LSTA-2015-0069 (Accepted)
A New Robust and Most Powerful Test in the Presence of Local Misspecification*

Anil K. Bera†
University of Illinois

Gabriel Montes-Rojas‡
Universidad de San Andrés - CONICET and City University London

Walter Sosa-Escudero§
Universidad de San Andrés - CONICET

March 30, 2016

Abstract

This paper proposes a new test that is consistent, achieves correct asymptotic size and is locally most powerful under local misspecification, and when any $\sqrt{n}$-estimator of the nuisance parameters is used. The new test can be seen as an extension of the Bera and Yoon (1993)

*The authors would like to thank two anonymous referees for their valuable comments and constructive suggestions. However, we retain the responsibility for any remaining errors. Computer programs for the implementation of the proposed tests and simulation results for other sample sizes ara available from the authors.

†Department of Economics, University of Illinois, Urbana-Champaign, 1407 W. Gregory Drive, Urbana, IL 61801. Phone: +1 (217) 333-4596; Fax: +1 (217) 244-6678; e-mail: abera@illinois.edu.

‡Corresponding author: CONICET-Universidad de San Andrés - CONICET and Department of Economics, City University London, Northampton Square, London EC1V 0HB, UK, email: Gabriel.Montes-Rojas.1@city.ac.uk

§Department of Economics, Universidad de San Andrés and CONICET, email: wsosa@udesa.edu.ar

procedure that deals with non-ML estimation, while preserving its optimality properties. Similarly, the proposed test extends Neyman's (1959) $C(\alpha)$ test to handle locally misspecified alternatives. A Monte Carlo study investigates the finite sample performance in terms of size, power and robustness to misspecification.

JEL classification: C12; C52

Keywords: Specification testing; Rao's score test; Local misspecification; Neyman's C(α).

1 Introduction

A standard practice in applied econometrics is to start by estimating a small model and then checking whether departures away from it are supported or not by the data. Rao's (1948) score (henceforth, RS) or Lagrange multiplier tests are convenient since, unlike likelihood ratio and Wald tests, they require estimation of only the restricted model under the null hypothesis.

The performance of RS tests depends on how the model is estimated and on whether the alternative hypothesis is correctly specified. Consider a model consisting of a probability distribution characterized by three vectors of parameters: θ_1, θ_2 and θ_3. Suppose that the primary interest is to test $H_0^2 : \theta_2 = \theta_{20}$ in a situation where θ_1 can be easily estimated under the joint null $H_0^{23} : \theta_2 = \theta_{20},\ \theta_3 = \theta_{30}$. The properties of a test for H_0^2 derived in such context depend on 1) how θ_1 is estimated and 2) whether $H_0^3 : \theta_3 = \theta_{30}$ holds.

When θ_1 is estimated by maximum likelihood (ML) under the joint null H_0^{23} the RS test for H_0^2 is consistent, has correct asymptotic size and is locally most powerful when the alternative model is correctly specified, i.e., when H_0^3 holds and thus the only deviation away from the joint null is due to H_0^2

being false (see Rao (1948); Rao and Poti (1946), Cox and Hinkley (1974) and Bera and Bilias (2001a)). If any other $\sqrt{n}-$consistent estimator of θ_1 under H_0^{23} is used, Neyman's (1959) $C(\alpha)$ test is asymptotically equivalent to the RS and hence inherits alls its optimality properties (see Smith (1987) and Bera and Bilias (2001b)).

When the alternative hypothesis is misspecified ($H_0^3 : \theta_3 \neq \theta_{30}$), both RS and $C(\alpha)$ tests reject H_0^2 spuriously, as shown by Davidson and MacKinnon (1987) and Saikkonen (1989). That is, they reject H_0^2 not because of being false but due to the fact that H_0^3 does not hold. For example, Bera, Sosa-Escudero and Yoon (2001) find that the standard Breusch and Pagan (1979) test for random effects in the error component model spuriously rejects its null under the presence of serial correlation. Bera and Yoon (1993) (henceforth, BY) propose a modification of the RS test for H_0^2 that is still based on the ML estimation of θ_1 under H_0^{23}, but unlike RS and $C(\alpha)$ tests, is consistent and has correct asymptotic size under local misspecification. The BY test can be shown to be asymptotically equivalent to a $C(\alpha)$ test and hence it is also locally most powerful. The BY principle has been successfully implemented in many econometric 'model search' problems, for instance see Anselin, Bera, Florax and Yoon (1996), Godfrey and Veall (2000), Bera, Sosa-Escudero and Yoon (2001), Baltagi and Li (2001) and Montes-Rojas (2010, 2011).

The use of an ML estimator is an obvious restriction on the applicability of BY tests. Bera, Montes-Rojas and Sosa-Escudero (2010) (henceforth, BMS) extended the BY principle to the GMM framework, proposing a test that is consistent and has correct asymptotic size for any initial GMM estimator and under locally misspecified alterantives.

In Box's (1953) characterization, the $C(\alpha)$ and the BY tests possess the *robustness of efficiency* property (see Welsh (1996, pp. 242-243)), in the sense that both, size and power, are preserved with respect to the original RS test. On the contrary, the test suggested by BMS is only *validity robust*, since it preserves consistency and correct asymptotic size but not necessarily efficiency.

In this paper we propose a new test that is still based on any $\sqrt{n}$-consistent estimator of θ_1 and has the robustness of efficiency property under local misspecification. Consequently, the proposed test improves upon three existing strategies by a) allowing for non-ML estimation in the BY test, b) allowing for locally misspecified alternatives in the classic $C(\alpha)$ procedure, and c) restoring asymptotic efficiency of BMS test. Intuitively, the new test is derived by applying a double $C(\alpha)$-style correction that deals simultaneously with the non-ML estimation and locally misspecified alternatives.

The practical relevance of the proposed tests relates to situations where simple estimators for relevant parameters are readily available, as compared to fully ML estimators. Linear panel data error components models are one example of such scenario, where method-of-moments estimators of the variance components are much simpler to compute than ML estimators. For example, Baltagi, Song and Jung (2001, 2002) consider a nested error components model $y_{ijt} = x'_{ijt}\beta + u_{ijt}$ with $u_{ijt} = \mu_i + \nu_{ij} + \epsilon_{ijt}$. Normality of the error components is assumed to develop a testing framework for the appropiate nested variance structure. A RS test for the presence of the random effect μ_i (or ν_{ij}) being present requires the estimation of β and the variance of ν_{ij} (or μ_i) and ϵ_{ijt}. Baltagi et al. (2001) suggest that, even though a fully ML estimator is available under normality, much simpler method-of-moments estimators of the nuisance parameters are very good competitors. Moreover, tests for the presence of either μ_i or ν_{ij} are also constructed as BY robust test for local misspecification of the random component in the level not being tested (Baltagi, Song and Jung (2002)). This is a clear example of a situation where the tests proposed in this paper can be very useful in practice, since they can be based on any consistent estimate, bypassing the need of initial ML estimation. We discuss a second example of least-squares and quantile regression models in the context of our Monte Carlo study (in Section 4) where the finite sample size and power of the tests are studied.

