Cahier 2003-10

Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments

DUFOUR, Jean-Marie
TAAMOUTI, Mohamed

Université
de Montréal

Projection-based statistical inference in linear structural models with possibly weak instruments [*]

Jean-Marie Dufour [†]
Université de Montréal

Mohamed Taamouti [‡]
INSEA and Université de Montréal

First version: August 1999
This version: May 2003

[†] 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

It is well known that standard asymptotic theory is not valid or is extremely unreliable in models with identification problems or weak instruments [Dufour (1997, Econometrica), Staiger and Stock (1997, Econometrica), Wang and Zivot (1998, Econometrica), Stock and Wright (2000, Econometrica), Dufour and Jasiak (2001, International Economic Review)]. One possible way out consists here in using a variant of the Anderson-Rubin (1949, Ann. Math. Stat.) procedure. The latter, however, allows one to build exact tests and confidence sets only for the full vector of the coefficients of the endogenous explanatory variables in a structural equation, which in general does not allow for individual coefficients. This problem may in principle be overcome by using projection techniques [Dufour (1997, Econometrica), Dufour and Jasiak (2001, International Economic Review)]. AR-types are emphasized because they are robust to both *weak instruments* and *instrument exclusion*. However, these techniques can be implemented only by using costly numerical techniques. In this paper, we provide a complete analytic solution to the problem of building projection-based confidence sets from Anderson-Rubin-type confidence sets. The latter involves the geometric properties of "quadrics" and can be viewed as an extension of usual confidence intervals and ellipsoids. Only least squares techniques are required for building the confidence intervals. We also study by simulation how "conservative" projection-based confidence sets are. Finally, we illustrate the methods proposed 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 production functions in the U.S. economy.

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

RÉSUMÉ

L'une des questions les plus étudiés récemment en économétrie est celle des modèles présentant des problèmes de quasi non-identification ou d'instruments faibles. L'une des conséquences importantes de ce problème est la non validité de la théorie asymptotique standard [Dufour (1997, Econometrica), Staiger et Stock (1997, Econometrica), Wang et Zivot (1998, Econometrica), Stock et Wright (2000, Econometrica), Dufour et Jasiak (2001, International Economic Review)]. Le défi majeur dans ce cas consiste à trouver des méthodes d'inférence robustes à ce problème. Une solution possible consiste à utiliser la statistique d'Anderson-Rubin (1949, Ann. Math. Stat.). Nous mettons l'emphase sur les procédures de type Anderson-Rubin, car celles-ci sont robustes tant à la présence d'*instruments faibles* et à l'*exclusion d'instruments*. Cette dernière ne fournit cependant des tests exacts que pour les hypothèses spécifiant le vecteur entier des coefficients des variables endogènes dans un modèle structurel, et de façon correspondante, que des régions de confiance simultanées pour ces coefficients. Elle ne permet pas de tester des hypothèses spécifiant des coefficients individuels ou sur des transformations de ces coefficients. Ce problème peut être résolu en principe par des techniques de projection [Dufour (1997, Econometrica), Dufour et Jasiak (2001, International Economic Review)]. Cependant , ces techniques ne sont pas toujours faciles à appliquer et requièrent en général l'emploi de méthodes numériques.

Dans ce texte, nous proposons une solution explicite complète au problème de la construction de régions de confiance par projection basées sur des statistiques de type Anderson-Rubin. Cette solution exploite les propriétés géométriques des "quadriques" et peut s'interpréter comme une extension des intervalles et ellipsoïdes de confiance usuels. Le calcul de ces régions ne requièrent que des techniques de moindres carrés. Nous étudions également par simulation le degré de conservatisme des régions de confiance obtenues par projection. Enfin, nous illustrons les méthodes proposées par trois applications différentes: la relation entre l'ouverture commerciale et la croissance, le rendement de l'éducation et une étude sur les rendement d'échelles dans l'économie américaine.

Mots clés : équations simultanées ; modèle structurel ; variable instrumentale ; instruments faibles; intervalle de confiance ; test ; projection ; inférence simultanée ; inférence exacte; théorie asymptotique.

