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

Jean-Marie Dufour † Université de Montréal

Mohamed Taamouti ‡ Université de Montréal and INSEA

First version: August 1999
Revised: May 2003, April 2004, June 2005, August 2005
Compiled: September 7, 2005, 3:45am

A shorter version of this paper is forthcoming in the *Journal of Econometrics*.

* The authors thank David Jaeger for providing his data on returns to education, Craig Burnside for his data on production, 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).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

‡ INSEA, Rabat and CIREQ, Université de Montréal. Mailing address: INSEA., B.P. 6217, Rabat-Instituts, Rabat, Morocco. TEL: 212 7 77 09 26; FAX: 212 7 77 94 57. e-mail: taamouti@insea.ac.ma.

ABSTRACT

In this paper, we provide several results on inference in linear simultaneous regression models (or IV regressions) when instruments can be weak. We define a family of Anderson-Rubin-type (AR-type) procedures based on a general class of auxiliary instruments and for which a finite-sample distributional theory is supplied. The setup considered allows for arbitrary collinearity among the instruments and model endogenous variables, including the presence of accounting relations and singular disturbance covariance matrices. We show that such procedures, in addition to being robust to weak instruments, are also robust to the exclusion of possibly relevant instruments and, more generally, to the distribution of explanatory endogenous variables, a property not shared by several alternative procedures. Using a closed-form solution to the problem of computing linear projections from a general possibly singular quadric surface, we derive computationally simple finite-sample confidence sets for linear combinations of structural parameters based on generalized AR-type procedures in the extended setup considered. We discuss the relation between projection-based confidence sets, Scheffé-type simultaneous confidence intervals and k-class estimators. The performance of projection-based confidence sets as well as the importance of robustness to excluded instruments are studied in a simulation experiment. Finally, the feasibility and usefulness of projection-based confidence sets is illustrated by applying them to three different examples: the relationship between trade and growth in a cross-section of countries, returns to education, and a study of returns to scale and externalities in U.S. production functions.

Key words : simultaneous equations; structural model; instrumental variable; weak instrument; collinearity; missing instrument; confidence interval; testing; projection; simultaneous inference; exact inference; asymptotic theory.

Journal of Economic Literature classification: C3, C13, C12, O4, O1, I2, J2, D2.

RÉSUMÉ

Dans cet article , nous présentons plusieurs résultats sur l'inférence statistique dans les modèles à équations simultanées (régressions avec variables instrumentales) quand les instruments sont ou peuvent être faibles. Sur la base d'une classe générale d'instruments auxilliaires, nous définissons une famille de procédures de type Anderson-Rubin (AR) pour laquelle nous développons une théorie distributionnelle à distance finie. Le cadre considéré supporte une collinéarité arbitraire entre les instruments et les variables endogènes du modèle, y compris la présence de relations comptables et/ou d'une matrice de covariances singulière sur les perturbations. Nous montrons qu'en plus d'être robustes aux instruments faibles, ces procédures sont également robustes à l'exclusion d'instruments pertinents et plus généralement à la distribution de variables explicatives endogènes, une propriété que ne possède pas plusieurs autres procédures. Utilisant une solution explicite du problème de calcul des projections d'une surface quadrique avec matrice singulière, nous dérivons de façon simple des régions de confiance exactes pour des combinaisons linéaires des paramètres structuraux basés sur des procédures AR généralisées. Nous discutons la relation entre les régions de confiance basées sur la projection, les régions de confiance simultanées de type Scheffé et les estimateurs k-class. La performance des intervalles de confiance basés sur la projection et la propriété de robustesse à l'exclusion d'instruments sont ensuite étudiés par simulation. Enfin, la faisabilité et l'utilité des régions de confiance par projection sont illustrées par trois applications empiriques, à savoir: la relation entre l'ouverture commerciale et la croissance économique, le rendement de l'éducation et une application aux rendements d'échelle et les effets d'externalité dans l'économie américaine.

Mots clés words : équations simultanées; modèle structurel; variable instrumentale; instrument faible; collinéarité; instrument omis; région de confiance; tests d'hypothèse; projection; inférence simultanée; inférence exacte; théorie asymptotique.

Classification du Journal of Economic Literature: C3, C13, C12, O4, O1, I2, J2, D2.

Contents

List of propositions and theorems

List of Tables

List of Figures

1. Introduction

Models where different values of the parameter vector may lead to observationally equivalent data distributions are quite widespread in statistics and econometrics. Important examples of such models include: (1) regression models when the matrix of the regressors does not have full rank (multicollinearity); (2) linear simultaneous equations models; (3) errors-in-variables and latent variable models; (4) ARMA and VARMA models; (5) dynamic models with cointegrating relations; (5) models with mixture distributions; etc.[1] Further, inference on such models often lead to complex problems, even when "identifying restrictions" are imposed. 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 such contexts, several papers have documented by simulation and approximate asymptotic methods the poor performance of standard asymptotically justified procedures [Nelson and Startz (1990*a*, 1990*b*), Buse (1992), Bound, Jaeger and Baker (1993, 1995), Hall et al. (1996), Staiger and Stock (1997), Zivot et al. (1998), Dufour and Jasiak (2001)]. The fact that standard asymptotic theory can be arbitrarily inaccurate in finite samples (of any size) is also shown rigorously in Dufour (1997), where it is observed that valid confidence intervals in a standard linear structural equations model must be unbounded with positive probability and Wald-type statistics have distributions which can deviate arbitrarily from their large-sample distribution (even when identification holds). The fact that both finite-sample and large-sample distributions exhibit strong dependence upon nuisance parameters has also been demonstrated by other methods, such as finite-sample distributional theory [see Choi and Phillips (1992)] and local to nonidentification asymptotics [see Staiger and Stock (1997) and Wang and Zivot (1998)].

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 (so that the resulting inference becomes unreliable) or whether

[1] For general expositions of the theory of identification in econometrics and statistics, the reader may consult Rothenberg (1971), Fisher (1976), Hsiao (1983), Prakasa Rao (1992), Bekker, Merckens and Wansbeek (1994) and Manski (1995, 2003).

[2] See, for example, Nelson and Startz (1990*a*, 1990*b*), Buse (1992), Maddala and Jeong (1992), Bound, Jaeger and Baker (1993, 1995), Angrist and Krueger (1995), Hall, Rudebusch and Wilcox (1996), Dufour (1997), Shea (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Startz, Nelson and Zivot (1999), Perron (1999), Chao and Swanson (2000), Stock and Wright (2000), Dufour and Jasiak (2001), Hahn and Hausman (2002*a*, 2002*b*), Hahn, Hausman and Kuersteiner (2001), Kleibergen (2002, 2004, 2005), Moreira (2001, 2003*a*, 2003*b*), Moreira and Poi (2001), Stock and Yogo (2002, 2003), Stock, Wright and Yogo (2002), Perron (2003), Wright (2003, 2002), Bekker and Kleibergen (2003) Hall and Peixe (2003), Forchini and Hillier (2003), Andrews, Moreira and Stock (2004), Dufour and Taamouti (2005), 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.

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. The fact that tests should be based on statistics whose distributions can be bounded and that confidence sets should be derived from pivotal statistics is, of course, a requirement of basic statistical theory [see Lehmann (1986)]. In the framework of linear simultaneous equations and in view of weak instrument problems, the importance of using pivotal functions for statistical inference has been recently reemphasized by several authors [see Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot et al. (1998), Startz et al. (1999), Dufour and Jasiak (2001), Stock and Wright (2000), Kleibergen (2002, 2004), Moreira (2001, 2003*a*), and Stock et al. (2002)]. In particular, 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).

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 close to being the only procedure 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. Of course, tests and confidence sets based on the AR method are asymptotically pivotal under much weaker distributional assumptions [see Dufour and Jasiak (1993, 2001), Staiger and Stock (1997)].

Other potential pivots aimed at being robust to weak instruments have recently been suggested by Wang and Zivot (1998), Kleibergen (2002) and Moreira (2003*a*). 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 thus may lead to power gains under the assumptions considered. But this will typically be at the expense of robustness. Further, only asymptotic distributional theories have been supplied for these statistics, so that the level of the procedures may not be controlled in finite samples. [4]

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. The practical importance of this point is demonstrated in a simulation experiment where alternative procedures exhibit strong size distortions, while AR-type tests are not affected (as expected from theory).

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

[4] Finite-sample conservative bounds have, however, been proposed by Dufour (1997) for LR statistics and by Bekker and Kleibergen (2003) for Kleibergen's statistic.

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.[5]

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 nonseparability*: 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. So, not surprisingly, the AR-procedure succeeds at achieving pivotality by considering tests for hypotheses of the form $H_0 : \beta = \beta_0$, where the vector β contains the coefficients of *all* the endogenous explanatory variables in a linear structural equation.

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 method has the interesting feature that the level of the resulting confidence sets for transformed coefficients is at least as large as the one of the original joint confidence set from which the projection is made, so in the case of an exact AR-type confidence set the corresponding projection-based confidence sets are also exact, in 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)].[6] In Dufour and Jasiak (2001), under a more restricted setup, such confidence sets were actually computed by using nonlinear optimization procedures, whose computational cost can be high. Exploiting the fact that AR confidence sets can be represented by *quadric* surfaces, we also showed in Dufour and Taamouti (2005) that projection-based confidence sets for linear transformations of model coefficients can be obtained in much simpler way (which does not require nonlinear optimization) in the *special case* where the quadratic part of the quadric involves a *full-rank matrix* (the *concentration matrix*). Here we

[5]Multicollinearity is one of the most basic form of identification failure, which has led to the classical theory of estimable fucntions. For further discussion, see Magnus and Neudecker (1991, Chapter 13), Rao (1973, Chapter 4), Rao and Mitra (1971, Chapter 7) and Scheffé (1959, Chapters 1-2).

[6]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.

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), Savin (1984) 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. Correspondingly, an arbitrary number of hypotheses on β can be tested without ever losing control of the overall level of the multiple tests, *i.e.* the probability of rejecting at least one true null hypothesis on β is not larger than the level α. This can provide an important check on data mining.

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, page 649)] where the "standard error" is corrected in a way that depends on the level of the test. The confidence interval for a linear combination of the parameters, say $w'\beta$ takes the usual form $[w'\hat{\beta} - \hat{\sigma} z_{\alpha},\ w'\hat{\beta} + \hat{\sigma} z_{\alpha}]$ with $\hat{\beta}$ a k-class type estimator of β.

Sixth, the methods discussed in this work are evaluated and compared on the basis of Monte Carlo simulations. In particular, we study how conservative projection-based confidence sets are as well as their robustness to weak and missing instruments.

Seventh, in order to illustrate the projection approach, we present three 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, Jaeger and Baker (1995), while in the third example we study returns to scale and externalities in various industrial sectors of the U.S. economy, using a production function specification previously considered by Burnside (1996).

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 Anderson-Rubin procedures in section 3. The projection approach and its simultaneity properties are discussed in section 4. The general closed-form solution to the problem of building projection-based confidence sets from a general quadric confidence set is presented in section 5. The relation between projection-based confidence sets, Scheffé confidence intervals and k-class estimators is discussed in section 6. In section 7, we report the results of our Monte Carlo simulations, while section 8 presents the empirical applications. We conclude in section 9.

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) and Stock et al. (2002)]:

$$y = Y\beta + X_1\gamma + u\,, \tag{2.1}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + V, \tag{2.2}$$

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables $(G \geq 1)$, 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:

$$X = [X_1,\, X_2] \text{ is a full-column rank } T \times k \text{ matrix, where } k = k_1 + k_2\,; \tag{2.3}$$

$$u \text{ and } X \text{ are independent;} \tag{2.4}$$

$$u \sim N\big[0,\, \sigma_u^2\, I_T\big]\,. \tag{2.5}$$

We consider the problem of building tests and confidence sets on β and γ. In view of the fact that these parameters may not be identified (which occurs when the matrix Π_2 has rank less than G), it is especially important that such procedures be based on proper pivotal (or boundedly pivotal functions); see Dufour (1997). In particular, Wald-type statistics are not pivotal in such a setup. More generally, test statistics in this context tend to depend heavily on various unknown nuisance parameters.