The rest of the paper is organized as follows. In Section 2, we review the loss of efficiency associated with non-ML estimation of θ_1, the $C(\alpha)$ approach (that preserves size and power of RS tests) and a new intermediate 'modified RS' test, that only restores size. We then show in Section 3 that, as in the case of the RS test, though being able to accommodate non-ML estimators of θ_1, both strategies are negatively affected when the alternative hypothesis becomes misspecified. We thus introduce the new tests, that are resistant to non-ML estimators and locally misspecified alternatives. We complement our theoretical analysis with a Monte Carlo experiment in Section 4, that investigates the small sample performance of the tests. Section 5 concludes.

2 Testing with non-maximum likelihood estimators

Consider the following parametric model for independent and identically distributed (iid) random samples.

Assumption 1. *Parametric model:*
(i) Let $\{z_i\}_{i=1}^n$ *be a random sample of iid random vectors* $z_i \in \mathcal{Z} \subset \Re^K$*;*
(ii) let the parametric family of models for the density of z *be given by* $\{f(.|\theta) : \theta \in \Theta\}$ *where* $\Theta \subset \Re^p$ *is a compact set that can be partitioned*

as $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$*, subsets of* $\Re^{p_1}$*,* $\Re^{p_2}$ *and* $\Re^{p_3}$*,* $p = p_1 + p_2 + p_3$*, respectively with typical elements* $\theta \equiv (\theta_1', \theta_2', \theta_3')'$ *and* $f(.|\theta)$ *is a density function to the measure* $v(dz)$ *for all* $\theta \in \Theta$*;*
(iii) for some $\theta_0 \in int(\Theta)$*,* $\theta_0 = argmax_{\theta\in\Theta} E[l(z,\theta)]$ *is unique, where* $E[\cdot] \equiv \int_{\mathcal{Z}} \cdot f(z,\theta_0)v(dz)$ *and* $\ell(z,\theta) \equiv lnf(z|\theta)$*;*
(iv) for each $\theta \in \Theta$*,* $\ell(.,\theta)$ *is a Borel measureable function on* $\mathcal{Z}$*, and for each* $z \in \mathcal{Z}$*,* $\ell(z,.)$ *is a continous function on* Θ*.*

Define $\ell(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \ell(z_i,\theta)$ as the log-likelihood. Let $d(z,\theta) \equiv \partial l(z,\theta)/\partial\theta$ and $d(\theta) \equiv \partial\ell(\theta)/\partial\theta$ denote the score vector (we will use $d_j(z,\theta)$ and $d_j(\theta)$ to denote the corresponding $p_j \times 1$ subvector $\partial l(\theta)/\partial\theta_j$, with $j = 1, 2, 3$). Moreover, let $J(z,\theta) \equiv -\partial^2 l(z,\theta)/\partial\theta\partial\theta'$ be a $p \times p$ matrix of second partial derivatives, and

$$J(\theta) \equiv -E\left[\frac{\partial^2 l(z,\theta)}{\partial\theta\partial\theta'}\right] \equiv \begin{bmatrix} J_{11}(\theta) & J_{12}(\theta) & J_{13}(\theta) \\ J_{21}(\theta) & J_{22}(\theta) & J_{23}(\theta) \\ J_{31}(\theta) & J_{32}(\theta) & J_{33}(\theta) \end{bmatrix}$$

denote the information matrix. For notational convenience we write $J(\theta_0) \equiv J$, i.e., we omit the dependence on θ when the functionals are evaluated at θ_0.

Assumption 2. *Scores and information matrix:*
(i) $\ell(z,.)$ *is twice continuously differentiable on* $int(\Theta)$*;*
(ii) all elements in $\ell(z,\theta)$*,* $d(z,\theta)$*,* $d(z,\theta)d(z,\theta)'$*,* $J(z,\theta)$ *are bounded in absolute value by a function* $b(z)$ *with* $E[b(z)] < \infty$ *for all* $\theta \in \Theta$*;*
(iii) J *is positive definite.*

Assumptions 1 and 2 provide sufficient conditions for identification, $\sqrt{n}$-consistency and asymptotic normality of a ML estimator (MLE) for iid random samples. These correspond to the assumptions of theorems 13.1 and 13.2 in Wooldridge (2010) and assumptions 1-9 for score functions in Newey (1985).

Consider first the problem of testing $H_0^2 : \theta_2 = \theta_{20}$ under the local alternative $H_A^2 : \theta_2 = \theta_{20} + \delta_2/\sqrt{n}$, $0 < \delta_2 < \infty$, and when $H_0^3 : \theta_3 = \theta_{30}$ holds. In this case the alternative hypothesis is said to be *correctly specified*, in the sense that H_0^3 holds, i.e., the only departure away from the joint null $H_0^{23} : \theta_2 = \theta_{20},\ \theta_3 = \theta_{30}$ is due to θ_2 being different from θ_{20}. Under this set up, the form of the optimal RS test statistic is given by

$$RS_{2\cdot 1}(\theta) = n\ d_2(\theta)'\ J_{2\cdot 1}^{-1}\ d_2(\theta), \tag{1}$$

where $J_{2\cdot 1} = J_2 - J_{21} J_{11}^{-1} J_{12}$. Let $\hat{\theta} = (\hat{\theta}_1', \theta_{20}', \theta_{30}')'$, where $\hat{\theta}_1$ is the restricted MLE of θ_1 under the joint null H_0^{23}. A standard result is that under H_A^2, and H_0^3, $RS_{2\cdot 1}(\hat{\theta}) \xrightarrow{d} \chi_{p_2}^2(\lambda_{2\cdot 1})$, as $n \to \infty$, where the non-centrality parameter is $\lambda_{2\cdot 1} = \delta_2' J_{2\cdot 1} \delta_2$. Thus under H_0^2, $RS_{2\cdot 1}(\hat{\theta})$ has, asymptotically, a *central* chi-squared distribution, ensuring its correct asymptotic size. Also, as mentioned in the Introduction, $RS_{2\cdot 1}(\hat{\theta})$ is locally most powerful.

In certain contexts it might be difficult to obtain the MLE $\hat{\theta}_1$, while a $\sqrt{n}$-consistent estimator $\tilde{\theta}_1$ may be easily available. However, the use of a $\sqrt{n}$-consistent estimator other than the MLE affects the asymptotic properties of the RS test. Assume that an M-estimator $\tilde{\theta} = (\tilde{\theta}_1', \theta_{20}', \theta_{30}')'$ is available, defined as $\tilde{\theta}_1 = \operatorname{argmin}_{\theta_1 \in \Theta_1} \sum_{i=1}^n q(z_i, \theta_1, \theta_2, \theta_3)$ for $q(z,\theta)$ an objective function of the random vector z. Assume that a general estimating function $h_1(\theta) = \frac{1}{n} \sum_{i=1}^n h_1(z_i, \theta)$ exists, where $h_1(z,\theta) \equiv \partial q(z,\theta)/\partial \theta_1$, and that $\tilde{\theta}_1$ is the unique zero of $h_1(.)$ for all n and for all $(\theta_2, \theta_3) \in (\Theta_2 \times \Theta_3)$. For example, the restricted MLE corresponds to $h_1(\theta_1, \theta_{20}, \theta_{30}) = d_1(\theta_1, \theta_{20}, \theta_{30})$, so in this case $\tilde{\theta}_1 = \hat{\theta}_1$. Define $H_1(\theta) = E\left[h_1(z,\theta) h_1(z,\theta)'\right]$ and $B_1(\theta) = E\left[\partial h_1(z,\theta)/\partial \theta_1\right]$. For notational convenience we omit the dependence on

