Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors [1]

Jean-Marie Dufour [2] and Joanna Jasiak [3]
Université de Montréal and York University

January 2000

[1]This work was supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds FCAR (Government of Québec). The authors thank Lise Salvas and Claude Montmarquette for providing the data as well as Gordon Fisher, Denis Fougère, Christian Gouriéroux, Helmut Lütkepohl, Edmond Malinvaud, Alain Monfort, Garry Phillips, Jim Stock, Mohamed Taamouti, Herman van Dijk, Eric Zivot, the Editor Frank Diebold and three anonymous referees for helpful comments. This paper is a revised version of Dufour and Jasiak (1993).

[2]C.R.D.E, Université de Montréal, C.P. 6128, Succursale Centre Ville, Montréal, Québec H3C 3J7, Canada. TEL: 1 514 343 6557; FAX: 1 514 343 5831; e-mail: dufour@crde.umontreal.ca.

[3]Department of Economics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. TEL: 1 416 736 2100 extension 77045; FAX: 1 416 736 5987; e-mail: jasiakj@yorku.ca.

ABSTRACT

We propose finite sample tests and confidence sets for models with unobserved and generated regressors as well as various models estimated by instrumental variables method. We study two distinct approaches for various models considered by Pagan (1984). The first one is an instrument substitution method which generalizes an approach proposed by Anderson and Rubin (1949) and Fuller (1987) for different (although related) problems, while the second one is based on splitting the sample. The instrument substitution method uses the instruments directly, instead of generated regressors, in order to test hypotheses about the "structural parameters" of interest and build confidence sets. The second approach relies on "generated regressors", which allows a gain in degrees of freedom, and a sample split technique. A distributional theory is obtained under the assumptions of Gaussian errors and strictly exogenous regressors. We show that the various tests and confidence sets proposed are (locally) "asymptotically valid" under much weaker assumptions. The properties of the tests proposed are examined in simulation experiments. In general, they outperform the usual asymptotic inference methods in terms of both reliability and power. Finally, the techniques suggested are applied to a model of Tobin's q and to a model of academic performance.

Key words: generated regressor; simultaneous equations; structural model; pivotal function; sample-split; Anderson-Rubin method; finite-sample inference; exact test; confidence region; instrumental variables; Tobin's q; academic performance.

RÉSUMÉ

Nous proposons des tests et régions de confiance exactes pour des modèles comportant des variables inobservées ou des régresseurs estimés de même que pour divers modèles estimés par la méthode des variables instrumentales. De façon plus spécifique, nous étudions deux approches différentes pour divers modèles considérés par Pagan (1984). La première est une méthode de substitution d'instruments qui généralise des techniques proposées par Anderson et Rubin (1949) et Fuller (1984) pour des problèmes différents, tandis que la seconde méthode est fondée sur une subdivision de l'échantillon. La méthode de substitution d'instruments utilise directement les instruments disponibles, plutôt que des régresseurs estimés, afin de tester des hypothèses et construire des régions de confiance sur les "paramètres structuraux" du modèle. La seconde méthode s'appuie sur des régresseurs estimés, ce qui permet un gain de degrés de liberté, ainsi que sur une technique de subdivision de l'échantillon. Nous fournissons une théorie distributionnelle exacte sous une hypothèse de normalité des perturbations et de régresseurs strictement exogènes. Nous montrons que les tests et régions de confiance ainsi obtenus sont aussi (localement) "asymptotiquement valides" sous des hypothèses distributionnelles beaucoup plus faibles. Nous étudions les propriétés des tests proposés dans le cadre d'une expérience de simulation. En général, celles-ci sont plus fiables et ont une meilleure puissance que les techniques traditionnelles. Finalement, les techniques proposées sont appliquées à un modèle du q de Tobin et à un modèle de performance scolaire.

Mots-clefs: régresseur estimé; équations simultanées; modèle structurel; fonction pivotale; subdivision d'échantillon; méthode d'Anderson-Rubin; variables instrumentales; inférence à distance finie; test exact; région de confiance; variables instrumentales; q de Tobin; performance scolaire.

Contents

1 Introduction

A frequent problem in econometrics and statistics consists in making inferences on models which contain unobserved explanatory variables, such as expectational or latent variables and variables observed with error; see, for example, Barro (1977), Pagan (1984, 1986) and the survey of Oxley and McAleer (1993). A common solution to such problems is based on using instrumental variables to replace the unobserved variables by proxies obtained from auxiliary regressions (*generated regressors*). It is also well known that using such regressors raises difficulties for making tests and confidence sets, and it is usually proposed to replace ordinary least squares (OLS) standard errors by instrumental variables (IV) based standard errors; see Pagan (1984, 1986) and Murphy and Topel (1985). In any case, all the methods proposed to deal with such problems only have an asymptotic justification, which means that the resulting tests and confidence sets can be extremely unreliable in finite samples. In particular, such difficulties occur in situations involving "weak instruments", a problem which has received considerable attention recently; see, for example, Nelson and Startz (1990a, b), Buse (1992), Maddala and Jeong (1992), Bound, Jaeger and Baker (1993, 1995), Angrist and Krueger (1994), Hall, Rudebusch and Wilcox (1996), Dufour (1997), Staiger and Stock (1997) and Wang and Zivot (1997) [for some early results relevant to the same issue, see also Nagar (1959), Richardson (1968) and Sawa (1969)].

In this paper, we treat these issues from a finite sample perspective and we propose finite sample tests and confidence sets for models with unobserved and generated regressors. We also consider a number of related problems in the more general context of linear simultaneous equations. To get reliable tests and confidence sets, we emphasize the derivation of truly pivotal (or boundedly pivotal) statistics, as opposed to statistics which are only asymptotically pivotal; for a general discussion of the importance of such statistics for inference, see Dufour (1997). We study two distinct approaches for various models considered by Pagan (1984). The first one is an instrument substitution method which generalizes an approach proposed by Anderson and Rubin (1949) and Fuller (1987) for different (although related) problems, while the second one is based on splitting the sample. The instrument substitution method uses the instruments directly, instead of generated regressors, in order to test hypotheses and build confidence sets about "structural parameters". The second approach relies on "generated regressors", allowing a gain in degrees of freedom, and a sample split technique. Depending on the problem considered, we derive either exact *similar* tests (and confidence sets) or conservative procedures. The hypotheses for which we obtain similar tests (and correspondingly similar confidence sets) include: a) hypotheses which set the value of the unobserved (expected) variable coefficient vector [as in Anderson and Rubin (1949) and Fuller (1987)]; b) analogous restrictions taken jointly with general linear constraints on the coefficients of the (observed) exogenous variables in the equation of interest; and c) hypothesis about the coefficients of "surprise" variables when such variables are included in the equation. Tests for these hypotheses are based on Fisher-type statistics, but the confidence sets typically involve nonlinear (although quite tractable) inequalities. In particular, when only one unobserved variable (or endogenous explanatory variable) appears in the model, the confidence interval for the associated coefficient can be computed easily on finding the roots of a quadratic polynomial. Note that Anderson-Rubin-type methods have not previously been suggested in the context of the general Pagan (1984) setup; further, problems such as those described in b) and c) above have not apparently been considered at all from this perspective.

In the case of the instrument substitution method, the tests and confidence sets so obtained can be interpreted as likelihood ratio (LR) procedures (based on appropriately chosen reduced form alternatives), or equivalently as profile likelihood techniques [for further discussion of such techniques, see Bates and Watts (1988, Chapter 6), Meeker and Escobar (1995) and Chen and Jennrich (1996)].

The exact distributional theory is obtained under the assumptions of Gaussian errors and strictly exogenous regressors, which ensures that we have well-defined testable models. Although we stress here applications to models with unobserved regressors, the extensions of Anderson–Rubin (AR) procedures that we discuss are also of interest for inference in various structural models which are estimated by instrumental variable methods (e.g., simultaneous equations models). Furthermore, we observe that the tests and confidence sets proposed are (locally) "asymptotically valid" under much weaker distributional assumptions (which may involve non-Gaussian errors and weakly exogenous instruments).

It is important to note that the confidence sets obtained by the methods described above, unlike Wald-type confidence sets, are unbounded with non-zero probability. As emphasized from a general perspective in Dufour (1997), this is a necessary property of any valid confidence set for a parameter that may not be identifiable on some subset of the parameter space. As a result, confidence procedures that do not have this property have true level zero, and the sizes of the corresponding tests (like Wald-type tests) must deviate arbitrarily from their nominal levels. It is easy to see that such difficulties occur in models with unobserved regressors, models with generated regressors, simultaneous equations models, and different types of the error-in-variables models. In the context of the first type of model, we present below simulation evidence that strikingly illustrates these difficulties. In particular, our simulation results indicate that tests based on instrument substitution methods have good power properties with respect to Wald-type tests, a feature previously pointed out for the AR tests by Maddala (1974) in a comparative study for simultaneous equations [on the power of AR tests, see also Revankar and Mallela (1972)]. Furthermore, we find that generated regressors sample-split tests perform better when the generated regressors are obtained from a relatively small fraction of the sample (*e.g.*, 10% of the sample) while the rest of the sample is used for the main regression (in which generated regressors are used).

An apparent shortcoming of the similar procedures proposed above, and probably one of the reasons why AR tests have not become widely used, is the fact that they are restricted to testing hypotheses which specify the values of the coefficients of all the endogenous (or unobserved) explanatory variables, excluding the possibility of considering a subset of coefficients (e.g., individual coefficients). We show that inference on individual parameters or subvectors of coefficients is however feasible by applying a projection technique analogous to the ones used in Dufour (1989, 1990), Dufour and Kiviet (1996, 1998) and Kiviet and Dufour (1997). We also show that such techniques may be used for inference on general possibly nonlinear transformations of the parameter vector of interest.

The plan of the paper is as follows. In Section 2, we describe the main model which may contain several unobserved variables (analogous to the "anticipated" parts of those variables), and we introduce the instrument substitution method for this basic model with various tests and confidence sets for the coefficients of the unobserved variables. In Section 3, we propose the sample split method for the same model with again the corresponding tests and confidence sets. In Section 4, we study the problem of testing joint hypotheses about the coefficients of the unobserved variables and various linear restrictions on the coefficients of other (observed) regressors in the model. Section 5 extends these results to a model which also contains error terms of the unobserved variables (the "unanticipated" parts of these variables). In Section 6, we consider the problem of making inference about general nonlinear transformations of model coefficients. Then, in Section 7, we discuss the "asymptotic validity" of the proposed procedures proposed under weaker distributional assumptions. Section 8 presents the results of simulation experiments in which the performance of our methods is compared with some widely used asymptotic procedures. Section 9 presents applications of the proposed methods to a model of Tobin's q and to an economic model of educational performance. The latter explains the relationship between students' academic performance, their personal characteristics and some socio-economic factors. The first example illustrates inference in presence of good instruments, while in the second example only poor instruments are available. As expected, confidence intervals for Tobin's q based on the Wald-type procedures largely coincide with those resulting from our methods. On the contrary, large discrepancies arise between the confidence intervals obtained from the asymptotic and the exact inference methods when poor instruments are used. We conclude in Section 10.

2 Exact inference by instrument substitution

In this section, we develop finite sample inference methods based on instrument substitution methods for models with unobserved and generated regressors. We first derive general formulae for the test statistics and then discuss the corresponding confidence sets. We consider the following basic setup which includes as special cases Models 1 and 2 studied by Pagan (1984):

$$y = Z_*\delta + X\gamma + e\,, \tag{2.1}$$

$$Z_* = WB + U_*\;,\; Z = Z_* + V_* \tag{2.2}$$

where y is a $T\times 1$ vector of observations on a dependent variable, Z_* is a $T\times G$ matrix of unobserved variables, X is a $T \times K$ matrix of exogenous explanatory variables in the structural model, Z is a $T \times G$ matrix of observed variables, W is a $T \times q$ matrix of variables related to Z_*, while $e = (e_1, \dots, e_T)'$, $U_* = [u'_{*1}, \dots, u'_{*T}]'$ and $V_* = [v'_{*1}, \dots, v'_{*T}]'$ are $T \times 1$ and $T \times G$ matrices of disturbances. The matrices of unknown coefficients δ, γ, and B have dimensions respectively $G \times 1$, $K \times 1$ and $q \times G$. In order to handle common variables in both equations (2.1) and (2.2), like for example the constant term, we allow for the presence of common columns in the matrices W and X. In the setup of Pagan (1984), U_* is assumed to be identically zero $(U_* = 0)$, e_t and v_{*t} are uncorrelated $[E(e_t v_{*t}) = 0]$, and the exogenous regressors X are excluded from the "structural" equation (2.1). In some cases below, we will need to reinstate some of the latter assumptions.