As pointed out in Dufour (1997) and Staiger and Stock (1997), a possible solution consists in exploiting a procedure suggested long ago by Anderson and Rubin (1949). This method is based on the simple idea that if β is specified, model (2.1) - (2.2) can be reduced to a simple linear regression equation. More precisely, if we consider the hypothesis $H_0 : \beta = \beta_0$ in equation (2.1), we can write:

$$y - Y\beta_0 = X_1\Delta_1 + X_2\Delta_2 + \varepsilon \tag{2.6}$$

where $\Delta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\Delta_2 = \Pi_2(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. We can test H_0 by testing $H_0' : \Delta_2 = 0$ using the standard F-statistic for H_0' [denoted $AR(\beta_0)$]. Under the assumptions (2.1) - (2.5) and H_0, equation (2.6) satisfies all the conditions of the linear regression model and we have:

$$AR(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2,\, T - k) \tag{2.7}$$

where for any full rank matrix B, $M(B) = I - B(B'B)^{-1}B'$. The distributional result in (2.7) holds irrespective of the rank of the matrix Π_2, which means that tests based on $AR(\beta_0)$ are *robust to weak instruments*.

The latter yields a confidence set with level $1-\alpha$ for β :

$$C_\beta(\alpha)=\{\beta_0: AR(\beta_0)\le F_\alpha(k_2,\,T-k)\} \tag{2.8}$$

where $F_\alpha(k_2,\,T-k)$ is the $1-\alpha$ quantile of the F distribution with k_2 and $T-k$ degrees of freedom. This confidence set is exact and does not require any identification assumption. When $G=1$, this set has an explicit form solution involving a quadratic inequation – i.e. $C_\beta(\alpha)=\{\beta_0: a\beta_0^2+b\beta_0+c\le 0\}$ where a, b and c are simple functions of the data and the critical value $F_\alpha(k_2,\,T-k)$ – and $C_\beta(\alpha)$ is unbounded if $F(\Pi_2=0)<F_\alpha$, where $F(\Pi_2=0)$ is the F-test for $H_0: \Pi_2=0$ in equation (2.2); see Dufour and Jasiak (2001) and Zivot et al. (1998) for details.

In model (2.1) - (2.2), the "identifying" instruments X_2 that are excluded from the structural equation (2.1) 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:

$$Y=X_1\Pi_1+X_2\Pi_2+X_3\Pi_3+V \tag{2.9}$$

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 (2.1), and could be unobservable.[7] 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, equation (2.2) is replaced by (2.9), but inference proceeds as if (2.2) were the actual equation.

Under the generating process (DGP) represented by (2.1) and (2.9), the variable $y-Y\beta_0$ used as the dependent variable by the AR procedure satisfies the equation:

$$y-Y\beta_0=X_1\Delta_1+X_2\Delta_2+X_3\Delta_3+\varepsilon \tag{2.10}$$

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 $AR(\beta_0)$ is $F(k_2,\,T-k)$ [under the assumptions (2.1), (2.3) - (2.5) and (2.9)], even if X_3 is excluded from the regression as in (2.6). The finite-sample validity of the test based on $AR(\beta_0)$ 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 $AR(\beta_0)$, so that X_3 does not have to be strictly exogenous.

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 nonlinear model of the form:

$$Y=g(X_1,\,X_2,\,X_3,\,V,\,\Pi) \tag{2.11}$$

[7] Clearly, this depends on the interpretation of the structural equation (2.1) 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 betwen Y_i and X_{3i} need not exist.

where $g(\cdot)$ is a possibly unspecified nonlinear 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 (2.6) 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 (2.2), 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 (2003*a*). In section 7.2, we present simulation evidence which clearly illustrates these difficulties.

3. A generalized Anderson-Rubin 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 (2.1) where the different symbols are defined as in (2.1). However, we shall make the following modified assumptions:

$$0 \leq \operatorname{rank}(X_1) = \nu_1 \leq k_1 \,, \tag{3.1}$$

$$\bar{X}_2 \text{ is a } T \times \bar{k}_2 \text{ matrix such that } 0 \leq \operatorname{rank}(\bar{X}_2) = \nu_2 \leq \bar{k}_2 \,, \tag{3.2}$$

$$u \,|\, \bar{X} \sim N\big[0,\, \sigma_u^2(\bar{X}) I_T\big] \text{ where } \bar{X} = [X_1,\, \bar{X}_2]. \tag{3.3}$$

Here (3.1) 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 (3.3) states that, conditional on $\bar{X}$, the disturbances in the structural equation (2.1) are *i.i.d.* normal. Of course, (3.1) - (3.3) cover the more usual assumptions (2.3) - (2.5) 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 (2.2), (2.9) or (2.11) 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

$$X_1 = [X_{11},\, X_{12}]\,,\ \gamma = (\gamma_1',\, \gamma_2')'\,, \tag{3.4}$$

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 $0 \leq k_{11} \leq k_1$. 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:

$$H_0(\beta_0,\, \gamma_{10}) : (\beta,\, \gamma_1) = (\beta_0,\, \gamma_{10}) \tag{3.5}$$

where, by convention, this reduces to $H_0 : \beta = \beta_0$, if $k_{11} = 0$. Under the null hypothesis, we have

$$y - Y\beta_0 - X_{11}\gamma_{10} = X_{12}\gamma_2 + u \tag{3.6}$$

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

$$y - Y\beta_0 - X_{11}\gamma_{10} = X_{11}\varDelta_{11} + X_{12}\varDelta_{12} + \bar{X}_2\varDelta_2 + u = \bar{X}\theta + u \tag{3.7}$$

where $\bar{X} \equiv [X_1,\, \bar{X}_2] = [X_{11},\, X_{12},\, \bar{X}_2]$, and then testing the restrictions

$$H_0^*(\beta_0,\, \gamma_{10}) : \varDelta_{11} = 0 \text{ and } \varDelta_2 = 0 \tag{3.8}$$

under which (3.7) becomes equivalent to the null model (3.6). Again, if $k_{11} = 0$, X_{11} simply drops from the left-hand side of (3.7), and $H_0^*(\beta_0,\, \gamma_{10})$ reduces to $H_0^*(\beta_0) : \varDelta_2 = 0$.

A Fisher-type test may still be applied here, provided corrected degrees of freedom are used. Let

$$\nu_2 = \mathrm{rank}(X_{12}) \quad \text{and} \quad \nu = \mathrm{rank}(\bar{X}) = \mathrm{rank}([X_{11},\, X_{12},\, \bar{X}_2])\,, \tag{3.9}$$

be the ranks of the regressor matrix respectively under the null hypothesis (3.6) and the alternative (3.7). The Fisher statistic for testing $H_0^*(\beta_0,\, \gamma_{10})$ is then:

$$AR(\beta_0,\, \gamma_{10};\, \bar{X}_2) = \frac{u(\beta_0,\, \gamma_{10})'[M(X_{12}) - M(\bar{X})]u(\beta_0,\, \gamma_{10})/(\nu - \nu_2)}{u(\beta_0,\, \gamma_{10})'M(\bar{X})u(\beta_0,\, \gamma_{10})/(T - \nu)} \tag{3.10}$$

where $u(\beta_0,\, \gamma_{10}) \equiv y - Y\beta_0 - X_{11}\gamma_{10}$. 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 the assumptions (3.1) - (3.3) and the null hypothesis $H_0^*(\beta_0,\, \gamma_{10})$, all the conditions of the classical linear model are satisfied and we can conclude that:

$$AR(\beta_0,\, \gamma_{10};\, \bar{X}_2) \sim F(\nu - \nu_2,\, T - \nu)\,; \tag{3.11}$$

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 $\nu - \nu_2 \leq rank([X_{11},\, \bar{X}_2])$, where a strict inequality is possible. Further the distribution and the rank of the Y matrix are irrelevant.

In view of (3.11), a confidence set with level $1 - \alpha$ for the vector $(\beta',\, \gamma_1')'$ can be obtained by inverting the statistic $AR(\beta_0,\, \gamma_{10};\, \bar{X}_2)$:

$$C_{(\beta,\gamma_1)}(\alpha) = \{(\beta_0',\, \gamma_{10}')' : AR(\beta_0,\, \gamma_{10};\, \bar{X}_2) \leq F_\alpha(\nu - \nu_2,\, T - \nu)\}\,. \tag{3.12}$$

Using an argument similar to the one in Dufour and Taamouti (2005), this set can be rewritten in the form

$$C_{(\beta,\,\gamma_1)}(\alpha) = \{(\beta_0',\,\gamma_{10}')' : (\beta_0',\,\gamma_{10}')A(\beta_0',\,\gamma_{10}')' + b'(\beta_0',\,\gamma_{10}')' + c \le 0\} \tag{3.13}$$

where $A = [Y,\,X_{11}]'H[Y,\,X_{11}]$, $b = -2[Y,\,X_{11}]'Hy$, $c = y'Hy$, and

$$H = M(X_{12}) - \left[1 + \frac{\nu - \nu_2}{T - \nu} F_\alpha(\nu - \nu_2,\, T - \nu)\right] M(\bar{X})\,. \tag{3.14}$$

We call A the *concentration matrix at level* α (or the α-concentration *matrix*) associated with $(\beta',\,\gamma_1')'$. The quadratic-linear form in (3.13) defines a quadric surface [see Shilov (1961, Chapter 11) and 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:

$$C_\beta(\alpha) = \{\beta_0 : a\beta_0^2 + b\beta_0 + c \le 0\} \tag{3.15}$$

where a, b and c are simple functions of the data and the critical value $F_\alpha(\nu - \nu_2,\, T - \nu)$. 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); details on the different possible cases are, however, the same except that the case where $a = 0$ may have a non-zero probability in problems where $[Y,\,X_{11}]$ does not have full-column rank. When $(\beta',\,\gamma_1')'$ contains more than one parameter, we face the problem of building confidence sets and tests for individual elements of $(\beta',\,\gamma_1')'$, which we now tackle through projection techniques.

4. The projection approach and simultaneous inference

The projection technique is a general approach that may be applied in different contexts. Given a confidence set $C_\theta(\alpha)$ with level $1-\alpha$ for the parameter vector θ, this method enables one to deduce confidence sets for general transformations g in $\mathbb{R}^m$ of this vector. For example, we may have $\theta = \beta$ or $\theta = (\beta',\,\gamma_1')'$. Since $x \in E \Rightarrow g(x) \in g(E)$ for any set E, we have

$$\mathsf{P}[\theta \in C_\theta(\alpha)] \ge 1 - \alpha \Rightarrow \mathsf{P}\big[g(\theta) \in g\,[C_\theta(\alpha)]\,\big] \ge 1 - \alpha \tag{4.1}$$

where $g\,[C_\theta(\alpha)] = \{x \in \mathbb{R}^m : \exists\,\theta \in C_\theta(\alpha),\ g(\theta) = x\}$. Hence $g\,[C_\theta(\alpha)]$ is a conservative confidence set for $g(\theta)$ with level $1-\alpha$.

