gouriéroux, and E. Renault. Efficient derivative pricing by the extended method of moments. *Econometrica*, 79(4):1181–1232, 2011.

C. Gouriéroux and A. Monfort. *Statistics and econometric models*, volume 1. Cambridge University Press, 1995.

P. Hall. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, 20(2):675–694, 1992.

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

J. L. Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531, 1992.

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

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

V. Kamat. On nonparametric inference in the regression discontinuity design. *Econometric Theory*, 34(3):694–703, 2018.

J. Kim and D. Pollard. Cube root asymptotics. *Annals of Statistics*, 18(1):191–219, 1990.

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

L. Le Cam. On the asymptotic theory of estimation and testing hypotheses. In *Contribution to the Theory of Statistics*. University of California Press, 1956.

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

D. S. Lee. Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697, 2008.

D. S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

A. Lewbel. A local generalized method of moments estimator. *Economics Letters*, 94(1):124–128, 2007.

C. F. Manski. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228, 1975.

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

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

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

J. Neyman. Sur une famille de tests asymptotiques des hypothéses statistiques composées. *Trabajos de Estadistica*, 5(2):161–168, 1954. ISSN 0490-219X.

J. Neyman. Optimal asymptotic tests of composite hypotheses. *Probability and Statsitics*, pages 213–234, 1959.

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

J. Neyman and E. S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20A(1/2):263–294, 1928.

C. Pal. Higher order $c(\alpha)$ tests with applications to mixture models. *Journal of Statistical Planning and Inference*, 113(1):179–187, 2003. ISSN 0378-3758.

S. R. Paul and R. K. Barnwal. 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):19–24, 1990. ISSN 00390526.

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

C. R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge University Press, 1948.

R. M. Ray. Maxmin $C(\alpha)$ tests against two-sided alternatives. *Annals of Statistics*, 2(6):1175–1188, 1974. ISSN 00905364.

J. P. Romano. On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scandinavian Journal of Statistics*, 31(4):567–584, 2004.

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

M. H. Seo and T. Otsu. Local m-estimation with discontinuous criterion for dependent and limited observations. *Annals of Statistics*, 46(1):344–369, 2018.

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

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

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

M. Takano. *Essays on Nonparametric and High-Dimensional Econometrics*. Ph.D. Thesis, Department of Economics, McGill University, Montréal, Canada, 2022.

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

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

R. E. Tarone and J. J. Gart. On the robustness of combined tests for trends in proportions. *Journal of the American Statistical Association*, 75(369):110–116, 1980. ISSN 01621459.

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

A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482, 1943.

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

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

J. M. Wooldridge. A unified approach to robust, regression-based specification tests. *Econometric Theory*, 6(1):17–43, 1990.

K.-L. Xu. Inference of local regression in the presence of nuisance parameters. *Journal of Econometrics*, 218(2):532–560, 2020.