The finite sample approach we adopt in this paper requires additional assumptions, especially on the distributional properties of the error term. Since (2.2) entails $Z = WB+V$ where $V = U_*+V_*$, we will suppose the following conditions are satisfied:

$$X \text{ and } W \text{ are independent of } e \text{ and } V_*\,; \tag{2.3}$$

$$\operatorname{rank}(X) = K\;, \quad 1 \leq \operatorname{rank}(W) = q < T\;, \quad G \geq 1\,, \quad 1 \leq K + G < T\,; \tag{2.4}$$

$$(e_t,\, v'_{*t})' \overset{ind}{\sim} N[0,\, \Omega]\,, \quad t = 1, \dots, T\,; \tag{2.5}$$

$$\det(\Omega) > 0\,. \tag{2.6}$$

If $K = 0$, X is simply dropped from equation (2.1). Note that no assumption on the distribution of U_* is required. Assumptions (2.3) – (2.6) can be relaxed if they are replaced by assumptions on the asymptotic behavior of the variables as $T \to \infty$. Results on the asymptotic "validity" of the various procedures proposed in this paper are presented in Section 7.

Let us now consider the null hypothesis:

$$H_0 : \delta = \delta_0\,. \tag{2.7}$$

The instrument substitution method is based on replacing the unobserved variable by a set of instruments. First, we substitute (2.2) into (2.1):

$$y = (Z - V_*)\delta + X\gamma + e = Z\delta + X\gamma + (e - V_*\delta)\,. \tag{2.8}$$

Then subtracting $Z\delta_0$ on both sides of (2.8), we get:

(2.9) $$y - Z\delta_0 = WB(\delta - \delta_0) + X\gamma + u$$

where $u = e - V_*\delta_0 + U_*(\delta - \delta_0)$. Now suppose that W and X have K_2 columns in common $(0 \le K_2 < q)$ while the other columns of X are linearly independent of W:

(2.10) $$W = [W_1,\, X_2]\,, \quad X = [X_1,\, X_2]\,, \quad \text{rank}\,[W_1,\, X_1,\, X_2] = q_1 + K < T$$

where W_1, X_1 and X_2 are $T \times q_1$, $T \times K_1$ and $T \times K_2$ matrices, respectively ($K = K_1 + K_2$, $q = q_1 + K_2$). We can then rewrite (2.9) as

(2.11) $$y - Z\delta_0 = W_1\delta_{1*} + X\gamma_* + u$$

where $\delta_{1*} = B_1(\delta - \delta_0)$, $\gamma_{2*} = \gamma_2 + B_2(\delta - \delta_0)$, $\gamma_* = (\gamma_1',\, \gamma_{2*}')'$, B_i is a $K_i \times G$ matrix ($i = 1$, 2) and $B = [B_1', B_2']'$.

It is easy to see that model (2.11) under H_0 satisfies all the assumptions of the classical linear model. Furthermore, since $\delta_{1*} = 0$ when $\delta = \delta_0$, we can test H_0 by a standard F-test of the null hypothesis

(2.12) $$H_{0*} : \delta_{1*} = 0\,.$$

This F-statistic has the form

(2.13) $$F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)'P(M(X)W_1)\,(y - Z\delta_0)/q_1}{(y - Z\delta_0)'M([W_1,\, X])\,(y - Z\delta_0)/(T - q_1 - K)}$$

where $P(A) = A(A'A)^{-1}A'$ and $M(A) = I_T - P(A)$ for any full column rank matrix A. When $\delta = \delta_0$, we have $F(\delta_0;\, W_1) \sim F(q_1, T - q_1 - K)$, so that $F(\delta_0;\, W_1) > F(\alpha;\, q_1,\, T - q_1 - K)$ is a critical region with level α for testing $\delta = \delta_0$, where $P[F(\delta_0;\, W_1) \le F(\alpha;\, q_1,\, T - q_1 - K)] = 1 - \alpha$. The essential ingredient of the test is the fact that $q_1 \ge 1$, *i.e.* some instruments must be excluded from X in (2.1). On the other hand, the usual order condition for "identification" $(q_1 \ge G)$ is not necessary for applying this procedure. In other words, it is possible to test certain hypotheses about δ even if the latter vector is not completely identifiable. It is then straightforward to see that the set

(2.14) $$C_\delta(\alpha) = \{\delta_0 : F(\delta_0;\, W_1) \le F(\alpha;\, q_1,\, T - q_1 - K)\}$$

is a confidence set with level $1 - \alpha$ for the coefficient δ. The tests based on the statistic $F(\delta_0;\, W_1)$ and the above confidence set generalizes the procedure proposed by Fuller (1987, pp. 16–17) for a model with one unobserved variable $(G = 1)$ and X limited to a constant variable $(K = 1)$.

Consider now the case where Z is a $T \times 1$ vector and X is a $T \times K$ matrix. In this case, the confidence set (2.14) for testing $H_0 : \delta = \delta_0$ has the following general form:

(2.15) $$C_\delta(\alpha) = \left\{\delta_0 : \frac{(y - Z\delta_0)'A_1(y - Z\delta_0)}{(y - Z\delta_0)'A_2(y - Z\delta_0)} \times \frac{\nu_2}{q_1} \le F_\alpha\right\}$$

where $F_\alpha = F(\alpha; q_1, T - q_1 - K)$ and $\nu_2 = T - q_1 - K$ and the matrices $A_1 = P(M(X)W_1)$, $A_2 = M([W_1,\, X])$. Since (ν_2/q_1) only takes positive values, the inequality in (2.15) is equivalent to the quadratic inequality:

(2.16) $$a\delta_0^2 + b\delta_0 + c \le 0$$

where $a = Z'CZ$, $b = -2y'CZ$, $c = y'Cy$, $C = A_1 - G_\alpha A_2$ and $G_\alpha = (q_1/\nu_2)F_\alpha$.

TABLE 1
CONFIDENCE SETS BASED ON THE QUADRATIC INEQUALITY $a\delta_0^2 + b\delta_0 + c \le 0$

		$\Delta \ge 0$ (real roots)	$\Delta < 0$ (complex roots)
$a > 0$		$[\delta_{1*},\ \delta_{2*}]$	Empty
$a < 0$		$(-\infty,\ \delta_{1*}] \cup [\delta_{2*},\ \infty)$	$(-\infty, +\infty)$
$a = 0$	$b > 0$	$\left(-\infty,\ -\frac{c}{b}\right]$	
	$b < 0$	$\left[-\frac{c}{b},\ \infty\right)$	
	$b = 0,\ \ c > 0$	Empty	
	$b = 0,\ \ c \le 0$	$(-\infty, +\infty)$	

In empirical work, some problems may arise due to the high dimensions of the matrices $M(X)$ and $M([W_1,\, X])$. A simple way to avoid this difficulty consists in using vectors of residuals from appropriate OLS regressions. Consider the coefficient $a = Z'CZ$. We may replace it by the expression $Z'A_1Z - G_\alpha Z'A_2Z$ and then rewrite both terms as follows:

$$
\begin{aligned}
Z'A_1Z &= (Z'M(X))\,(M(X)W_1)\,[(M(X)W_1)'(M(X)W_1)]^{-1}(M(X)W_1)'(M(X)Z)\,,\\
Z'A_2Z &= Z'M([W_1,\, X])Z = [M([W_1,\, X])Z]'[M([W_1,\, X])Z]\,.
\end{aligned}
$$

In the above expressions, $M(X)Z$ is the vector of residuals obtained by regressing Z on X, $M(X)W_1$ is the vector of residuals from the regression of W_1 on X, and finally $M([W_1,\, X])Z$ is a vector of residuals from the regression of Z on X and W_1. We can proceed in the same way to compute the two other coefficients of the quadratic inequality (2.16). This will require only two additional regressions: y on X, and y on both X and W_1.

It is easy to see that the confidence set (2.16) is determined by the roots of the second order polynomial in (2.16). The shape of this confidence set depends on the signs of a and $\Delta = b^2 - 4ac$. All possible options are summarized in Table 1 where δ_{1*} denotes the smaller root and by δ_{2*} the larger root of the polynomial (when both roots are real).

Note that the confidence set $C_\delta(\alpha)$ may be empty or unbounded with a non-zero probability. Since the reduced form for y can be written

$$y = W_1\pi_1 + X_1\pi_{12} + X_2\pi_{22} + v_y \tag{2.17}$$

where $\pi_1 = B_1\delta$, $\pi_{21} = \gamma_1$, $\pi_{22} = \gamma_2 + B_2\gamma$ and $v_y = e + U_*\delta$, we see that the condition $\pi_1 = B_1\delta$ may be interpreted as an overidentifying restriction. Jointly with $\delta = \delta_0$, this condition entails the hypothesis $H_{0*} : B_1(\delta - \delta_0) = 0$ which is tested by the statistic $F(\delta_0;\, W_1)$. Thus an empty confidence set means the condition $B_1(\delta - \delta_0) = 0$ is rejected for any value of δ_0 and so indicates that the overidentifying restrictions entailed by the structural model (2.1) - (2.2) are not supported by the data, *i.e.* the specification is rejected. However, if the model is correctly specified, the probability of obtaining an empty confidence set is not greater than α. On the other hand, the possibility of an unbounded confidence set is a *necessary* characteristic of any valid confidence set in the present context, because the structural parameter δ may not be identifiable [see Dufour (1997)]. Unbounded confidence sets are most likely to occur when δ is not identified or close to being unidentified, for then all values of δ are almost observationally equivalent. Indeed an unbounded confidence set obtains when $a < 0$ or (equivalently) when $F(\Pi_1 = 0) < F_\alpha$, where $F(\Pi_1 = 0)$ is the F-statistic for testing $\Pi_1 = 0$ in the regression

$$Z = W_1\Pi_1 + X\Pi + v_Z\,. \tag{2.18}$$

In other words, the confidence interval (2.15) is unbounded if and only if the coefficients of the exogenous regressors in W_1 [which is excluded from the structural equation (2.1)] are not significantly related to Z at level α : *i.e.*, W_1 can be interpreted as a matrix of "weak instruments" for Z. In contrast, Wald-type confidence sets for δ are typically bounded with probability one, so their true level must be zero. Note finally that an unbounded confidence set can be informative: *e.g.*, the set $(-\infty,\, \delta_{1*}] \cup [\delta_{2*},\, \infty)$ may exclude economically important values of δ ($\delta = 0$ for example).

3 Inference with generated regressors

Test statistics similar to those of the previous section may alternatively be obtained from linear regressions with generated regressors. To obtain finite sample inferences in such contexts, we propose to compute adjusted values from an independent sample. In particular, this can be done by applying a sample split technique.

Consider again the model described by (2.1) to (2.6). In (2.9), a natural thing to do would consist in replacing WB by $W\hat{B}$, where $\hat{B}$ is an estimator of B. Take $\hat{B} = (W'W)^{-1}W'Z$, the least squares estimate of B based on (2.2). Then we have:

$$y - Z\delta_0 = W\hat{B}(\delta - \delta_0) + X\gamma + [u + W(B - \hat{B})\,(\delta - \delta_0)] = \hat{Z}\delta_{0*} + X\gamma + u_* \tag{3.1}$$

where $\delta_{0*} = \delta - \delta_0$ and $u_* = e - V_*\delta_0 + [U_* + W(B - \hat{B})\,](\delta - \delta_0)$. Again, the null hypothesis $\delta = \delta_0$ may be tested by testing $H_{0*} : \delta_{0*} = 0$ in model (3.1). Here the standard F statistic for H_{0*} is obtained by replacing W_1 by $\hat{Z}$ in (2.13), *i.e.*

$$F(\delta_0;\, \hat{Z}) = \frac{(y - Z\delta_0)'P(M(X)\hat{Z})\,(y - Z\delta_0)/G}{(y - Z\delta_0)'M([\hat{Z},\, X])\,(y - Z\delta_0)/(T - G - K)}\ ; \tag{3.2}$$

if $K = 0$ [no X matrix in (2.1)], we conventionally set $M(X) = I_T$ and $[\hat{Z},\, X] = \hat{Z}$. However, to get a null distribution for $F(\delta_0;\, \hat{Z})$, we will need further assumptions. For example, in addition to the assumptions (2.1) to (2.6), suppose, as in the original Pagan (1984) setup, that

$$e \text{ and } V \equiv U_* + V_* \text{ are independent.} \tag{3.3}$$

In this case, when $\delta = \delta_0 = 0$, $\hat{Z}$ and u_* are independent and, conditional on $\hat{Z}$, model (3.1) satisfies all the assumptions of the classical linear model (with probability 1). Thus the null distribution of the statistic $F(0;\, \hat{Z})$ for testing $\delta_0 = 0$ is $F(G,\, T - G - K)$. Unfortunately, this property does not extend to the more general statistic $F(\delta_0;\, \hat{Z})$ where $\delta_0 \neq 0$ because $\hat{Z}$ and u_* are not independent in this case. A similar observation (in an asymptotic context) was made by Pagan (1984).