θ when the functionals are evaluated at θ_0. We will consider the following assumptions:

Assumption 3. *M-estimators:*
(i) $\theta_{10} = argmin_{\theta_1 \in \Theta_1} E[q(z, \theta_1, \theta_{20}, \theta_{30})]$ *is unique and* $E[h_1(z, \theta_1, \theta_{20}, \theta_{30})] = 0$ *only if* $\theta_1 = \theta_{10}$*;*
(ii) for each $\theta \in \Theta$*,* $(q(.,\theta), h_1(.,\theta))$ *are Borel measureable functions on* $\mathcal{Z}$*, and for each* $z \in \mathcal{Z}$*,* $(q(z,.), h_1(z,.))$ *are continous functions on* Θ*;*
(iii) $q(z,\theta)$ *is twice continuously differentiable on* $int(\Theta)$*;*
(iv) all elements in $q(z,\theta)$*,* $h_1(z,\theta)$*,* $h_1(z,\theta)h_1(z,\theta)'$*,* $h_1(z,\theta)d(z,\theta)'$*,* $\partial h_1(z,\theta)/\partial\theta$ *are bounded in absolute value by a function* $b(z)$ *with* $E[b(z)] < \infty$ *for all* $\theta \in \Theta$*;*
(v) B_1 *is positive definite;*
(vi) let $w(z,\theta) = [h_1(z,\theta)' \ \ d(z,\theta)']'$*,* $E[w(z,\theta_0)w(z,\theta_0)']$ *is positive definite.*

Assumption 3, together with 1-2, guarantees identification, $\sqrt{n}$-consistency and asymptotic normality of $\tilde{\theta}_1$ under H_0^{23}, given by $\sqrt{n}(\tilde{\theta}_1 - \theta_{10}) \xrightarrow{d} N(0_{p_1}, B_1^{-1}H_1B_1^{-1})$, as $n \to \infty$. See Wooldridge (2010, ch.12) for a general discussion on M-estimators. Assumptions 1-3 correspond to assumptions 1-9 in Newey (1985), in which case $\tilde{\theta}_1$ is defined as a Z-estimator based on the estimating function $h_1(z,\theta)$. Dependent random vectors (i.e. time-series) can be addressed with the use of the statistical model of Newey and West (1987). Moreover, the assumptions can be relaxed for non-smooth log-likelihood or q-objective functions (eg. quantile regression models) following Newey and McFadden (1994). For the sake of brevity we do not discuss the standard regularity conditions for consistency of estimators for J, B_1 and H_1, and we assume that all matrices that need to be inverted in the construction of the statistics in this paper are non-singular.

$RS_{2\cdot 1}(\tilde{\theta})$ is no longer asymptotically chi-squared distributed, since it is based on an incorrect variance. The correct variance of $d_2(\tilde{\theta})$ is $V_{2\cdot 1} = J_2 - J_{21}B_1^{-1}H_1B_1^{-1}J_{12}$, which can be easily derived as in Newey and McFadden (1994) using the delta method. Consider the following modified RS test using the correct variance of the score function:

$$\widetilde{RS}_{2\cdot 1}(\theta) = n\ d_2(\theta)'\ V_{2\cdot 1}^{-1}\ d_2(\theta). \tag{2}$$

The following result establishes the consistency and asymptotic validity of this test, where θ is now replaced by a non-MLE $\tilde{\theta}$.

Theorem 1. *Consider Assumptions 1-3. When H_0^3 is true and H_A^2 holds and $n \to \infty$,*

$$\widetilde{RS}_{2\cdot 1}(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\tilde{\lambda}_{2\cdot 1}),$$

with $\tilde{\lambda}_{2\cdot 1} = \delta_2' V_{2\cdot 1}\delta_2$.

Proof: In the Appendix.

Though consistent and with correct asymptotic size, $\widetilde{RS}_{2\cdot 1}(\tilde{\theta})$ is less powerful than $RS_{2\cdot 1}(\hat{\theta})$. Note that $\lambda_{2\cdot 1} - \tilde{\lambda}_{2\cdot 1} = \delta_2'\Big(J_{2\cdot 1} - V_{2\cdot 1}\Big)\delta_2$. The asympotic efficiency of the MLE of θ_1 implies that $J_1^{-1} - B_1^{-1}H_1B_1^{-1}$ is negative semi-definite, thus $J_{2\cdot 1} - V_{2\cdot 1}$ is positive semi-definite, and hence $\lambda_{2\cdot 1} - \tilde{\lambda}_{2\cdot 1} \geq 0$.

An optimal test for H_0^2 when any $\sqrt{n}$-consistent estimator of θ_1 under H_0^{23} is used, can be based on Neyman's (1959) $C(\alpha)$ test statistic

$$C_{2\cdot 1}(\theta) = n\ d_{2\cdot 1}(\theta)'\ J_{2\cdot 1}^{-1}\ d_{2\cdot 1}(\theta), \tag{3}$$

where $d_{2\cdot 1}(\theta) = d_2(\theta) - J_{21}J_{11}^{-1}d_1(\theta)$ is known as the *effective score*. A well known result is that $C_{2\cdot 1}(\tilde{\theta})$ is asymptotically equivalent to $RS_{2\cdot 1}(\hat{\theta})$, and hence it has correct asymptotic size and is also locally most powerful. Intuitively, the $C(\alpha)$ test replaces the score of the test parameters θ_2 by its projection on the orthogonal complement of the space spanned by the score of the nuisance parameters θ_1, evaluated at $\tilde{\theta}$. And it does so in such way that replacing the MLE $\hat{\theta}$ by $\tilde{\theta}$ does not lead to any loss in asymptotic power. It is relevant to remark that $C_{2.1}(\hat{\theta}) = RS_{2.1}(\hat{\theta})$, since $d_{2.1}(\hat{\theta}) = d_2(\hat{\theta})$ due to $d_1(\hat{\theta}) = 0$.

3 Testing under local misspecification

Suppose that H_0^2 is true but the alternative hypothesis is locally misspecified, that is, $H_A^3 : \theta_3 = \theta_{30} + \delta_3/\sqrt{n}$, $0 < \delta_3 < \infty$ holds. Davidson and MacKinnon (1987) and Saikkonen (1989) show that in such case $RS_{2\cdot 1}(\hat{\theta}) \xrightarrow{d} \chi^2_{p_2}(\lambda_{2/3\cdot 1})$, where $\lambda_{2/3\cdot 1} = \delta_3' J_{32\cdot 1} J_{2\cdot 1}^{-1} J_{23\cdot 1} \delta_3$ with $J_{23\cdot 1} \equiv J_{23} - J_{21} J_{11}^{-1} J_{13} = J_{32\cdot 1}'$. That is, even when H_0^2 is true, $RS_{2\cdot 1}(\hat{\theta})$ has a *non-central* chi-squared distribution due to $\theta_3 \neq \theta_{30}$, and hence leading to spurious rejections of H_0^2 due to misspecification and not to its falseness. Naturally this result affects Neyman's C(α) test alike, since it is asymptotically equivalent to $RS_{2\cdot 1}(\hat{\theta})$.

