ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Further results on projection-based inference in IV regressions with weak, collinear or missing instruments

Jean-Marie Dufour[a,*], Mohamed Taamouti[b,c]

[a]*CIRANO, CIREQ, and Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7*

[b]*CIREQ, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7*

[c]*INSEA, Rabat, Morocco*

Abstract

We study a general family of Anderson–Rubin-type procedures, allowing for arbitrary collinearity among the instruments and endogenous variables. Using finite-sample distributional theory, we show that the proposed procedures, besides being robust to weak instruments, are also robust to the exclusion of relevant instruments and to the distribution of endogenous regressors. A solution to the problem of computing linear projections from general possibly singular quadric surfaces is derived and used to build finite-sample confidence sets for individual structural parameters. The importance of robustness to excluded instruments is studied by simulation. Applications to the trade-growth relationship and to education returns are presented.

JEL classification: C3; C13; C12; O4; O1; I2; J2

Keywords: Simultaneous equations; Weak instrument; Collinearity; Missing instrument; Projection

*Corresponding author. Tel.: +1 514 343 2400; fax: +1 514 343 5831.
E-mail addresses: jean.marie.dufour@umontreal.ca (J.-M. Dufour), taamouti@insea.ac.ma (M. Taamouti).
URL: http://www.fas.umontreal.ca/SCECO/Dufour.

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.06.008

1. Introduction

Models where different values of the parameter vector may lead to observationally equivalent data distributions are quite widespread in statistics and econometrics. Further, inference on such models often lead to complex problems, even when "identifying restrictions" are imposed.[1] A context where these difficulties have been extensively explored is the one of simultaneous equations or instrumental variable (IV) regressions when the instruments are poorly correlated with endogenous explanatory variables and, more generally, when structural parameters are close to not being identifiable. The literature on so-called "weak instruments" problems is now considerable.[2]

In view of these difficulties, a basic problem is to develop procedures that are *robust to weak instruments*. Other features we shall also consider here is robustness to the exclusion of possibly relevant instruments (*robustness to missing instruments*), and more generally robustness to the distribution of explanatory endogenous variables (*robustness to endogenous explanatory variable distribution*).[3] We view all these features as important because it is typically difficult to know whether a set of instruments is globally weak or whether relevant instruments have been excluded (which seems highly likely in most practical situations).

In such contexts, it is particularly important that tests and confidence sets be based on properly pivotal (or boundedly pivotal) functions, as well as to study inference procedures from a finite-sample perspective. This suggests that confidence sets should be built by inverting likelihood ratio (LR) and Lagrange multiplier (LM) type statistics, as opposed to the more usual method which consists in inverting Wald-type statistics (such as asymptotic t-ratios); see Dufour (1997).

We focus here on extensions of the procedure originally proposed by Anderson and Rubin (1949, henceforth AR). There are two basic reasons for that. First, it is completely robust to weak instruments. Second, it is one of the few procedures for which a truly finite-sample distributional theory has been supplied under standard parametric assumptions (error Gaussianity, instrument strict exogeneity), which is based on the classical linear model. In view of the non-uniformity of large-sample approximations, we view this feature as the best starting point for the development of procedures that are robust to the presence of weak instruments.

Other potential pivots aimed at being robust to weak instruments have recently been suggested by Wang and Zivot (1998), Kleibergen (2002) and Moreira (2003a). These methods are closer to being full-information methods—in the sense that they rely on a relatively specific formulation of the model for the endogenous explanatory variables—and

[1]For general expositions of the theory of identification in econometrics and statistics, the reader may consult Hsiao (1983) and Bekker et al. (1994).

[2]See, for example, Nelson and Startz (1990a, b), Buse (1992), Hall et al. (1996), Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot et al. (1998), Startz et al. (1999), Chao and Swanson (2000), Stock and Wright (2000), Dufour and Jasiak (2001), Hahn and Hausman (2002a, b), Kleibergen (2002, 2004, 2005), Moreira (2003a, b), Moreira and Poi (2001), Stock and Yogo (2002, 2003), Stock et al. (2002), Perron (2003), Wright (2003, 2002), Bekker and Kleibergen (2003), Hall and Peixe (2003), Forchini and Hillier (2003), Andrews et al. (2004), Dufour and Taamouti (2004), and the reviews of Stock et al. (2002) and Dufour (2003).

[3]We borrow the terminology "robust to weak instruments" from Stock et al. (2002, p. 518). Robustness to instrument exclusion appears to have been little discussed in the literature on weak instruments.

thus may lead to power gains under the assumptions considered. But this will typically be at the expense of robustness.

In this paper, we study a number of issues associated with the use of AR-type procedures and we provide a number of extensions. More precisely, we show *first* that AR-type tests and confidence sets enjoy remarkably strong robustness properties because they allow one to produce valid inference in finite samples despite the presence of weak instruments, missing relevant instruments, and indeed irrespective of the data generating process (DGP) which determines the behavior of the endogenous explanatory variables in the structural equation of interest. In contrast, alternative procedures that exploit more specific models for the latter variables are much more fragile.

Second, we study a theoretical setup broader than the one under which finite-sample validity of AR tests is usually derived, and we propose an extended class of AR-type procedures based on a general class of *auxiliary instruments*. Arbitrary *collinearity* among the instruments and model endogenous variables is allowed, and the auxiliary instruments may not include all the exogenous variables which determine the endogenous explanatory variables. Accounting relations and singular covariance matrices between model disturbances are included as special cases of this setup. The extended AR procedure deals in a transparent way with situations where the exogenous variables and the instruments may be linearly dependent (as can happen easily if the latter contain dummy variables), without reparametrizations that can modify the interpretation of model coefficients. This provides a unified treatment of two basic cases of identification failure: namely, inference in a structural model which may be underidentified as well as regressions with collinear regressors.[4]

Third, we consider the problem of building tests and confidence sets for individual parameters and, more generally, for linear transformations of structural parameters. A central feature of models where parameters may fail to be identified is *parametric non-separability*: in general, individual coefficients may not be empirically meaningful without information on other parameters in the model (which may be viewed as *nuisance parameters*). Reliable informative inference on certain model coefficients may not be feasible, but inference on parameter vectors can often be achieved. This suggests a ‘‘joint’’ approach where we start with inference on vectors of model parameters and then see what can be inferred on individual coefficients.

To produce inference on transformations of model parameters, we consider the *projection* technique described in Dufour (1990, 1997), Wang and Zivot (1998), Dufour and Jasiak (2001) and Dufour and Taamouti (2005). This technique produces exact confidence sets in the sense that the probability of covering the true parameter value is at least as large as the stated level (in accordance with the standard definition of Lehmann 1986, Sections 3.1 and 3.5).[5] Exploiting the fact that AR confidence sets can be represented

[4]Multicollinearity is one of the most basic form of identification failure, which has led to the classical theory of estimable functions. For further discussion, see Magnus and Neudecker (1991, Chapter 13) and Scheffé (1959, Chapters 1–2).

[5]This problem was also considered by Choi and Phillips (1992), Stock and Wright (2000) and Kleibergen (2004). While Choi and Phillips (1992) did not propose an operational method for dealing with the problem, the methods considered by Stock and Wright (2000) and Kleibergen (2004) rely on the assumption that the structural parameters not involved in the restrictions are well identified and rely on large-sample approximations (which become invalid when the identification assumptions made do not hold). Consequently they are not robust to weak instruments. For these reasons, we shall focus here on the projection approach.

by *quadric* surfaces, we showed in Dufour and Taamouti (2004) that projection-based confidence sets for linear transformations of model coefficients can be easily obtained in the *special case* where the quadratic part of the quadric involves a *full-rank matrix* (the *concentration matrix*). Here, we extend this result by giving a completely general closed-form solution to the problem of building projection-based confidence sets for linear combinations of parameters when the joint confidence set belongs to the quadric class. In particular, this solution applies to the generalized AR-type confidence sets introduced above (where the concentration matrix can easily be singular) and leads to confidence sets which are as easy to compute as standard two-stage least squares (2SLS) confidence intervals. The solution of this mathematical problem may also be of independent interest.

Fourth, we show that the confidence sets obtained in this way enjoy another important property, namely *simultaneity* in the sense discussed by Miller (1981) and Dufour (1989). More precisely, projection-based confidence sets (or confidence intervals) can be viewed as Scheffé-type simultaneous confidence sets—which are widely used in analysis of variance—so that the probability that any number of the confidence statements made (for different functions of the parameter vector) hold jointly is controlled.

Fifth, we show that when the projection-based confidence intervals are bounded, they may be interpreted as confidence intervals based on k-class estimators (for a discussion of k-class estimators, see Davidson and MacKinnon, 1993, p. 649) where the "standard error" is corrected in a way that depends on the level of the test.

Sixth, in order to illustrate the projection approach, we present two empirical applications. In the first one, we study the relationship between standards of living and openness in the context of an equation previously considered by Frankel and Romer (1999). The second application deals with the famous problem of measuring returns to education using the model and data considered by Angrist and Krueger (1995) and Bound et al. (1995).

The paper is organized as follows. The problem of robustness to excluded instruments and the endogenous regressor model is discussed in Section 2. We describe the general setup that we consider and the corresponding generalized AR procedures in Section 3. The general closed-form solution to the problem of building projection-based confidence sets from a general quadric confidence set is presented in Section 4. The relation between projection-based confidence sets, Scheffé confidence intervals and k-class estimators is discussed in Section 5. In Section 6, we report the results of our Monte Carlo simulations, while Section 7 presents the empirical applications. We conclude in Section 8.