To deal with more general hypotheses, suppose now that an estimate $\tilde{B}$ of B such that

$$\tilde{B} \text{ is independent of } e \text{ and } V_* \tag{3.4}$$

is available, and replace $\hat{Z} = W\hat{B}$ by $\tilde{Z} = W\tilde{B}$ in (3.1). We then get

$$y - Z\delta_0 = \tilde{Z}\delta_{0*} + X\gamma + u_{**} \tag{3.5}$$

where $u_{**} = e - V_*\delta_0 + [U_* + W(B - \tilde{B})]\,(\delta - \delta_0)$. Under the assumptions (2.1) – (2.6) with $\delta = \delta_0$ and conditional on $\tilde{Z}$ (or $\tilde{B}$), model (3.5) satisfies all the assumptions of the classical linear model and the usual F-statistic for testing $\delta_{0*} = 0$,

$$F(\delta_0;\, \tilde{Z}) = \frac{(y - Z\delta_0)'P(M(X)\tilde{Z})(y - Z\delta_0)/G}{(y - Z\delta_0)'M([\tilde{Z},\, X])(y - Z\delta_0)/(T - G - K)} \tag{3.6}$$

where the usual notation has been adopted, follows an $F(G,\, T-G-K)$ distribution. Consequently, the critical region $F(\delta_0; \tilde{Z}) > F(\alpha;\, G,\, T-G-K)$ has size α. Note that condition (3.3) is not needed for this result to hold. Furthermore

$$\tilde{C}_\delta(\alpha) = \{\delta_0 :\ F(\delta_0;\, \tilde{Z}) \leq F(\alpha;\, G,\, T-G-K)\} \tag{3.7}$$

is a confidence set for δ with size $1-\alpha$. For scalar δ $(G = 1)$, this confidence set takes a form similar to the one in (2.15), except that $A_1 = P(M(X)\tilde{Z})$ and $A_2 = M([\tilde{Z},\, X])$.

A practical problem here consists in finding the independent estimate $\tilde{B}$. Under the assumptions (2.1) – (2.6), this can be done easily by splitting the sample. Let $T = T_1 + T_2$, where $T_1 > G + K$ and $T_2 \geq q$, and write: $y = (y'_{(1)}\,,\, y'_{(2)})'$, $X = (X'_{(1)}\,,\, X'_{(2)})'$, $Z = (Z'_{(1)}\,,\, Z'_{(2)})'$, $W = (W'_{(1)}\,,\, W'_{(2)})'$, $e = (e'_{(1)}\,,\, e'_{(2)})'$, $V_* = (V'_{*(1)}\,,\, V'_{*(2)})'$ and $(U'_{*(1)}\,,\, U'_{*(2)})'$, where the matrices $y_{(i)}$, $X_{(i)}$, $Z_{(i)}$, $W_{(i)}$, $e_{(i)}$, $V_{*(i)}$ and $U_{*(i)}$ have T_i rows $(i = 1, 2)$. Consider now the equation

$$y_{(1)} - Z_{(1)}\delta_0 = \tilde{Z}_{(1)}\delta_{0*} + X_{(1)}\gamma + u_{(1)**} \tag{3.8}$$

where $\tilde{Z}_{(1)} = W_{(1)}\tilde{B}$, $\tilde{B} = [W'_{(2)}W_{(2)}]^{-1}W'_{(2)}Z_{(2)}$ is obtained from the second sample, and $u_{(1)**} = e_{(1)} - V_{*(1)}\delta_0 + [U_{*(1)} + W_{(1)}(B - \hat{B})\,](\delta - \delta_0)$. Clearly $\tilde{B}$ is independent of $e_{(1)}$ and $V_{*(1)}$, so the statistic $F(\delta_0;\, \tilde{Z}_{(1)})$ based on equation (3.8) follows a $F(G,\, T_1 - K - G)$ distribution when $\delta = \delta_0$.

A sample split technique has also been suggested by Angrist and Krueger (1994) to build a new IV estimator, called Split Sample Instrumental Variables (SSIV) estimator. Its advantage over the traditional IV method is that SSIV yields an estimate biased toward zero, rather than toward the probability limit of the OLS estimator in finite sample if the instruments are weak. Angrist and Krueger show that an unbiased estimate of the attenuation bias can be calculated and, consequently, an asymptotically unbiased estimator (USSIV) can be derived. In their approach, Angrist and Krueger rely on splitting the sample in half, *i.e.*, setting $T_1 = T_2 = \frac{T}{2}$ when T is even. However, in our setup, different choices for T_1 and T_2 are clearly possible. Alternatively, one could select at random the observations assigned to the vectors $y_{(1)}$ and $y_{(2)}$. As we will show later (see Section 8) the number of observations retained for the first and the second subsample have a direct impact on the power of the test. In particular, it appears that one can get a more powerful test once we use a relatively small number of observations for computing the adjusted values and keep more observations for the estimation of the structural model. This point is illustrated below by simulation experiments. Finally, it is of interest to observe that sample splitting techniques can be used in conjunction with the Boole-Bonferroni inequality to obtain finite-sample inference procedures in other contexts, such as seemingly unrelated regressions and models with moving average errors; for further discussion, the reader may consult Dufour and Torrès (1998).

4 Joint tests on δ and γ

The instrument substitution and sample split methods described above can easily be adapted to test hypotheses on the coefficients of both the latent variables and the exogenous regressors. In this section, we derive F-type tests for general linear restrictions on the coefficient vector. Consider again model (2.1) – (2.6), which after substituting the term $(Z - V_*)$ for the latent variable yields the following equation:

$$y = (Z - V_*)\delta + X\gamma + e = Z\delta + X\gamma + (e - V_*\delta)\,. \tag{4.1}$$

We first consider a hypothesis which fixes simultaneously δ and an arbitrary set of linear transformations of γ:

$$H_0: \ \delta = \delta_0, \quad R_1\gamma = \nu_{10}$$

where R_1 is a $r_1 \times K$ fixed matrix such that $1 \le \text{rank}(R_1) = r_1 \le K$. The matrix R_1 can be viewed as a submatrix of a $K \times K$ matrix $R = [R_1', \, R_2']'$ where $\det(R) \ne 0$, so that we can write

$$R\gamma = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \gamma = \begin{bmatrix} R_1\gamma \\ R_2\gamma \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}. \tag{4.2}$$

Let $X_R = XR^{-1} = [X_{R_1}, \, X_{R_2}]$ where X_{R_1} and X_{R_2} are $T \times r_1$ and $T \times r_2$ matrices $(r_2 = K - r_1)$. Then we can rewrite (4.1) as

$$y = Z\delta + X_{R_1}\nu_1 + X_{R_2}\nu_2 + (e - V_*\delta). \tag{4.3}$$

Subtracting $Z\delta_0$ and $X_{R_1}\nu_{10}$ on both sides, we get

$$\begin{aligned} y - Z\delta_0 - X_{R_1}\nu_{10} \ &= \ [W_1B_1 + X_2B_2](\delta - \delta_0) + X_{R_1}(\nu_1 - \nu_{10}) \\ &\quad + X_{R_2}\nu_2 + [e - V_*\delta_0 + U_*(\delta - \delta_0)]. \end{aligned} \tag{4.4}$$

Suppose now that W and X have K_2 columns in common (with $0 \le K_2 < q$), while the other columns of X are linearly independent of W as in (2.10). Since $X = [X_1, \, X_2] = X_R R = X_{R_1}R_1 + X_{R_2}R_2$, we can write $X = [X_1, \, X_2] = [X_{R_1}R_{11} + X_{R_2}R_{21}, \ X_{R_1}R_{12} + X_{R_2}R_{22}]$, where $R_1 = [R_{11}, \, R_{12}]$, $R_2 = [R_{21}, \, R_{22}]$ and R_{ij} is a $r_i \times K_j$ matrix $(i, \, j = 1, \, 2)$. Then replace X_2 by $X_{R_1}R_{12} + X_{R_2}R_{22}$ in (4.4):

$$y - Z\delta_0 - X_{R_1}\nu_{10} = W_1\delta_1^* + X_{R_1}\gamma_1^* + X_{R_2}\gamma_2^* + u \tag{4.5}$$

where $\delta_1^* = B_1(\delta - \delta_0), \ \gamma_1^* = R_{12}B_2(\delta - \delta_0) + (\nu_1 - \nu_{10}), \ \gamma_2^* = R_{22}B_2(\delta - \delta_0) + \nu_2,$ and $u = e - V_*\delta_0 + U_*(\delta - \delta_0)$. Consequently, we can test H_0 by testing $H_0': \ \delta_1^* = 0, \ \gamma_1^* = 0$,in (4.5), which leads to the statistic:

$$F(\delta_0, \, \nu_{10}; \, W_1, \, X_{R_1}) = \frac{\{y\,(\delta_0, \, \nu_{10})'P(M(X_{R_2})W_{R_1})\,y\,(\delta_0, \, \nu_{10})/(q_1 + r_1)\}}{\{y\,(\delta_0, \, \nu_{10})'M([W_1, \, X])\,y\,(\delta_0, \, \nu_{10})/(T - q_1 - K)\}} \tag{4.6}$$

where $y\,(\delta_0, \, \nu_{10}) = y - Z\delta_0 - X_{R_1}\nu_{10}$ and $W_{R_1} = [W_1, \, X_{R_1}]$; if $r_2 = 0$, we set $M(X_{R_2}) = I_T.$. Under H_0, $F(\delta_0, \, \nu_{10}; \ W_1, \, X_{R_1}) \sim F(q_1 + r_1, \ T - q_1 - K)$ and we reject H_0 at level α when $F(\delta_0, \, \nu_{10}; \ W_1, \, X_{R_1}) > F(\alpha; \ q_1 + r_1, \ T - q_1 - K)$. Correspondingly, $\{(\delta_0', \, \nu_{10}')' : F(\delta_0, \, \nu_{10}; \ W_1, \, X_{R_1}) \le F(\alpha; \ q_1 + r_1, \ T - q_1 - K)\}$ is a confidence set with level $1 - \alpha$ for δ and $\nu_1 = R_1\gamma_1$.

Suppose now we employ the procedure with generated regressors using an estimator $\tilde{B}$ independent of u and V. We can then proceed in the following way. Setting $\tilde{Z} = W\tilde{B}$ and $\hat{V} = Z - \tilde{Z}$, we have:

$$y - Z\delta_0 - X_{R_1}\nu_{10} = \tilde{Z}\delta_1^* + X_{R_1}\nu_1^* + X_{R_2}\nu_2 + u_{**} \tag{4.7}$$

where $\delta_1^* = \delta - \delta_0, \ \nu_1^* = \nu_1 - \nu_{10}$ and $u_{**} = e - V_*\delta_0 + [U_* + W(B - \tilde{B})](\delta - \delta_0)$. In this case we will simply test the hypothesis $H_0: \ \delta_1^* = 0, \ \nu_1^* = 0$. The F statistic for H_0 takes the form:

$$F(\delta_0, \, \nu_{10}; \ \tilde{Z}, \, X_{R_1}) = \frac{\{y\,(\delta_0, \, \nu_{10})'P(M(X_{R_2})\tilde{Z}_{R_1})\,y\,(\delta_0, \, \nu_{10})/(G + r_1)\}}{\{y\,(\delta_0, \, \nu_{10})'M([\tilde{Z}, \, X])\,y\,(\delta_0, \, \nu_{10})/(T - G - K)\}} \tag{4.8}$$

where $y(\delta_0, \, \nu_{10}) = y - Z\delta_0 - X_{R_1}\nu_{10}$, and $\tilde{Z}_{R_1} = [\tilde{Z}, \, X_{R_1}]$. Under H_0, $F(\delta_0, \, \nu_{10}; \ \tilde{Z}, \, X_{R_1}) \sim F(G + r_1, \ T - G - K)$. The corresponding critical region with level α is given by $F(\delta_0, \, \nu_{10}; \ \tilde{Z}, \, X_{R_1}) > F(\alpha; \ G + r_1, \ T - G - r_1)$, and the confidence set at level $1 - \alpha$ is thus $\{(\delta_0', \, \nu_{10}')' : F(\delta_0, \, \nu_{10}; \ \tilde{Z}, \, X_{R_1}) \le F(\alpha; \ G + r_1, \ T - G - K\}$.