Bera and Yoon (1993) proposed the following modified test

$$RS^*_{2\cdot 1}(\theta) = n\ d_2^*(\theta)'\ J^{-1}_{2(3)\cdot 1}\ d_2^*(\theta), \tag{4}$$

where $d_2^*(\theta) = d_2(\theta) - J_{23\cdot 1} J_{3\cdot 1}^{-1} d_3(\theta)$ and $J_{2(3)\cdot 1} = J_{2\cdot 1} - J_{23\cdot 1} J_{3\cdot 1}^{-1} J_{32\cdot 1}$. Their key result is that under H_0^2 and when H_A^3 holds, $RS^*_{2\cdot 1}(\hat{\theta}) \xrightarrow{d} \chi^2_{p_2}(0)$. That is, the BY test has asymptotic *centered* chi-squared distribution even when H_0^3 is false (in a local sense), hence it does not led to spurious rejections induced by misspecification. It is relevant to remark that both $RS^*_{2\cdot 1}(\hat{\theta})$ and $RS_{2\cdot 1}(\hat{\theta})$ are based on $\hat{\theta}$, the MLE of θ under the joint null H_0^{23}, and hence the use of the robustified test statistic shares all the computational advantages of the standard RS test. See Bera, Montes-Rojas and Sosa-Escudero (2009) for a geometrical interpretation of these results.

A quick inspection of the expressions of the $C(\alpha)$ and the BY test statistics respectively, in (3) and (4), suggests strong similarities between them, specially in terms of orthogonalization, i.e., in calculating the effective score. The most interesting fact is that the structure of orthogonalization is the same for replacing an MLE by a $\sqrt{n}-$consistent estimator of θ_1, and for taking account of local misspecification relating to the parameter θ_3.

Regarding power, the asymptotic distribution of $RS^*_{2\cdot 1}(\hat{\theta})$ under H^2_A is non-central chi-squared with non-centrality parameter $\lambda^*_{2\cdot 1} = \delta_2' J_{2(3)\cdot 1}\delta_2$. Note that when H^2_A and H^3_A are true, $\lambda^*_{2\cdot 1} = \lambda_{2\cdot 13} + o_p(1/\sqrt{n})$, where $\lambda_{2\cdot 13}$ is the non-centrality parameter of a RS test for H^2_0 when both (θ_1, θ_3) are estimated by MLE. Consequently, the BY test restores consistency and correct asymptotic size under misspecified alternatives, with no power loss compared to the standard RS that estimates θ_1 and θ_3 by MLE. Similarly, note that $(\hat{\theta}_1, \theta_{30})$ is trivially a $\sqrt{n}$-consistent estimator of (θ_1, θ_3) under H^2_0 and H^3_A, hence $RS^*_{2\cdot 1}(\hat{\theta}) \stackrel{a}{=} C_{2\cdot 13}(\hat{\theta})$, where $C_{2\cdot 13}(\theta)$ is defined analogously as in (3).

Nevertheless, the BY test requires the use of the MLE for θ_1. A simple modification that can handle any $\sqrt{n}$-consistent estimator for θ_1 based on $h_1(\theta)$ is as follows. Define $B_{j\cdot 1} = J_j - J_{j1}B_1^{-1}H_1B_1^{-1}J_{1j}, \quad j = 2, 3, 23, 32,$ where the subindex 23 (similarly 32) is used to label the redefined parameter $\theta_{23} = (\theta_2', \theta_3')'$. In order to account for the effect of H^3_A consider the adjusted score $\tilde{d}^*_{2\cdot 1}(\theta) = d_2(\theta) - B_{23\cdot 1}B_{3\cdot 1}^{-1}d_3(\theta)$. Now, following BY, consider the adjusted RS statistic:

$$\widetilde{RS}^*_{2\cdot 1}(\theta) = n\, \tilde{d}^*_{2\cdot 1}(\theta)' V^{-1}_{2(3)\cdot 1}\tilde{d}^*_{2\cdot 1}(\theta), \tag{5}$$

where $V_{2(3)\cdot 1} = B_{2\cdot 1} - B_{23\cdot 1}B_{3\cdot 1}^{-1}B_{32\cdot 1}$ is the variance of $\tilde{d}^*_{2\cdot 1}(\theta)$.

The next theorem establishes the properties of a locally size-robust 'modified' BY test under non-MLE estimation of θ_1.

Theorem 2. *Consider Assumptions 1-3.*

(i) When H_0^2 is true, but H_A^3 holds, as $n \to \infty$

$$\widetilde{RS}_{2\cdot 1}(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\tilde{\lambda}_{2/3\cdot 1}),$$

with $\tilde{\lambda}_{2/3\cdot 1} = \delta_3' \left(J_{23} - J_{21}B_1^{-1}H_1B_1^{-1}J_{13}\right)' V_{2\cdot 1}^{-1} \left(J_{23} - J_{21}B_1^{-1}H_1B_1^{-1}J_{13}\right) \delta_3.$
(ii) When H_A^2 and H_A^3 hold, as $n \to \infty$

$$\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\tilde{\lambda}^*_{2\cdot 1}),$$

where $\tilde{\lambda}^*_{2\cdot 1} = \delta_2' V_{2(3)\cdot 1}\delta_2.$

Proof: In the Appendix.

The main result of this paper is that a fully modified size and power robust test can be derived to accommodate non-ML estimators and misspecified alternatives. Define $d^*_{2\cdot 1}(\theta) = d_{2\cdot 1}(\theta) - J_{23\cdot 1}J_{3\cdot 1}^{-1}d_{3\cdot 1}(\theta)$ and

$$C^*_{2\cdot 1}(\theta) = n\, d^*_{2\cdot 1}(\theta)' J^{-1}_{2(3)\cdot 1} d^*_{2\cdot 1}(\theta), \tag{6}$$

where $d_{3.1}(\theta) = d_3(\theta) - J_{31}J_{11}^{-1}d_1(\theta)$ analogously as $d_{2.1}(\theta)$ in $C_{2.1}(\theta)$ in (3). The asymptotic properties of this new test are established in the following theorem.

Theorem 3. *Consider Assumptions 1-3. When H_A^2 and H_A^3 hold and $n \to \infty$*

$$C^*_{2\cdot 1}(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\lambda^*_{2\cdot 1}).$$

Proof: In the Appendix.

The optimality of the new procedure is due to the fact the theorem implies that $C^*_{2\cdot 1}(\tilde{\theta})$ is asymptotically equivalent to $RS^*_{2\cdot 1}(\hat{\theta})$. This equivalence is analog to that between $RS_{2\cdot 1}(\hat{\theta})$ and $C_{2\cdot 1}(\tilde{\theta})$ in Section 2 when the alternative hypothesis is correctly specified. Consequently, this new test has both the robustness of validity and efficiency properties when a non-MLE of θ_1 is used and when the alternative hypothesis is locally misspecified. Also note that $C^*_{2\cdot1}(\hat{\theta}) = RS^*_{2.1}(\hat{\theta})$. The improvement from $RS^*_{2\cdot1}(\theta)$ to $C^*_{2\cdot1}(\theta)$ is achieved by starting with $d_{2\cdot1}(\theta)$ and $d_{3\cdot1}(\theta)$ instead of $d_2(\theta)$ and $d_3(\theta)$, respectively, to take account of the fact that for the non-MLE $d_1(\tilde{\theta}) \neq 0$. We can also view $C^*_{2\cdot1}(\theta)$ as a modification of our initial C(α) statistic $C_{2\cdot1}(\theta)$ in (3), by further orthogonalizing $d_{2.1}(\theta)$, now with respect to $d_{3.1}(\theta)$ to incorporate the fact that $d_3(\tilde{\theta}) \neq 0$. This duality goes back to our earlier observation that two orthogonalizations for taking care of the $\sqrt{n}$-consistent estimation of θ_1 (as in C(α)) and for allowing for the possible local presence of θ_3 (as in BY) are structurally the *same*.