2. Robustness to missing instruments and endogenous regressor model

Let us consider first the following common simultaneous equation framework, which has been the basis of many recent papers on inference in models with possibly weak instruments (see Dufour, 2003; Stock et al., 2002):

$$y = Y\beta + X_1\gamma + u, \tag{1}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + V, \tag{2}$$

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables ($G \geqslant 1$), X_1 and X_2 are $T \times k_1$ and $T \times k_2$ matrices of exogenous variables, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients, Π_1 and Π_2 are $k_1 \times G$ and $k_2 \times G$ matrices of unknown coefficients, $u = (u_1, \ldots, u_T)'$ is a vector of structural disturbances, and $V = [V_1, \ldots, V_T]'$ is a $T \times G$ matrix of disturbances. Further, in order to allow for a finite-sample distributional theory, we suppose that:

$$X = [X_1, X_2] \text{ is a full-column rank } T \times k \text{ matrix, where } k = k_1 + k_2, \tag{3}$$

$$u \text{ and } X \text{ are independent}, \tag{4}$$

$$u \sim \mathrm{N}[0, \sigma_u^2 I_T]. \tag{5}$$

We consider the problem of building tests and confidence sets on β and γ. Anderson and Rubin (1949) test for the hypothesis $\mathrm{H}_0 : \beta = \beta_0$ in Eq. (1) involves considering the transformed equation

$$y - Y\beta_0 = X_1 \Delta_1 + X_2 \Delta_2 + \varepsilon, \tag{6}$$

where $\Delta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\Delta_2 = \Pi_2(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. H_0 can then be assessed by testing $\mathrm{H}_0' : \Delta_2 = 0$ using the standard F-statistic for H_0' (denoted $\mathrm{AR}(\beta_0)$). Under H_0, we have: $\mathrm{AR}(\beta_0) \sim F(k_2, T - k)$. This distributional result holds irrespective of the rank of the matrix Π_2, which means that tests based on $\mathrm{AR}(\beta_0)$ are *robust to weak instruments*.

In model (1)–(2), the "identifying" instruments X_2 that are excluded from the structural equation (1) may be quite uncertain. In particular, we may wonder what happens if instruments are "left out" of the analysis. A way to look at this problem consists in considering a situation where Y depends on a third set of instruments X_3 which are not used within the inference

$$Y = X_1 \Pi_1 + X_2 \Pi_2 + X_3 \Pi_3 + V, \tag{7}$$

where X_3 is a $T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous). In particular, X_3 may include any variable that could be viewed as independent of the structural disturbance u in (1), and could be unobservable.[6] We view this situation as important because, in practice, it is quite rare that one can consider all the relevant instruments that could be used. In other words, Eq. (2) is replaced by (7), but inference proceeds as if (2) were the actual equation.

Under the generating process (DGP) represented by (1) and (7), the variable $y - Y\beta_0$ used as the dependent variable by the AR procedure satisfies the equation

$$y - Y\beta_0 = X_1 \Delta_1 + X_2 \Delta_2 + X_3 \Delta_3 + \varepsilon, \tag{8}$$

where $\Delta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\Delta_2 = \Pi_2(\beta - \beta_0)$, $\Delta_3 = \Pi_3(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Since $\Delta_2 = 0$ and $\Delta_3 = 0$ under H_0, it is easy to see that the null distribution of $\mathrm{AR}(\beta_0)$ is $F(k_2, T - k)$ (under assumptions (1), (3)–(5) and (7)), even if X_3 is excluded from the regression as in (6). The finite-sample validity of the test based on $\mathrm{AR}(\beta_0)$ is unaffected by the fact that potentially relevant instruments are not taken into account. For this reason, we will say it is robust to missing instruments (or *instrument exclusion*). Furthermore, the distribution of X_3 is irrelevant to the null distribution of $\mathrm{AR}(\beta_0)$, so that X_3 does not have to be strictly exogenous.

---

[6] Clearly, this depends on the interpretation of the structural equation (1) and its parameters, which is itself affected by both explicit and implicit conditionings. These features are, of course, context-specific. Note also that the rows $X_{3i}, i = 1, \ldots, T$, of X_3 may have heterogeneous distributions—in which case the observations Y_i (the rows of Y) would typically also be heterogeneous—and a stable relationship between Y_i and X_{3i} need not exist.

It is also interesting to observe that the distribution of V need not be otherwise restricted; in particular, the vectors $V_1, \ldots, V_T$ may not follow a Gaussian distribution and may be heteroskedastic. Even more generally, we could assume that Y obeys a general non-linear model of the form:

$$Y = g(X_1,\ X_2,\ X_3,\ V,\ \Pi), \tag{9}$$

where $g(\cdot)$ is a possibly unspecified non-linear function, Π is an unknown parameter matrix and V follows an arbitrary distribution. Since, under H_0. both Δ_2 and Δ_3 in the regression (8) must be zero, the null distribution of the AR statistic $AR(\beta_0)$ is still $F(k_2, T-k)$: it is unaffected by the distribution of explanatory endogenous variables. We call this feature *robustness to endogenous explanatory variable distribution*. It is clear that this type of robustness includes robustness to instrument exclusion as a special case.

By contrast, any procedure which exploits the special form of model (2), entailing the exclusion of X_3 from the variables that determine Y, will not typically enjoy the same robustness features. For example, if relevant regressors X_3 are missing, the covariance matrix Σ of V_t typically cannot be consistently estimated, and any method that relies on this possibility will be affected. Clearly, such problems can affect the procedures recently proposed by Wang and Zivot (1998), Kleibergen (2002) and Moreira (2003a). In Section 6, we present simulation evidence which clearly illustrates these difficulties.

3. A generalized AR procedure

The above observations suggest that AR-type procedures may easily be adapted to deal with a much wider array of troublesome situations than the model for which it was originally proposed. Specifically, let us consider again the structural equation (1) where the different symbols are defined as in (1). However, we shall make the following modified assumptions:

$$0 \leqslant \text{rank}(X_1) = v_1 \leqslant k_1, \tag{10}$$

$$\bar{X}_2 \text{ is a } T \times \bar{k}_2 \text{ matrix such that } 0 \leqslant \text{rank}(\bar{X}_2) = v_2 \leqslant \bar{k}_2, \tag{11}$$

$$u \mid \bar{X} \sim N[0, \sigma_u^2(\bar{X}) I_T] \quad \text{where } \bar{X} = [X_1, \bar{X}_2]. \tag{12}$$

Here (10) allows X_1 to have an arbitrary rank (compatible with its dimension), $\bar{X}_2$ is a general "instrument matrix" whose rank may not be full, while (12) states that, conditional on $\bar{X}$, the disturbances in the structural equation (1) are i.i.d. normal. Of course, (10)–(12) cover the more usual assumptions (3)–(5) as a special case. No additional assumption on the DGP of Y will be needed at this stage. In particular, any model of the type (2), (7) or (9) is allowed. Further, the matrix $\bar{X}_2$ may include any subset of columns from X_1, X_2 and X_3, as well as any other instrument (which may be weak). From the power viewpoint, the choice of $\bar{X}_2$ may (and should) be influenced by whatever model we have in mind for Y, but we will see below that it is irrelevant to size control. Note also that no rank assumption is made on Y; in particular, the latter matrix may not have full column rank because the variables in Y satisfy accounting identities.

Let $X_1 = [X_{11}, X_{12}]$, $\gamma = (\gamma_1', \gamma_2')'$, where X_{1i} is a $T \times k_{1i}$ matrix, γ_i is $k_{1i} \times 1$ vector $(i = 1, 2)$, with $k_{11} + k_{12} = k_1$ and $0 \leqslant k_{11} \leqslant k_1$. By convention, we consider that a matrix is simply not present if its number of columns is equal to zero. Consider now the problem of testing an hypothesis of the form: $\mathrm{H}_0(\beta_0,\gamma_{10}):(\beta,\gamma_1)=(\beta_0,\gamma_{10})$, where, by convention, this reduces to $\mathrm{H}_0:\beta=\beta_0$, if $k_{11}=0$. Under the null hypothesis, we have

$$y-Y\beta_0-X_{11}\gamma_{10}=X_{12}\gamma_2+u, \tag{13}$$

where γ_2 is a free parameter. An extension of the AR procedure is then obtained by considering a regression of the form

$$y-Y\beta_0-X_{11}\gamma_{10}=X_{11}\Delta_{11}+X_{12}\Delta_{12}+\bar{X}_2\Delta_2+u=\bar{X}\theta+u, \tag{14}$$

where $\bar{X}\equiv[X_1,\bar{X}_2]=[X_{11},X_{12},\bar{X}_2]$, and then testing the restrictions $\mathrm{H}_0^*(\beta_0,\gamma_{10}):\Delta_{11}=0$ and $\Delta_2=0$, under which (14) becomes equivalent to the null model (13). Again, if $k_{11}=0$, X_{11} simply drops from the left-hand side of (14), and $\mathrm{H}_0^*(\beta_0,\gamma_{10})$ reduces to $\mathrm{H}_0^*(\beta_0):\Delta_2=0$.