5 Inference with a surprise variable

In many economic models we encounter so-called "surprise" terms among the explanatory variables. These reflect the differences between the expected values of latent variables and their realizations. In this section we study a model which contains the unanticipated part of Z [Pagan (1984, model 4)] as an additional regressor beside the latent variable, namely:

$$y = Z_*\delta + (Z - Z_*)\gamma + X\beta + e = Z\delta + V_*\gamma + X\beta + e - V_*\delta\,, \tag{5.1}$$

$$Z = Z_* + V_* = WB + (U_* + V_*) = WB + V\,, \tag{5.2}$$

where the general assumptions (2.3) – (2.6) still hold. The term $(Z - Z_*)$ represents the unanticipated part of Z. This setup raises more difficult problems especially for inference on γ. Nevertheless we point out here that the procedures described in the preceding sections for inference on δ and γ remain applicable essentially without modification, and we show that similar procedures can be obtained as well for inference on γ provided we make the additional assumption (3.3).

Consider first the problem of testing the hypothesis $H_0 :\ \delta = \delta_0$. Applying the same procedure as before, we get the equation:

$$y - Z\delta_0 = WB(\delta - \delta_0) + X\beta + V_*\gamma + (e - V_*\delta_0) \tag{5.3}$$

hence, assuming that W and X have K_2 columns in common,

$$y - Z\delta_0 = W_1B_1(\delta - \delta_0) + X_1\beta_1 + X_2\beta_2^* + e + V_*(\gamma - \delta_0) = W_1\delta_{1*} + X\beta_* + u \tag{5.4}$$

where $\delta_{1*} = B_1(\delta - \delta_0)$, $\beta_2^* = \beta_2 + B_2(\delta - \delta_0)$, $\beta_* = (\beta_1',\ \beta_2^{*\prime})'$ and $u = e + V_*(\gamma - \delta_0)$. Then we can test $\delta = \delta_0$ by using the F-statistic for $\delta_{10} = 0$:

$$F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)'P(M(X)W_1)\,(y - Z\delta_0)/q_1}{(y - Z\delta_0)'M[X(W_1)]\,(y - Z\delta_0)/(T - q_1 - K)}\,. \tag{5.5}$$

When $\delta = \delta_0$, $F(\delta_0;\, W_1) \sim F(q_1,\ T - q_1 - K)$. It follows that $F(\delta_0;\, W_1) > F(\alpha;\, q - K_2,\ T - q_1 - K)$ is a critical region with level α for testing $\delta = \delta_0$ while $\{\delta_0 :\ F(\delta_0;\, W_1) \leq F(\alpha;\, q_1,\ T - q_1 - K)\}$ is a confidence set with level $1 - \alpha$ for δ. Thus, the procedure developed for the case where no surprise variable is present applies without change. If generated regressors are used, we can write:

$$y - Z\delta_0 = W\hat{B}(\delta - \delta_0) + X\beta + e + V_*(\gamma - \delta_0) + \hat{V}(\delta - \delta)\,. \tag{5.6}$$

Replacing $W\hat{B}$ by $\tilde{Z} = W\tilde{B}$, where $\tilde{B}$ is an estimator independent of e and V, we get

$$y - Z\delta_0 = \tilde{Z}\delta_* + X\beta + u \tag{5.7}$$

where $\delta_* = \delta - \delta_0\,, u = e + V_*(\gamma - \delta_0) + \tilde{V}(\delta - \delta_0)$ and $\tilde{V} = Z - \tilde{Z}$. Here the hypothesis $\delta = \delta_0$ entails $H_0' : \delta_* = 0$. The F-statistic $F(\delta_0;\, \tilde{Z})$ defined in (3.6) follows an $F(G,\, T - G - K)$ distribution when $\delta = \delta_0$. Consequently, the tests and confidence set procedures based on $F(\delta_0;\, \tilde{Z})$ apply in the same way. Similarly, it is easy to see that the joint inference procedures described in Section 4 also apply without change.

Let us now consider the problem of testing an hypothesis on the coefficient of the surprise term, *i.e.* $H_0 :\ \gamma = \gamma_0$. In this case, it appears more difficult to obtain a finite-sample test under the assumptions (2.1) – (2.6). So we will assume that the following conditions, which are similar to assumptions made by Pagan (1984) setup, hold:

(5.8) a) $U_* = 0$; b) e and V are independent.

Then we can write:

(5.9) $$y = Z_* \delta + (Z - Z_*)\gamma + X\beta + e = Z\gamma + W_1 \delta_1^* + X\beta_* + e \, .$$

Subtracting $Z\gamma_0$ on both sides yields

(5.10) $$y - Z\gamma_0 = Z\gamma_* + W_1 \delta_{1*} + X\beta_* + e$$

where $\gamma_* = \gamma - \gamma_0$. We can thus test $\gamma = \gamma_0$ by testing $\gamma_* = 0$ in (5.10), using

(5.11) $$F(\gamma_0;\, Z) = \frac{(y - Z\gamma_0)' P(M([W_1,\, X])Z)\, (y - Z\gamma_0)/G}{(y - Z\gamma_0)' M([W_1,\, Z,\, X])\, (y - Z\gamma_0)/(T - G - q_1 - K)} \, .$$

When $\gamma = \gamma_0$, $F(\gamma_0;\, Z) \sim F(G,\ T - G - q_1 - K)$ so that $F(\gamma_0;\, Z) \geq F(\alpha;\ G,\ T - G - q_1 - K)$ is a critical region with level α for $\gamma = \gamma_0$ and

(5.12) $$\{\gamma_0 :\ F(\gamma_0;\, Z) \leq F(\alpha;\ G,\ T - G - q_1 - K)\}$$

is a confidence set with level $1 - \alpha$ for γ. When γ is a scalar, this confidence set can be written as:

(5.13) $$\left\{\gamma_0 : \frac{(y - Z\gamma_0)' D (y - Z\gamma_0)}{(y - Z\gamma_0)' E (y - Z\gamma_0)} \times \frac{\nu_2}{\nu_1} \leq F_\alpha \right\}$$

where $\nu_1 = G = 1$, $\nu_2 = T - G - q_1 - K$, $D = P(M([W_1,\, X]))$, $E = M([W_1,\, Z,\, X])$. Since the ratio ν_2/ν_1 always takes positive values, the confidence set is obtained by finding the values γ_0 that satisfy the inequality $a\gamma_0^2 + b\gamma_0 + c \leq 0$, where $a = Z'LZ$, $b = -2Z'Ly$, $c = y'Ly$, $L = D - H_\alpha E$ and $H_\alpha = (\nu_1/\nu_2)F_\alpha$. Finally it is straightforward to see that the problem of testing a joint hypothesis of the type $H_0 :\ \gamma = \gamma_0,\, R_1\beta = \nu_{10}$ can be treated by methods similar to the ones presented in Section 4.

6 Inference on general parameter transformations

The finite sample tests presented in this paper are based on extensions of Anderson–Rubin statistics. An apparent limitation of Anderson–Rubin type tests comes from the fact that they are designed for hypothesis fixing the complete vector of the endogenous (or unobserved) regressor coefficients. In this section, we propose a solution to this problem which is based on applying a projection technique. Even more generally, we study inference on general nonlinear transformations of δ in (2.1), or more generally of $(\delta', \nu_1')'$ where $\nu_1 = R_1\gamma$ is a linear transformation of γ, and we propose finite sample tests of general restrictions on subvectors of δ or $(\delta', \nu_1')'$. For a similar approach, see Dufour (1989, 1990) and Dufour and Kiviet (1998).

Let $\theta = \delta$ or $\theta = (\delta', \nu_1')'$ depending on the case of interest. In the previous sections, we derived confidence sets for θ which take the general form

(6.1) $$C_\theta(\alpha) = \{\theta_0 : F(\theta_0) \leq F_\alpha\}$$

where $F(\theta_0)$ is a test statistic for $\theta = \theta_0$ and F_α is a critical value such that $P[\theta \in C_\theta(\alpha)] \geq 1 - \alpha$. If $\theta = \theta_0$, we have

(6.2) $$P[\theta_0 \in C_\theta(\alpha)] \geq 1 - \alpha \, , \; P[\theta_0 \notin C_\theta(\alpha)] \leq \alpha \, .$$

Consider a (possibly nonlinear) transformation $\eta = f(\theta)$ of θ. Then it is easy to see that

(6.3) $$C_\eta(\alpha) \equiv \{\eta_0 : \eta_0 = f(\theta) \text{ for some } \theta \in C_\theta(\alpha)\}$$

is a confidence set for η with level at least $1-\alpha$, *i.e.*

(6.4) $$P[\eta \in C_\eta(\alpha)] \geq P[\theta \in C_\theta(\alpha)] \geq 1-\alpha$$

hence

(6.5) $$P[\eta \notin C_\eta(\alpha)] \leq \alpha\,.$$

Thus, by rejecting $H_0 : \eta = \eta_0$ when $\eta_0 \notin C_\eta(\alpha)$, we get a test of level α. Further

(6.6) $$\eta_0 \notin C_\eta(\alpha) \Leftrightarrow \eta_0 \neq f(\theta_0)\ ,\ \forall \theta_0 \in C_\theta(\alpha)$$

so that the condition $\eta_0 \notin C_\eta(\alpha)$ can be verified by minimizing $F(\theta_0)$ over the set $f^{-1}(\eta_0) = \{\theta_0 : f(\theta_0) = \eta_0\}$ and checking whether the infimum is greater than F_α.

When $\eta = f(\theta)$ is a scalar, it is easy to obtain a confidence interval for η by considering variables $\eta_L = \inf\{\eta_0 : \eta_0 \in C_\eta(\alpha)\}$ and $\eta_U = \sup\{\eta_0 : \eta_0 \in C_\eta(\alpha)\}$ obtained by minimizing and maximizing η_0 subject to the restriction $\eta_0 \in C_\eta(\alpha)$. It is then easy to see that

(6.7) $$P[\eta_L \leq \eta \leq \eta_U] \geq P[\eta \in C_\eta(\alpha)] \geq 1-\alpha$$

so that $[\eta_L, \eta_U]$ is a confidence interval with level $1-\alpha$ for η. Further, if such confidence intervals are built for several parametric functions, say $\eta_i = f_i(\theta)$, $i = 1, \ldots, m$, from the same confidence set $C_\theta(\alpha)$, the resulting confidence intervals $[\eta_{iL}, \eta_{iU}]$, $i = 1, \ldots, m$, are simultaneous at level $1-\alpha$, in the sense that the corresponding $m-$dimensional confidence box contains the true vector $(\eta_1, \ldots, \eta_m)$ with probability (at least) $1-\alpha$; for further discussion of simultaneous confidence sets, see Miller (1981), Savin (1984) and Dufour (1989). When a set of confidence intervals are not simultaneous, we will call them "marginal intervals".

Consider the special case where $\theta = \delta = (\delta_1, \delta_2')'$ and $\eta = \delta_1$, *i.e.* η is an element of δ. Then the confidence set $C_\eta(\alpha)$ takes the form:

(6.8) $$C_\eta(\alpha) = C_{\delta_1}(\alpha) = \{\delta_{10} : (\delta_{10}, \delta_2')' \in C_\delta(\alpha),\ \text{for some } \delta_2\}.$$

Consequently we must have:

(6.9) $$P[\delta_1 \in C_{\delta_1}(\alpha)] \geq 1-\alpha\,,\ P[\delta_{10} \notin C_{\delta_1}(\alpha)] \leq \alpha\,.$$

Further if we consider the random variables $\delta_1^L = \inf\{\delta_{10} : \delta_{10} \in C_{\delta_1}(\alpha)\}$ and $\delta_1^U = \sup\{\delta_{10} : \delta_{10} \in C_{\delta_1}(\alpha)\}$ obtained by minimizing and maximizing δ_{10} subject to the restriction $\delta_{10} \in C_{\delta_1}(\alpha)$, $[\delta_1^L, \delta_1^U]$ is a confidence interval with level $1-\alpha$ for δ_1. The test which rejects $H_0 : \delta_1 = \delta_{10}$ when $\delta_{10} \notin C_{\delta_1}(\alpha)$ has level not greater than α. Furthermore,

(6.10) $$\delta_{10} \notin C_{\delta_1}(\alpha) \Leftrightarrow F\left((\delta_{10}',\, \delta_2')'\right) > F_\alpha\,, \forall \delta_2\,.$$

Condition (6.10) can be checked by minimizing the $F\left((\delta_{10}',\, \delta_2')'\right)$ statistic with respect to δ_2 and comparing the minimal value with F_α. The hypothesis $\delta_1 = \delta_{10}$ is rejected if the infimum of $F\left((\delta_{10}',\, \delta_2')'\right)$ is greater than F_α. In practice, the minimizations and maximizations required by the above procedures can be performed easily through standard numerical techniques.