Even if $g(\theta)$ is scalar, the projection-based confidence set is not necessarily an interval. However, it is easy to see that

$$\mathsf{P}[g^L(\alpha) \le g(\theta) \le g^U(\alpha)] \geqslant 1 - \alpha \tag{4.2}$$

where $g^L(\alpha) = \inf\{g(\theta_0), \theta_0 \in C_\theta(\alpha)\}$ and $g^U(\alpha) = \sup\{g(\theta_0), \theta_0 \in C_\theta(\alpha)\}$; see Dufour (1997), Abdelkhalek and Dufour (1998) or Dufour and Jasiak (2001). Thus $I_U(\alpha) = \big[g^L(\alpha),\, g^U(\alpha)\big]\backslash\{-\infty,\, +\infty\}$ is a confidence interval with level $1-\alpha$ for $g(\theta)$, where it is assumed that $-\infty$ and $+\infty$ are not admissible. This interval is not bounded when $g^L(\alpha)$ or $g^U(\alpha)$ is infinite.

It is worth noting that we obtain in this way *simultaneous confidence sets* for any number of transformations of θ: $g_1(\theta),\, g_2(\theta),\, \ldots\, ,\, g_n(\theta)$. The set $C_{g_1(\theta)}(\alpha) \times C_{g_2(\theta)}(\alpha) \times \cdots \times C_{g_n(\theta)}(\alpha)$ where $C_{g_i(\theta)}(\alpha)$ is the projection-based confidence set for $g_i(\theta)$, $i = 1, \ldots , n$, is a simultaneous confidence set for the vector $\big(g_1(\theta),\, g_2(\theta),\, \ldots\, ,\, g_n(\theta)\big)'$ with level greater than or equal to $1 - \alpha$. More generally, if $\{g_a(\theta) : a \in A\}$ is a set of functions of θ,where A is some index set, then

$$\mathsf{P}\big[g_a(\theta) \in g_a\left[C_\theta(\alpha)\right] \text{ for all } a \in A\big] \geq 1 - \alpha\,. \tag{4.3}$$

If these confidence intervals are used to test different hypotheses, an unlimited number of hypotheses can be tested without losing control of then overall level. The confidence sets obtained in this way are *simultaneous* in the sense of Scheffé. For further discussion of simultaneous inference, the reader may consult Miller (1981), Savin (1984), and Dufour (1989).

If the aim is to test $H_0 : g(\theta) = 0$, we can easily deduce from $C_\theta(\alpha)$ a conservative test. The latter consists in rejecting H_0 when all the vectors θ_0 that satisfy H_0 are rejected by the AR test, or equivalently when the minimum of $AR(\theta_0)$ subject to the constraint (s.c.) $g(\theta) = 0$ is larger than $F_\alpha(k_2, T-k)$, *i.e.*when $\min\{AR(\theta) : g(\theta) = 0\} \geq F_\alpha(k_2, T-k)$.

5. 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:

$$C_\theta = \{\theta_0 : \theta_0' A\theta_0 + b'\theta_0 + c \leq 0\} \tag{5.1}$$

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:

$$C_{w'\theta} \equiv g[C_\theta] = \{\delta_0 : \delta_0 = w'\theta_0 \text{ where } \theta_0' A\theta_0 + b'\theta_0 + c \leq 0\}. \tag{5.2}$$

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 nonsingular transformation of θ :

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} w'\theta \\ R_2\theta \end{bmatrix} = R\theta\,, \quad R = \begin{bmatrix} w' \\ R_2' \end{bmatrix} = \begin{pmatrix} w_1 & w_2' \\ 0 & I_{p-1} \end{pmatrix}, \tag{5.3}$$

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 (5.1) may be written:

$$\theta' A\theta + b'\theta + c = \delta'\bar{A}\delta + \bar{b}'\delta + c \equiv \bar{Q}(\delta) \tag{5.4}$$

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

$$C_{w'\theta} = C_{\delta_1} = \{\delta_1 : \delta = (\delta_1,\, \delta_2')' \text{ satisfies } \bar{Q}(\delta) \le 0\}\,. \tag{5.5}$$

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

$$\bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{A}_{21}' \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \end{pmatrix}, \tag{5.6}$$

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:

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \delta_2'\bar{A}_{22}\delta_2 + [2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_2 \tag{5.7}$$

where, by convention, the two last terms of (5.7) simply disappear when $p = 1$. For $p \ge 1$, let $r_2 = \operatorname{rank}(\bar{A}_{22})$, where $0 \le r_2 \le p-1$, and consider the spectral decomposition:

$$\bar{A}_{22} = P_2 D_2 P_2'\,, \quad D_2 = \operatorname{diag}(d_1,\, \ldots\,,\, d_{p-1}) \tag{5.8}$$

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

$$\begin{aligned} d_i \;\; &\ne 0, \text{ if } 1 \le i \le r_2\,, \\ &= 0, \text{ if } i > r_2\,. \end{aligned} \tag{5.9}$$

Let us also define (whenever the objects considered exist)

$$\tilde{\delta}_2 = P_2'\delta_2\,, \quad \tilde{A}_{21} = P_2'\bar{A}_{21}\,, \quad \tilde{b}_2 = P_2'\bar{b}_2\,,\; D_{2*} = \operatorname{diag}(d_1,\, \ldots\,,\, d_{r_2})\,, \tag{5.10}$$

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:

$$\tilde{\delta}_{2*} = P_{21}'\delta_2\,, \quad \tilde{A}_{21*} = P_{21}'\bar{A}_{21}\,, \quad \tilde{b}_{2*} = P_{21}'\bar{b}_2\,, \quad P_2 = [P_{21},\, P_{22}] \tag{5.11}$$

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 5.1 PROJECTION-BASED CONFIDENCE SETS WITH A POSSIBLY SINGULAR CONCENTRATION MATRIX. *Under the assumptions and notations* (5.4) - (5.11), *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*

$$C_{w'\theta} = \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0\} \cup S_1 \tag{5.12}$$

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*

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

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

$$C_{w'\theta} = \{\delta_1 : \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c \leq 0\} \cup S_2 \tag{5.13}$$

where

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

(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 the 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 **5.1** 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 (5.12) and (5.13). Specifically, it is easy to see that the quadratic set $\widetilde{C}_{w'\theta} = \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0\}$ in (5.12) 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 (5.13).

The results in this paper generalize those provided in Dufour and Taamouti (2005) by allowing A to have an arbitrary rank. In (3.13), 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 (5.12) reduces to

$$C_{w'\theta} = \left[w'\tilde{\theta} - \sqrt{d\,(w'A^{-1}w)}\ ,\ w'\tilde{\theta} + \sqrt{d\,(w'A^{-1}w)}\ \right] \tag{5.14}$$

where $\tilde{\theta} = -\frac{1}{2}A^{-1}b$, and $d = \frac{1}{4}b'A^{-1}b - c \geq 0$ (if $d < 0$, $C_{w'\theta}$ is empty). If, furthermore, $w = e_i = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{pi})'$, with $\delta_{ji} = 1$ if $j = i$ and $\delta_{ji} = 0$ otherwise, the set $C_{w'\theta}$ is a confidence interval for the component θ_i and is given by:

$$C_{\theta_i} = \left[\ \tilde{\theta}_i - \sqrt{d\,(A^{-1})_{ii}}\ ,\ \tilde{\theta}_i + \sqrt{d\,(A^{-1})_{ii}}\ \right] \tag{5.15}$$

where $\tilde{\theta}_i = -(A^{-1})_{i.}b/2$ is the i-th element of $\tilde{\theta} = -\frac{1}{2}'A^{-1}b$, $(A^{-1})_{i.}$ is the i-th row of A^{-1}, $(A^{-1})_{ii}$ is the i-th element of the diagonal of A^{-1}, and $(A^{-1})_{ii} > 0$.

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

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

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

It is interesting to notice the relationship of the above results with Scheffé-type confidence sets in the context of model (2.1) - (2.2). 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 $w'\beta$.

In the case where A is nonsingular, has exactly one negative eigenvalue, $w'A^{-1}w < 0$, and $d < 0$, the confidence set for $w'\beta$ reduces to

$$C_{w'\beta} = \left]-\infty\,,\ w'\tilde{\beta} - \sqrt{d\,(w'A^{-1}w)}\right] \cup \left[w'\tilde{\beta} + \sqrt{d\,(w'A^{-1}w)}\,,\ +\infty\right[\,. \tag{6.1}$$

Note here that $C_{w'\beta}$ can remain informative, even if it is unbounded. In particular, if we want to test $H_0 : w'\beta = r$ and consider as a decision rule which rejects H_0 when $r \notin C_{w'\beta}$, H_0 will be rejected for all values of r in the interval $\left(w'\tilde{\beta} - \sqrt{d(w'A^{-1}w)}\,,\ w'\tilde{\beta} + \sqrt{d(w'A^{-1}w)}\right)$. In this case, $g^L(\alpha) = -\infty$ and $g^U(\alpha) = \infty$, so that $I_U(\alpha) = \mathbb{R}$ an uninformative set, while in fact the true projection-based confidence set is a proper subset of $\mathbb{R}$.

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 (5.14)] 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 $\left|(w'\tilde{\beta} - w'\beta)/[\hat{\sigma}^2(w'A^{-1}w)]^{1/2}\right|$. Replacing A and b by their expressions, the estimator $\tilde{\beta}$ may be written:

$$\tilde{\beta} = (Y'HY)^{-1}Y'Hy\,.$$

$\tilde{\beta}$ may be interpreted as an instrumental variables estimator. Indeed, on multiplying (2.1) by $(HY)'$, we get

$$Y'Hy = Y'HY\beta + Y'Hu\,,$$

which yields the IV estimator

$$\tilde{\beta}_{IV} = (Y'HY)^{-1}Y'Hy = \tilde{\beta}\,.$$

If $rank(\mathit{\Pi}_2) = G$ and the following usual assumptions hold,

$$\left(\frac{X'X}{T},\, \frac{X'u}{T},\, \frac{X'V}{T}\right) \underset{T\to\infty}{\xrightarrow{\mathsf{p}}} (Q_{XX},\, 0,\, 0)\,, \quad \frac{X'u}{\sqrt{T}} \underset{T\to\infty}{\xrightarrow{L}} N\big[0,\, \sigma_u^2 Q_{XX}\big]\,, \tag{6.2}$$

then HY is asymptotically uncorrelated with the disturbances u and

$$\sqrt{T}\,(\tilde{\beta} - \beta) \underset{T\to\infty}{\xrightarrow{L}} N\big[0, \sigma_u^2 \underset{T\to\infty}{\text{plim}}\big(\frac{1}{T}A\big)^{-1}\big] \tag{6.3}$$

where $\underset{T\to\infty}{\text{plim}}\, \frac{1}{T}A = \mathit{\Pi}_2'\big[\,Q_{X_2X_2} - Q_{X_2X_1}Q_{X_1X_1}^{-1}Q_{X_2X_1}'\big]\mathit{\Pi}_2$ 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:

$$\tilde{\beta} = \{Y'[M(X_1) - (1+f_\alpha)M(X)]Y\}^{-1}Y'[M(X_1) - (1+f_\alpha)M(X)]y. \tag{6.4}$$

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 two stage least squares estimator [see Davidson and MacKinnon (1993, page 649)]. The later may be written:

$$\hat{\beta}_{2SLS} = \{Y'[M(X_1) - M(X)]Y\}^{-1}Y'[M(X_1) - M(X)]y.$$

Hence, when $\mathit{\Pi}_2$ 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):

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

7. Simulation study

In this section, we study by Monte Carlo methods the properties of AR-type and projection-based confidence procedures. We focus on two main issues. First, we evaluate how conservative projection-based confidence sets are and compare the confidence sets based on different test statistics. The tests considered are the exact Anderson-Rubin test based on (2.7), the asymptotic version of this test using the $\chi^2(k_2)/k_2$ distribution, as well as the LR and LM tests proposed by Wang and Zivot (1998). Second, we study the robustness to instrument exclusion on the finite sample behavior of the statistics considered above and two other statistics proposed recently in the literature, namely, Kleibergen's (2002) K-test and the conditional LR test of Moreira (2003*a*).