Contents

List of Propositions and Theorems

List of Tables

List of Figures

1. Introduction

One of the classic problems of econometrics consists in making inference on the coefficients of structural models. Such models typically involve endogenous explanatory variables (which can lead to endogeneity biases), the need to use instrumental variables, and the possibility that "structural parameters of interest" may not be identifiable. Recently, the statistical problems raised by structural modelling have received new attention in view of the observation that proposed instruments are often "weak", *i.e.* poorly correlated with the relevant endogenous variables, which correspond to situations where the structural parameters are close to being not identifiable (through the instruments used). The literature on so-called "weak instruments" problems is now considerable; see, for example, Nelson and Startz (1990a, 1990b), 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), Hall and Peixe (2000), Stock and Wright (2000), Dufour and Jasiak (2001), Hahn and Hausman (2002a, 2002b), Kleibergen (2001, 2002), Moreira (2001, 2002), Stock and Yogo (2002) and Stock, Wright, and Yogo (2002)].

In such contexts, several papers have documented by simulation and approximate asymptotic methods the poor performance of standard asymptotically justified procedures [Nelson and Startz (1990a, 1990b), Buse (1992), Bound, Jaeger, and Baker (1993, 1995), Hall, Rudebusch, and Wilcox (1996), Staiger and Stock (1997), Zivot, Startz, and Nelson (1998), Dufour and Jasiak (2001)]. The main difficulty here is that the finite-sample distributions of the relevant statistics (in particular, test statistics) are very sensitive to unknown nuisance parameters; indeed, they can exhibit an arbitrary large sensitivity to such parameters. Further, limiting distributions are non-standard when identification conditions do not hold, while usual large-sample approximations do not converge uniformly, so that the latter may be arbitrarily inaccurate in finite samples even when identification and standard regularity conditions obtain. The fact that standard asymptotic theory can be arbitrarily inaccurate in finite samples (of any size) is 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)].

As a result, it appears especially important in such problems to build tests and confidence sets 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 obtained 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, Startz, and Nelson (1998), Startz, Nelson, and Zivot (1999), Dufour and Jasiak (2001), Stock and Wright (2000), Kleibergen (2001, 2002), Moreira (2001, 2002), and Stock, Wright, and Yogo (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).

In this paper, we wish to concentrate on procedures for which finite-sample pivotality obtains under standard assumptions. In view of the results in Dufour (1997), we consider this approach as the best guide to selecting test and confidence set procedures (even though asymptotic validity will hold under weaker distributional assumptions). Useful pivotal functions are however difficult to find in structural models. The oldest one appears to be the statistic proposed by Anderson and Rubin (1949, henceforth AR). The latter is a limited-information method which allows one to test an hypothesis setting the value of the full vector of the endogenous explanatory variable coefficients in a linear structural equation; under usual parametric assumptions (error Gaussianity, instrument strict exogeneity) the distribution of the statistic is a central Fisher distribution, while under weaker (standard) assumptions it is asymptotically chi-square, irrespective of the presence of weak instruments.

Limited-information methods typically involve an efficiency loss with respect to *full-information* methods, but do allow for a less complete specification of the model and more robustness. Indeed, the AR statistic enjoys several remarkable *invariance* (or *robustness*) properties. Namely it is completely robust (in finite samples) to the presence of weak instruments (*robustness to weak instruments*), to 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*).[1] More precisely, its finite-sample distribution (under the null hypothesis) is completely unaffected by the presence of "weak instruments", the exclusion of relevant instruments, and the error distribution in the reduced form for the explanatory endogenous variables. 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 relevant instruments have been excluded (which seems highly likely in most practical situations). As a result, tests and confidence sets based on the AR statistic remains valid irrespective whether instruments are weak or relevant instruments have been excluded. Extensions of the AR statistics with the same basic robustness properties and a finite-sample distributional theory are also proposed in Dufour and Jasiak (1993, 2001).