A Fisher-type test may still be applied here, provided corrected degrees of freedom are used. The Fisher statistic for testing $\mathrm{H}_0^*(\beta_0,\gamma_{10})$ is then

$$\mathrm{AR}(\beta_0,\gamma_{10};\bar{X}_2)=\frac{u(\beta_0,\gamma_{10})'[M(X_{12})-M(\bar{X})]u(\beta_0,\gamma_{10})/(v-v_2)}{u(\beta_0,\gamma_{10})'M(\bar{X})u(\beta_0,\gamma_{10})/(T-v)}, \tag{15}$$

where $u(\beta_0,\gamma_{10})\equiv y-Y\beta_0-X_{11}\gamma_{10}$, $v_2=\mathrm{rank}(X_{12})$ and $v=\mathrm{rank}(\bar{X})$. For any matrix B, $M(B)=I-P(B)$, $P(B)=B(B'B)^-B'$ is the projection matrix on the space spanned by the columns of B and $(B'B)^-$ is any generalized inverse of $B'B$ ($M(B)$ is invariant to the choice of generalized inverse). Under assumptions (10)–(12) and the null hypothesis $\mathrm{H}_0^*(\beta_0,\gamma_{10})$, all the conditions of the classical linear model are satisfied and we can conclude that: $\mathrm{AR}(\beta_0,\gamma_{10};\bar{X}_2)\sim F(v-v_2,T-v)$; see Dufour (1982) and Scheffé (1959, Sections 2.5–2.6). The only features of the distribution which are affected by rank deficiencies are the degrees of freedom. Note that $v-v_2\leqslant\mathrm{rank}([X_{11},\bar{X}_2])$, where a strict inequality is possible. Further the distribution and the rank of the Y matrix are irrelevant.

Further, a confidence set with level $1-\alpha$ for the vector $(\beta',\gamma_1')'$ can be obtained by inverting the statistic $\mathrm{AR}(\beta_0,\gamma_{10};\bar{X}_2)$:

$$\begin{aligned}C_{(\beta,\gamma_1)}(\alpha)&=\{(\beta_0',\gamma_{10}')':\mathrm{AR}(\beta_0,\gamma_{10};\bar{X}_2)\leqslant F_\alpha(v-v_2,T-v)\}\\&=\{(\beta_0',\gamma_{10}')':(\beta_0',\gamma_{10}')A(\beta_0',\gamma_{10}')'+b'(\beta_0',\gamma_{10}')'+c\leqslant 0\},\end{aligned} \tag{16}$$

where $A=[Y,X_{11}]'H[Y,X_{11}]$, $H=M(X_{12})-\kappa(\alpha,v-v_2,T-v)M(\bar{X})$, $b=-2[Y,X_{11}]'Hy$, $c=y'Hy,\kappa(\alpha,\ n_1,n_2)=1+(n_1/n_2)F_\alpha(n_1,n_2)$, and $F_\alpha(n_1,n_2)$ is the $1-\alpha$ quantile of $F(n_1,n_2)$. We call A the *concentration matrix at level* α (or the α-concentration *matrix*) associated with $(\beta',\gamma_1')'$. The quadratic-linear form in (16) defines a quadric surface (see Shilov, 1961, Chapter 11; Pettofrezzo and Marcoantonio, 1970, Chapters 9–10). In the special case where $(\beta',\gamma_1')'$ reduces to a single parameter (i.e., $G=1$ and $k_{11}=0$), the set $C_{(\beta,\gamma_1)}(\alpha)$ has a closed-form solution involving a quadratic inequality: $C_\beta(\alpha)=\{\beta_0:a\beta_0^2+b\beta_0+c\leqslant 0\}$, where a, b and c are simple functions of the data and thecritical value. The set $C_\beta(\alpha)$ can be viewed as an extension of the quadratic forms described in Dufour and Jasiak (2001) and Zivot et al. (1998).

4. Projection-based confidence sets for scalar linear transformations

We will now consider the problem of building a projection-based confidence set for a scalar linear transformation $g(\theta)=w'\theta$, where w is a non-zero $p\times 1$ vector, from a confidence set defined by a general quadric form:

$$C_\theta = \{\theta_0 : \theta_0' A\theta_0 + b'\theta_0 + c \leqslant 0\}, \tag{17}$$

where A is a symmetric $p \times p$ matrix (possibly singular), b is a $p \times 1$ vector, and c is a real scalar. By definition, the associated projection-based confidence set for $w'\theta$ is

$$C_{w'\theta} \equiv g[C_\theta] = \{\delta_0 : \delta_0 = w'\theta_0 \text{ where } \theta_0' A\theta_0 + b'\theta_0 + c \leqslant 0\}. \tag{18}$$

Since $w \neq 0$, we can assume without loss of generality that the first component of w (denoted w_1) is different from zero. It will be convenient to consider a non-singular transformation of θ

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} w'\theta \\ R_2\theta \end{bmatrix} = R\theta, \quad R = \begin{bmatrix} w' \\ R_2 \end{bmatrix} = \begin{pmatrix} w_1 & w_2' \\ 0 & I_{p-1} \end{pmatrix}, \tag{19}$$

where $w' = [w_1, w_2']$ and $R_2 = [0, I_{p-1}]$ is a $(p-1) \times p$ matrix. If $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$, it is clear from this notation that $\delta_2 = (\theta_2, \ldots, \theta_p)'$. We study the problem of building a confidence set for δ_1.

The quadric form which defines C_θ in (17) may be written

$$\theta' A\theta + b'\theta + c = \delta'\bar{A}\delta + \bar{b}'\delta + c \equiv \bar{Q}(\delta), \tag{20}$$

where $\bar{A} = (R^{-1})' A R^{-1}$, $\bar{b} = (R^{-1})' b$, so that

$$C_{w'\theta} = C_{\delta_1} = \{\delta_1 : \delta = (\delta_1, \delta_2')' \text{ satisfies } \bar{Q}(\delta) \leqslant 0\}. \tag{21}$$

On partitioning A, $\bar{A}$ and $\bar{b}$ conformably with $\delta = (\delta_1, \delta_2')'$, we have

$$A = \begin{pmatrix} a_{11} & A_{21}' \\ A_{21} & A_{22} \end{pmatrix}, \quad \bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{A}_{21}' \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \end{pmatrix}, \tag{22}$$

where $\bar{A}_{22}$ has dimension $(p-1) \times (p-1)$ and, by convention, we set $\bar{A} = [\bar{a}_{11}]$ and $b = [\bar{b}_1]$ when $p = 1$. It is easy to see that: $\bar{a}_{11} = a_{11}/w_1^2$, $\bar{A}_{21} = [w_1 A_{21} - a_{11} w_2]/w_1^2$,

$$\bar{A}_{22} = \frac{1}{w_1^2}[a_{11} w_2 w_2' - w_1 A_{21} w_2' - w_1 w_2 A_{21}' + w_1^2 A_{22}], \quad \bar{b} = \frac{1}{w_1} \begin{pmatrix} b_1 \\ -b_1 w_2 + w_1 b_2 \end{pmatrix}.$$

We can then write

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \delta_2'\bar{A}_{22}\delta_2 + [2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_2, \tag{23}$$

where, by convention, the two last terms of (23) simply disappear when $p = 1$. For $p \geqslant 1$, let $r_2 = \operatorname{rank}(\bar{A}_{22})$, where $0 \leqslant r_2 \leqslant p-1$, and consider the spectral decomposition

$$\bar{A}_{22} = P_2 D_2 P_2', \quad D_2 = \operatorname{diag}(d_1, \ldots, d_{p-1}), \tag{24}$$

where $d_1, \ldots, d_{p-1}$ are the eigenvalues of $\bar{A}_{22}$ and P_2 is an orthogonal matrix. Without loss of generality, we can assume that

$$\begin{aligned} d_i &\neq 0 \quad \text{if } 1 \leqslant i \leqslant r_2, \\ &= 0 \quad \text{if } i > r_2. \end{aligned} \tag{25}$$

Let us also define (whenever the objects considered exist)

$$\tilde{\delta}_2 = P_2'\delta_2, \quad \tilde{A}_{21} = P_2'\bar{A}_{21}, \quad \tilde{b}_2 = P_2'\bar{b}_2, \ D_{2*} = \operatorname{diag}(d_1, \ldots, d_{r_2}), \tag{26}$$

and denote by $\tilde{\delta}_{2*}$, $\tilde{A}_{21*}$ and $\tilde{b}_{2*}$ the vectors obtained by taking the first r_2 components of $\tilde{\delta}_2$, $\tilde{A}_{21}$ and $\tilde{b}_2$ respectively

$$\tilde{\delta}_{2*} = P'_{21}\delta_2, \quad \tilde{A}_{21*} = P'_{21}\bar{A}_{21}, \quad \tilde{b}_{2*} = P'_{21}\bar{b}_2, \quad P_2 = [P_{21}, P_{22}], \tag{27}$$

where P_{21} and P_{22} have dimensions $(p-1)\times r_2$ and $(p-1)\times(p-1-r_2)$, respectively. The form of the set $C_{w'\theta} = C_{\delta_1}$ is given by the following theorem:

Theorem 4.1 (*Projection-based confidence sets with a possibly singular concentration matrix*). *Under assumptions and notations* (19)–(27), *the set* $C_{w'\theta}$ *takes one of the three following forms*:

(a) *if* $p>1$ *and* $\bar{A}_{22}$ *is positive semidefinite with* $\bar{A}_{22}\neq 0$, *then*

$$C_{w'\theta} = \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0\} \cup S_1, \tag{28}$$

where $\tilde{a}_1 = \bar{a}_{11} - \bar{A}'_{21}\bar{A}^{+}_{22}\bar{A}_{21}$, $\tilde{b}_1 = \bar{b}_1 - \bar{A}'_{21}\bar{A}^{+}_{22}\bar{b}_2$, $\tilde{c}_1 = c - \frac{1}{4}\bar{b}'_2\bar{A}^{+}_{22}\bar{b}_2$, $\bar{A}^{+}_{22}$ *is the Moore–Penrose inverse of* $\bar{A}_{22}$, *and*

$$S_1 = \begin{cases} \emptyset & \text{if } \mathrm{rank}(\bar{A}_{22}) = p-1, \\ \{\delta_1 : P'_{22}(2\bar{A}_{21}\delta_1 + \bar{b}_2)\neq 0\} & \text{if } 1\leqslant \mathrm{rank}(\bar{A}_{22}) < p-1; \end{cases}$$

(b) *if* $p=1$ *or* $\bar{A}_{22}=0$, *then*

$$C_{w'\theta} = \{\delta_1 : \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c \leqslant 0\} \cup S_2, \tag{29}$$

where

$$S_2 = \begin{cases} \emptyset & \text{if } p=1, \\ \{\delta_1 : 2\bar{A}_{21}\delta_1 + \bar{b}_2 \neq 0\} & \text{if } p>1 \text{ and } \bar{A}_{22}=0; \end{cases}$$

(c) *if* $p>1$ *and* $\bar{A}_{22}$ *is not positive semidefinite*, *then* $C_{w'\theta} = \mathbb{R}$.

The proof of this theorem is given in Appendix. In all the cases covered by the latter theorem the joint confidence set C_θ is unbounded if A is singular. However, we can see from Theorem 4.1 that confidence intervals for some parameters (or linear transformations of θ) can be bounded. This depends on the values of the coefficients of the second-order polynomials in (28) and (29). Specifically, it is easy to see that the quadratic set $\tilde{C}_{w'\theta} = \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0\}$ in (28) can take several basic forms; for convenience, the latter are summarized in Table 1. Of course, a similar result holds for the quadratic set in (29).

The results in this paper generalize those provided in Dufour and Taamouti (2004) by allowing A to have an arbitrary rank. In (16), A is almost surely singular when X_{11} does not have full column rank or when identities hold between the variables in Y. Other cases are, of course, possible. When A is positive definite, the confidence interval in (28) reduces to

$$C_{w'\theta} = \left[w'\tilde{\theta} - \sqrt{d(w'A^{-1}w)},\, w'\tilde{\theta} + \sqrt{d(w'A^{-1}w)} \right], \tag{30}$$

where $\tilde{\theta} = -\frac{1}{2}A^{-1}b$, and $d = \frac{1}{4}b'A^{-1}b - c \geqslant 0$ (if $d<0$, $C_{w'\theta}$ is empty).

Table 1
Alternative forms of confidence set $\tilde{C}_{w'\theta} = \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0\}$, $\tilde{\Delta}_1 \equiv \tilde{b}_1^2 - 4\tilde{a}_1\tilde{c}_1$

$$\tilde{C}_{w'\theta} = \begin{cases} \left[\dfrac{-\tilde{b}_1 - \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}, \dfrac{-\tilde{b}_1 + \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}\right] & \text{if } \tilde{a}_1 > 0 \text{ and } \tilde{\Delta}_1 \geqslant 0, \\ \left]-\infty, \frac{-\tilde{b}_1 + \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}\right] \cup \left[\frac{-\tilde{b}_1 - \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}, \infty\right[& \text{if } \tilde{a}_1 < 0 \text{ and } \tilde{\Delta}_1 \geqslant 0, \\ ]-\infty, -\tilde{c}_1/\tilde{b}_1] & \text{if } \tilde{a}_1 = 0 \text{ and } \tilde{b}_1 > 0, \\ [-\tilde{c}_1/\tilde{b}_1, \infty[& \text{if } \tilde{a}_1 = 0 \text{ and } \tilde{b}_1 < 0, \\ \mathbb{R} & \text{if } (\tilde{a}_1 < 0 \text{ and } \tilde{\Delta}_1 < 0) \\ & \quad \text{or } (\tilde{a}_1 = \tilde{b}_1 = 0 \text{ and } \tilde{c}_1 \leqslant 0), \\ \emptyset & \text{if } (\tilde{a}_1 > 0 \text{ and } \tilde{\Delta}_1 < 0) \\ & \quad \text{or } (\tilde{a}_1 = \tilde{b}_1 = 0 \text{ and } \tilde{c}_1 > 0). \end{cases}$$

5. Scheffé confidence intervals, k-class estimators, and projections

It is interesting to note the relationship of the above results with Scheffé-type confidence sets in the context of model (1)–(2). The confidence set for β is based on the F-test of $H_0 : \Delta_2 = \Pi_2(\beta - \beta_0) = 0$ in the regression equation

$$y - Y\beta_0 = X_1\Delta_1 + X_2\Delta_2 + \varepsilon.$$

Following Scheffé (1959), this F-test is equivalent to the test which does not reject H_0 when all hypotheses of the form $H_0(a) : a'\Delta_2 = 0$ are not rejected by the criterion $|t(a)| > S(\alpha)$, for all $k_2 \times 1$ non-zero vectors a, where $t(a)$ is the t-statistic for $H_0(a)$ and $S(\alpha) = \sqrt{k_2\, F_\alpha(k_2, T - k)}$; see also Savin (1984). Since $a'\Delta_2 = w'(\beta - \beta_0)$ where $w = \Pi_2' a$, this entails that no hypothesis of the form $H_0'(w) : w'\beta = w_0'\beta$, is rejected. The projection-based confidence set for $w'\beta$ can be viewed as a Scheffé-type simultaneous confidence interval for $w'\beta$.

When the eigenvalues of the matrix A are positive and the projection-based confidence set for $w'\beta$ is bounded, it is interesting to note that the form of this confidence set (see (30)) is similar to the standard form: $[\hat{\beta} - \hat{\sigma}z(\alpha), \hat{\beta} + \hat{\sigma}z(\alpha)]$. Since $\beta = w'\beta$, the corresponding estimator of β is $\tilde{\beta} = -(1/2)A^{-1}b$. The estimated variance of the estimator should be a scalar (say $\hat{\sigma}^2$) times the matrix A^{-1}, $\hat{\sigma}^2 A^{-1}$, and since the confidence interval has level greater than or equal to $1 - \alpha$, $\sqrt{d}/\hat{\sigma}$ should correspond to a quantile of an order greater than or equal to $1 - \alpha$ of the statistic $|(w'\tilde{\beta} - w'\beta)/[\hat{\sigma}^2(w'A^{-1}w)]^{1/2}|$. Replacing A and b by their expressions, the estimator $\tilde{\beta}$ may be written: $\tilde{\beta} = (Y'HY)^{-1}Y'Hy$, which can be interpreted as the IV estimator obtained by taking HY as the instrument for Y. If $\operatorname{rank}(\Pi_2) = G$ and the following usual assumptions hold,

$$\left(\frac{X'X}{T}, \frac{X'u}{T}, \frac{X'V}{T}\right) \xrightarrow[T\to\infty]{\mathrm{p}} (Q_{XX}, 0, 0), \qquad \frac{X'u}{\sqrt{T}} \xrightarrow[T\to\infty]{L} \mathrm{N}[0, \sigma_u^2 Q_{XX}], \tag{31}$$

we see easily that $\tilde{\beta}$ is asymptotically uncorrelated with the disturbances u and

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow[T\to\infty]{L} \mathrm{N}\left[0, \sigma_u^2 \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}A\right)^{-1}\right] \tag{32}$$

where

$$\underset{T\to\infty}{\text{plim}}\ \frac{1}{T} A = \Pi_2'[Q_{X_2X_2} - Q_{X_2X_1}Q_{X_1X_1}^{-1}Q_{X_2X_1}']\Pi_2$$

and

$$Q_{X_iX_j} = \underset{T\to\infty}{\text{plim}}\ \frac{1}{T} X_i'X_j.$$

On developing the expression of $\tilde{\beta}$, we may also write

$$\tilde{\beta} = \{Y'[M(X_1) - (1+f_\alpha)M(X)]Y\}^{-1}Y'[M(X_1) - (1+f_\alpha)M(X)]y. \tag{33}$$

This is the expression of the well-known Theil's k-class estimator with $k = 1 + f_\alpha$, and since f_α tends to 0 when T becomes large, $\tilde{\beta}$ is asymptotically equivalent to the 2SLS estimator of β (see Davidson and MacKinnon, 1993, p. 649). Hence, when Π_2 is of full rank and the eigenvalues of A are positive, the projection-based confidence set for $w'\beta$ may be interpreted as a Wald-type confidence interval based on the statistic (which is asymptotically pivotal)

$$\tilde{t}(w'\beta) = (w'\tilde{\beta} - w'\beta)/\sqrt{\hat{\sigma}_u^2(w'A^{-1}w)}\ . \tag{34}$$