7.1. Performance of projection-based confidence sets

To study the properties of projection-based confidence sets, we consider the following data generating process:

$$y = Y_1\beta_1 + Y_2\beta_2 + X_1\gamma + u\,, \tag{7.1}$$

$$(Y_1,\, Y_2) = X_1\varPi_1 + X_2\varPi_2 + (V_1,\, V_2)\,, \tag{7.2}$$

$$(u_t,\, V_{1t},\, V_{2t})' \overset{i.i.d.}{\sim} N(0,\, \varSigma)\,, \quad \varSigma = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .3 \\ .8 & .3 & 1 \end{pmatrix}, \tag{7.3}$$

where X_1 is a $T \times 1$ column of ones and X_2 is a $T \times k_2$ (fixed) matrix of instruments. The elements of X_2 were generated as i.i.d. $N(1,\, 1)$ random variables, but they are kept fixed over the simulation. The parameters values are set at $\beta_1 = \frac{1}{2}$, $\beta_2 = 1$, $\gamma = 2$, and $\varPi_1 = (0.1, 0.5)$. The correlation coefficient r between u and V_i $(i = 1, 2)$ is set equal to 0.8, the variables Y_1 and Y_2 are endogenous and the instrumental variables X_2 are necessary. The matrix $\varPi_2$ is such that $\varPi_2 = C/\sqrt{T}$. We consider three different sample sizes $T = 50,\ 100,\ 200$. The number of instruments (k_2) varies from 2 to 40. All simulations are based on 10000 replications.

Table 2 presents results on the performance of Wald-type 2SLS-based confidence sets, while the three following tables report results on the other procedures for three basic cases: (1) in Table 3, $C = 0$ (complete unidentification); (2) in Table 4, the components c_{ij} of the matrix C satisfy $1 < c_{ij} < 5$ (weak identification); (3) in Table 5, we have $10 \leq c_{ij} \leq 20$ (strong identification). The nominal level of the confidence procedures is 95%.

Let us consider first the behavior of the Wald procedure (Table 2). As expected from the results in Dufour (1997), its real coverage rate may reach 0 when the instruments are very poor. The only case where it behaves well is when identification holds and the number of instruments is small compared to the sample size. This shows how crucial is the need for alternative valid pivotal statistics.

For the exact AR statistic, no size distortion, even very small, is observed. The main observation is that the coverage rate of the projection-based confidence sets for β_1 decreases as k_2 increases and gets closer to the exact confidence level $1 - \alpha$ of the confidence set for β.[8] *Thus the projection-*

[8] Recall that theoretically, this rate is always greater than or equal to the confidence level of the set from which the projection is done.

Table 2. Empirical coverage rate of 2SLS-based Wald confidence sets

T	k	$C_{ij}=0$	$1 \leq C_{ij} \leq 5$	$10 \leq C_{ij} \leq 20$
50	2	56.13	97.40	94.46
	3	25.10	94.05	93.71
	4	9.19	89.06	93.68
	5	3.82	84.49	93.65
	10	0.03	78.28	93.33
	15	0.00	78.99	93.16
	20	0.00	77.14	92.88
	30	0.00	68.47	93.35
	40	0.00	67.84	92.30
100	2	55.22	97.68	95.07
	3	24.53	94.33	94.43
	4	10.52	89.45	95.16
	5	3.81	87.16	94.16
	10	0.03	83.88	94.44
	15	0.00	81.40	94.12
	20	0.00	72.29	94.19
	30	0.00	61.47	93.20
	40	0.00	45.48	93.66
200	2	55.53	97.85	95.32
	3	24.55	94.68	94.80
	4	10.32	90.33	94.95
	5	4.10	89.19	94.64
	10	0.04	83.99	94.75
	15	0.00	81.28	94.14
	20	0.00	71.71	94.32
	30	0.00	62.26	93.89
	40	0.00	54.99	93.76

Table 3. Characteristics of AR and ARS projection-based confidence sets – $C_{ij} = 0$

		AR					ARS				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
	2	95.06	100.00	99.98	0.00	99.94	94.07	99.99	99.98	0.00	99.89
	3	95.51	99.98	99.97	0.00	99.86	94.43	99.97	99.96	0.00	99.72
	4	94.61	99.97	99.94	0.00	99.81	93.01	99.88	99.89	0.00	99.62
	5	94.69	99.92	99.94	0.00	99.70	92.77	99.86	99.87	0.00	99.30
50	10	94.74	99.88	99.90	0.00	99.61	91.28	99.74	99.72	0.00	98.91
	15	94.77	99.94	99.93	0.00	99.64	89.08	99.59	99.61	0.00	98.32
	20	95.12	99.91	99.94	0.00	99.66	86.79	99.28	99.43	0.02	97.22
	30	95.16	99.89	99.89	0.00	99.48	80.98	98.18	98.19	0.12	93.52
	40	95.01	99.81	99.85	0.00	99.37	69.41	94.78	94.33	0.68	82.79
	2	94.73	99.98	99.98	0.00	99.88	94.38	99.98	99.97	0.00	99.83
	3	94.79	99.95	99.96	0.00	99.78	94.15	99.94	99.95	0.00	99.75
	4	95.15	99.96	99.95	0.00	99.74	94.34	99.89	99.88	0.00	99.63
	5	95.30	99.98	99.98	0.00	99.80	94.32	99.97	99.97	0.00	99.75
100	10	95.34	99.88	99.87	0.00	99.68	93.71	99.82	99.84	0.00	99.50
	15	94.75	99.92	99.90	0.01	99.68	92.15	99.81	99.83	0.01	99.22
	20	94.72	99.96	99.93	0.00	99.62	91.35	99.75	99.78	0.00	98.87
	30	95.19	99.94	99.94	0.00	99.73	90.45	99.67	99.70	0.00	98.67
	40	94.45	99.92	99.89	0.00	99.56	87.95	99.41	99.38	0.03	97.72
	2	94.99	99.99	99.98	0.00	99.83	94.81	99.98	99.98	0.00	99.83
	3	95.10	99.94	99.93	0.00	99.76	94.91	99.93	99.93	0.00	99.74
	4	94.99	99.94	99.93	0.00	99.68	94.61	99.94	99.91	0.00	99.65
	5	94.95	99.93	99.90	0.00	99.67	94.43	99.92	99.89	0.00	99.61
200	10	95.08	99.92	99.94	0.00	99.66	94.22	99.90	99.93	0.00	99.57
	15	94.98	99.95	99.92	0.00	99.70	94.00	99.89	99.90	0.00	99.58
	20	95.08	99.95	99.93	0.00	99.66	93.55	99.90	99.89	0.00	99.44
	30	94.88	99.95	99.93	0.00	99.63	92.79	99.92	99.91	0.00	99.30
	40	95.14	99.88	99.86	0.00	99.49	92.27	99.75	99.72	0.00	99.00

Table 3 (continued). Characteristics of LR and LM projection-based confidence sets – $C_{ij} = 0$

		LM					LR				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
50	2	95.08	100.00	99.98	0.00	99.94	94.05	99.99	99.98	0.00	99.89
	3	95.77	99.98	99.97	0.00	99.91	95.06	99.97	99.97	0.00	99.80
	4	95.10	99.97	99.98	0.00	99.92	94.76	99.96	99.99	0.00	99.86
	5	95.48	99.94	99.95	0.00	99.92	95.36	99.94	99.94	0.00	99.82
	10	96.57	99.97	99.95	0.00	99.95	97.16	99.99	99.99	0.00	99.92
	15	97.76	99.99	99.99	0.00	99.99	97.64	100.00	99.99	0.00	99.94
	20	98.86	100.00	99.99	0.00	99.99	97.92	100.00	100.00	0.00	99.96
	30	99.97	100.00	100.00	0.00	10.00	96.99	99.97	99.95	0.00	99.83
	40	100.00	100.00	100.00	0.00	100.00	89.49	99.50	99.50	0.00	97.47
100	2	94.74	99.98	99.98	0.00	99.88	94.38	99.98	99.97	0.00	99.83
	3	94.99	99.96	99.96	0.00	99.87	95.10	99.97	99.97	0.00	99.88
	4	95.38	99.96	99.95	0.00	99.85	96.15	99.98	99.97	0.00	99.85
	5	95.64	99.98	99.98	0.00	99.93	96.38	100.00	100.00	0.00	99.90
	10	96.20	99.90	99.89	0.00	99.86	98.20	99.98	99.95	0.00	99.93
	15	95.99	99.95	99.93	0.00	99.90	98.68	100.00	100.00	0.00	99.99
	20	96.75	100.00	99.99	0.00	99.99	99.37	100.00	100.00	0.00	100.00
	30	98.10	99.99	99.98	0.00	99.98	99.65	100.00	100.00	0.00	100.00
	40	98.77	100.00	99.99	0.00	99.99	99.60	100.00	100.00	0.00	99.99
200	2	94.99	99.99	99.98	0.00	99.83	94.81	99.98	99.98	0.00	99.83
	3	95.16	99.94	99.93	0.00	99.86	95.41	99.96	99.94	0.00	99.82
	4	95.11	99.94	99.93	0.00	99.79	96.14	99.98	99.97	0.00	99.84
	5	95.11	99.95	99.91	0.00	99.81	96.59	99.98	99.98	0.00	99.88
	10	95.58	99.95	99.95	0.00	99.90	98.47	99.98	99.99	0.00	99.96
	15	95.74	99.97	99.95	0.00	99.92	99.20	100.00	100.00	0.00	100.00
	20	96.18	99.97	99.96	0.00	99.93	99.66	100.00	100.00	0.00	100.00
	30	96.27	99.98	99.97	0.00	99.96	99.84	100.00	100.00	0.00	100.00
	40	97.19	99.95	99.93	0.00	99.92	99.98	100.00	100.00	0.00	100.00

Table 4. Characteristics of AR and ARS projection-based confidence sets – $1 \leq C_{ij} \leq 5$

		AR					ARS				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
50	2	95.14	98.71	98.50	0.00	94.21	94.16	98.39	98.20	0.00	93.20
	3	95.08	97.99	98.08	0.00	92.16	93.86	97.40	97.62	0.00	90.05
	4	95.03	97.70	98.10	0.02	90.67	93.41	96.89	97.21	0.03	87.87
	5	95.10	97.47	87.48	0.11	70.46	93.36	96.36	83.78	0.2	64.22
	10	95.24	96.82	57.95	0.59	34.75	91.52	94.34	46.57	1.40	24.35
	15	95.05	96.38	12.27	1.72	4.10	89.37	91.89	5.62	4.49	1.42
	20	95.20	96.33	28.75	1.47	15.85	86.90	89.30	13.13	5.72	5.55
	30	95.22	96.08	16.86	2.05	6.55	80.72	83.25	2.66	12.11	0.55
	40	94.74	95.33	46.47	1.70	24.17	68.03	70.18	4.77	22.67	0.86
100	2	94.94	98.65	98.46	0.00	94.03	94.48	98.38	98.31	0.00	93.55
	3	94.98	97.89	98.03	0.00	91.95	94.31	97.62	97.79	0.00	91.15
	4	94.92	97.72	97.40	0.00	90.61	94.06	97.25	96.90	0.00	88.95
	5	95.06	97.63	85.87	0.09	68.36	94.32	97.26	83.93	0.12	65.18
	10	94.76	96.96	22.03	1.16	9.29	93.11	95.75	17.99	1.82	6.98
	15	95.00	96.40	6.02	1.81	2.35	92.75	94.56	4.17	3.03	1.26
	20	94.87	96.32	3.64	2.13	0.81	91.84	93.82	2.02	3.55	0.31
	30	94.71	95.80	3.90	2.39	1.10	89.49	91.44	1.65	5.54	0.40
	40	94.99	95.84	3.82	2.74	0.48	88.40	90.03	1.05	7.14	0.07
200	2	94.90	98.60	98.64	0.00	94.49	94.65	98.58	98.57	0.00	94.25
	3	95.03	98.13	98.19	0.00	92.68	94.71	97.99	98.10	0.00	92.21
	4	95.00	97.86	97.71	0.00	91.20	94.68	97.70	97.57	0.00	90.54
	5	95.02	97.37	88.50	0.08	73.88	94.65	97.10	87.63	0.10	72.76
	10	95.11	97.15	28.79	0.83	14.55	94.40	96.68	26.76	0.96	16.14
	15	94.87	96.50	11.24	1.61	4.63	93.74	95.66	9.57	1.98	3.84
	20	95.00	96.25	11.78	1.80	4.56	93.60	95.29	9.64	2.39	3.47
	30	95.03	96.10	1.04	2.44	0.19	92.91	94.31	0.65	3.59	0.08
	40	95.44	96.33	0.21	2.51	0.01	92.66	94.01	0.09	4.07	0.01