Finally, it is worthwhile noting that, even though the simultaneous confidence set $C_\theta(\alpha)$ for θ may be interpreted as a confidence set based on inverting LR-type tests for $\theta = \theta_0$ [or a profile likelihood confidence set (see Meeker and Escobar, 1995, or Chen and Jennrich, 1996)], projection-based confidence sets, such as $C_\eta(\alpha)$, are not (strictly speaking) LR confidence sets.

7 Asymptotic validity

In this section we show that the finite sample inference methods described above remain valid under weaker assumptions provided the number of observations is sufficiently large. Consider again the model described by (2.1) – (2.6) and (2.10), which yields the following equations:

$$y = Z\delta + X\gamma + u \ , \tag{7.1}$$

$$Z = W_1 B_1 + X_2 B_2 + V \ , \tag{7.2}$$

where $u = e - V\delta$. If we are prepared to accept a procedure which is only asymptotically "valid", we can relax the finite-sample assumptions (2.3) – (2.6) since the normality of error terms and their independence are no longer necessary. To do this, let us focus on the statistic $F(\delta_0;\, W_1)$ defined in (2.13). Then, under general regularity conditions, we can show:
a) under the null hypothesis $\delta = \delta_0$ the F-statistic in (2.13),

$$F(\delta_0;\, W_1) = \frac{(y - Z\delta_0)' M(X) W_1 [W_1' M(X) W_1]^{-1} W_1' M(X)\, (y - Z\delta_0)/q_1}{(y - Z\delta_0)' M([X,\, W_1])\, (y - Z\delta_0)/(T - q_1 - K)} \ , \tag{7.3}$$

follows a $\chi^2_{q_1}/q_1$ distribution asymptotically (as $T \to \infty$);
b) under the fixed alternative $\delta = \delta_1$, provided $B_1(\delta_1 - \delta_0) \neq 0$, the value of (2.13) tends to get infinitely large as T increases, *i.e.* the test based on $F(\delta_0;\, W_1)$ is consistent.

Assume that the following limits hold jointly:

$$\left(\frac{u'u}{T},\ \frac{u'V}{T},\ \frac{V'V}{T}\right) \ \to\ \left(\sigma_u^2,\ \Sigma_{uV},\ \Sigma_V\right) \ , \tag{7.4}$$

$$\left(\frac{X'X}{T},\ \frac{X'W_1}{T},\ \frac{W_1'W_1}{T}\right) \ \to\ \left(\Sigma_{XX},\ \Sigma_{XW_1},\ \Sigma_{W_1W_1}\right) \ , \tag{7.5}$$

$$\left(T^{-\frac{1}{2}}X'u,\ T^{-\frac{1}{2}}W_1'u,\ T^{-\frac{1}{2}}X'V,\ T^{-\frac{1}{2}}W_1'V\right) \Rightarrow \Phi \equiv \left(\Phi_{Xu},\ \Phi_{W_1u},\ \Phi_{XV},\ \Phi_{W_1V}\right) \ , \tag{7.6}$$

where $\to$ and $\Rightarrow$ denote respectively convergence in probability and convergence in distribution as $T \to \infty$, and the joint distribution of the random variables in Φ is multinormal with the covariance matrix of $(\Phi_{Xu}',\ \Phi_{W_1u}')'$ given by

$$\Sigma = V \begin{bmatrix} \Phi_{Xu} \\ \Phi_{W_1u} \end{bmatrix} = \begin{bmatrix} \sigma^2\Sigma_{XX} & \sigma\Sigma_{XW_1} \\ \sigma\Sigma_{W_1X} & \Sigma_{W_1W_1} \end{bmatrix}$$

where $\Sigma_{XW_1} = \Sigma_{W_1X}'$ and $\det(\Sigma) \neq 0$. We know from equation (2.11) that

$$y - Z\delta_0 = W_1 B_1 (\delta - \delta_0) + X\gamma_* + u \, .$$

Under the null hypothesis $\delta = \delta_0$, the numerator of $F(\delta_0;\, W_1)$ is equal to

$$\begin{aligned} N &= u' M(X) W_1 [W_1' M(X) W_1]^{-1} W_1' M(X) u / q_1 \\ &= u'(I - P) W_1 [W_1'(I - P) W_1]^{-1} W_1'(I - P) u / q_1 \\ &= \left[T^{-\frac{1}{2}} W_1'(I - P) u\right]' \left[\tfrac{1}{T} W_1'(I - P) W_1\right]^{-1} \left[T^{-\frac{1}{2}} W_1'(I - P) u\right] / q_1 \end{aligned}$$

where $P = P(X) = X(X'X)^{-1}X'$. Under the assumptions (7.4) to (7.6), we have the following convergence:

$$\begin{aligned} T^{-\frac{1}{2}} W_1'(I - P) u \ &= \ T^{-\frac{1}{2}} W_1' u - \left(\tfrac{1}{T} W_1' X\right) \left(\tfrac{1}{T} X'X\right)^{-1} \left(T^{-\frac{1}{2}} X'u\right) \\ &\Rightarrow \ \Phi_{W_1|X} \ \equiv \ \Phi_{W_1u} - \Sigma_{W_1X}\, \Sigma_{XX}^{-1}\, \Phi_{Xu} \end{aligned}$$

where

$$
\begin{aligned}
V[\Phi_{W_1|X}] &= V[\Phi_{W_1 u}] + \Sigma_{W_1 X}\, \Sigma_{XX}^{-1}\, V[\Phi_{Xu}]\, \Sigma_{XX}^{-1}\, \Sigma_{XW_1} \\
&\quad - E[\Phi_{W_1 u}\, \Phi'_{Xu}]\, \Sigma_{XX}^{-1}\, \Sigma_{XW_1} - \Sigma_{W_1 X}\, \Sigma_{XX}^{-1}\, E[\Phi_{Xu}\, \Phi'_{W_1 u}] \\
&= \Sigma_{W_1 W_1} - \Sigma_{W_1 X}\, \Sigma_{XX}^{-1}\, \Sigma_{XW_1}
\end{aligned}
$$

and

$$
\begin{aligned}
\tfrac{1}{T}\, W_1'(I-P)W_1 &= \tfrac{1}{T} W_1' W_1 - \tfrac{1}{T}\, W_1' X \left(\tfrac{1}{T} X'X\right)^{-1} \left(\tfrac{1}{T}\, X' W_1\right) \\
&\to \Sigma_{W_1 W_1} - \Sigma_{W_1 X}\, \Sigma_{XX}^{-1}\, \Sigma_{XW_1}.
\end{aligned}
$$

Consequently

$$
N \Rightarrow \Phi'_{W_1|X} \left(\Sigma_{W_1 W_1} - \Sigma_{W_1 X}\, \Sigma_{XX}^{-1}\, \Sigma'_{XW_1}\right)^{-1} \Phi_{W_1|X}/q_1 \sim \chi^2(q_1)/q_1 .
$$

This means that we can define the confidence intervals as the sets of points δ_0 for which the statistic (7.3) fails to reject, using the asymptotic $\chi^2_{q_1}/q_1$ critical values or the somewhat stronger (and probably more accurate) critical values of the Fisher distribution. Furthermore, it is easy to see that, both under the null and the alternative, the denominator D converges to σ_u^2 as $T \to \infty$:

$$
\begin{aligned}
D &= u'M([X,\, W_1])u/T \\
&= \tfrac{u'u}{T} \;-\; \tfrac{u'[X,\, W_1]\{[X,\, W_1]'[X,\, W_1]\}^{-1}[X,\, W_1]u}{T} \;\to\; \sigma_u^2 .
\end{aligned}
$$

Consider now a fixed alternative $\delta = \delta_1$. When $\delta = \delta_1$, we have

$$
\begin{aligned}
N &= [W_1 B_1(\delta_1 - \delta_0) + u]' M(X) W_1 [W_1' M(X) W_1]^{-1} W_1' M(X)\, [W_1 B_1(\delta_1 - \delta_0) + u]/q_1 \\
&= T^{-\frac{1}{2}} \left[(W_1' M(X) W_1) B_1(\delta_1 - \delta_0) + W_1' M(X) u)\right]' \left[\tfrac{W_1' M(X) W_1}{T}\right]^{-1} \\
&\quad \times T^{-\frac{1}{2}} \left[(W_1' M(X) W_1) B_1(\delta_1 - \delta_0) + W_1' M(X) u)\right]/q_1 .
\end{aligned}
$$

The behavior of the variable N depends on the convergence limits of the terms on the right-hand side of the last equation. It means that we can find the limit of N by showing the convergence of the individual components. The major building block of the expression for N is

$$
\begin{aligned}
T^{-\frac{1}{2}}[W_1' M(X) W_1 B_1(\delta_1 - \delta_0) + W_1' M(X) u] &= T^{\frac{1}{2}} \left(\tfrac{W_1' M(X) W_1}{T}\right) B_1(\delta_1 - \delta_0) \\
&\quad + \; T^{-\frac{1}{2}} W_1' M(X) u \, .
\end{aligned}
$$

As we have shown, $T^{-\frac{1}{2}} W_1' M(X) u$ converges in distribution to a random variable $\Phi_{W_1|X}$ and the term $T^{\frac{1}{2}} \left(\frac{W_1' M(X) W_1}{T}\right) B_1(\delta_1 - \delta_0)$ diverges in probability as T gets large. Consequently, under a fixed alternative, the whole expression goes to infinity, and the test is consistent. It is easy to prove similar asymptotic results for the other tests proposed in this paper.

8 Monte Carlo study

In this section, we present the results of a small Monte Carlo experiment comparing the performance of the exact tests proposed above with other available (asymptotically justified) procedures, especially Wald-type procedures.

A total number of one thousand realizations of an elementary version of the model (2.1)–(2.2), equivalent to Model 1 discussed by Pagan (1984), were simulated for a sample of size $T = 100$. In this particular specification, only one latent variable Z is present. The error terms in e and V (where

e and V are vectors of length 100) are independent with $N(0,\,1)$ distributions. We allow for the presence of only one instrumental variable W in the simulated model, which was also independently drawn from a $N(0,\,1)$ distribution. Following Pagan's original specification, there is no constant term or any exogenous variables included.

The explanatory power of the instrumental variable W depends on the value of the parameter B. Hence, we let B take the following values: 0, 0.05, 0.1, 0.5 and 1. When B is close or equal to zero, W has little or no explanatory power, *i.e.* W is a bad instrument for the latent variable Z. For each value of B we consider five null hypotheses:

$$H_0 : \delta = \delta_0\ , \quad \text{for } \delta_0 = 0,\ 1,\ 5,\ 10 \text{ and } 50\ ,$$

each one being tested against four alternative hypotheses of the form

$$H_1 : \delta = \delta_1\ , \quad \text{for } \delta_1 = \delta_0 + p^* I(\delta_0)\ .$$

The alternative H_1 is constructed by adding an increment to the value of δ_0 where $p^* =$ 0, 0.5, 1, 2 and 4, and $I(\delta_0) = 1$ for $\delta_0 = 0$, and $I(\delta_0) = \delta_0$ otherwise.

Table 2 summarizes the results. In the first 3 columns, we report the values of B, δ_0 and the alternative δ_1. When the entries in columns II and III are equal, we have $\delta_0 = \delta_1$, and the corresponding row reports the levels of the tests. The next three columns (IV, V and VI) show the performance of the Wald-type IV-based test [as proposed by Pagan (1984)], which consists in correcting the understated standard errors of a two stage procedure by replacing them by a 2SLS standard error. We report the corresponding results in column IV [asymptotic (As.)]. In cases where the level of Pagan's test exceeds 5%, we consider two correction methods. The first method is based on the critical value of the test at the 5% level for specific values of δ_0 and B in each row of the table [locally size-corrected tests; column V (C.L.)]. The critical value is obtained from an independent simulation with 1000 realizations of the model. Another independent simulation allows us to compute the critical value at 5% level in an extreme case when the instrumental variable is very bad, *i.e.* by supposing $B = 0$ also for each value of δ_0 [globally size-corrected tests; column VI (C.G.)]. This turns out to yield larger critical values and is thus closer to the theoretically correct critical value to be used here (on the assumption that B is actually unknown). In column VII, we present the power of the exact test based on the instrument substitution method. In the following four columns (VIII to XI) we show the performance of the exact test based on splitting the sample, where the numbers of observations used to estimate the structural equation are, respectively, 25, 50, 75 and 90 over 100 observations. Finally, we report the level and power of a naive two-stage test as well as the results of a test obtained by replacing the latent variable Z_* in the structural equation by the observed value Z.