4 Monte Carlo experiments

We present the results of a simple but illustrative empirical exploration of the costs and benefits of the alternative robustification strategies discussed earlier. Consider the following regression model:

$$y_i = \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i} + u_i, \quad i = 1, 2, ..., n \tag{7}$$

with

$$x_{1i} = a_i + e_{1i}, \;\; x_{2i} = a_i + e_{2i}, \;\; x_{3i} = a_i + e_{3i},$$

and

$$u_i, a_i, e_{1i}, e_{2i}, e_{3i} \sim \text{ iid } N(0,1).$$

We use $\theta_1 = 1$, $n = 100$ and, we consider 1000 replications. Results for other sample sizes are very similar qualitatively, and are available from the authors. All tests are based on a nominal size of 0.05.

Using the framework discussed in the previous sections, the joint null $H_0^{23} : \theta_2 = 0,\ \theta_3 = 0$ corresponds to a simple regression model, i.e. $y_i = \theta_1 x_{1i} + u_i$. The restricted MLE, $\hat{\theta} = (\hat{\theta}_1, 0, 0)$, is the ordinary least-squares (OLS) estimator of θ_1 that regresses y on x_1. In order to evaluate the performance of the tests under alternative consistent estimators, we have considered the 0.1 quantile regression estimator of θ_1, $\tilde{\theta} = (\tilde{\theta}_1, 0, 0)$. The error term u is generated independently of x_1, x_2 and x_3, and identically across all observations, which implies a simple location-shift model. Consequently, the quantile regression estimator for any quantile is consistent for θ_1. We use the 0.1 quantile in order to produce a consistent though inefficient non-MLE. Note that any quantile could have been selected, and that this particular estimator will be asymptotically efficient if u follows an asymmetric Laplace distribution with location parameter at the 0.1 quantile of its distribution. In fact the reverse analysis can be implemented, that is, when the data is generated using the asymmetric Laplace distribution and then a consistent but inefficient estimator would be the OLS estimator. The score functions and the tests implemented below would then be based on the influence function of the quantile regression estimator at 0.1-quantile. The availability of a multitude of $\sqrt{n}$-consistent estimators can be viewed as a drawback of the $C(\alpha)$ test. While all will lead to asymptotic equivalent tests, their finite sample behavior could be quite different.

In this setup, the correlation between any pair of explanatory variables is 0.5, therefore, a test for $H_0^2 : \theta_2 = 0$ based on either $\hat{\theta}$ or $\tilde{\theta}$ will be affected by misspecification in θ_3 (i.e. $\theta_3 \neq 0$). This is a simple omitted variable setup, where leaving x_3 out of the model affects both the estimate of θ_1 and the test for θ_2. A simple way to see this is to consider a Wald test statistic for H_0^2, which is based on the OLS estimate of θ_2. This non-robustness can also be seen from the non-singularity of the matrix $J_{23\cdot 1}$.

The results that evaluate the performance of alternative tests, for different estimators and values of θ_2 and θ_3, are presented in Table 1. For part (a) we generated data using the joint null $\theta_2 = \theta_3 = 0$; for part (b) we considered $\theta_2 > 0, \theta_3 = 0$, and finally, part (c) is based on data with $\theta_2 = 0, \theta_3 > 0$. The first four columns present tests for the single hypothesis H_0^2 without any correction for whether H_0^3 is valid or not. $RS_{2\cdot 1}(\hat{\theta})$ is constructed using the restricted MLE; $RS_{2\cdot 1}(\tilde{\theta})$ and $\widetilde{RS}_{2\cdot 1}(\tilde{\theta})$ use a non-MLE; and $C_{2\cdot 1}(\tilde{\theta})$ is the $C(\alpha)$ test using a non-MLE. Note that $C_{2\cdot 1}(\hat{\theta}) = RS_{2\cdot 1}(\hat{\theta})$ by definition of MLE. The last four columns present tests for the single hypothesis H_0^2 but correcting for local departures from H_0^3. $RS^*_{2\cdot 1}(\hat{\theta})$ is the BY test using the restricted MLE; $RS^*_{2\cdot 1}(\tilde{\theta})$ and $\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta})$ are the BY tests using a non-MLE; and $C^*_{2\cdot 1}(\tilde{\theta})$ is our proposed fully robust test using a non-MLE. All test statistics are based on the score functions derived from the Gaussian log-likelihood. Therefore each score is of the form $d_j(\theta) = \frac{1}{n}\sum_{i=1}^n x'_{ji}u_i(\theta)$, $j = 1, 2, 3$, where $u_i(\theta) = y_i - \theta_1 x_{1i} - \theta_2 x_{2i} - \theta_3 x_{3i}$. Each element in J_{jh}, $j, h = 1, 2, 3$ is estimated by the outer product of gradients $\frac{1}{n}\sum_{i=1}^n d_{ji}(z_i, \theta) d_{hi}(z_i, \theta)'$ where $d_{ji}(z_i, \theta) = x'_{ji}u_i(\theta)$, $j = 1, 2, 3$, $z_i = (y_i, x_{1i}, x_{2i}, x_{3i})$. Finally, $B_1^{-1}H_1B_1^{-1}$ is given by the variance of the 0.1-quantile regression estimator.

When $\theta_2 = \theta_3 = 0$ holds (part (a)), as expected, $RS_{2\cdot 1}(\hat{\theta})$, $\widetilde{RS}_{2\cdot 1}(\tilde{\theta})$ and $C_{2\cdot 1}(\tilde{\theta})$ have correct empirical size, while $RS_{2\cdot 1}(\tilde{\theta})$ has an empirical size that is more than twice of the nominal size and much larger than that of its counterparts implemented with the correct variance. Similar results are found for the BY statistics. That is, the size of $RS^*_{2\cdot 1}(\tilde{\theta})$ is also quite high while that of $RS^*_{2\cdot 1}(\hat{\theta})$, $\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta})$ and $C^*_{2\cdot 1}(\tilde{\theta})$ is approximately correct.

Under correctly specified alternatives (part (b)), the highest power is achieved by the optimal RS test, $RS_{2\cdot 1}(\hat{\theta})$, followed very closely by Neyman's

$C_{2\cdot 1}(\tilde{\theta})$. The tests robust to misspecification of θ_3, $RS^*_{2\cdot 1}(\hat{\theta})$ and $C^*_{2\cdot 1}(\tilde{\theta})$, show less power than those of $RS_{2\cdot 1}(\hat{\theta})$ and $C_{2\cdot 1}(\tilde{\theta})$, consistent with the fact that $\lambda_{2\cdot 1} \geq \lambda^*_{2\cdot 1}$. This is the 'robustification cost' for unnecesarily using a modified test. Nevertheless, it is interesting to highlight that, in this case, the power loss is minimal. A comparison of $RS^*_{2\cdot 1}(\hat{\theta})$ and $C^*_{2\cdot 1}(\tilde{\theta})$ shows that, as predicted by the theory, they have very similar power, suggesting that the power of the BY procedure can be successfully restored through a properly modified test based on a consistent, non-MLE. Moreover, $\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta})$ has less power than $C^*_{2\cdot 1}(\tilde{\theta})$.