Table 4 (continued). Characteristics of LR and LM projection-based confidence sets – $1 \leq C_{ij} \leq 5$

		LM					LR				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
50	2	95.17	98.71	98.50	0.00	94.21	94.16	98.38	98.20	0.00	93.17
	3	97.38	99.17	98.16	0.00	94.78	95.46	98.20	98.12	0.00	92.36
	4	98.62	99.52	98.27	0.00	94.96	96.10	98.22	98.57	0.00	91.99
	5	99.34	99.78	88.69	0.00	80.11	97.75	99.08	92.09	0.00	78.53
	10	100.00	100.00	64.74	0.00	52.91	99.75	99.87	80.05	0.00	61.17
	15	100.00	100.00	21.07	0.00	13.61	99.97	99.99	44.22	0.00	27.38
	20	100.00	100.00	53.08	0.00	45.17	99.97	99.97	68.65	0.00	53.63
	30	100.00	100.00	87.30	0.00	82.73	99.98	99.99	55.57	0.00	38.66
	40	100.00	100.00	100.00	0.00	100.00	99.78	99.85	59.82	0.00	40.39
100	2	94.94	98.65	98.46	0.00	94.03	94.48	98.38	98.31	0.00	93.55
	3	97.18	99.00	98.08	0.00	94.46	95.66	98.31	98.50	0.00	93.09
	4	98.28	99.35	97.48	0.00	94.29	96.80	98.63	98.30	0.00	93.14
	5	99.35	99.81	86.37	0.00	76.28	98.46	99.37	92.18	0.00	79.41
	10	99.92	100.00	23.82	0.00	14.40	99.92	99.97	52.26	0.00	33.90
	15	100.00	100.00	7.74	0.00	4.39	99.99	99.99	36.95	0.00	23.87
	20	100.00	100.00	5.64	0.00	2.54	100.00	100.00	36.80	0.00	19.25
	30	100.00	100.00	7.88	0.00	4.64	100.00	100.00	53.71	0.00	36.68
	40	100.00	100.00	11.78	0.00	6.84	100.00	100.00	62.24	0.00	40.59
200	2	94.90	98.60	98.64	0.00	94.49	94.65	98.58	98.56	0.00	94.25
	3	97.12	99.05	98.20	0.00	94.98	96.03	98.55	98.66	0.00	94.09
	4	98.20	99.41	97.72	0.00	94.73	96.89	98.78	98.80	0.00	94.42
	5	99.37	99.86	88.75	0.00	80.45	98.33	99.39	93.98	0.00	84.38
	10	99.93	99.99	29.84	0.00	20.04	99.97	99.98	61.02	0.00	43.69
	15	100.00	10.00	12.28	0.00	7.64	99.99	100.00	50.27	0.00	35.27
	20	100.00	100.00	13.21	0.00	7.95	100.00	100.00	62.28	0.00	44.82
	30	100.00	100.00	1.42	0.00	0.65	100.00	100.00	40.73	0.00	23.27
	40	100.00	100.00	0.34	0.00	0.09	100.00	100.00	31.24	0.00	16.00

Table 5. Characteristics of AR and ARS projection-based confidence sets – $10 \leq C_{ij} \leq 20$

		AR					ARS				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
50	2	95.14	98.52	0.00	0.00	0.00	94.02	98.16	0.00	0.00	0.00
	3	94.92	97.89	0.00	0.59	0.00	93.74	97.34	0.00	0.69	0.00
	4	95.11	97.78	0.00	0.75	0.00	93.33	96.89	0.00	1.19	0.00
	5	95.14	97.45	0.00	1.21	0.00	93.29	96.29	0.00	1.67	0.00
	10	94.99	96.53	0.00	2.17	0.00	91.29	93.65	0.00	4.13	0.00
	15	94.80	96.19	0.00	2.85	0.00	88.68	91.24	0.00	6.43	0.00
	20	94.89	95.93	0.00	3.13	0.00	86.50	88.70	0.00	9.13	0.00
	30	94.96	95.79	0.00	3.56	0.00	80.43	82.38	0.00	15.78	0.00
	40	95.27	95.70	0.00	3.84	0.00	69.38	70.94	0.00	27.43	0.00
100	2	95.10	98.56	0.00	0.00	0.00	94.65	98.44	0.00	0.00	0.00
	3	95.02	97.81	0.00	0.48	0.00	94.52	97.60	0.00	0.58	0.00
	4	95.22	97.91	0.00	0.73	0.00	94.47	97.51	0.00	0.91	0.00
	5	94.69	97.11	0.00	1.23	0.00	93.76	96.57	0.00	1.51	0.00
	10	95.08	96.77	0.00	1.92	0.00	93.53	95.57	0.00	2.70	0.00
	15	94.68	96.12	0.00	2.59	0.00	92.17	94.25	0.00	3.99	0.00
	20	95.34	96.35	0.00	2.57	0.00	92.17	93.97	0.00	4.45	0.00
	30	94.72	95.78	0.00	3.37	0.00	89.85	91.50	0.00	6.88	0.00
	40	94.71	95.55	0.00	3.74	0.00	87.60	89.07	0.00	9.29	0.00
200	2	94.67	98.60	0.00	0.00	0.00	94.43	98.54	0.00	0.00	0.00
	3	95.05	98.11	0.00	0.41	0.00	94.76	97.96	0.00	0.49	0.00
	4	95.04	97.56	0.00	0.93	0.00	94.57	97.42	0.00	1.02	0.00
	5	95.15	97.49	0.00	1.10	0.00	94.75	97.27	0.00	1.23	0.00
	10	94.96	96.83	0.00	1.88	0.00	94.22	96.28	0.00	2.23	0.00
	15	94.59	96.20	0.00	2.62	0.00	93.52	95.29	0.00	3.24	0.00
	20	94.87	96.10	0.00	2.96	0.00	93.53	94.93	0.00	3.66	0.00
	30	95.17	96.10	0.00	2.92	0.00	93.18	94.40	0.00	4.12	0.00
	40	95.34	96.15	0.00	3.27	0.00	92.53	93.17	0.00	5.20	0.00

Table 5 (continued). Characteristics of LR and LM projection-based confidence sets – $10 \leq C_{ij} \leq 20$

		LM					LR				
T	k	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = $\mathbb{R}$
50	2	95.15	98.53	0.00	0.00	0.00	93.98	98.14	0.00	0.00	0.00
	3	98.14	99.45	0.00	0.00	0.00	97.40	99.26	0.00	0.00	0.00
	4	99.30	99.84	0.00	0.00	0.00	98.75	99.64	0.00	0.00	0.00
	5	99.69	99.97	0.00	0.00	0.00	99.30	99.86	0.00	0.00	0.00
	10	99.99	100.00	0.00	0.00	0.00	99.99	100.00	0.00	0.00	0.00
	15	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	20	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	30	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	40	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
100	2	95.10	98.56	0.00	0.00	0.00	94.64	98.44	0.00	0.00	0.00
	3	98.08	99.39	0.00	0.00	0.00	97.72	99.28	0.00	0.00	0.00
	4	99.19	99.81	0.00	0.00	0.00	98.97	99.77	0.00	0.00	0.00
	5	99.65	99.89	0.00	0.00	0.00	99.44	99.84	0.00	0.00	0.00
	10	99.99	100.00	0.00	0.00	0.00	99.99	100.00	0.00	0.00	0.00
	15	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	20	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	30	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	40	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
200	2	94.67	98.60	0.00	0.00	0.00	94.43	98.54	0.00	0.00	0.00
	3	97.98	99.45	0.00	0.00	0.00	97.79	99.39	0.00	0.00	0.00
	4	99.11	99.76	0.00	0.00	0.00	98.96	99.74	0.00	0.00	0.00
	5	99.71	99.95	0.00	0.00	0.00	99.60	99.91	0.00	0.00	0.00
	10	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	15	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	20	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	30	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00
	40	100.00	100.00	0.00	0.00	0.00	100.00	100.00	0.00	0.00	0.00

Table 6. Comparison between AR and LR projection-based confidence sets when they are bounded

		$1 \leq C_{ij} \leq 5$			$10 \leq C_{ij} \leq 20$		
T	k_2	AR shorter than LR (%)	CI mean length AR	CI mean length LR	AR shorter than LR (%)	CI mean length AR	CI mean length LR
50	2	0.00	9.80	13.28	0.00	0.53	0.51
	3	38.65	25.85	15.59	45.37	0.43	0.45
	4	59.68	20.89	31.69	68.54	0.59	0.65
	5	71.75	82.79	62.85	80.47	0.49	0.57
	10	91.32	17.96	23.62	95.65	0.44	0.58
	15	96.24	6.83	11.22	97.39	0.35	0.49
	20	94.14	16.07	17.61	97.59	0.35	0.51
	30	87.98	7.30	14.94	93.66	0.35	0.51
	40	53.66	13.66	11.54	67.12	0.49	0.59
100	2	0.00	13.05	12.88	0.00	0.62	0.61
	3	44.21	16.37	15.93	59.74	0.49	0.52
	4	69.88	17.00	23.77	82.57	0.58	0.66
	5	85.97	16.48	16.16	92.01	0.43	0.50
	10	99.20	6.04	14.87	99.65	0.36	0.48
	15	99.79	4.71	10.85	99.92	0.28	0.40
	20	100.00	4.78	23.20	99.98	0.33	0.50
	30	99.96	3.85	31.25	100.00	0.28	0.46
	40	100.00	8.67	17.75	100.00	0.27	0.47
200	2	0.00	13.59	43.78	0.00	0.53	0.52
	3	56.82	33.94	18.59	70.54	0.49	0.52
	4	88.33	41.99	259.61	91.35	0.55	0.62
	5	95.67	21.27	15.42	96.71	0.40	0.47
	10	99.86	7.82	14.02	99.93	0.32	0.43
	15	100.00	7.90	14.17	100.00	0.28	0.40
	20	100.00	5.30	24.65	100.00	0.23	0.35
	30	100.00	2.14	11.72	100.00	0.24	0.41
	40	100.00	1.61	20.78	100.00	0.24	0.43

Table 7. Power of tests induced by projection-based confidence sets – $H_0 : \beta_1 = 0\,; T = 100$