Let us first discuss the reliability of the asymptotic procedures. The level of the IV test proposed by Pagan exceeds 5% essentially always when the parameter B is less then 0.5, sometimes by very wide margins. The tests based on the two-stage procedure or replacing the latent variable by the vector of observed values are both extremely unreliable no matter the value of the parameter B. The performance of Pagan's test improves once we move to higher values of the parameter B, *i.e.* when the quality of the instrument increases. The improvement is observed both in terms of level and power. It is however important to note that Pagan's test has, in general, the same or less power than the exact tests. The only exception is the sample split test reported in column VIII, where only 25 observations were retained to estimate the structural equation. For B higher then 0.5, the two other asymptotic tests are still performing worse then the other tests. They are indeed extremely unreliable. In the same range of B, the exact tests behave very well. They show the best power properties compared to the asymptotically based procedures and in general outperform the other tests.

TABLE 2

SIMULATION STUDY OF TEST PERFORMANCE FOR A MODEL WITH UNOBSERVED REGRESSORS

Parameter Values			Rejection Frequencies									
B	δ_0	δ_1	Wald-type			IS	Split-sample				2S	OLS
			As.	C.L.	C.G.		25	50	75	90		
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
0.00	0.0	0.0	0.1	·	·	5.1	5.1	6.1	5.2	5.4	5.1	·
0.00	0.0	0.5	0.0	·	·	4.7	5.1	4.4	4.1	3.9	4.7	·
0.00	0.0	1.0	0.0	·	·	5.6	4.8	5.5	5.7	5.4	5.6	·
0.00	0.0	2.0	0.0	·	·	4.2	4.5	4.5	3.8	4.5	4.2	·
0.00	0.0	4.0	0.0	·	·	5.2	5.3	5.9	4.3	5.0	5.2	·
0.00	1.0	1.0	7.3	5.1	5.1	5.0	4.6	4.9	4.8	5.2	15.7	4.7
0.00	1.0	1.5	6.8	5.5	5.5	4.4	4.8	4.4	5.4	6.1	15.7	6.8
0.00	1.0	2.0	7.6	5.9	5.9	5.0	4.3	4.8	4.8	5.1	17.9	6.5
0.00	1.0	3.0	8.6	6.6	6.6	6.3	5.0	4.9	5.0	5.8	19.9	7.0
0.00	1.0	5.0	6.6	4.9	4.9	4.4	4.3	4.6	5.5	4.6	18.1	5.1
0.00	5.0	5.0	54.1	5.5	5.5	5.1	5.5	4.2	5.2	4.9	70.5	69.3
0.00	5.0	7.5	52.8	5.4	5.4	4.9	6.1	4.9	5.1	4.6	69.7	69.0
0.00	5.0	10.0	56.5	5.7	5.7	4.8	4.5	6.1	5.0	4.8	71.7	71.5
0.00	5.0	15.0	50.7	4.6	4.6	4.8	4.5	4.3	4.5	3.8	66.6	67.0
0.00	5.0	25.0	52.7	5.2	5.2	4.6	4.5	4.6	5.6	5.0	67.8	68.8
0.00	10.0	10.0	69.0	4.5	4.5	4.9	5.3	6.0	4.9	5.1	84.5	85.0
0.00	10.0	15.0	68.4	5.7	5.7	5.9	4.7	5.0	5.6	4.5	84.3	83.9
0.00	10.0	20.0	68.6	5.0	5.0	5.7	4.3	4.9	4.7	5.2	84.6	84.3
0.00	10.0	30.0	70.2	4.9	4.9	4.5	5.4	5.2	5.0	5.2	85.4	84.4
0.00	10.0	50.0	68.7	5.3	5.3	4.8	4.2	5.1	5.6	5.0	83.6	83.1
0.00	50.0	50.0	86.5	6.4	6.4	5.4	4.4	5.0	5.1	5.4	96.9	96.5
0.00	50.0	75.0	85.2	6.7	6.7	6.2	3.9	5.0	6.6	6.7	95.1	96.1
0.00	50.0	100.0	87.4	5.2	5.2	4.6	6.5	5.0	4.5	5.5	96.8	96.4
0.00	50.0	150.0	85.8	6.5	6.5	5.8	5.0	5.3	5.9	5.9	97.1	97.1
0.00	50.0	250.0	86.7	6.8	6.8	5.9	4.8	6.0	6.2	5.8	97.1	97.3
0.05	0.0	0.0	0.0	·	·	4.8	5.0	3.6	3.6	5.3	4.8	·
0.05	0.0	0.5	0.2	·	·	4.9	5.1	5.5	4.8	5.2	4.9	·
0.05	0.0	1.0	0.0	·	·	7.4	5.4	5.7	6.2	7.6	7.4	·
0.05	0.0	2.0	0.3	·	·	16.6	8.7	11.7	14.7	15.7	16.6	·
0.05	0.0	4.0	1.0	·	·	47.8	16.4	26.9	38.1	44.0	47.8	·

TABLE 2 (continued)

0.05	1.0	1.0	6.9	5.2	5.6	4.7	4.8	4.4	4.8	5.5	16.9	7.9
0.05	1.0	1.5	6.0	4.6	4.7	5.4	6.0	6.0	5.4	5.2	16.9	7.5
0.05	1.0	2.0	4.7	3.9	3.9	5.3	5.7	4.6	5.1	5.2	18.1	7.6
0.05	1.0	3.0	4.0	2.7	2.7	9.9	6.3	7.4	8.4	10.5	25.3	7.4
0.05	1.0	5.0	2.6	2.1	2.1	27.0	9.0	14.9	23.2	25.4	51.1	5.6
0.05	5.0	5.0	33.8	4.6	1.6	4.6	5.8	5.3	5.2	4.8	71.7	72.7
0.05	5.0	7.5	21.0	2.3	0.2	6.3	4.8	4.6	5.3	6.0	69.7	71.4
0.05	5.0	10.0	12.4	0.4	0.1	8.7	4.8	5.6	7.6	8.5	71.9	69.9
0.05	5.0	15.0	5.1	0.1	0.0	14.8	6.1	8.6	11.7	13.2	81.2	66.9
0.05	5.0	25.0	3.9	0.0	0.0	47.1	15.3	26.2	39.1	43.0	93.6	59.0
0.05	10.0	10.0	34.9	7.6	0.2	6.3	6.6	6.3	6.4	6.5	84.8	84.0
0.05	10.0	15.0	22.9	1.3	0.0	6.4	4.4	5.8	5.8	5.9	85.8	78.9
0.05	10.0	20.0	14.1	0.6	0.0	8.6	5.1	6.1	6.7	7.6	88.9	79.0
0.05	10.0	30.0	5.1	0.0	0.0	14.5	6.7	10.4	13.3	13.9	90.0	74.2
0.05	10.0	50.0	4.4	0.1	0.0	52.5	18.6	30.1	40.8	49.1	97.5	62.2
0.05	50.0	50.0	32.7	5.1	0.0	4.7	4.7	6.0	5.2	4.5	97.5	92.0
0.05	50.0	75.0	21.2	1.7	0.0	6.4	4.5	4.9	5.3	6.2	96.9	89.2
0.05	50.0	100.0	14.3	0.6	0.0	8.5	5.8	7.0	7.2	7.3	97.7	86.5
0.05	50.0	150.0	6.4	0.3	0.0	17.6	7.0	11.1	15.1	15.8	97.0	79.8
0.05	50.0	250.0	3.2	0.0	0.0	51.3	16.0	28.3	38.7	46.1	99.8	65.3
0.10	0.0	0.0	0.0	·	·	4.8	4.2	4.9	4.5	5.0	4.8	·
0.10	0.0	0.5	0.2	·	·	8.2	6.8	7.1	6.9	7.4	8.2	·
0.10	0.0	1.0	0.1	·	·	15.8	7.1	8.9	13.9	13.5	15.8	·
0.10	0.0	2.0	2.4	·	·	49.4	16.9	29.3	40.7	46.0	49.4	·
0.10	0.0	4.0	8.8	·	·	97.1	47.7	78.9	93.2	95.9	97.1	·
0.10	1.0	1.0	7.3	4.4	5.6	4.7	5.3	5.1	4.5	4.7	15.2	14.0
0.10	1.0	1.5	4.4	2.9	3.8	6.6	4.4	5.6	6.3	6.2	19.8	16.2
0.10	1.0	2.0	3.0	1.9	2.3	10.6	6.6	7.3	9.5	10.0	25.8	14.3
0.10	1.0	3.0	0.9	0.7	0.9	28.3	9.3	18.7	23.8	26.6	49.5	10.9
0.10	1.0	5.0	0.6	0.3	0.5	80.1	26.4	49.4	66.1	74.1	92.4	7.4
0.10	5.0	5.0	17.4	4.6	0.6	5.2	5.2	4.7	4.8	5.4	71.5	78.9
0.10	5.0	7.5	5.8	1.1	0.0	7.2	6.0	6.4	7.4	7.5	73.7	74.4
0.10	5.0	10.0	2.3	0.2	0.0	16.5	7.9	11.1	14.0	16.0	81.6	73.0
0.10	5.0	15.0	1.0	0.0	0.0	50.5	15.4	27.2	38.7	45.7	94.8	65.2
0.10	5.0	25.0	0.4	0.0	0.0	97.0	45.5	76.6	89.4	95.0	100.0	46.9

TABLE 2 (continued)

0.10	10.0	10.0	17.1	5.6	0.0	4.7	4.6	4.7	6.0	5.7	84.6	86.0
0.10	10.0	15.0	6.0	1.5	0.0	7.0	6.4	7.0	8.0	6.7	85.0	84.8
0.10	10.0	20.0	2.7	0.1	0.0	14.1	6.5	10.4	11.3	13.2	90.7	79.4
0.10	10.0	30.0	0.8	0.0	0.0	51.9	18.0	28.8	40.9	47.9	97.8	68.9
0.10	10.0	50.0	0.5	0.1	0.0	96.5	49.5	77.6	91.6	94.1	100.0	49.3
0.10	50.0	50.0	19.8	4.8	0.0	5.9	4.5	5.1	5.1	4.8	97.0	89.6
0.10	50.0	75.0	6.5	0.8	0.0	7.7	5.5	5.7	6.6	6.6	97.4	86.1
0.10	50.0	100.0	3.5	0.5	0.0	17.7	9.4	12.3	15.7	17.3	97.7	82.2
0.10	50.0	150.0	0.9	0.0	0.0	45.9	16.4	27.7	39.5	43.5	99.6	73.1
0.10	50.0	250.0	0.8	0.0	0.0	97.2	48.9	78.5	94.0	95.6	100.0	49.7
0.50	0.0	0.0	2.7	·	·	4.6	5.4	4.3	4.8	4.4	4.6	·
0.50	0.0	0.5	60.3	·	·	67.7	24.1	41.8	55.0	63.8	67.7	·
0.50	0.0	1.0	98.8	·	·	99.9	68.7	92.8	99.1	99.6	99.9	·
0.50	0.0	2.0	99.6	·	·	100.0	98.4	100.0	100.0	100.0	100.0	·
0.50	0.0	4.0	99.0	·	·	100.0	100.0	100.0	100.0	100.0	100.0	·
0.50	1.0	1.0	5.3	4.8	4.2	5.0	4.7	5.1	4.9	4.6	17.6	98.4
0.50	1.0	1.5	8.5	5.2	2.6	41.4	15.5	24.4	32.4	39.3	64.4	92.8
0.50	1.0	2.0	68.0	58.1	47.4	93.4	39.7	68.6	84.3	90.6	98.4	62.6
0.50	1.0	3.0	98.7	98.2	97.5	100.0	90.3	99.8	100.0	100.0	100.0	1.7
0.50	1.0	5.0	99.8	99.7	99.6	100.0	100.0	100.0	100.0	100.0	100.0	0.1
0.50	5.0	5.0	7.4	5.6	0.0	5.1	4.2	5.0	4.4	5.3	69.6	100.0
0.50	5.0	7.5	9.7	1.7	0.0	66.6	18.4	39.4	54.5	61.6	97.7	99.9
0.50	5.0	10.0	92.6	69.1	0.0	99.7	63.9	90.5	97.9	99.4	100.0	99.2
0.50	5.0	15.0	99.1	97.9	0.0	100.0	98.8	100.0	100.0	100.0	100.0	5.4
0.50	5.0	25.0	99.6	99.1	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.1
0.50	10.0	10.0	6.9	5.2	0.0	5.1	5.5	5.2	4.2	5.6	83.5	100.0
0.50	10.0	15.0	8.6	1.0	0.0	67.9	21.7	39.9	55.4	62.0	99.6	99.7
0.50	10.0	20.0	92.1	74.2	0.0	99.7	66.6	93.2	98.7	99.8	100.0	99.1
0.50	10.0	30.0	99.5	99.0	0.0	100.0	99.4	100.0	100.0	100.0	100.0	5.6
0.50	10.0	50.0	99.5	99.1	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.0
0.50	50.0	50.0	8.3	6.7	0.0	4.6	3.9	4.5	4.4	4.5	96.3	100.0
0.50	50.0	75.0	8.9	3.7	0.0	69.8	21.8	39.1	56.1	64.7	99.9	100.0
0.50	50.0	100.0	94.3	88.8	0.0	99.6	63.2	92.3	98.5	99.5	100.0	99.4
0.50	50.0	150.0	98.8	98.3	0.0	100.0	99.4	100.0	100.0	100.0	100.0	5.2
0.50	50.0	250.0	99.5	99.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	0.3