Other potential pivots aimed at being robust to weak instruments have recently been suggested by Wang and Zivot (1998), Kleibergen (2002) and Moreira (2002). 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 of the model _ 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. Indeed, it is easy to see that none of these

---

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

statistics is pivotal in finite samples (*i.e.*, their finite-sample distributions involve unknown nuisance parameters) or robust to the exclusion of relevant instruments. It is clear that these statistics do not qualify as pivotal in finite samples.

An important practical shortcoming of the above methods is that they are designed to test hypotheses of the form $H_0 : \beta = \beta_0$, where β is the coefficient vector for *all* the endogenous explanatory variables. In particular these statistics do not allow to test linear and nonlinear restrictions on the vector β. A general solution to this problem is the *projection* technique described in Dufour (1990, 1997), Wang and Zivot (1998) and Dufour and Jasiak (2001).[2] This problem was also considered by Choi and Phillips (1992), Stock and Wright (2000) and Kleibergen (2001). 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 (2001) 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.

The basic idea behind the projection technique is simple. Let θ be a multidimensional parameter vector for which we can build a confidence set $C_\theta(\alpha)$ with level $1 - \alpha$: $\mathsf{P}[\theta \in C_\theta(\alpha)] \geq 1 - \alpha$. Now consider a transformation of interest $g(\theta)$ which takes its values in $\mathbb{R}^m$. For example, $g(\theta)$ could be one of the components of θ. Then it is easy to see that the image set $g[C_\theta(\alpha)] = \{g(\theta) \in \mathbb{R}^m : \theta \in C_\theta(\alpha)\}$ is a confidence set with level $1 - \alpha$ for $g(\theta)$, *i.e.* $\mathsf{P}\big[g(\theta) \in g[C_\theta(\alpha)]\big] \geqslant 1 - \alpha$. Such methods are also exploited in Abdelkhalek and Dufour (1998) and Dufour and Kiviet (1998) for completely different models. In general, however, the calculation of $g[C_\theta(\alpha)]$ is not simple and may require the use of costly numerical methods [as done, for example, in Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998) or Dufour and Jasiak (2001)].

In this paper, we study some general geometric features of AR-type confidence sets and we provide a complete explicit solution to the problem of building projection-based confidence sets from such confidence sets. We *first* observe that AR-type confidence sets can be described as **quadrics** [see Shilov (1961, Chapter 11) and Pettofrezzo and Marcoantonio (1970)], a class of geometric figures which covers as special cases the usual confidence intervals and ellipsoids but also includes hyperboloids and paraboloids. In particular, we give a simple necessary and sufficient condition under which such confidence sets are bounded (which indicates that the parameters considered are identifiable). We use the projection technique to build confidence sets for components of the vector of unknown parameters and for linear combinations of these components.

Second, using these results, we then derive simple explicit expressions for projection-based

---

[2]Another shortcoming of AR-type tests comes from the fact that power may decline as the number of instruments increases, especially if they have little relevance. This indicates that the number of instruments should be kept as small as possible. Because AR statistics are robust to the exclusion of instruments (even if they are relevant), 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 (2001b, 2001a)]. In the present paper, we focus on the problem of building projection-based confidence sets, for a *given* set of instruments. For results relevant to instrument selection, the reader may also consult Cragg and Donald (1993), Hall, Rudebusch, and Wilcox (1996), Shea (1997), Staiger and Stock (1997), Chao and Swanson (2000), Donald and Newey (2001), Hall and Peixe (2000), Hahn and Hausman (2002a, 2002b), Stock and Yogo (2002).

confidence intervals in the case of individual structural coefficients (or linear transformations of these coefficients). Consequently, no search by nonlinear methods is anymore required. The explicit calculation of the confidence sets thus makes the projection approach very attractive. 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.

Thirdly, 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, multiple hypotheses on β can be tested without ever losing control the overall level of the 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.

Fourth, the methods discussed in this work are evaluated and compared on the basis of Monte Carlo simulations. In particular we analyze the conservatism of the projection-based confidence sets.

Fifth, 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).