Part (c) studies the effects of misspecification through θ_3. As expected, all the non-robust versions, $RS_{2\cdot 1}(\hat{\theta})$, $RS_{2\cdot 1}(\tilde{\theta})$, $\widetilde{RS}_{2\cdot 1}(\tilde{\theta})$ and $C_{2\cdot 1}(\tilde{\theta})$, have unwanted rejection for H^2_0, as θ_3 increases, which is compatible with $\lambda_{2/3\cdot 1} > 0$. Nevertheless, the robustified versions ($RS^*_{2\cdot 1}(\hat{\theta})$, $\widetilde{RS}^*_{2\cdot 1}(\hat{\theta})$ and $C^*_{2\cdot 1}(\tilde{\theta})$) have rejection probabilities close to 0.05 or less. The empirical size of $RS^*_{2\cdot 1}(\hat{\theta})$ and $C^*_{2\cdot 1}(\tilde{\theta})$ reduce gradually as θ_3 increases, possibly due to the fact that adjustments are designed only for local misspecifications, i.e., for θ_3 values close to 0. We offer some intuitive explanation. In our set up $\theta_3 = \delta_3/\sqrt{n}$. For $n = 100$, choosing θ_3 between 0.1 and 1.0, δ_3 is allowed to vary from 1.0 to 10.0. Let us consider the case of our suggested test $C^*_{2\cdot 1}(\theta)$ which takes account of the presence of θ_3 by indirectly estimating it through $d_3(\theta)$, evaluated at $\tilde{\theta}$. Since in our Monte Carlo design the explanatory variables have positive correlation (0.5), the components of the information matrix $J(\theta)$ will be positive. Thus the effective score $d^*_{2\cdot 1}(\theta)$ can be expected to be lower than $d_{2\cdot 1}(\theta)$ which again can be expected to be lower than the raw score $d_2(\theta)$. Thus for non-local misspecification there could be some overcorrection for our Monte Carlo set up.

5 Final remarks

This paper proposes a new test that is consistent, achieves correct asymptotic size and is locally most powerful under local misspecification, and when any $\sqrt{n}$-estimator of the nuisance parameters is used. The new test can be seen as an extension of the Bera and Yoon (1993) procedure in order to deal with non-ML estimation, while preserving its optimality properties. Similarly, the procedure can be viewed as extending the standard $C(\alpha)$ test (that by construction admits non-ML estimators) to handle locally misspecified alternatives. In many practical situations non-ML strategies are favored to handle initial, restricted models, such as the case of dynamic panels and spatial panel models, where GMM estimators are usually preferred.

Appendix

Proof of Theorem 1

The asymptotic distribution follows from an application of Newey (1985, theorem 2.3). Note that Assumptions 1-3 correspond to assumptions 1-9 in Newey (1985). Define the vector of functions $w(z,\theta) = [h_1(z,\theta)' \ d_2(z,\theta)']'$ and $w(\theta) = [h_1(\theta)' \ d_2(\theta)']'$. Also define the matrices $\Gamma = [\iota_{p_1} \ 0_{p_2}]$ and $\Pi = [0_{p_1} \ \iota_{p_2}]$, where $\iota_.$ is a vector of 1s and $0_.$ a vector of 0s. The estimating equations for θ_1 can then be rewritten as

$$\Gamma E\left[w(z,\theta_1,\theta_{20},\theta_{30})\right] = 0 \text{ only if } \quad \theta_1 = \theta_{10}.$$

The specification test can be based on the testing equations

$$\Pi E\left[w(z,\theta_{10},\theta_2,\theta_{30})\right] = 0 \text{ only if } \quad \theta_2 = \theta_{20}.$$

We follow the notation in Bera, Montes-Rojas and Sosa-Escudero (2010). Let $K = E\left[\partial w(z,\theta)/\partial\theta_1\right]_{\theta=\theta_0} = [B_1' \quad -J_{21}']'$, where the equality $E\left[\partial d_2(z,\theta)/\partial\theta_1\right]_{\theta=\theta_0} =$

$-J_{21}$ follows from the information matrix equality,

$$V = E\left[w(z,\theta_0)\ w(z,\theta_0)'\right] = \left[\begin{array}{cc} H_1 & V_{12} \\ V_{21} & J_{22} \end{array}\right],$$

where $V_{12} = E[h_1(z,\theta_0)d_2(z,\theta_0)'] = V_{21}'$, $D = E\left[w(z,\theta_0)d_2(z,\theta_0)'\right] = [V_{12}'\ J_{22}']'$, and $P = I - K(\Gamma K)^{-1}\Gamma$.

Then under H_A^2, $\sqrt{n}\Pi w(\tilde{\theta}) \quad \xrightarrow{d} \quad N\left(\Pi P D\delta_2, \Pi P V P'\Pi'\right)$, as $n \to \infty$, hence

$$n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta}) \ \xrightarrow{d}\ \chi^2_{p_2}(\tilde{\lambda}_{2\cdot 1}),$$

as $n \to \infty$ with $\widetilde{\lambda}_{2\cdot 1} = (\Pi P D\delta_2)'(\Pi P V P'\Pi')^{-1}(\Pi P D\delta_2)$.

After some matrix algebra we obtain

$$\Pi P = \left[J_{21}\ B_1^{-1}\ \iota_{p_2}\right],$$

$$\Pi P V = \left[J_{21}B_1^{-1}H_1 + V_{21}\ J_{21}B_1^{-1}V_{12} + J_{22}\right],$$

$$\Pi P V P'\Pi = J_2 + J_{21}B_1^{-1}V_{12} + V_{21}B_1^{-1}J_{12} + J_{21}B_1^{-1}H_1B_1^{-1}J_{12}.$$

Thus, $\Pi P V P'\Pi' = V_{2\cdot 1}$. Moreover, $\Pi P D = J_{22} + J_{21}B_1^{-1}V_{12}$. Then, $\tilde{\lambda}_{2\cdot 1} = \delta_2'(J_{22} + J_{21}B_1^{-1}V_{12})'V_{2\cdot 1}^{-1}(J_{22} + J_{21}B_1^{-1}V_{12})\delta_2$. Finally note that by an application fo the the generalized information matrix equality (Newey and McFadden, 1994, p. 2163) $V_{12} = E[h_1(z,\theta_0)d_2(z,\theta_0)'] = -E[\partial h_1(z,\theta)/\partial\theta_2']_{\theta=\theta_0} = -E[h_1(z,\theta_0)h_1(z,\theta_0)']E[\partial h_1(z,\theta)/\theta_1]^{-1}_{\theta=\theta_0}E[d_1(z,\theta_0)d_2(z,\theta_0)'] = -H_1B_1^{-1}J_{12} = V_{21}'$. Thus, $\tilde{\lambda}_{2\cdot 1} = \delta_2'V_{2\cdot 1}\delta_2$. Finally, note that $\widetilde{RS}_{2\cdot 1}(\tilde{\theta}) = n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta}).QED$