TABLE 2 (continued)

1.00	0.0	0.0	5.1	·	·	5.6	4.9	5.0	5.6	5.8	5.6	·
1.00	0.0	0.5	99.5	·	·	99.5	64.9	91.2	98.5	99.2	99.5	·
1.00	0.0	1.0	100.0	·	·	100.0	99.2	100.0	100.0	100.0	100.0	·
1.00	0.0	2.0	100.0	·	·	100.0	100.0	100.0	100.0	100.0	100.0	·
1.00	0.0	4.0	100.0	·	·	100.0	100.0	100.0	100.0	100.0	100.0	·
1.00	1.0	1.0	6.8	7.2	3.8	6.3	5.4	7.0	6.9	6.8	17.9	99.7
1.00	1.0	1.5	87.9	89.2	82.2	93.3	39.5	68.3	84.7	90.1	98.1	33.7
1.00	1.0	2.0	100.0	100.0	100.0	100.0	89.9	99.8	100.0	100.0	100.0	0.7
1.00	1.0	3.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	57.3
1.00	1.0	5.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1
1.00	5.0	5.0	4.8	4.4	0.0	4.1	5.5	4.4	4.7	4.8	67.2	100.0
1.00	5.0	7.5	98.8	98.3	0.0	99.6	62.5	91.5	98.0	99.4	100.0	67.6
1.00	5.0	10.0	100.0	100.0	0.0	100.0	99.0	100.0	100.0	100.0	100.0	1.3
1.00	5.0	15.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	65.9
1.00	5.0	25.0	100.0	100.0	7.3	100.0	100.0	100.0	100.0	100.0	100.0	98.3
1.00	10.0	10.0	5.1	4.4	0.0	6.0	6.2	5.8	6.9	6.3	85.3	100.0
1.00	10.0	15.0	98.8	98.5	0.0	99.6	63.1	91.1	97.7	99.4	100.0	69.5
1.00	10.0	20.0	100.0	100.0	0.0	100.0	99.0	100.0	100.0	100.0	100.0	0.6
1.00	10.0	30.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	66.5
1.00	10.0	50.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	99.2
1.00	50.0	50.0	5.2	5.0	0.0	5.5	5.5	5.3	5.2	6.9	96.8	100.0
1.00	50.0	75.0	99.0	98.7	0.0	99.9	65.8	91.4	98.3	99.3	100.0	68.1
1.00	50.0	100.0	100.0	100.0	0.0	100.0	98.8	100.0	100.0	100.0	100.0	0.6
1.00	50.0	150.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	67.0
1.00	50.0	250.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0

Notes:

I: value of parameter B;
II: null hypothesis;
III: alternative hypothesis;
IV: Pagan's test;
V: Pagan's test locally size-corrected (B known);
VI: Pagan's test globally size-corrected ($B = 0$);
VII: instrument substitution test (IS);
VIII: sample split test using 25 observations for the structural equation;
IX: sample split using 50 observations;
X: sample split using 75 observations;
XI: sample split using 90 observations;
XII: two-stage test ($2S$);
XIII: test with latent variable replaced by observed vector (OLS).

9 Empirical illustrations

In this section, we present empirical results on inference in two distinct economic models with latent regressors. The first example is based on Tobin's marginal q model of investment (Tobin, 1969), with fixed assets used as the instrumental variable for q. The second model stems from educational economics and relates students' academic achievements to a number of personal characteristics and other socioeconomic variables. Among the personal characteristics, we encounter a variable defined as "self–esteem" which is viewed as an imperfect measure of a latent variable and is instrumented by measures of the prestige of parents' professional occupation. The first example is one where we have good instruments, while the opposite holds for the second example.

Consider first Tobin's marginal q model of investment (Tobin, 1969). Investment of an individual firm is defined as an increasing function of the shadow value of capital, equal to the present discounted value of expected marginal profits. In Tobin's original setup, investment behavior of all firms is similar and no difference arises from the degree of availability of external financing. In fact, investment behavior varies across firms and is determined to a large extent by financial constraints some firms are facing in the presence of asymmetric information. For those firms, external financing may either be too costly or not provided for other reasons. Thus investment depends heavily on the firm's own source of financing, namely the cash flow. To account for differences in investment behavior implied by financial constraints, several authors [Abel (1979), Hayashi (1985), Abel and Blanchard (1986), Abel and Eberly (1993)] introduced the cash flow as an additional regressor to Tobin's q model. It can be argued that another explanatory variable controlling the profitability of investment is also required. For this reason, one can argue that the firm's income has to be included in the investment regression as well. The model is thus

$$I_i = \gamma_0 + \delta Q_i + \gamma_1 CF_i + \gamma_2 R_i + e_i \tag{9.1}$$

where I_i denotes the investment expenses of an individual firm i, CF_i and R_i its cash flow and income respectively, while Q_i is Tobin's q measured by equity plus debt and approximated empirically by adding data on current debt, long term debt, deferred taxes and credit, minority interest and equity less inventory; δ and $\gamma = (\gamma_0,\ \gamma_1,\ \gamma_2)'$ are fixed coefficients to be estimated. Given the compound character of Q_i, which is constructed from several indexes, fixed assets are used as an explanatory variable for Q_i in the regression which completes the model:

$$Q_i = \beta_0 + \beta_1 F_i + v_i\,. \tag{9.2}$$

For the purpose of building finite-sample confidence intervals following the instrument substitution method, the latter equation may be replaced (without any change to the results) by the more general equation (called below the "full instrumental regression"):

$$Q_i = \beta_0 + \beta_1 F_i + \beta_3 CF_i + \beta_4 R_i + v_i\,. \tag{9.3}$$

Our empirical work is based on "Stock Guide Database" containing data on companies listed at the Toronto and Montreal stock exchange markets between 1987 and 1991. The records consist of observations on economic variables describing the firms' size and performance, like fixed capital stock, income, cash flow, stock market price, etc. All data on the individual companies have previously been extracted from their annual, interim and other reports. We retained a subsample of 9285 firms whose stocks were traded on the Toronto and Montreal stock exchange markets in 1991.

Since we are interested in comparing our inference methods to the widely used Wald-type tests, we first consider the approach suggested by Pagan (1984). Since usual estimators of coefficient

TABLE 3

TOBIN'S Q MODEL, $N = 9285$

A) 2SLS estimators of investment equation (9.1)

Dependent variable: INVESTMENT (I)				
Explanatory variable	Estimated coefficient	Standard error	t statistic	p-value
Constant	0.0409	0.0064	6.341	0.0000
Q	0.0052	0.0013	3.879	0.0001
CF	0.8576	0.0278	30.754	0.0000
R	0.0002	0.0020	0.109	0.9134

B) Instrumental OLS regressions _ Dependent variable: Q

	Full instrumental regression				Equation (9.2)			
Regressor	Estimated coefficient	Stand. error	t	p-value	Estimated coefficient	Stand. error	t	p-value
Constant	0.6689	0.0919	7.271	0.0000	1.0853	0.1418	7.650	0.0000
F	-2.7523	0.0527	-52.195	0.0000	2.4063	0.0400	60.100	0.0000
CF	21.2102	0.3188	66.517	0.0000				
R	1.2273	0.0291	42.111	0.0000				

C) Confidence intervals

Marginal confidence intervals for δ		Projection-based simultaneous confidence intervals (instrument substitution)	
Method	Interval	Coefficient	Interval
2SLS	$[0.0026\ ,\ 0.0078]$	γ_0	$[0.0257\ ,\ 0.0564]$
Augmented two-stage	$[0.0025\ ,\ 0.0079]$	δ	$[0.0037\ ,\ 0.0072]$
Two-stage	$[-0.0091\ ,\ -0.0029]$	γ_1	$[0.7986\ ,\ 0.9366]$
Instrument substitution	$[0.0025\ ,\ 0.0078]$	γ_2	$[0.0033\ ,\ 0.0042]$
Sample split 50%	$[0.0000\ ,\ 0.0073]$		
Sample split 75%	$[0.0017\ ,\ 0.0077]$		
Sample split 90%	$[0.0023\ ,\ 0.0078]$		

variances obtained from the OLS estimation of equation (9.1) with Q_i replaced by $\hat{Q}_i$ are inconsistent [for a proof, see Pagan (1984)], Pagan proposed to use standard two-stage least squares (2SLS) methods, which yield in the present context (under appropriate regularity conditions) asymptotically valid standard errors and hypothesis tests. For the 2SLS estimation of model (9.1)–(9.2), the dependent variable I_i is first regressed on all the exogenous variables of the system, *i.e.*, the constant, CF_i, R_i and F_i, where F_i is the identifying instrument for Q_i, and then the fitted values $\hat{Q}_i$ are substituted for Q_i in the second stage regression.

The results are summarized in Tables 3A, while the instrumental OLS regressions appear in 3B. From the latter, we see that the identifying instrument for Q is strongly significant and so appears to be a "good" instrument. Table 3C presents 95% (marginal) confidence intervals for Tobin's q parameter based on various methods, as well as projection-based simultaneous confidence intervals for the coefficients of equation (9.1). The three first intervals are obtained from, respectively, 2SLS, two-stage and augmented two-stage methods by adding or subtracting 1.96 times the standard error to/from the estimated parameter value.[1] Below we report the exact confidence intervals (instrument substitution and sample split) based on the solution of quadratic equations as described in Sections 2 and 3. Recall that the precision of the confidence intervals depends, in the case of the sample split method, on the number of observations retained for the estimation of the structural equation. We thus show the results for, respectively, 50%, 75% and 90% of the entire sample (selected randomly). The simultaneous confidence intervals for the elements of the vector $\theta = (\gamma_0, \delta, \gamma_1, \gamma_2)'$ are obtained by first building a simultaneous confidence set $C_\theta(\alpha)$, with level $1 - \alpha = 0.95$ for θ according to the instrument substitution method described in Section 4 and then by both minimizing and maximizing each coefficient subject to the restriction $\theta \in C_\theta(\alpha)$ [see Section 6]. The program used to perform these constrained optimizations is the subroutine NCONF from the IMSL mathematical library. The corresponding four-dimensional confidence box has level 95% (or possibly more), *i.e*. we have simultaneous confidence intervals (at level 95%).

From these results, we see that all the confidence intervals for δ, except for the two-stage interval (which is not asymptotically valid), are quite close to each other. Among the finite-sample intervals, the ones based on the instrument substitution and the 90% sample split method appear to be the most precise. It is also worthwhile noting that the projection-based simultaneous confidence intervals all appear to be quite short. This shows that the latter method works well in the present context and can be implemented easily.

Let us now consider another example where, on the contrary, important discrepancies arise between the intervals based on the asymptotic and the exact inference methods. Montmarquette and Mahseredjian (Montmarquette and Mahseredjian, 1989; Montmarquette, Houle, Crespo and Mahseredjian, 1989) studied students' academic achievements as a function of personal and socioeconomic explanatory variables. Students' school results in French and mathematics are measured by the grade, taking values on the interval $0 - 100$. The grade variable is assumed to depend on personal characteristics, such as age, intellectual ability (IQ) observed in kindergarten and "self–esteem" measured on an adapted children self–esteem scale ranging from 0 to 40. Other explanatory variables include parents' income, father's and mother's education measured in number of years of schooling, the number of siblings, student's absenteeism, his own education and experience as well as the class size. We examine the significance of self–esteem, which is viewed as an imperfectly measured latent variable to explain the first grader's achievements in mathematics. The self esteem of younger children was measured by a French adaptation of the McDaniel–Piers scale. Noting the measurement scale may not be equally well adjusted to the age of all students and due to the high degree of arbitrariness in the choice of this criterion, the latter was instrumented by Blishen indices reflecting the prestige of father's and mother's professional occupations in order to take account of eventual mismeasurement.