6. Simulation study: the effect of instrument exclusion

In this section, we present a small study on the finite sample behavior of different tests aimed at being robust to weak instruments when some of the relevant instruments are omitted.[7] The tests considered are the exact AR test (based on (15) with $k_{11} = 0$), the asymptotic version of this test based on the $\chi^2(k_2)/k_2$ distribution (ARS), the LR and LM tests proposed by Wang and Zivot (1998), Kleibergen's (2002) K-test, and the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003a). The DGP is

$$y = Y_1\beta_1 + Y_2\beta_2 + u, \quad (Y_1, Y_2) = X_2\Pi_2 + X_3\delta + (V_1, V_2), \tag{35}$$

$$(u_t, V_{1t}, V_{2t})' \overset{\text{i.i.d.}}{\sim} \text{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .3 \\ .8 & .3 & 1 \end{pmatrix}, \tag{36}$$

where X_2 is a $T \times k_2$ matrix of included instruments and X_3 is a $T \times 1$ omitted instrument vector which is not taken into account when computing the different statistics. We took $X_3 = M(X_2)\tilde{X}_3$, where the elements of X_2 and $\tilde{X}_3$ were generated as i.i.d. N(0, 1) variables, so that X_3 is orthogonal to X_2. Both X_2 and X_3 are kept fixed over the simulation experiment.[8] The parameters values are set at $\beta_1 = \frac{1}{2}$, $\beta_2 = 1, \delta = \lambda(1, 1)'$ and λ takes the values 0, 1 or 10. The correlation coefficient r between u and V_i $(i = 1, 2)$ is set equal to 0.8,

[7]The discussion paper version of this article (Dufour and Taamouti, 2005) also contains a study of the performance of projection-based confidence sets in the setup considered, in particular how conservative such sets are. The results indicate that, besides being the only provably valid confidence sets for individual coefficients (when $G>1$), the projection-based confidence sets are not overly conservative and are sufficiently precise to be useful in practice.

[8]In Dufour and Taamouti (2005), we also study the case where the missing instruments are regenerated at each replication. The results are qualitatively the same.

Table 2
Instrument exclusion and the size of tests robust to weak instruments

k_2	AR	ARS	K	LM	LR	LR1	LR2	AR	ARS	K	LM	LR	LR1	LR2
	(a) $\delta = 0$ and $\rho = 0.01$							(b) $\delta = 0$ and $\rho = 1$						
2	5.4	6.2	6.2	5.4	5.9	5.9	6.2	4.8	5.0	5.0	4.6	5.0	5.0	5.0
3	4.4	4.8	5.0	3.9	3.9	5.1	5.1	5.0	6.1	6.2	2.0	2.9	6.3	6.3
4	5.1	6.0	6.6	4.5	4.2	6.0	6.1	5.4	6.0	4.9	0.6	0.8	5.1	5.4
5	3.2	3.6	4.7	2.9	1.8	3.7	3.7	5.0	5.7	5.6	0.7	0.8	5.4	5.7
10	4.9	6.5	7.8	3.9	1.7	6.3	6.9	6.6	7.7	5.5	0.0	0.0	4.7	5.7
20	3.9	7.6	7.6	2.1	0.4	7.7	8.0	4.9	8.7	5.3	0.0	0.0	5.4	5.7
40	5.6	11.8	17.7	1.0	0.4	15.9	15.1	4.5	10.5	7.7	0.0	0.0	7.1	8.2
	(c) $\delta = 1$ and $\rho = 0.01$							(d) $\delta = 1$ and $\rho = 1$						
2	5.0	5.4	5.4	4.8	5.4	5.4	5.4	5.4	5.8	5.8	5.4	5.7	5.7	5.8
3	5.7	6.3	8.0	5.4	6.3	6.4	7.0	4.7	5.3	5.0	1.9	2.3	4.7	4.9
4	6.2	7.3	11.6	5.7	7.1	7.2	7.4	5.5	6.5	4.9	0.8	1.3	4.8	5.0
5	5.0	5.8	14.5	3.8	5.7	6.0	6.1	5.1	6.0	4.4	0.1	0.3	4.3	4.3
10	5.1	6.1	36.5	4.1	6.3	6.6	6.1	6.0	8.4	6.3	0.0	0.0	6.5	6.9
20	3.7	7.2	57.6	2.1	9.8	10.7	7.5	5.1	8.3	6.3	0.0	0.0	6.3	6.8
40	5.9	13.3	80.2	1.0	31.8	35.5	14.4	4.9	10.8	11.2	0.0	0.0	12.0	12.6
	(e) $\delta = 10$ and $\rho = 0.01$							(f) $\delta = 10$ and $\rho = 1$						
2	5.2	5.6	5.6	5.2	5.6	5.6	5.6	4.4	4.9	4.9	4.1	4.8	4.8	4.9
3	3.8	4.3	10.0	3.7	4.2	4.4	4.5	4.8	5.5	4.9	2.3	4.6	5.2	5.4
4	4.8	5.5	17.2	4.1	5.1	5.8	5.9	5.4	6.2	6.6	1.0	5.4	6.5	6.6
5	6.2	6.8	28.7	5.3	6.8	7.2	7.4	5.2	6.1	7.0	0.4	5.5	6.3	6.4
10	5.2	7.6	72.4	4.2	7.9	8.4	7.7	3.6	5.1	11.5	0.0	4.4	5.5	5.3
20	6.8	10.1	95.1	3.6	13.1	14.0	10.1	5.4	8.2	42.9	0.0	10.5	12.9	9.2
40	6.0	15.7	97.7	1.2	38.7	41.9	16.7	5.8	13.2	69.6	0.0	33.5	36.9	14.5

Nominal size $= 0.05$. Results are given in percentages.

the variables Y_1 and Y_2 are endogenous and the IVs X_2 are necessary. The matrix Π_2 is such that $\Pi_2 = \rho\Pi/\sqrt{T}$, where ρ takes the values 0.01 or 1, and Π is obtained from the identity matrix by keeping the first k_2 lines and the first G columns. The number of instruments k_2 varies from 2 to 40. The sample size is $T = 100$. The number of replications is $N = 1000$ and the conditional LR critical values are computed using the same number of replications.

For each statistic, we computed the empirical rejection probability of the null hypothesis $H_0 : \beta = \beta_0$ when β_0 is the true value of the parameter. The nominal level of the tests is 5%. Six basic cases are considered. In cases (a) and (b), we have $\delta = 0$, which means that there is no omitted instrument: this is a benchmark for comparison with other cases. In cases (c) and (d), we have $\delta = 1$, which means that there is an omitted instrument. In cases (e) and (f), we have $\delta = 10$, which means that the omitted instrument is a very relevant one. For each value of δ, we consider a design with weak identification ($\rho = 0.01$) and a design where identification is strong ($\rho = 1$). The results are presented in Table 2.

The main observation from these results is that the sizes of the tests K, LR1 and LR2 can be seriously affected by the omission of a relevant instrument, with empirical rejection frequencies as high as 97% (rather than 5%). The more relevant the omitted instrument is, the larger the distortion. The conditional LR (LR1 and LR2) tests are clearly more robust than the K test, but sizeable size distortions are also observable. The distortion persists even if the included instruments are relevant. On the other hand, the AR and ARS tests are completely robust to instrument exclusion (as expected from the theory). The slight distortion in ARS size is due to the fact that the chi-square critical value is used rather than the Fisher critical value.

7. Empirical illustrations

In this section we illustrate the statistical inference methods discussed in the previous sections through two empirical applications related to important issues in the macroeconomic and labor economics literature. The first one concerns the relation between growth and trade examined through cross-country data on a large sample of countries, while the second one considers the widely studied problem of returns to education.

7.1. Trade and growth

A large number of cross-country studies in the macroeconomics literature have looked at the relationship between standards of living and openness. The recent literature includes Irwin and Tervio (2002), Frankel and Romer (1996, 1999), Harrison (1996), Mankiw et al. (1992) and the survey of Rodrik (1995). Despite the great effort that has been devoted to studying this issue, there is little persuasive evidence concerning the effect of openness on income even if many studies conclude that openness has been conductive to higher growth.

Estimating the impact of openness on income through a cross-country regression raises two basic difficulties. The first one consists in finding an appropriate indicator of openness. The most commonly used one is the trade share (ratio of imports or exports to GDP). The second problem is the endogeneity of this indicator. Frankel and Romer (1999) argue that the trade share should be viewed as an endogenous variable, and similarly for the other indicators such as trade policies.

As a solution to this problem, Frankel and Romer (1999) proposed to use IV methods to estimate the income-trade relationship. The equation studied is given by

$$y_i = a + bT_i + c_1 N_i + c_2 A_i + u_i, \tag{37}$$

where y_i is log income per person in country i, T_i the trade share (measured as the ratio of imports and exports to GDP), N_i the logarithm of population, and A_i the logarithm of country area. The trade share T_i can be viewed as endogenous, and to take this into account, the authors used an instrument constructed on the basis of geographic characteristics (see Frankel and Romer, 1999, Eq. (6), p. 383).

The data used include for each country the trade share in 1985, the area and population (1985), and per capita income (1985).[9] The authors focus on two samples. The first is the full 150 countries covered by the Penn World Table, and the second sample is the 98-country sample considered by Mankiw et al. (1992). In this paper, we consider the sample of 150 countries. For this sample, it is not clear how “weak” the instruments

[9]The data set and its sources are given in the appendix of Frankel and Romer (1999).

are. The F-statistic of the first stage regression

$$T_i = \alpha + \beta Z_i + \gamma_1 N_i + \gamma_2 A_i + \varepsilon_i \tag{38}$$

is about 13; see Frankel and Romer (1999, Table 2, p. 385).