In Section 2, we present the background model and statistical inference methods on the coefficient vector of the explanatory endogenous variables. Section 3 presents the simultaneous confidence sets. In Section 4, we discuss some general properties of quadric confidence sets and provide a simple necessary and sufficient condition under which such sets are bounded. Section 5 provides explicit projection-based confidence intervals for individual structural parameters and linear transformations of these parameters. We also discuss the simultaneity property of these confidence intervals. In Section 6, we report the results of our Monte Carlo simulations, while Section 7 presents the empirical applications. Finally, Section 8 concludes.

2. Framework

We consider here the standard simultaneous equations model (SEM):

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

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

4

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables, 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 reduced-form disturbances. Further,

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

where $k = k_1 + k_2$. Finally, to get a finite-sample distributional theory for the test statistics, we shall use the following assumption on the conditional distribution of u given X :

$$u \,|\, X \sim N\big[0,\, \sigma_u^2(X) I_T\big] \tag{2.4}$$

where $\sigma_u^2(X)$ is a positive scalar parameter which may depend on X (but not on β or γ). This means that, conditional on X, the disturbances $u_1, \ldots, u_T$ are i.i.d. Gaussian. In particular, it is clear (2.4) holds under the following standard assumptions:

$$u \text{ and } X \text{ are independent;} \tag{2.5}$$

$$u \sim N\big[0,\, \sigma_u^2\, I_T\big]. \tag{2.6}$$

(2.5) may be interpreted as the strict exogeneity of X with respect to u.

Note that the distribution of V is not otherwise restricted; in particular, the vectors $V_1, \ldots, V_T$ need not follow a Gaussian distribution and may be heteroskedastic. Below, we shall also consider the situation where the reduced-form equation for Y includes a third set of instruments X_3 which are not used in the estimation:

$$Y = X_1\Pi_1 + X_2\Pi_2 + X_3\Pi_3 + V \tag{2.7}$$

where X_3 is a $T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous); in particular, X_3 may be unobservable. We view this situation as important because, in practice, it is quite rare that one can consider all the relevant instruments have been or should be used.

In such a model, we are generally interested in making inference on β and γ. In Dufour (1997), it is shown that if the model is unidentified (the matrix Π_2 does not have its maximal rank) any valid confidence set for β or γ must be unbounded with positive probability. This is due to the fact that such a model may be unidentified and holds indeed even if identification restrictions are imposed. This result explains many recent findings about the performance of standard asymptotic statistics when the instruments X_2 are weakly correlated with the endogenous explanatory variables Y. The usual approach, which consists in inverting Wald-type statistics to obtain confidence sets, is not valid in these situations since the resulting confidence sets are bounded with probability 1. This is related to the fact that finite-sample distributions of such statistics are not pivotal and follow distributions which depend heavily on nuisance parameters.

Choi and Phillips (1992) considered the same model where they suppose that a subset of parameters are not identified. They derive exact and asymptotic distributions of the instrumental variables

estimator and the Wald statistic. The analytic expressions obtained are complex and differ from commonly known ones. Staiger and Stock (1997) considered the same model but assumed that the elements of the matrix Π_2 tend to 0 as T increases ($\Pi_2 = C/\sqrt{T}$, where C is a fixed matrix). They derive the asymptotic distributions of different statistics, including two-stage least squares (2SLS) limited information maximum likelihood (LIML) and the Wald-type statistics based on these estimators. In conformity with the results in Dufour (1997), these distributions depend on nuisance parameters and are not pivotal. Wang and Zivot (1998) derived [under the same assumption as Staiger and Stock, *i.e.* $\Pi_2 = C/\sqrt{T}$] the asymptotic distributions of likelihood ratio (LR) and Lagrange multiplier (LM) type statistics based on maximum likelihood and GMM estimation methods. As before, these distributions depend on nuisance parameters and are not pivotal. These derivations provide useful insights for understanding the poor performance of asymptotic approximations reported in previous work, but they do not solve the statistical inference problem in these models.