The data stem from a 1981–1982 survey of first graders attending Montreal francophone public elementary schools. The sample consists of 603 observations on students' achievements in mathematics. The model considered is:

$$\begin{aligned}\text{(9.4)} \quad \mathrm{LMAT}_i = \; & \beta_0 + \delta\,\mathrm{SE}_i + \beta_1\,\mathrm{IQ}_i + \beta_2\,\mathrm{I}_i + \beta_3\,\mathrm{FE}_i + \beta_4\,\mathrm{ME}_i + \beta_5\,\mathrm{SN}_i \\ & + \beta_6\,\mathrm{A}_i + \beta_7\,\mathrm{ABP}_i + \beta_8\,\mathrm{EX}_i + \beta_9 \mathrm{ED}_i + \beta_{10}\,\mathrm{ABS}_i + \beta_{11}\,\mathrm{CS}_i + e_i\end{aligned}$$

where (for each individual i) $\mathrm{LMAT} = \ell n$(grade/(100 $-$ grade)), $\mathrm{SE} = \ell n$(self esteem test result/(40 $-$ self esteem test result)), IQ is a measure of intelligence (observed in kindergarten), I is

[1]The augmented two-stage method uses all the available instruments to compute the generated regressors (full instrumental regression), rather than the restricted instrumental equation (9.2). As with the two-stage method, OLS-based coefficient standard errors obtained in this way are inconsistent; see Pagan (1984) for further discussion.

parents' income, FE and ME are father's and mother's years of schooling, SN denotes the sibling's number, A is the age of the student, ABP is a measure of teacher's absenteeism, EX indicates the years of student's work experience, ED measures his education in years, ABS is student's absenteeism and CS denotes the class size. Finally, the instrumental regression is:

$$\mathrm{SE}_i = \gamma_0 + \gamma_1 \, \mathrm{FP}_i + \gamma_2 \, \mathrm{MP}_i + v_i \tag{9.5}$$

where FP and MP correspond to the prestige of the father and mother's profession expressed in terms of Blishen indices. We consider also the more general instrumental regression which includes all the explanatory variables on the right-hand side of (9.4) except SE. The 2SLS estimates and projection-based simultaneous confidence are reported in Table 4A while the results of the instrumental regressions appear in Table 4B.

Standard (bounded) Wald-type confidence intervals are of course entailed by the 2SLS estimation. For δ however, the instrument substitution method yields the confidence interval defined by the inequality: $-31.9536\,\delta_0^2 - 84.7320\,\delta_0 - 850.9727 \leq 0$. Since the roots of this second order polynomial are complex and $a < 0$, this confidence interval actually covers the whole real line. Indeed, from the full instrumental regression and using t-tests as well as the relevant F-test (Table 4B), we see that the coefficients of FP and MP are not significantly different from zero, *i.e.* the latter appear to be poor instruments. So the fact that we get here an unbounded confidence interval for δ is expected in the light of the remarks at the end of Section 2. The projection-based confidence intervals (Table 4A) yield the same message for δ, although it is of interest to note that the intervals for the other coefficients of the model can be quite short despite the fact that δ may be difficult to identify. As in the case of multicollinearity problems in linear regressions, inference about some coefficients of a model remains feasible even if the certain parameters are not identifiable.

10 Conclusions

The inference methods presented in this paper are applicable to a variety of models, such as regressions with unobserved explanatory variables or structural models which can be estimated by instrumental variable methods (e.g., simultaneous equations models). They may be considered as extensions of Anderson-Rubin procedures where the major improvement consists of providing tests of hypotheses on subsets or elements of the parameter vector. This is accomplished via a projection technique allowing for inference on general possibly nonlinear transformations of the parameter vector of interest. We emphasized that our test statistics, being pivotal or at least boundedly pivotal functions, yield valid confidence sets which are unbounded with a non-zero probability. The unboundedness of confidence sets is of particular importance when the instruments are poor and the parameter of interest is not identifiable or close to being unidentified. Accordingly, a valid confidence set should cover the entire set of real numbers since all values are observationally equivalent [see Dufour (1997) and Gleser and Hwang (1987)]. Our empirical results indicate that inference methods based on Wald-type statistics are unreliable in the presence of poor instruments since such methods typically yield bounded confidence sets with probability one. The results in this paper thus underscore another shortcoming of Wald-type procedures which is quite distinct from other problematic properties, such as non-invariance to reparameterizations [see Dagenais and Dufour (1991)].

In general, non-identifiability of parameters results either from low quality instruments or, more fundamentally, from a poor model specification. A valid test yielding an unbounded confidence set becomes thus a relevant indicator of problems involving the econometric setup. The power properties of exact and Wald-type tests were compared in a simulation-based experiment. The test

TABLE 4

MATHEMATICS ACHIEVEMENT MODEL _ $N = 603$

2SLS estimators of achievement equation (9.4)					Projection-based
Dependent variable: LMAT					95% confidence intervals
Explanatory variable	Estimated coefficient	Standard error	t statistic	p-value	(instrument substitution)
Constant	-4.1557	0.9959	-4.173	0.0000	[-4.8601 , -3.7411]
SE	0.2316	0.3813	0.607	0.5438	$(-\infty, +\infty)$
IQ	0.0067	0.0015	4.203	0.0000	[0.006600 , 0.006724]
I	0.0002	0.3175	0.008	0.9939	[-0.09123 , 0.10490]
FE	0.0015	0.0089	0.172	0.8636	[-0.00914 , 0.01889]
ME	0.0393	0.0117	3.342	0.0009	[0.02868 , 0.05762]
SN	-0.0008	0.0294	-0.029	0.9767	[-0.1546 , 0.1891]
A	0.0144	0.0070	2.050	0.0408	[0.01272 , 0.01877]
ABP	-0.0008	0.0005	-1.425	0.1548	[-0.003778 , 0.000865]
EX	-0.0056	0.0039	-1.420	0.1561	[-0.01307 , 0.00333]
ED	-0.0007	0.0206	-0.035	0.9718	[-0.0123 , 0.2196]
ABS	-0.0001	0.0002	-0.520	0.6033	[-0.0001764 , 0.0000786]
CS	-0.0184	0.0093	-1.964	0.0500	[-0.03003 , -0.009790]
Marginal 95% quadratic confidence interval for δ					$(-\infty, +\infty)$

Instrumental OLS regressions _ Dependent variable: SE

	Full instrumental regression				Equation (9.5)			
Regressor	Estimated coefficient	Stand. error	t	p-value	Estimated coefficient	Stand. error	t	p-value
Constant	-1.2572	1.0511	-1.1960	0.232	0.8117	0.1188	6.830	0.0000
FP	0.5405	0.3180	1.7000	0.090	0.5120	0.2625	1.951	0.0516
FM	0.3994	0.3327	1.2004	0.230	0.6170	0.2811	2.194	0.0286
IQ	0.003822	0.000611	6.2593	0.000				
I	0.02860	0.03161	0.9049	0.366	F-statistic for significance of FP and FM in full instrumental regression: $F(2, 589) = 2.654$ (p-value = 0.078)			
FE	-0.01352	0.01136	-1.1899	0.235				
ME	-0.004028	0.01517	-0.2655	0.791				
SN	-0.01439	0.03325	-0.4326	0.665				
A	0.003216	0.008161	0.3941	0.694				
ABP	0.000698	0.000577	1.2108	0.226				
EX	-0.002644	0.004466	-0.5920	0.554				
ED	-0.02936	0.02080	-1.4117	0.159				
ABS	0.000426	0.000194	2.1926	0.029				
CS	0.01148	0.009595	1.1966	0.232				

performances were examined by simulations on a simple model with varying levels of instrument quality and the extent to which the null hypotheses differ from the true parameter value. We found that the tests proposed in this paper were preferable to more usual IV-based Wald-type methods from the points of view of level control and power. This seems to occur despite the fact that AR-type procedures involve "projections onto a high-dimensional subspace which could result in reduced power and thus wide confidence regions" [Staiger and Stock (1997, p. 570)]. However, it is important to remember that size-correcting Wald-type procedures requires one to use huge critical values that can easily destroy power. Wald-type procedures can be made useful only at the cost introducing important and complex restrictions on the parameter space that one is not generally prepare to impose; for further discussion of these difficulties, see Dufour (1997, Section 6).

It is important to note that although the simulations were performed under the normality assumption, our tests yield valid inferences in more general cases involving non-Gaussian errors and weakly exogenous instruments. This result has a theoretical justification and is also confirmed by our empirical examples. Since the inference methods we propose are as well computationally easy to perform, they can be considered as a reliable and a powerful alternative to more usual Wald-type procedures.

References

ABEL, A. (1979): *Investment and the Value of Capital*. New York: Garland Publishing.

ABEL, A. AND O. J. BLANCHARD (1986): "The Present Value of Profits and Cyclical Movements in Investment," *Econometrica*, 54, 249–273.

ABEL, A. AND J. EBERLY (1993): "A Unified Model of Investment under Uncertainty," Technical Report 4296, National Bureau of Economic Research, Cambridge, MA.

ANDERSON, T. W. AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

ANGRIST, J. D. AND A. B. KRUEGER (1994): "Split Sample Instrumental Variables," Technical Working Paper 150, N.B.E.R., Cambridge, MA.

BARRO, R. J. (1977): "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review*, 67, 101–115.

BATES, D. M. AND D. G. WATTS (1988): *Nonlinear Regression Analysis and its Applications*. New York: John Wiley & Sons.

BOUND, J., D. A. JAEGER, AND R. BAKER (1993): "The Cure can be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," Technical Working Paper 137, National Bureau of Economic Research, Cambridge, MA.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BUSE, A. (1992): "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

CHEN, J.-S. AND R. I. JENNRICH (1996): "The Signed Root Deviance Profile and Confidence Intervals in Maximum Likelihood Analysis," *Journal of the American Statistical Association*, 91, 993–999.

DAGENAIS, M. G. AND J.-M. DUFOUR (1991): "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica*, 59, 1601–1615.

DUFOUR, J.-M. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.

——(1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.

——(1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

DUFOUR, J.-M. AND J. JASIAK (1993): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," Technical report, C.R.D.E., Université de Montréal.

DUFOUR, J.-M. AND J. F. KIVIET (1996): "Exact Tests for Structural Change in First-Order Dynamic Models," *Journal of Econometrics*, 70, 39–68.

——(1998): "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrica*, 66, 79–104.

DUFOUR, J.-M. AND O. TORRÈS (1998): "Union-Intersection and Sample-Split Methods in Econometrics with Applications to SURE and MA Models," in *Handbook of Applied Economic Statistics*, ed. by D. E. A. Giles and A. Ullah. New York: Marcel Dekker, pp. 465–505.

FULLER, W. A. (1987): *Measurement Error Models*. New York: John Wiley & Sons.

GLESER, L. J. AND J. T. HWANG (1987): "The Nonexistence of $100(1-\alpha)$ Confidence Sets of Finite Expected Diameter in Errors in Variables and Related Models," *The Annals of Statistics*, 15, 1351–1362.

HALL, A. R., G. D. RUDEBUSCH, AND D. W. WILCOX (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–298.

HAYASHI, F. (1982): "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.

KIVIET, J. AND J.-M. DUFOUR (1997): "Exact Tests in Single Equation Autoregressive Distributed Lag Models," *Journal of Econometrics*, 80, 325–353.

MADDALA, G. S. (1974): "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 42, 841–851.

MADDALA, G. S. AND J. JEONG (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

MEEKER, W. Q. AND L. A. ESCOBAR (1995): "Teaching About Approximate Confidence Regions Based on Maximum Likelihood Estimation," *The American Statistician*, 49, 48–53.

MILLER, R. G., JR. (1981): *Simultaneous Statistical Inference (Second Edition)*. New York: Springer-Verlag.

MONTMARQUETTE, C., R. HOULE, M. CRESPO, AND S. MAHSEREDJIAN (1989): *Les interventions scolaires en milieu défavorisé: estimation et évaluation*. Montréal: Les Presses de l'Université de Montréal.

MONTMARQUETTE, C. AND S. MAHSEREDJIAN (1989): "Could Teacher Grading Practices Account for Unexplained Variation in School Achievements?," *Economics of Education Review*, 8, 335–343.

MURPHY, K. M. AND R. H. TOPEL (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370–379.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.

NELSON, C. R. AND R. STARTZ (1990a): "The Distribution of the Instrumental Variable Estimator and its t-ratio When the Instrument is a Poor One," *Journal of Business*, 63, 125–140.

——(1990b): "Some Further Results on the Exact Small Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

OXLEY, L. AND M. MCALEER (1993): "Econometric Issues in Macroeconomic Models with Generated Regressors," *Journal of Economic Surveys*, 7, 1–39.

PAGAN, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

——(1986): "Two Stage and Related Estimators and their Applications," *Review of Economic Studies*, 53, 517–538.

RICHARDSON, D. H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2,* ed. by Z. Griliches and M. D. Intrilligator. Amsterdam: North-Holland, pp. 827–879.

SAWA, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–937.

STAIGER, D. AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.