	$k_2 = 2$				$k_2 = 4$				$k_2 = 10$				$k_2 = 20$			
β_1	AR	ARS	LM	LR	AR	ARS	LM	LR	AR	ARS	LM	LR	AR	ARS	LM	LR
-0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.0	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.1	0.95	0.95	0.95	0.95	0.97	0.97	0.94	0.95	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
0.2	0.76	0.77	0.76	0.77	0.78	0.80	0.67	0.70	0.96	0.97	0.74	0.82	1.00	1.00	0.91	0.98
0.3	0.38	0.39	0.38	0.39	0.38	0.40	0.23	0.26	0.58	0.63	0.15	0.23	0.87	0.91	0.12	0.33
0.4	0.09	0.10	0.09	0.10	0.09	0.10	0.02	0.03	0.12	0.15	0.00	0.01	0.21	0.28	0.00	0.00
0.5	0.01	0.01	0.01	0.01	0.02	0.03	0.00	0.00	0.04	0.05	0.00	0.00	0.04	0.07	0.00	0.00
0.6	0.09	0.09	0.09	0.09	0.08	0.09	0.03	0.03	0.11	0.14	0.00	0.01	0.18	0.26	0.00	0.00
0.7	0.35	0.36	0.35	0.36	0.29	0.32	0.188	0.19	0.48	0.54	0.12	0.14	0.75	0.82	0.10	0.18
0.8	0.71	0.72	0.71	0.72	0.63	0.66	0.52	0.52	0.88	0.90	0.61	0.62	0.99	0.99	0.76	0.82
0.9	0.91	0.91	0.91	0.91	0.87	0.89	0.82	0.81	0.99	0.99	0.93	0.93	1.00	1.00	0.99	1.00
1.0	0.98	0.98	0.98	0.98	0.97	0.97	0.95	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.1	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

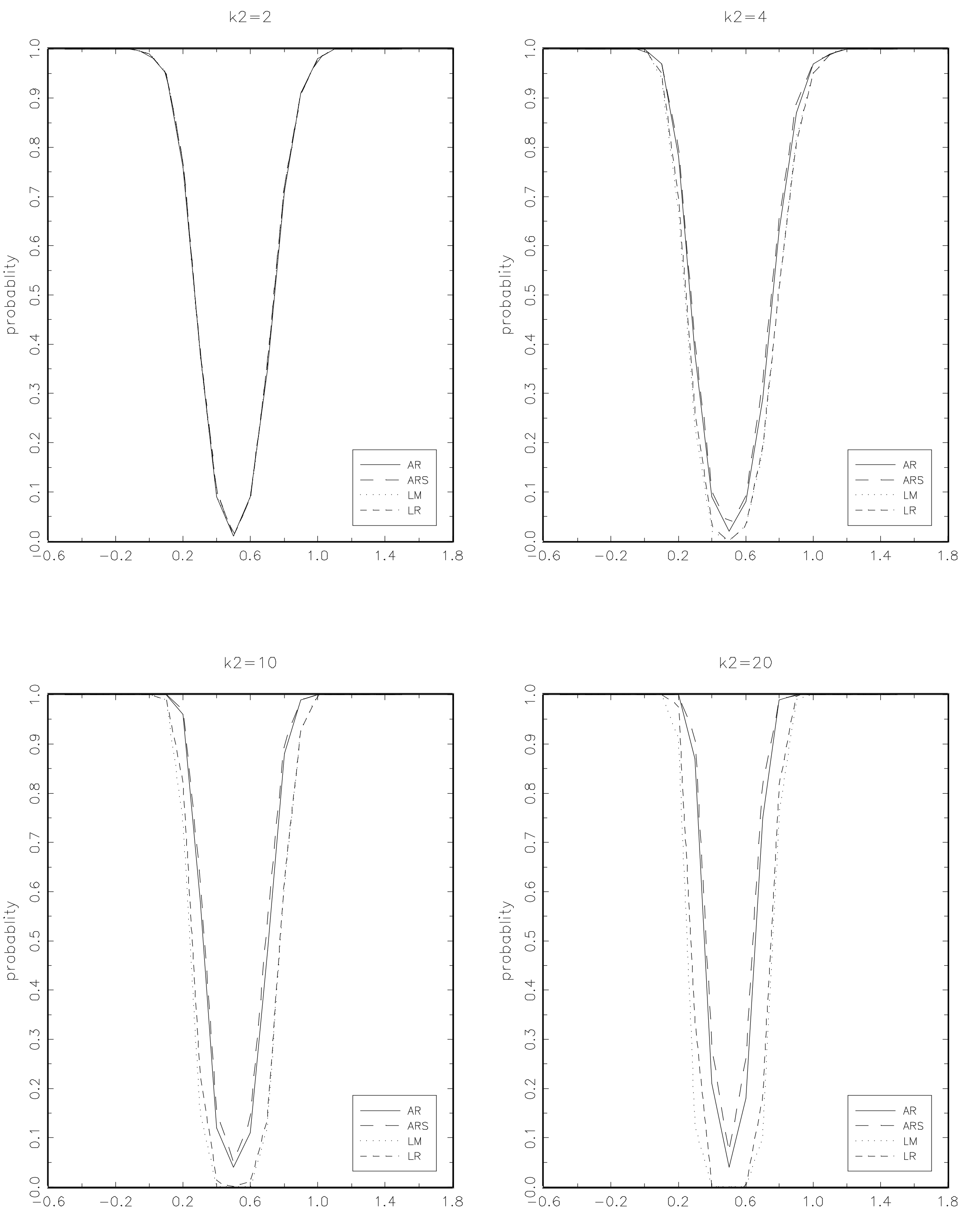

Figure 1. Power of tests induced by projection-based confidence sets
$H_0 : \beta_1 = 0.5$

based confidence sets become less conservative as the number of relevant instruments increases. This suggests using of a number of relevant instruments as large as possible. But on the other hand, as noted by Dufour and Taamouti (2001*b*) and Kleibergen (2002), a large number of instruments will induce loss of power for the Anderson-Rubin test for β.

The proportions of unbounded confidence sets and confidence sets equal to the real line are nearly zero when identification holds (Table 5). When we approach nonidentification (tables 4 and 3), these proportions become large but decrease as the number of instruments increases. This is predictable according to the results in Dufour (1997). It is natural when the components of Π_2 approach 0 to get an unbounded confidence set, for β is not identified in this case and the set of possible values is large.

The ARS test behaves in the same way as AR, except when the sample size is small with respect to the number of instruments. In this case we observe a size distortion, in the sense that the empirical coverage rate for β becomes smaller than the nominal level (95%).

For the LR and LM tests, the main observation is that they produce confidence sets much more conservative than those based on AR or ARS, and unlike the AR test, the conservative character of the resulting confidence sets increases with the number of instruments k_2. The coverage rate of the confidence sets based on the LM and LR statistics are always greater than 98.5% and approaches rapidly 100% as k_2 increases. This is predictable since the LM and LR based confidence sets are doubly conservative, by majorization of their distribution and by projection. Even in the strongly identified case, the LR test exhibits a downward size distortion.

In Table 6, we report comparisons between alternative confidence sets from the viewpoint of their length (in identified cases, conditional on getting a bounded interval). We see from these results that AR-based confidence sets tend to be shorter than confidence sets based on the LR statistic. This may be due that the latter procedure is based on a conservative critical value even when the full β vector is tested.

As we may expect the high coverage rate of the LM and LR-based confidence sets induces power loss for the test that rejects $H_0 : \beta_1 = \beta_1^0$ when the projection-based confidence set for β_1 excludes β_1^0. This is shown in Table 7 and Figure 1 where we present estimates of $\mathsf{P}[$rejecting $H_0 : \beta_1 = 0.5 \,|\, \beta_1 = \beta_1^i]$ with a decision rule consisting of rejecting H_0 if 0.5 is excluded from the confidence set for β_1. The theoretical size is 95%. The value of the alternative varies from -0.5 to 1.5 with increments of 0.1.We see from these results that, for $k_2 = 2$, the three tests have the same power. But, as k_2 increases, the LM and LR based tests are undersized and exhibit less power.

7.2. The effect of instrument exclusion

In this subsection, 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. We consider the statistics AR, ARS, LM, and LR described above, to which we add Kleibergen's (2002) K-test and the two versions of the conditional LR test (LR1 and LR2) of Moreira (2003*a*). The reduced form equation (7.2) is then replaced by

$$(Y_1,\, Y_2) = X_2\Pi_2 + X_3\delta + (V_1,\, V_2),$$

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. Two cases were considered: (a) both X_2 and X_3 are kept fixed over the simulation experiment; (b) X_2 and X_3 are regenerated at each replication (random missing instruments). 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 number of replications is $N = 1000$ and the conditional LR critical values are computed using the same number of replications. The matrix C is set equal to: $C = \rho\, \Pi$ 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. k_2 is the number of instruments.

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 8 and 9.

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.

8. Empirical illustrations

In this section we illustrate the statistical inference methods discussed in the previous sections through three 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, the second one considers the widely studied problem of returns to education, and the third application is about the returns to scale and externality spillovers in U.S. industry.

8.1. 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, Romer and Weil (1992) and the survey of Rodrik (1995). Despite the great effort that has been devoted to studying this issue,

Table 8. Instrument exclusion and the size of tests robust to weak instruments.
Nominal size $= 0.05$. Results are given in percentages.

	AR	ARS	K	LM	LR	LR1	LR2	AR	ARS	K	LM	LR	LR1	LR2
k_2	(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

Table 9. Instrument exclusion and the size of tests robust to weak instruments
Random missing instruments
Nominal size $= 0.05$. Results are given in percentages.

	AR	ARS	K	LM	LR	LR1	LR2	AR	ARS	K	LM	LR	LR1	LR2
k_2	(a) $\delta = 0$ and $\rho = 0.01$							(b) $\delta = 0$ and $\rho = 1$						
2	5.0	5.2	5.2	4.8	5.1	5.1	5.2	5.5	5.9	5.9	5.0	5.8	5.8	5.9
3	3.8	4.6	5.6	3.5	3.6	4.5	4.5	5.0	6.2	5.6	2.0	1.7	5.8	5.8
4	5.4	5.7	5.7	4.9	4.1	5.4	5.6	4.8	5.6	5.5	1.3	1.1	5.6	5.5
5	6.6	7.7	5.9	5.6	3.9	7.4	7.7	4.3	5.0	4.6	0.4	0.4	4.9	5.1
10	4.3	5.6	6.0	4.1	1.7	6.0	6.2	4.2	5.6	4.6	0.0	0.0	4.2	4.3
20	5.5	9.0	8.4	3.0	0.5	9.1	9.2	4.9	7.7	4.8	0.0	0.0	5.3	5.5
40	4.8	12.4	16.5	0.9	0.0	14.6	14.9	4.1	11.0	5.8	0.0	0.0	6.3	6.2
	(c) $\delta = 1$ and $\rho = 0.01$							(d) $\delta = 1$ and $\rho = 1$						
2	4.9	5.5	5.5	4.9	5.3	5.3	5.5	4.4	4.8	4.8	4.2	4.8	4.8	4.8
3	5.0	5.5	7.4	4.6	5.3	5.7	5.7	4.4	4.9	5.1	1.8	2.5	5.0	5.0
4	5.0	5.7	11.5	4.5	5.7	5.8	5.9	5.2	6.3	4.7	0.6	0.8	4.6	4.7
5	5.4	6.3	15.7	4.7	5.9	6.6	6.7	5.1	6.2	5.2	0.4	0.8	5.7	6.0
10	4.9	7.2	34.5	3.8	7.7	8.0	7.8	4.8	6.7	6.4	0.1	0.1	6.6	6.7
20	4.7	7.2	56.9	2.9	9.3	10.7	7.8	4.8	7.7	6.6	0.0	0.0	6.7	7.0
40	4.2	11.8	77.3	1.0	29.8	33.5	12.9	5.3	12.5	11.9	0.0	0.0	14.4	15.6
	(e) $\delta = 10$ and $\rho = 0.01$							(f) $\delta = 10$ and $\rho = 1$						
2	4.4	4.7	4.7	4.2	4.5	4.5	4.7	5.0	5.4	5.4	4.9	5.2	5.2	5.4
3	4.3	4.4	9.6	4.0	4.4	4.6	4.8	4.8	5.6	5.0	1.8	4.6	6.1	6.3
4	3.3	3.9	15.9	3.1	3.8	3.9	4.0	5.0	6.0	6.6	0.8	5.2	6.1	6.4
5	5.3	5.7	28.9	4.6	5.6	5.8	5.9	4.4	4.9	6.1	0.4	4.4	5.2	5.5
10	5.2	7.0	74.7	4.2	7.5	8.0	7.6	5.0	6.7	15.0	0.1	6.0	7.8	7.4
20	5.1	7.9	94.6	2.6	11.7	12.5	8.9	4.5	7.1	39.8	0.0	8.9	10.7	7.7
40	5.0	10.8	97.9	0.7	33.5	36.2	12.8	5.2	12.4	73.6	0.0	30.5	34.7	14.1