A first solution to this problem [see Dufour (1997) and Staiger and Stock (1997)] consists in using the Anderson-Rubin statistic [Anderson and Rubin (1949)]. This test 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:

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{2.8}$$

where $\theta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\theta_2 = \Pi_2(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Equation (2.8) satisfies all the conditions of the linear regression model. We can test H_0 by testing $H_0' : \theta_2 = 0$ using the standard F-statistic H_0' [denoted $AR(\beta_0)$]. With the additional assumptions (2.3) - (2.4), we have under H_0 :

$$AR(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2, T - k) \tag{2.9}$$

where for any full rank matrix B, $M(B) = I - P(B)$ and $P(B) = B(B'B)^{-1}B'$ is the projection matrix on the space spanned by the columns of B. The distributional result in (2.9) holds irrespective on the rank of the matrix Π_2, which means that tests based on $AR(\beta_0)$ are *robust to weak instruments*. It is also interesting to note that this distribution is not affected by the distribution of V; in other words, $AR(\beta_0)$ is *robust to the distribution of the endogenous explanatory variables Y*.

Another important feature of $AR(\beta_0)$ comes from the fact that (2.9) also obtains under the wider model (2.7), because in this case:

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + X_3\theta_3 + \varepsilon \tag{2.10}$$

where $\theta_1 = \gamma + \Pi_1(\beta - \beta_0)$, $\theta_2 = \Pi_2(\beta - \beta_0)$, $\theta_3 = \Pi_3(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Since $\theta_2 = 0$ and $\theta_3 = 0$ under H_0, it is straightforward to see that the null distribution of $AR(\beta_0)$ is $F(k_2, T - k)$ [under the assumptions (2.1), (2.7), (2.3) and (2.4)]. As a result, the 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 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. Even more generally, we could also assume that Y obeys a general nonlinear model of the form:

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

where $g(\cdot)$ is a possibly unspecified nonlinear function and Π is an unknown parameter matrix. Since, under H_0,

$$y - Y\beta_0 = X_1\theta_1 + \varepsilon \,,$$

the coefficient θ_2 in the regression (2.8) must be zero, and (2.9) still holds.

A confidence set for β with level $1 - \alpha$ can be obtained by inverting the statistic $AR(\beta_0)$:

$$C_\beta(\alpha) = \{\beta_0 : AR(\beta_0) \leq F_\alpha(k_2, T - k)\} \tag{2.12}$$

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 an identification assumption. When $G = 1$, this set has an explicit form solution involving a quadratic inequation _ i.e. $C_\beta(\alpha) = \{\beta_0 : a\beta_0^2 + b\beta_0 + c \leq 0\}$ 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, Startz, and Nelson (1998) for details. Further, Monte Carlo simulations [Maddala (1974), Dufour and Jasiak (2001)] indicate that the AR-based test behaves well in terms of power (as long as the number of k_2 of additional instruments is not unduly large). This test also remains asymptotically valid under weaker distributional assumptions, in the sense that the asymptotic null distribution of $AR(\beta_0)$ is $\chi^2(k_2)/k_2$ [see Dufour and Jasiak (2001) and Staiger and Stock (1997)].

Below, we shall also consider two alternative statistics proposed by Wang and Zivot (1998). The first one is an $LR-$type statistic and the second is an $LM-$type statistic. Under the assumptions (2.1)-(2.6) and additional regularity conditions on the asymptotic behavior of the instruments [described by Wang and Zivot (1998)], these two statistics follow $\chi^2(k_2)$ distributions asymptotically when the model is exactly identified ($k_2 = G$), and are bounded by a $\chi^2(k_2)$ distribution when the model is over-identified ($k_2 > G$). To test $H_0 : \beta = \beta_0$, these statistics are:

$$LR_{\text{LIML}}(\beta_0) = T[\ln(k(\beta_0)) - \ln[k(\widehat{\beta}_{LIML})] \,, \tag{2.13}$$

$$LM_{\text{2SLS}}(\beta_0) = \frac{T(y - Y\beta_0)'P[P[M(X_1)X_2]Y](y - Y\beta_0)}{(y - Y\beta_0)'M(X_1)(y - Y\beta_0)} \,, \tag{2.14}$$

where

$$k(\beta_0) = \frac{(y - Y\beta_0)'M(X_1)(y - Y\beta_0)}{(y - Y\beta_0)'M(X)(y - Y\beta_0)} \,.$$

Asymptotic and conservative confidence sets for β can be obtained by inverting the latter tests. However, it is easy to see that these statistics are not generally robust to instrument exclusion.[3]

A common shortcoming of all these tests is that they require one to specify the entire vector β.