To draw inference on the coefficients of structural equation (37), we can use the AR method in two ways. First if we are interested only in the coefficient of trade share, we can invert the AR test for $H_0 : b = b_0$ to obtain a quadratic confidence set for b. On the other hand, if we wish to build confidence sets for the other parameters of (37), we must first use the AR test to obtain a joint confidence set for b and each one of the other parameters and then use the projection approach to obtain confidence sets for each one of these parameters.[10] As assumed in the literature, the observations are considered to be homoskedastic and uncorrelated but not necessarily normal, we use the asymptotic AR test with a χ^2 distribution. The results are as follows:

The 95% quadratic confidence set for the coefficient of trade share b is given by

$$C_b(\alpha) = \{b : 0.963b^2 - 4.754b + 1.274 \leqslant 0\} = [0.284, 4.652]. \tag{39}$$

The p-value of the AR test for $H_0 : b = 0$ is 0.0244, this means a significant positive impact of trade on income at the usual 5% level. The IV estimation of this coefficient is 1.97 with a standard error of 0.99, yielding the confidence interval $[\hat{b}_{\mathrm{IV}} - 2\hat{\sigma}_{\hat{b}_{\mathrm{IV}}}, \hat{b}_{\mathrm{IV}} + 2\hat{\sigma}_{\hat{b}_{\mathrm{IV}}}] = [-0.01, 3.95]$, which is not very different from the AR-based confidence set. In particular, in contrast with $C_b(\alpha)$ in (39), it does not exclude zero and may suggest that b is not significantly different from zero.

The joint confidence sets obtained by applying the method developed in this paper to each pair obtained by putting the trade share coefficient and each one of the other coefficients in (37) are given in Table 3A. All the confidence sets are bounded, a natural outcome since we do not have a serious problem of identification in this model. From these confidence sets we can obtain projection-based confidence intervals for each one of the parameters; see Table 3B. Even if zero is covered by the confidence intervals for the openness coefficient, the intervals almost entirely consist of positive values. AR-projection-based confidence sets are conservative so when the level of the joint confidence set is 95% it is likely that the level of the projection is close to 98% (see the simulations in Dufour and Taamouti, 2004), but if we compare them to those obtained from t-statistics, they are not notably larger.

7.2. Education and earnings

The second application considers the well known problem of estimating returns to education. Since the work of Angrist and Krueger (1991), a lot of research has been done on this problem; see, for example, Angrist and Krueger (1995), Angrist et al. (1999), and Bound et al. (1995). The central equation in this work is a relationship where the log weekly earning is explained by the number of years of education and several other covariates (age, age squared, year of birth, region,...). Education can be viewed as an endogenous variable, so Angrist and Krueger (1991) proposed to use the birth quarter as an instrument, since individuals born in the first quarter of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near

[10]We cannot use the AR test to build directly confidence sets for the coefficients of the exogenous variables.

Table 3
Confidence sets for the coefficients of the Frankel–Romer income-trade equation
A. Bivariate joint confidence sets (*confidence level* = 95%)

θ	Joint confidence set (95%)
(b, c_1)	$\theta'\begin{pmatrix} 1.78 & -16.36 \\ -16.36 & 257.85 \end{pmatrix}\theta + \begin{pmatrix} -2.23, & -34.50 \end{pmatrix}\theta + 0.19 \leqslant 0$
(b, c_2)	$\theta'\begin{pmatrix} 3.83 & -34.58 \\ -34.58 & 386.87 \end{pmatrix}\theta + \begin{pmatrix} -10.6, & 69.17 \end{pmatrix}\theta + 2.13 \leqslant 0$
(b, a)	$\theta'\begin{pmatrix} 38.41 & 33.34 \\ 33.35 & 29.52 \end{pmatrix}\theta + \begin{pmatrix} -611.55, & -537.47 \end{pmatrix}\theta + 2445.58 \leqslant 0$

B. Projection-based individual confidence intervals (*confidence level* ⩾95%)

Coefficient	Projection-based confidence sets	IV-based Wald-type confidence sets
Openness	$[-0.21, 6.18]$	$[-0.01, 3.95]$
Population	$[-0.01, 0.52]$	$[-0.01, 0.37]$
Area	$[-0.14, 0.49]$	$[-0.11, 0.29]$
Constant	$[2.09, 9.38]$	$[0.56, 9.36]$

the end of the year. Consequently, individuals born at the beginning of the year are likely to earn less than those born during the rest of the year. Other versions of this IV regression take as instruments interactions between the birth quarter and regional and/or birth year dummies.

It is well documented that the instrument set used by Angrist and Krueger (1991) is weak and explains very little of the variation in education; see Bound et al. (1995). Consequently, standard IV-based inference is quite unreliable. We shall now apply the methods developed in this paper to this relationship. The model considered is the following:

$$y = \beta_0 + \beta_1 E + \sum_{i=1}^{k_1} \gamma_i X_i + u, \quad E = \pi_0 + \sum_{i=1}^{k_2} \pi_i Z_i + \sum_{i=1}^{k_1} \phi_i X_i + v,$$

where y is log-weekly earnings, E is the number of years of education (possibly endogenous), X contains the exogenous covariates (age, age squared, marital status, race, standard metropolitan statistical area (SMSA), 9 dummies for years of birth, and 8 dummies for division of birth). Z contains 30 dummies obtained by interacting the quarter of birth with the year of birth. β_1 measures the return to education. The data set consists of the 5% public-use sample of the 1980 US census for men born between 1930 and 1939. The sample size is 329 509 observations.

Since the instruments are likely to be weak, it appears important to use a method which is robust to weak instruments. We consider here the AR procedure. If we were only interested in the coefficient of education, we could compute the quadratic confidence set for β_1. But if we wish to evaluate the other coefficients, for example the age coefficient (say, γ_1), the only way to get a confidence interval is to compute the AR joint confidence set for (β_1, γ_1) and then deduce by projection a confidence set for γ_1. Since the instruments are

Table 4
Projection-based confidence sets for the coefficients of the exogenous covariates in the income-education equation (size = 95%)

Covariate	CS for education	CS for covariate	Wald CS covariate
Constant	[−0.86076934, 0.77468002]	[−4.4353178, 16.836347]	[4.121, 5.600]
Age	[−0.86076841, 0.77467914]	[−0.12099477, 0.06963698]	[−0.031, 0.002]
Age squared	[−.86076865, 0.77467917]	[−0.00772368, 0.00748569]	[−0.001, 0.002]
Marital status	$\mathbb{R}$	$\mathbb{R}$	[0.234, 0.263]
SMSA	$\mathbb{R}$	$\mathbb{R}$	[0.120, 0.240]
Race	$\mathbb{R}$	$\mathbb{R}$	[−0.352, −0.173]
Year 1	[−0.86076899, 0.77467898]	[−0.72434684, 1.1399276]	[−0.002, 0.187]
Year 2	[−0.86076919, 0.7746792]	[−0.64290291, 1.0246588]	[0.003, 0.172]
Year 3	[−0.86076854, 0.77467918]	[−0.51469586, 0.84369807]	[0.008, 0.154]
Year 4	[−.86076758, 0.77467916]	[−0.4042831, 0.69265631]	[0.013, 0.141]
Year 5	[−0.86076725, 0.77467906]	[−0.28675828, 0.52165559]	[0.015, 0.123]
Year 6	[−0.8607684, 0.77467903]	[−0.2206811, 0.39879656]	[0.007, 0.0980]
Year 7	$\mathbb{R}$	$\mathbb{R}$	[0.008, 0.080]
Year 8	[−0.86768146, 0.78338792]	[−0.08312128, 0.17409244]	[0.005, 0.0581]
Year 9	[−0.86076735, 0.77467921]	[−0.04610583, 0.1050552]	[0.005, 0.038]
Division 1	$\mathbb{R}$	$\mathbb{R}$	[−0.150, −0.081]
Division 2	$\mathbb{R}$	$\mathbb{R}$	[−0.094, −0.015]
Division 3	$\mathbb{R}$	$\mathbb{R}$	[−0.048, 0.073]
Division 4	$\mathbb{R}$	$\mathbb{R}$	[−0.153, −0.067]
Division 5	$\mathbb{R}$	$\mathbb{R}$	[−0.205, −0.080]
Division 6	$\mathbb{R}$	$\mathbb{R}$	[−0.265, −0.074]
Division 7	$\mathbb{R}$	$\mathbb{R}$	[−0.161, −0.051]
Division 8	$\mathbb{R}$	$\mathbb{R}$	[−0.111, −0.075]

weak, we expect large, if not completely uninformative, intervals. Table 4 gives projection-based confidence sets for the coefficients of education and different covariates. For each covariate X_i, we computed the AR joint confidence set with education (a confidence set for (β_1, γ_i)) and then project to obtain a confidence set for β_1 (column 2) and a confidence set for γ_i (column 3). The last column gives Wald-based confidence sets for each covariate obtained by 2SLS estimation of the education equation. As expected many of the valid confidence sets are unbounded while Wald-type confidence sets are always bounded but unreliable.