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

$$y_i = a + bT_i + c_1 N_i + c_2 A_i + u_i \tag{8.1}$$

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, equation (6), page 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 are. The F-statistic of the first stage regression

$$T_i = \alpha + \beta Z_i + \gamma_1 N_i + \gamma_2 A_i + \varepsilon_i \tag{8.2}$$

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

To draw inference on the coefficients of the structural equation (8.1), we can use the Anderson-Rubin 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 (8.1), 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:

$$C_b(\alpha) = \{b : 0.963b^2 - 4.754b + 1.274 \leq 0\} = [0.284\,,\ 4.652]\,. \tag{8.3}$$

The p-value of the Anderson-Rubin 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 $\left[\hat{b}_{IV} - 2\hat{\sigma}_{\hat{b}_{IV}},\, \hat{b}_{IV} + 2\hat{\sigma}_{\hat{b}_{IV}}\right] =$

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

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

Table 10. 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)	$\theta' \begin{pmatrix} 1.78 & -16.36 \\ -16.36 & 257.85 \end{pmatrix} \theta + \begin{pmatrix} -2.23\,, & -34.50 \end{pmatrix} \theta + 0.19 \leq 0$
(b, c_2)	$\theta' \begin{pmatrix} 3.83 & -34.58 \\ -34.58 & 386.87 \end{pmatrix} \theta + \begin{pmatrix} -10.6\,, & 69.17 \end{pmatrix} \theta + 2.13 \leq 0$
(b, a)	$\theta' \begin{pmatrix} 38.41 & 33.34 \\ 33.35 & 29.52 \end{pmatrix} \theta + \begin{pmatrix} -611.55\,, & -537.47 \end{pmatrix} \theta + 2445.58 \leq 0$

B. Projection-based individual confidence intervals (confidence level $\geq 95\%$)

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	$[2.09\,,\ 9.38]$	$[0.56\,,\ 9.36]$

$[-0.01\,,\ 3.95]$, which is not very different from the AR-based confidence set. In particular, in contrast with $C_b(\alpha)$ in (8.3), 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 (8.1) are given in Table 10A. 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 10B. 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 section 7.1), but if we compare them to those obtained from t-statistics, they are not really larger.

8.2. 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, Imbens and Krueger (1999), 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 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 329509 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 $(\beta_1,\, \gamma_1)$ and then deduce by projection a confidence set for γ_1. Since the instruments are weak, we expect large, if not completely uninformative, intervals. Table 11 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 $W\_IC_\alpha(\beta_1) = [0.0031,\, 0.1167]$.

Table 11. 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	$\mathbb{R}$	$\mathbb{R}$	$[-0.111, -0.075]$

8.3. Returns to scale and externality spillovers in U.S. industry

One of the widely studied problems in recent macroeconomics literature is the extent of returns to scale and externalities in the U.S. industry. Recent work on these issues includes Hall (1990), Caballero and Lyons (1989, 1992), Basu and Fernald (1995, 1997) and Burnside (1996). The results of these researches and many others have important implications on many fields of macroeconomics, such as growth and business cycle models.

Burnside (1996) presents a short survey of different specifications of the production function adopted in this literature. One of these specifications considers the following equation:

$$Y_{it} = F(K_{it}, L_{it}, E_{it}, M_{it}) \tag{8.4}$$

where, for each industry i and each period t, Y_{it} is the gross output, K_{it} is the amount of capital services used, L_{it} is the amount of labor, E_{it} is energy used, and M_{it} is the quantity of materials. If we assume that F is a differentiable function and homogeneous of degree ρ, we get the following regression equation [see Burnside (1996)]:

$$\Delta y_{it} = \rho \Delta x_{it} + \Delta a_{it} \tag{8.5}$$

where Δy_{it} is the growth rate of the output, Δx_{it} is a weighted average of the inputs and Δa_{it} represents technological changes.[11] In this specification, ρ is the coefficient that measures the extent of returns to scale. Returns to scale are increasing, constant or decreasing depending on whether $\rho > 1$, $\rho = 1$ or $\rho < 1$.

To identify simultaneously the effects of externalities between industries, Caballero and Lyons (1992) added to the previous regression equation the aggregated industrial output as a measure of this effect. Burnside (1996) suggested a variable based on inputs rather than output, arguing by the fact that the first measure may induce spurious externalities for industries with a large output. Adopting the later suggestion, the previous regression equation becomes:

$$\Delta y_{it} = \rho \Delta x_{it} + \eta \Delta x_t + u_{it} \tag{8.6}$$

where Δx_t is the cost shares weighted average of the Δx_{it} [Burnside (1996, equation (2.8))] and $u_{it} = \Delta a_{it}$. The coefficient η measures the externalities effect.

To estimate this equation, Hall (1990) proposed a set of instruments that was used in most subsequent researches. These instruments include the growth rate of military purchases, the growth rate of world oil price, a dummy variable representing the political party of the President of Unites States and one lag of each of these variables. Estimation methods used include ordinary least squares, two stages least squares and three stages least squares.

The regressions are performed using panel data on two-digit SIC (Standard Industrial Classification) code level manufacturing industries. This classification includes 21 industries. The data set is described in detail by Jorgenson, Gollop and Fraumeni (1987) and contains information on gross output, labor input, stock of capital, energy use, and materials inputs.

These regressions are interesting as an application for the statistical inference methods developed in this paper because the instruments used appear to be weak and may induce identification problems. These instruments have been studied in detail by Burnside (1996) who showed on the basis of calculations of R^2 and partial R^2 [Shea (1997)], that these instruments are weak. A valid method to draw inference on ρ (returns to scale) and η (externalities) then consists in using an extension of the Anderson-Rubin approach [as suggested in Dufour and Jasiak (2001)] to build a joint confidence set for $(\rho, \eta)'$ and then build through projection individual confidence intervals for ρ and η.[12]

Given this identification problem, we expect unbounded confidence sets. Using the same data set as Burnside (1996), we obtained the results presented in Table 12. This table presents the 2SLS estimates and the confidence sets for the returns to scale coefficients and externalities coefficients in 21 U.S. manufacturing industries over the period 1953-1984. The projection based confidence sets

[11] The weights are the production cost shares of each input.

[12] As reported in Caballero and Lyons (1989), there is no evidence of serial correlation from either the Durbin-Watson statistic or the Ljung-Box Q statistic.

Table 12. Confidence sets for the returns to scale and externality coefficients in different U.S. industries (size $\geq 90\%$)

	Returns to scale		Externalities	
Industry	2SLS	Confidence set	2SLS	Confidence set
7: Food & kindred products	0.99	$\mathbb{R}$	-0.06	$\mathbb{R}$
8: Tobacco	1.06	$\mathbb{R}$	0.28	$\mathbb{R}$
9: Textile mill products	0.61	$]-\infty,\ 0.56] \cup [2.23,\ \infty[$	0.20	$\mathbb{R}$
10: Apparel	1.09	$\varnothing$	-0.05	$\varnothing$
11: Lumber & wood	0.86	$\mathbb{R}$	-0.08	$\mathbb{R}$
12: Furniture and fixtures	1.13	$]-\infty,\ 0.58] \cup [1.77,\ \infty[$	-0.01	$]-\infty,\ -0.73] \cup [0.55,\ \infty[$
13: Paper and allied	0.54	$]-\infty,\ 0.74] \cup [4.56,\ \infty[$	0.61	$]-\infty,\ -4.51] \cup [0.45,\ \infty[$
14: Printing; publishing	0.93	$[-1.2,\ 4.23]$	0.23	$[-0.11,\ 1.05]$
15: Chemicals	0.22	$[-7.36,\ 0.54]$	1.06	$[0.85,\ 11.7]$
16: Petroleum & coal products	0.34	$\mathbb{R}$	0.29	$\mathbb{R}$
17: Rubber & misc. plastics	1.29	$\mathbb{R}$	-0.31	$\mathbb{R}$
18: Leather	0.39	$\mathbb{R}$	0.01	$\mathbb{R}$
19: Stone, clay, glass	1.21	$[1,\ 3.34]$	-0.03	$[-3.16,\ 0.15]$
20: Primary metal	0.79	$[0.46,\ 1.01]$	0.42	$[-0.37,\ 1.51]$
21: Fabricated metal	0.80	$]-\infty,\ 2.25] \cup [1.15,\ \infty[$	0.30	$]-\infty,\ -0.13] \cup [4.21,\ \infty[$
22: Machinery, non-electrical	1.16	$[0.73,\ 1.81]$	0.02	$[-1.41,\ 0.76]$
23: Electrical machinery	1.17	$]-\infty,\ 0.29] \cup [2.47,\ \infty[$	0.05	$]-\infty,\ 1.16] \cup [1.72,\ \infty[$
24: Motor vehicles	1.23	$\mathbb{R}$	-0.12	$\mathbb{R}$
25: Transportation equipment	1.07	$[0.64,\ 1.55]$	0.10	$[-0.36,\ 1.6]$
26: Instruments	1.38	$[1.19, 3.29]$	-0.07	$[-1.5,\ 0.38]$
27: Misc. manufacturing	1.5	$]-\infty,\ -88.7] \cup [0.48,\ \infty[$	-0.51	$]-\infty,\ 0.12] \cup [102.1,\ \infty[$
Mean	0.94		0.11	

are obtained from joint confidence sets for $(\rho,\ \eta)$ of level 90%.[13]

The average estimation over all industries of the coefficients ρ and η are of the same order as those obtained by Burnside (1996).[14] Only 7 among 21 confidence sets are bounded. For industries 19 (stone, clay and glass) and 26 (instruments), the returns to scale are increasing. For industry 15 (chemicals), the returns to scale are decreasing. For industries 9 (textile mill products), 12 (furniture and fixtures), 13 (paper and allied), and 23 (electrical machinery) the hypothesis of constant returns to scale is rejected with a significance level smaller than or equal 10%. For industry 10 (apparel) the confidence set is empty which may be explained by the fact that the data does not support the model. For industries 7 (food and kindred products), 8 (tobacco), 11 (lumber and wood), 16 (petroleum and coal products), 17 (rubber and miscellaneous plastics), 18 (leather), and 24 (motor vehicles), the confidence sets are equal to $\mathbb{R}$ and thus provide no information on ρ and η.

[13] We used χ^2 as asymptotic distribution for the Anderson-Rubin statistic instead of the Fisher distribution valid under normality and independence assumption.

[14] The small differences may be due to the use of TSLS instead of 3SLS.

9. 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. In addition to being completely robust (in finite samples) to the presence of weak instruments, their validity is unaffected by the exclusion of possibly relevant instruments (*robustness to instrument exclusion*), and more generally to the distribution of explanatory endogenous variables (*robustness to endogenous explanatory variable distribution*). More precisely, the finite-sample distribution (under the null hypothesis) 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 (which includes the form of the associated DGP and the disturbance distribution). These features can be quite important and useful from a practical viewpoint. AR-type procedures constitute *limited-information* methods, which typically involve an efficiency loss with respect to *full-information* methods, but do allow for a less complete specification of the model. 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. Finally, we presented simulation results as well as three experimental examples which showed that projection-based AR-type confidence sets are indeed quite easy to implement and perform reasonably well in terms of accuracy.