Proof of Theorem 2

(i) The proofs follows from a modification of the proof of Theorem 1 where d_2 is replaced by $d_{23} = [d_2'\ d_3']'$. Consider a new partition of a three parameter space $(\theta_1, \theta_2, \theta_3)$ into $(\theta_1, (\theta_2, \theta_3))$. This is only notation to emphasize that the "block" (θ_2, θ_3) is taken together. Thus 23 denotes this new redefined parameter $\theta_{23} = (\theta_2', \theta_3')'$. Define the matrix

$$J = \begin{bmatrix} J_{11} & J_{1,23} \\ J_{23,1} & J_{23} \end{bmatrix}$$

and the vector of functions $w(z, \theta) = [h_1(z, \theta)' \; d_{23}(z, \theta)']'$ and $w(\theta) = [h_1(\theta)' \; d_{23}(\theta)']'$. Also define the matrices $\Gamma = [\iota_{p_1} \; 0_{p_2} \; 0_{p_3}]$ and $\Pi = [0_{p_1} \; \iota_{p_2} \; 0_{p_3}]$. Moreover, define $V_{1,23} = E[h_1(z, \theta_0) \; d_{23}(z, \theta_0)'] = V_{23,1}'$.

Following the notation in Bera, Montes-Rojas and Sosa-Escudero (2010), let $K = E\left[\partial w(z, \theta)/\partial\theta_1\right]_{\theta=\theta_0} = [B_1 \; - J_{23,1}]$,

$$V = E\left[w(z, \theta_0) w(z, \theta_0)'\right] = \begin{bmatrix} H_1 & V_{1,23} \\ V_{23,1} & J_{23,23} \end{bmatrix},$$

$D = E\left[w(z, \theta_0) \; d_{23}(z, \theta_0)'\right] = [V_{23,1} \; J_{23,23}]$, and $P = I - K(\Gamma K)^{-1}\Gamma$.

Then under H_0^2 and H_A^3, $\sqrt{n}\Pi w(\tilde{\theta}) \xrightarrow{d} N\left(\Pi P D[0_{p_2}\delta_3], \Pi P V P'\Pi'\right)$, as $n \to \infty$, hence

$$n \; w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\tilde{\lambda}_{2/3\cdot 1}),$$

as $n \to \infty$ with $\tilde{\lambda}_{2/3\cdot 1} = (\Pi P D\delta_3)'(\Pi P V P'\Pi')^{-1}(\Pi P D\delta_3)$.

After some algebra we obtain $\Pi P D = J_{23} + J_{21}B_1^{-1}V_{13}$ and $\Pi P V P'\Pi' = V_{2\cdot 1}$. Thus, $\tilde{\lambda}_{2/3\cdot 1} = \delta_3'\left(J_{23} + J_{21}B_1^{-1}V_{13}\right)' V_{2\cdot 1}^{-1}\left(J_{23} + J_{21}B_1^{-1}V_{13}\right)\delta_3$. Moreover, note that by an application fo the the generalized information matrix equality (Newey and McFadden, 1994, p. 2163) $V_{13} = E[h_1 d_3'] = -E[\partial h_1/\partial\theta_3']_{\theta=\theta_0} = -H_1 B_1^{-1} J_{13} = V_{31}'$. Finally, note that $\widetilde{RS}_{2\cdot 1}(\tilde{\theta}) = w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta})$.

(ii) The result follows from part (i) and Bera, Montes-Rojas and Sosa-Escudero (2010, Theorem 3). We need to modify the score function for θ_2, d_2, to account for the local misspecification in θ_3. This is done by

considering the adjusted score for the score d_2 in the function: $w(z,\theta) = [h_1(z,\theta)'\ \tilde{d}^*_{2\cdot 1}(z,\theta)'\ d_3(z,\theta)']'$ where $\tilde{d}^*_{2\cdot 1}(z,\theta) = d_2(z,\theta) - B_{23\cdot 1}B^{-1}_{3\cdot 1}d_3(z,\theta)$, and $w(\theta) = [h_1(\theta)'\ \tilde{d}^*_{2\cdot 1}(\theta)'\ d_3(\theta)']'$ where $\tilde{d}^*_{2\cdot 1}(\theta) = d_2(\theta) - B_{23\cdot 1}B^{-1}_{3\cdot 1}d_3(\theta)$. Define Γ and Π as in part (i) and obtain K, V, D and P with the same procedure for the newly defined $w(z,\theta)$. Then under H^2_0 and H^3_A, $\sqrt{n}\Pi w(\tilde{\theta}) \xrightarrow{d} N\left(0_{p_2}, \Pi P V P'\Pi'\right)$, that is, it recovers the zero mean of the testing function. Finally, after some algebra $\Pi P V P'\Pi' = V_{2(3)\cdot 1}$, where $V_{2(3)\cdot 1}$ accounts for the variance of $\widetilde{d}^*_{2\cdot 1}(\tilde{\theta})$, and is given by $V_{2(3)\cdot 1} = B_{2\cdot 1} - B_{23\cdot 1}B^{-1}_{3\cdot 1}B_{32\cdot 1}$, and the chi-squared distribution follows. Under H^2_A and H^3_A, $\sqrt{n}\Pi w(\tilde{\theta}) \xrightarrow{d} N\left(\Pi P D\delta_2, \Pi P V P'\Pi'\right)$, as $n\to\infty$, hence, $n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\tilde{\lambda}^*_{2\cdot 1})$, with
$\tilde{\lambda}^*_{2\cdot 1} = (\Pi P D\delta_2)'(\Pi P V P'\Pi')^{-1}(\Pi P D\delta_2) = \delta_2' V_{2(3)\cdot 1}\delta_2$. Finally, note that $\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta}) = n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta})$. *QED*

Proof of Theorem 3

Define $d^*_{2\cdot 1}(z,\theta) = d_{2\cdot 1}(z,\theta) - J_{23\cdot 1}J^{-1}_{3\cdot 1}d_{3\cdot 1}(z,\theta)$, $w(z,\theta) = [h_1(z,\theta)'\ d^*_{2\cdot 1}(z,\theta)'\ d_3(z,\theta)']'$, $d^*_{2\cdot 1}(\theta) = d_{2\cdot 1}(\theta) - J_{23\cdot 1}J^{-1}_{3\cdot 1}d_{3\cdot 1}(\theta)$, $w(\theta) = [h_1(\theta)'\ d^*_{2\cdot 1}(\theta)'\ d_3(\theta)']'$. Define Γ and Π as in Theorem 2, part (i), and obtain K, V, D and P with the same procedure for the newly defined $w(\theta)$.