For the coefficient β_1 measuring returns to education, the *AR*-based quadratic confidence interval of confidence level 95% is given by $AR\_IC_\alpha(\beta_1) = [-0.86, 0.77]$. It is bounded but too large to provide relevant information on the magnitude of returns to education. The 2SLS estimate for β_1 is 0.06 with a standard error of 0.023 yielding the Wald-type confidence interval $W\_IC_\alpha(\beta_1) = [0.0031, 0.1167]$.

8. Conclusion

In this paper, we have provided extensions of AR-type procedures based on a general class of *auxiliary instruments*, for which we supplied a finite-sample distributional theory. The new procedures allow for arbitrary *collinearity* among the instruments and model endogenous variables, including the presence of accounting relations and singular disturbance covariance matrices. For inference on parameter transformations, we used the projection approach to obtain finite-sample tests and closed-form confidence sets. The confidence sets so obtained have the additional feature of being simultaneous in the sense of Scheffé and when they take the form of a closed interval, they can be interpreted as Wald-type confidence intervals based on k-class estimators.

We also stressed that AR-type procedures enjoy remarkable invariance (or robustness) properties. The finite-sample distribution of AR-type test statistics is completely unaffected by the presence of "weak instruments", the exclusion of relevant instruments, and the distribution of the explanatory endogenous variables. These features can be quite important and useful from a practical viewpoint. The robustness of AR-type procedures and the non-robustness of alternative procedures aimed at being more robust to weak instruments was also documented in a simulation experiment. In several cases, the difference in reliability is spectacular.

Of course, the class of AR-type tests, especially in the generalized form introduced in this paper, is quite large. This raises the problem of selecting instruments. Further, one must be aware that power may decline as the number of instruments increases, especially if they have little relevance, which suggests that the number of instruments should be kept as small as possible. Because AR statistics are robust to the exclusion of instruments, this can be done relatively easily. We discuss the problem of selecting optimal instruments and reducing the number of instruments in two companion papers (Dufour and Taamouti, 2001a, b). For other results relevant to the instrument selection, the reader may consult Cragg and Donald (1993), Hall et al. (1996), Shea (1997), Chao and Swanson (2000), Donald and Newey (2001), Hall and Peixe (2003), Hahn and Hausman (2002a, b), and Stock and Yogo (2002).

Finally, we think that the analytical results presented here on quadric confidence sets can be useful in other contexts involving, for example, errors-in-variables models (see Dufour and Jasiak, 2001), non-linear models, and dynamic models. Such extensions would go beyond the scope of the present paper. We study such extensions in another companion paper (Dufour and Taamouti, 2001b).

Acknowledgments

The authors thank David Jaeger for providing his data on returns to education, as well as Laurence Broze, John Cragg, Jean-Pierre Florens, Christian Gouriéroux, Joanna Jasiak, Frédéric Jouneau, Lynda Khalaf, Nour Meddahi, Benoît Perron, Tim Vogelsang, Eric Zivot, two anonymous referees, and the Editor Geert Dhaene for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec). One of the authors (Taamouti) was also supported by a Fellowship of the Canadian International Development Agency (CIDA).

Appendix A. Proofs

Proof of Theorem 4.1. To simplify the notation, we write $C_{\delta_1} \equiv C_{w'\theta}$, as in (21). (a) Consider first the case where $p>1$ and $\bar{A}_{22}$ is positive semidefinite with $\bar{A}_{22}\neq 0$. To cover this situation, it will be convenient to distinguish between 2 subcases: (a.1) $r_2 = p-1$; (a.2) $1\leqslant r_2 < p-1$.

(a.1) If $r_2 = p-1$, $\bar{A}_{22}$ is positive definite. From (23), we can write $\bar{Q}(\delta) = \bar{Q}(\delta_1, \delta_2)$. Then, $\delta_1 \in C_{\delta_1}$ iff the following condition holds: (1) if $\bar{Q}(\delta_1, \delta_2)$ has a minimum with respect to δ_2, the minimal value is less than or equal to zero, and (2) if $\bar{Q}(\delta_1, \delta_2)$ does not have a minimum with respect to δ_2, the infimum is less than zero. To check this, we consider the problem of minimizing $\bar{Q}(\delta_1, \delta_2)$ with respect to δ_2. The first and second order derivatives of $\bar{Q}$ with respect to δ_2 are

$$\frac{\partial \bar{Q}}{\partial \delta_2} = 2\bar{A}_{22}\delta_2 + 2\bar{A}_{21}\delta_1 + \bar{b}_2 = 0, \quad \frac{\partial^2 \bar{Q}}{\partial \delta_2 \partial \delta_2'} = 2\bar{A}_{22}. \tag{40}$$

Here the Hessian $2\bar{A}_{22}$ is positive definite, so that there is a unique minimum obtained by setting $\partial\bar{Q}/\partial\delta_2 = 0$:

$$\tilde{\delta}_2 = -\tfrac{1}{2}\bar{A}_{22}^{-1}[2\bar{A}_{21}\delta_1 + \bar{b}_2] = -\bar{A}_{22}^{-1}\bar{A}_{21}\delta_1 - \tfrac{1}{2}\bar{A}_{22}^{-1}\bar{b}_2. \tag{41}$$

On setting $\delta_2 = \tilde{\delta}_2$ in $\bar{Q}(\delta_1, \delta_2)$, we get (after some algebra) the minimal value: $\bar{Q}(\delta_1, \tilde{\delta}_2) = \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1$, where $\tilde{a}_1 = \bar{a}_{11} - \bar{A}_{21}'\bar{A}_{22}^{-1}\bar{A}_{21}$, $\tilde{b}_1 = \bar{b}_1 - \bar{A}_{21}'\bar{A}_{22}^{-1}\bar{b}_2, \tilde{c}_1 = c - \frac{1}{4}\bar{b}_2'\bar{A}_{22}^{-1}\bar{b}_2$. In this case, we also have $\bar{A}_{22}^{-1} = \bar{A}_{22}^{+}$, and (28) holds with $S_1 = \emptyset$.

(a.2) If $1\leqslant r_2 < p-1$, we get, using (23) and (25)–(27)

$$\begin{aligned}\bar{Q}(\delta) &= \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \tilde{\delta}_2'D_2\tilde{\delta}_2 + [2\tilde{A}_{21}\delta_1 + \tilde{b}_2]'\tilde{\delta}_2 \\ &= \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \tilde{\delta}_{2*}'D_{2*}\tilde{\delta}_{2*} + [2\tilde{A}_{21*}\delta_1 + \tilde{b}_{2*}]'\tilde{\delta}_{2*} + [P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2)]'\tilde{\delta}_{22},\end{aligned}$$

where $\tilde{\delta}_{2*} = P_{21}'\delta_2$, $\tilde{\delta}_{22} = P_{22}'\delta_2$, and D_{2*} is a positive definite matrix. We will now distinguish between two further cases: (i) $P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) = 0$, and (ii) $P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2)\neq 0$.

(i) If $P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) = 0$, $\bar{Q}(\delta)$ takes the form:

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \tilde{\delta}_{2*}'D_{2*}\tilde{\delta}_{2*} + [2\tilde{A}_{21*}\delta_1 + \tilde{b}_{2*}]'\tilde{\delta}_{2*}. \tag{42}$$

By an argument similar to the one used for (a.1), we can see that: $\delta_1 \in C_{\delta_1}$ iff $\tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0$, where $\tilde{a}_1 = \bar{a}_{11} - \bar{A}_{21*}'D_{2*}^{-1}\bar{A}_{21*}$, $\tilde{b}_1 = \bar{b}_1 - \bar{A}_{21*}'\ D_{2*}^{-1}\bar{b}_{2*}$, $\tilde{c}_1 = c - \frac{1}{4}\bar{b}_{2*}'D_{2*}^{-1}\bar{b}_{2*}$. Further, since $\bar{A}_{22} = P_2D_2P_2'$, the Moore–Penrose inverse of $\bar{A}_{22}$ is (see Harville, 1997, Chapter 20):

$$\bar{A}_{22}^{+} = P_2\begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0\end{bmatrix}P_2' = [P_{21}, P_{22}]\begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0\end{bmatrix}[P_{21}, P_{22}]' = P_{21}D_{2*}^{-1}P_{21}', \tag{43}$$

hence $\bar{A}_{21*}'D_{2*}^{-1}\bar{A}_{21*} = \bar{A}_{21}'P_{21}D_{2*}^{-1}P_{21}'\bar{A}_{21} = \bar{A}_{21}'\bar{A}_{22}^{+}\bar{A}_{21}$, $\bar{A}_{21*}'D_{2*}^{-1}\bar{b}_{2*} = \bar{A}_{21}'P_{21}D_{2*}^{-1}P_{21}'\bar{b}_2 = \bar{A}_{21}'\bar{A}_{22}^{+}\bar{b}_2$, $\bar{b}_{2*}'D_{2*}^{-1}\bar{b}_{2*} = \bar{b}_2'P_{21}D_{2*}^{-1}P_{21}'\bar{b}_2 = \bar{b}_2'\bar{A}_{22}^{+}\bar{b}_2$.