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 (2001*b*, 2001*a*)]. 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 (2002*a*, 2002*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)], nonlinear 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 (2001*b*)].

A. Appendix: Proofs

PROOF OF THEOREM **5.1** To simplify the notation, we write $C_{\delta_1} \equiv C_{w'\theta}$ as in (5.5). **(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) $1 \leq r_2 < p - 1$.
(a.1) If $r_2 = p - 1$, $\bar{A}_{22}$ is positive definite. From (5.7), 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:

$$\frac{\partial \bar{Q}}{\partial \delta_2} = 2\bar{A}_{22}\delta_2 + 2\bar{A}_{21}\delta_1 + \bar{b}_2 = 0\,, \quad \frac{\partial^2 \bar{Q}}{\partial \delta_2 \partial \delta_2'} = 2\bar{A}_{22}\,. \tag{A.1}$$

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

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

On setting $\delta_2 = \tilde{\delta}_2$ in $\bar{Q}(\delta_1, \delta_2)$, we get (after some algebra) the minimal value:

$$\bar{Q}(\delta_1, \tilde{\delta}_2) = \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \tag{A.3}$$

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 (5.12) holds with $S_1 = \emptyset$.
(a.2) If $1 \leq r_2 < p - 1$, we get, using (5.7) and (5.9) - (5.11):

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

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:

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \tilde{\delta}_{2*}'D_{2*}\tilde{\delta}_{2*} + [2\tilde{A}_{21*}\delta_1 + \tilde{b}_{2*}]'\tilde{\delta}_{2*}\,. \tag{A.4}$$

By an argument similar to the one used for (a.1), we can see that

$$\delta_1 \in C_{\delta_1} \text{ iff } \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0 \tag{A.5}$$

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)]:

$$\bar{A}_{22}^{+} = P_2\begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0 \end{bmatrix}P_2' = [P_{21},\, P_{22}]\begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0 \end{bmatrix}[P_{21},\, P_{22}]' = P_{21}D_{2*}^{-1}P_{21}'\,, \tag{A.6}$$

hence

$$\bar{A}'_{21*}D_{2*}^{-1}\bar{A}_{21*} = \bar{A}'_{21}P_{21}D_{2*}^{-1}P'_{21}\bar{A}_{21} = \bar{A}'_{21}\bar{A}_{22}^{+}\bar{A}_{21}\,, \tag{A.7}$$

$$\bar{A}'_{21*}D_{2*}^{-1}\bar{b}_{2*} = \bar{A}'_{21}P_{21}D_{2*}^{-1}P'_{21}\bar{b}_{2} = \bar{A}'_{21}\bar{A}_{22}^{+}\bar{b}_{2}\,, \tag{A.8}$$

$$\bar{b}'_{2*}D_{2*}^{-1}\bar{b}_{2*} = \bar{b}'_{2}P_{21}D_{2*}^{-1}P'_{21}\bar{b}_{2} = \bar{b}'_{2}\bar{A}_{22}^{+}\bar{b}_{2}\,. \tag{A.9}$$

(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

$$\begin{aligned} C_{\delta_1} &= \{\delta_1 : P'_{22}(2\bar{A}_{21}\delta_1 + \bar{b}_2) = 0 \text{ and } \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0\} \cup \{\delta_1 : P'_{22}(2\bar{A}_{21}\delta_1 + \bar{b}_2) \neq 0\} \\ &= \{\delta_1 : \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0\} \cup \{\delta_1 : P'_{22}(2\bar{A}_{21}\delta_1 + \bar{b}_2) \neq 0\} \end{aligned} \tag{A.10}$$

and (5.12) 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:

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

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 $\bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c \leq 0$. 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 $\varDelta_0$, we have:

$$\bar{Q}(\delta_1,\, \varDelta_0\delta_{20}) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \varDelta_0^2\, q_0 + \varDelta_0[2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_{20}\,. \tag{A.12}$$

Since $q_0 < 0$, we can choose $\varDelta_0$ sufficiently large to have $\bar{Q}(\delta_1,\, \varDelta_0\delta_{20}) < 0$, 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

Abdelkhalek, T. and Dufour, J.-M. (1998), 'Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy', *Review of Economics and Statistics* **LXXX**, 520–534.

Anderson, T. W. and 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, D. W. K., Moreira, M. J. and Stock, J. H. (2004), Optimal invariant similar tests for instrumental variables regression, Technical report, Cowles Foundation for Research in Economics, Yale University and Department of Economics of Harvard University, Harvard University, New Haven, Connecticut.

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

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

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

Basu, S. and Fernald, J. G. (1995), 'Are apparent productive spillovers a figment of specification error?', *Journal of Monetary Economics* **36**, 165–188.

Basu, S. and Fernald, J. G. (1997), 'Returns to scale in U.S. production : Estimates and implications', *Journal of Political Economy* **105**(2), 249–283.

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

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

Bound, J., Jaeger, D. A. and Baker, R. (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., Jaeger, D. A. and 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.

Burnside, C. (1996), 'Production function regressions, returns to scale, and externalities', *Journal of Monetary Economics* **37**, 177–201.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Caballero, R. J. and Lyons, R. K. (1989), The role of external economies in US manufacturing, Technical Report 3033, National Bureau of Economic Research, Cambridge, Massachusetts.

Caballero, R. J. and Lyons, R. K. (1992), 'External effects in U.S. procyclical productivity', *Journal of Monetary Economics* **29**, 209–225.

Chao, J. and Swanson, N. R. (2000), Bias and MSE of the IV estimators under weak identification, Technical report, Department of Economics, University of Maryland.

Choi, I. and 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. and Donald, S. G. (1993), 'Testing identifiability and specification in instrumental variable models', *Econometric Theory* **9**, 222–240.

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

Donald, S. G. and 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. and Jasiak, J. (1993), Finite sample inference methods for simultaneous equations and models with unobserved and generated regressors, Technical report, C.R.D.E., Université de Montréal. 38 pages.

Dufour, J.-M. and 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. and Taamouti, M. (2001*a*), On methods for selecting instruments, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2001*b*), Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**(4), 1351–1365.

Fisher, F. M. (1976), *The Identification Problem in Econometrics*, Krieger Publishing Company, Huntington (New York).

Forchini, G. and Hillier, G. (2003), 'Conditional inference for possibly unidentified structural equations', *Econometric Theory* **19**(5), 707–743.

Frankel, J. A. and Romer, D. (1996), Trade and growth: An empirical investigation, Technical Report 5476, National Bureau of Economic Research, Cambridge, Massachusetts.

Frankel, J. A. and Romer, D. (1999), 'Does trade cause growth?', *The American Economic Review* **89**(3).

Hahn, J. and Hausman, J. (2002*a*), 'A new specification test for the validity of instrumental variables', *Econometrica* **70**, 163–189.

Hahn, J. and Hausman, J. (2002*b*), 'Notes on bias in estimators for simultaneous equation models', *Economics Letters* **75**, 237–241.

Hahn, J., Hausman, J. and Kuersteiner, G. (2001), Higher order MSE of jackknife 2SLS, Technical report, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hall, A. R. and 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. and Wilcox, D. W. (1996), 'Judging instrument relevance in instrumental variables estimation', *International Economic Review* **37**, 283–298.

Hall, R. E. (1990), Invariance properties of Solow's productivity residual, *in* P. Diamond, ed., 'Growth, Productivity, Employment', The MIT Press, Cambridge, Massachusetts, pp. 71–112.

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-Verlag, New York.

Hsiao, C. (1983), Identification, *in* Z. Griliches and M. D. Intrilligator, eds, 'Handbook of Econometrics, Volume 1', North-Holland, Amsterdam, chapter 4, pp. 223–283.

Irwin, A. D. and Tervio, M. (2002), 'Does trade raise income? evidence from the Twentieth Century', *Journal of International Economics* **58**, 1–18.

Jorgenson, D. W., Gollop, F. M. and Fraumeni, B. M. (1987), *Productivity and U.S. Economic Growth*, Harvard University Press, Cambridge, Massachusetts.

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* **forthcoming**.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses, 2nd edition*, John Wiley & Sons, New York.

Maddala, G. S. and Jeong, J. (1992), 'On the exact small sample distribution of the instrumental variable estimator', *Econometrica* **60**, 181–183.

Magnus, J. R. and Neudecker, H. (1991), *Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition*, John Wiley & Sons, New York.

Mankiw, G. N., Romer, D. and Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics* **107**(2), 407–37.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge and London.

Mansky, C. (2003), *Partial Identification of Probability Distributions*, Springer Series in Statistics, Springer-Verlag, New York.

Miller, Jr., R. G. (1981), *Simultaneous Statistical Inference*, second edn, Springer-Verlag, New York.

Moreira, M. J. (2001), Tests with correct size when instruments can be arbitrarily weak, Technical report, Department of Economics, Harvard University, Cambridge, Massachusetts.

Moreira, M. J. (2003*a*), 'A conditional likelihood ratio test for structural models', *Econometrica* **71**(4), 1027–1048.

Moreira, M. J. (2003*b*), A general theory of hypothesis testing in the simultaneous equations model, Technical report, Department of Economics, Harvard University, Cambridge, Massachusetts.

Moreira, M. J. and Poi, B. P. (2001), 'Implementing tests with correct size in the simultaneous equations model', *The Stata Journal* **1**(1), 1–15.

Nelson, C. R. and Startz, R. (1990*a*), '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. and Startz, R. (1990*b*), 'Some further results on the exact small properties of the instrumental variable estimator', *Econometrica* **58**, 967–976.

Perron, B. (1999), Semi-parametric weak instrument regressions with an application to the risk return trade-off, Technical Report 0199, C.R.D.E., Université de Montréal.

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. and Marcoantonio, M. L. (1970), *Analytic Geometry with Vectors*, Scott, Fosman and Company, Glenview, Illinois.

Prakasa Rao, B. L. S. (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Academic Press, New York.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Rao, C. R. and Mitra, S. K. (1971), *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, New York.

Rodrik, D. (1995), Trade and industrial policy reform, *in* J. Behrman and T. N. Srinivasan, eds, 'Handbook of Development Economics', Vol. 3A, Elsevier Science, Amsterdam.

Rothenberg, T. J. (1971), 'Identification in parametric models', *Econometrica* **39**, 577–591.

Savin, N. E. (1984), Multiple hypothesis testing, *in* Z. Griliches and M. D. Intrilligator, eds, 'Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, chapter 14, pp. 827–879.

Scheffé, H. (1959), *The Analysis of Variance*, John Wiley & Sons, 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, New Jersey.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

Startz, R., Nelson, C. R. and Zivot, E. (1999), Improved inference for the instrumental variable estimator, Technical report, Department of Economics, University of Washington.

Stock, J. H. and Wright, J. H. (2000), 'GMM with weak identification', *Econometrica* **68**, 1097–1126.

Stock, J. H., Wright, J. H. and 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.

Stock, J. H. and Yogo, M. (2002), Testing for weak instruments in linear IV regression, Technical Report 284, N.B.E.R., Harvard University, Cambridge, Massachusetts.

Stock, J. H. and Yogo, M. (2003), Asymptotic distributions of instrumental variables statistics with many weak instruments, Technical report, Department of Economics, Harvard University, Cambridge, Massachusetts.

Wang, J. and 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, Technical Report 732, International Finance Discussion Papers, Board of Governors of the Federal Reserve System, Washington, D.C.

Wright, J. H. (2003), 'Detecting lack of identification in GMM', *Econometric Theory* **19**(2), 322–330.

Zivot, E., Startz, R. and Nelson, C. R. (1998), 'Valid confidence intervals and inference in the presence of weak instruments', *International Economic Review* **39**, 1119–1144.