Then under H^2_0 and H^3_A, $\sqrt{n}\Pi w(\tilde{\theta}) \xrightarrow{d} N\left(0_{p_2}, \Pi P V P'\Pi'\right)$, as $n\to\infty$, that is, it recovers the asymptotic zero mean of the testing function. Moreover under H^2_A and H^3_A, $\sqrt{n}\Pi w(\tilde{\theta}) \xrightarrow{d} N\left(\Pi P D\delta_2, \Pi P V P'\Pi'\right)$, as $n\to\infty$, hence, $n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta}) \xrightarrow{d} \chi^2_{p_2}(\lambda^*_{2\cdot 1})$, as $n\to\infty$ with
$\lambda^*_{2\cdot 1} = (\Pi P D\delta_2)'(\Pi P V P'\Pi')^{-1}(\Pi P D\delta_2) = \delta_2' J_{2(3)\cdot 1}\delta_2$, where $J_{2(3)\cdot 1}$ accounts for the variance of $d^*_{2\cdot 1}(\tilde{\theta})$, and is given by $J_{2(3)\cdot 1} = J_{2\cdot 1} - J_{23\cdot 1}J^{-1}_{3\cdot 1}J_{32\cdot 1}$. Finally, note that $C^*_{2\cdot 1}(\tilde{\theta}) = n\ w(\tilde{\theta})'\Pi'(\Pi P V P'\Pi')^{-1}\Pi w(\tilde{\theta})$. *QED*

References

Anselin, L., Bera, A., Florax, R. and Yoon, M. (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics* 26, 77-104.

Baltagi, B. and Li, D. (2001) LM tests for functional form and spatial error correlation, *International Regional Science Review* 24, 194-225.

Baltagi, B. H., Song, S.H., and Jung, B.C. (2001) The unbalanced nested error component regression model, *Journal of Econometrics* 101, 357-381.

Baltagi, B. H., Song, S.H., and Jung, B.C. (2002) Simple LM tests for the unbalanced nested error component regression model, *Econometric Reviews* 21(2), 167-187.

Bera, A. and Bilias, Y. (2001a) On some optimality properties of Fisher-Rao score function in testing and estimation, *Communications in Statistics - Theory and Methods* 30, 1533-1559.

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

Bera, A., Montes-Rojas, G. and Sosa-Escudero, W. (2009) Testing under local misspecification and artificial regressions, *Economics Letters* 104, 66-68.

Bera, A., Montes-Rojas, G. and Sosa-Escudero, W. (2010) General specification testing with locally misspecified models, *Econometric Theory* 26, 1838-1845.

Bera, A., Sosa-Escudero, W. and Yoon, M. (2001) Tests for the error component model in the presence of local misspecification, *Journal of Econometrics* 101, 1-23.

Bera, A. and Yoon, M. (1993) Specification testing with locally misspecified alternatives, *Econometric Theory* 9, 649-658.

Breusch, T.S. and Pagan, A.R., 1979. A simple test for heteroskedasticity and random coefficient variation, *Econometrica* 47, 1287–1294.

Box, G. (1953) Non-normalities and tests on variances, *Biometrika* 40, 318-335.

Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*, Champan & Hall/CRC, New York.

Davidson, R. and MacKinnon, J. (1987) Implicit alternatives and the local power of test statistics, *Econometrica* 55, 1305-1329.

Godfrey, L. and Orme, C. (1996) On the behavior of conditional moment tests in the presence of unconsidered local alternatives, *International Economic Review* 37, 263-281.

Godfrey, L. and Veall, M.R. (2000) Alternative approaches to testing by variable addition, *Econometric Reviews* 19, 241-261.

Montes-Rojas, G. (2010) Testing for random effects and serial correlation in spatial autorregresive models, *Journal of Statistical Planning and Inference* 140, 1013-1020.

Montes-Rojas, G. (2011) Robust misspecification tests for the Heckman's two-step estimator, *Econometric Reviews* 30, 1-19.

Newey, W. (1985) Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047-1070.

Newey, W. and McFadden, D. (1994) Large sample estimation and hypothesis testing, in Engle, R. and McFadden, D. (ed.), *Handbook of Econometrics*, Volume 4, 2111-2245, North Holland, Amsterdam.

Newey, W. and West, K. (1987) Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777-787.

Neyman, J. (1959) Optimal asymptotic tests of composite statistical hypotheses, in *Probability and Statistics, the Harald Cramer Volume*, ed. by U. Grenander, New York, Wiley.

Rao, C.R. (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Procedings of Cambridge Philosophical Society* 44, 50-57.

Rao, C.R. and Poti, S.J. (1946) On locally most powerful tests when alternative are one sided, *Sankhya: The Indian Journal of Statistics* 7, 439.

Saikkonen, P. (1989) Asymptotic relative efficiency of the classical tests under misspecification, *Journal of Econometrics* 42, 351-369.

Smith, R. (1987) Alternative asymptotically optimal tests and their application to dynamic specification, *Review of Economic Studies* 54, 665-80.

Welsh, A. (1996) *Aspects of Statistical Inference*, Wiley, New York.

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data* (second ed.) MIT Press.

Table 1: Monte Carlo simulations

θ_2	θ_3	$RS_{2\cdot 1}(\hat{\theta})$	$RS_{2\cdot 1}(\tilde{\theta})$	$\widetilde{RS}_{2\cdot 1}(\tilde{\theta})$	$C_{2\cdot 1}(\tilde{\theta})$	$RS^*_{2\cdot 1}(\hat{\theta})$	$RS^*_{2\cdot 1}(\tilde{\theta})$	$\widetilde{RS}^*_{2\cdot 1}(\tilde{\theta})$	$C^*_{2\cdot 1}(\tilde{\theta})$
				(a) Size					
0.00	0.0	0.051	0.120	0.059	0.043	0.043	0.083	0.046	0.038
				(b) Power in the θ_2-direction					
0.05	0.0	0.112	0.167	0.103	0.105	0.108	0.150	0.085	0.097
0.10	0.0	0.221	0.280	0.183	0.204	0.196	0.232	0.146	0.180
0.15	0.0	0.458	0.462	0.362	0.411	0.397	0.402	0.321	0.370
0.20	0.0	0.651	0.623	0.527	0.612	0.562	0.563	0.470	0.521
0.25	0.0	0.824	0.774	0.699	0.801	0.750	0.716	0.628	0.714
0.30	0.0	0.925	0.856	0.799	0.903	0.866	0.827	0.760	0.827
0.40	0.0	0.997	0.972	0.956	0.991	0.984	0.971	0.947	0.972
0.50	0.0	1.000	0.997	0.989	0.999	0.996	0.991	0.982	0.992
0.60	0.0	1.000	1.000	0.998	1.000	0.998	0.998	0.995	0.998
0.70	0.0	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.999
				(c) Robustness to θ_3-misspecification					
0.00	0.1	0.082	0.153	0.094	0.075	0.051	0.099	0.067	0.053
0.00	0.2	0.132	0.212	0.123	0.121	0.041	0.088	0.047	0.033
0.00	0.3	0.209	0.275	0.181	0.187	0.037	0.069	0.060	0.033
0.00	0.4	0.323	0.355	0.258	0.296	0.017	0.044	0.048	0.011
0.00	0.5	0.399	0.408	0.304	0.354	0.012	0.050	0.049	0.012
0.00	0.6	0.483	0.483	0.396	0.442	0.009	0.042	0.047	0.008
0.00	0.7	0.543	0.563	0.461	0.514	0.008	0.042	0.062	0.005
0.00	0.8	0.632	0.613	0.535	0.588	0.009	0.039	0.041	0.007
0.00	0.9	0.690	0.635	0.553	0.634	0.015	0.058	0.053	0.013
0.00	1.0	0.715	0.670	0.591	0.665	0.009	0.037	0.043	0.007

Notes: Tests for $H_0^2 : \theta_2 = 0$. Robust tests consider potential local departures from $H_0^3 : \theta_3 = 0$. Empirical rejection rates based on a nominal size of 0.05. Sample size = 100, number of replications = 1000.