(ii) If $P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2)\neq 0$, then for any value of δ_1 we can choose $\tilde{\delta}_{22}$ so that $\bar{Q}(\delta_1, \delta_2) < 0$, which entails that $\delta_1 \in C_{\delta_1}$. Putting together the conclusions drawn in (i) and (ii) above, we see that

$$\begin{aligned}C_{\delta_1} &= \{\delta_1 : P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) = 0 \text{ and} \\ &\qquad \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0\} \cup \{\delta_1 : P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) \neq 0\} \\ &= \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leqslant 0\} \cup \{\delta_1 : P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) \neq 0\} \qquad (44)\end{aligned}$$

and (28) holds with $S_1 = \{\delta_1 : P_{22}'(2\bar{A}_{21}\delta_1 + \bar{b}_2) \neq 0\}$. This completes the proof of part (a) of the theorem.

(b) If $p = 1$ or $\bar{A}_{22} = 0$, we can write: $\bar{Q}(\delta_1, \delta_2) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + [2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_2$, where we set $\bar{A}_{21} = \bar{b}_2 = 0$ when $p = 1$. If $2\bar{A}_{21}\delta_1 + \bar{b}_2 = 0$, we see immediately that: $\delta_1 \in C_{\delta_1}$ iff $\bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c \leqslant 0$. Of course, this obtains automatically when $p = 1$. If $2\bar{A}_{21}\delta_1 + \bar{b}_2 \neq 0$, we can choose δ_2 so that $\bar{Q}(\delta_1, \delta_2) < 0$, irrespective of the value of δ_1. Part (b) of the theorem follows on putting together these two observations.

(c) If $p > 1$ and $\bar{A}_{22}$ is not positive semidefinite, this entails that $\bar{A}_{22} \neq 0$, and we can find a vector δ_{20} such that $\delta_{20}'\bar{A}_{22}\delta_{20} \equiv q_0 < 0$. Now, for any scalar Δ_0, we have

$$\bar{Q}(\delta_1, \Delta_0\delta_{20}) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \Delta_0^2 q_0 + \Delta_0[2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_{20}. \qquad (45)$$

Since $q_0 < 0$, we can choose Δ_0 sufficiently large to have $\bar{Q}(\delta_1, \Delta_0\delta_{20}) < 0$, irrespective of the value of δ_1. This entails that all values of δ_1 belong to C_{δ_1}, hence $C_{\delta_1} = \mathbb{R}$. □

References

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Andrews, Donald W.K., Moreira, M.J., Stock, J.H., 2004. Optimal Invariant Similar Tests for Instrumental Variables Regression, July. Cowles Foundation for Research in Economics, Yale University and Department of Economics of Harvard University, Harvard University, New Haven, CT.

Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earning? Quarterly Journal of Economics CVI, 979–1014.

Angrist, J.D., Krueger, A.B., 1995. Split-sample instrumental variables estimates of the return to schooling. Journal of Business and Economic Statistics 13, 225–235.

Angrist, J.D., Imbens, G.W., Krueger, A.B., 1999. Jackknife instrumental variables estimation. Journal of Applied Econometrics 14, 57–67.

Bekker, P.A., Kleibergen, F., 2003. Finite-sample instrumental variables inference using an asymptotically pivotal statistic. Econometric Theory 19 (5), 744–753.

Bekker, P.A., Merckens, A., Wansbeek, T.J., 1994. Identification, Equivalent Models, and Computer Algebra. Academic Press, Boston.

Bound, J., Jaeger, D.A., Baker, R.M., 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.

Chao, J., Swanson, N.R., 2000. Bias and MSE of the IV Estimators Under Weak Identification. Department of Economics, University of Maryland.

Choi, I., Phillips, P.C.B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. Journal of Econometrics 51, 113–150.

Cragg, J.G., Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. Econometric Theory 9, 222–240.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.
Dufour, J.-M., 1982. Generalized Chow tests for structural change: a coordinate-free approach. International Economic Review 23, 565–575.
Dufour, J.-M., 1989. Nonlinear hypotheses inequality restrictions, and non-nested hypotheses: exact simultaneous tests in linear regressions. Econometrica 57, 335–355.
Dufour, J.-M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58, 475–494.
Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.
Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36 (4), 767–808.
Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. International Economic Review 42, 815–843.
Dufour, J.-M., Taamouti, M., 2001a. On Methods for Selecting Instruments. C.R.D.E., Université de Montréal.
Dufour, J.-M., Taamouti, M., 2001b. Point-optimal Instruments and Generalized Anderson–Rubin Procedures for Nonlinear Models. C.R.D.E., Université de Montréal.
Dufour, J.-M., Taamouti, M., 2005. Further Results on Projection-based Inference in IV Regressions with Weak, Collinear or Missing Instruments. Département de sciences économiques, Université de Montréal.
Dufour, J.-M., Taamouti, M., 2004. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73 (4), 1351–1365.
Forchini, G., Hillier, G., 2003. Conditional inference for possibly unidentified structural equations. Econometric Theory 19 (5), 707–743.
Frankel, J.A., Romer, D., 1996. Trade and Growth: An Empirical Investigation. Cambridge, Massachusetts, National Bureau of Economic Research, Technical report 5476.
Frankel, J.A., Romer, D., 1999. Does trade cause growth? The American Economic Review 89(3), 379–399.
Hahn, J., Hausman, J., 2002a. A new specification test for the validity of instrumental variables. Econometrica 70, 163–189.
Hahn, J., Hausman, J., 2002b. Notes on bias in estimators for simultaneous equation models. Economics Letters 75, 237–241.
Hall, A.R., Peixe, F.P.M., 2003. A consistent method for the selection of relevant instruments. Econometric Reviews 2 (3), 269–287.
Hall, A.R., Rudebusch, G.D., Wilcox, D.W., 1996. Judging instrument relevance in instrumental variables estimation. International Economic Review 37, 283–298.
Harrison, A., 1996. Openness and growth: a time-series, cross-country analysis for developing countries. Journal of Developement Economics 48, 419–447.
Harville, D.A., 1997. Matrix Algebra from a Statistician's Perspective. Springer, New York.
Hsiao, C., 1983. Identification, In: Zvi, G., Michael D.I. (Eds.), Handbook of Econometrics, vol. 1. North-Holland, Amsterdam, 223–283 (Chapter 4).
Irwin, A.D., Tervio, M., 2002. Does trade raise income? Evidence from the twentieth century. Journal of International Economics 58, 1–18.
Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70 (5), 1781–1803.
Kleibergen, F., 2004. Testing subsets of structural coefficients in the IV regression model. Review of Economics and Statistics 86, 418–423.
Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica.
Lehmann, E.L., 1986. Testing Statistical Hypotheses, second ed. Wiley, New York.
Magnus, J.R., Neudecker, H., 1991. Matrix Differential Calculus with Applications in Statistics and Econometrics, revised ed. Wiley, New York.
Mankiw, G.N., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics 107 (2), 407–437.
Miller Jr., R.G., 1981. Simultaneous Statistical Inference, second ed. Springer, New York.
Moreira, M.J., 2003a. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.
Moreira, M.J., 2003b. A General Theory of Hypothesis Testing in the Simultaneous Equations Model. Department of Economics, Harvard University, Cambridge, MA.

Moreira, M.J., Poi, B.P., 2001. Implementing tests with correct size in the simultaneous equations model. The Stata Journal 1 (1), 1–15.
Nelson, C.R., Startz, R., 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.
Nelson, C.R., Startz, R., 1990b. Some further results on the exact small properties of the instrumental variable estimator. Econometrica 58, 967–976.
Perron, B., 2003. Semiparametric weak instrument regressions with an application to the risk return tradeoff. Review of Economics and Statistics 85 (2), 424–443.
Pettofrezzo, A.J., Marcoantonio, M.L., 1970. Analytic Geometry with Vectors. Scott, Fosman and Company, Glenview, IL.
Rodrik, D., 1995. Trade and industrial policy reform. In: Behrman, J., Srinivasan, T.N. (Eds.), Handbook of Development Economics, vol. 3A. Elsevier Science, Amsterdam.
Savin, N.E., 1984. Multiple hypothesis testing. In: Zvi G., Michael D.I. (Eds.), Handbook of Econometrics, vol. 2. North-Holland, Amsterdam, pp. 827–879 (Chapter 14).
Scheffé, H., 1959. The Analysis of Variance. Wiley, New York.
Shea, J., 1997. Instrument relevance in multivariate linear models: a simple measure. Review of Economics and Statistics LXXIX, 348–352.
Shilov, G.E., 1961. An Introduction to the Theory of Linear Spaces. Prentice-Hall, Englewood Cliffs, NJ.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.
Startz, R., Nelson, C.R., Zivot, E., 1999. Improved Inference for the Instrumental Variable Estimator. Department of Economics, University of Washington.
Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68, 1097–1126.
Stock, J.H., Yogo, M., 2002. Testing for Weak Instruments in Linear IV Regression. N.B.E.R., Harvard University, Cambridge, MA.
Stock, J.H., Yogo, M., 2003. Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments. Department of Economics, Harvard University, Cambridge, MA.
Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20 (4), 518–529.
Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66 (6), 1389–1404.
Wright, J.H., 2002. Testing the null of identification in GMM. International Finance Discussion Papers, Board of Governors of the Federal Reserve System, Washington, DC.
Wright, J.H., 2003. Detecting lack of identification in GMM. Econometric Theory 19 (2), 322–330.
Zivot, E., Startz, R., Nelson, C.R., 1998. Valid confidence intervals and inference in the presence of weak instruments. International Economic Review 39, 1119–1144.