---

[3]We do not study here the tests proposed by Kleibergen (2002) and Moreira (2002), because it does not appear that the associated confidence sets can be covered by the theory described in this paper (in terms of quadrics). Furthermore, these procedures are not robust to instrument exclusion.

In particular, they do not allow for general hypotheses of the form $H_0 : g(\beta) = 0$, where $g(\beta)$ may be any transformation of β, such as $g(\beta) = \beta^i - \beta_0^i$, where β^i is any scalar component of β.

In this paper, we deal with this problem by studying the characteristics of the confidence sets obtained by inverting such statistics, and we use them to derive confidence sets for the components of β or linear combinations of these components. We first show that the confidence sets based on the statistics AR, LR and LM can be expressed in terms of a quadratic-linear form involving a matrix A, a vector b and a scalar c. These sets (replacing the inequality by an equality) are known as **quadrics** [Shilov (1961, Chapter 11), Pettofrezzo and Marcoantonio (1970, Chapters 9-10)]. We will then study the different possible cases as functions of A, b and c, and we will derive analytic expressions for projection-based confidence intervals in the case of linear transformations of model parameters.

3. Quadric confidence sets

Let us first consider the AR statistic. A simple algebraic calculation shows that the inequality

$$AR(\beta_0) \leq F_\alpha(k_2, T - k)$$

may be written in the following simple form:

$$\beta_0' A \beta_0 + b' \beta_0 + c \leq 0 \tag{3.1}$$

where $A = Y'HY$, $b = -2Y'Hy$, $c = y'Hy$ and

$$H \equiv H_{AR} = M(X_1) - M(X)\left[1 + \frac{k_2 F_\alpha(k_2, T - k)}{T - k}\right]. \tag{3.2}$$

We can thus write:

$$C_\beta(\alpha) = \{\beta_0 : \beta_0' A \beta_0 + b' \beta_0 + c \leq 0\}. \tag{3.3}$$

If β is scalar, this set is the solution of a quadratic inequation:

$$C_\beta(\alpha) = \{\beta_0 : a\beta_0^2 + b\beta_0 + c \leq 0\}. \tag{3.4}$$

Depending on the values of a, b and c, this set may take several forms (a closed interval, a semi-open interval, a union of two semi-open intervals, the set $\mathbb{R}$ of all possible values, or the empty set); see Dufour and Jasiak (2001), and Zivot, Startz, and Nelson (1998).

If we use the statistic $LR_{LIML}(\beta_0)$ or $LM_{2SLS}(\beta_0)$ instead of AR, we get analogous confidence sets which only differ through the H matrix. For $LR_{LIML}(\beta_0)$, this matrix takes the form

$$H_{LR} = M(X_1) - M(X)\, k(\widehat{\beta}_{LIML}) \exp[\chi_\alpha^2(k_2)/T] \tag{3.5}$$

while, for $LM_{2SLS}(\beta_0)$, it is

$$H_{LM} = P\big[P[M(X_1)X_2]Y\big] - M(X_1)[\chi_\alpha^2(k_2)/T]. \tag{3.6}$$

For the AR and LR statistics, the matrix A can be written:

$$A = Y'M(X_1)Y - Y'M(X)Y(1 + f_\alpha)$$

where $f_\alpha = k_2 F_\alpha(k_2, T - k)/(T - k)$ for AR and $f_\alpha = \exp[\chi^2_\alpha(k_2)/T]k(\widehat{\beta}_{LIML}) - 1$ for the LR statistic. Clearly A is symmetric and a typical diagonal element of this matrix is

$$A_{ii} = Y_i'M(X_1)Y_i - Y_i'M(X)Y_i(1 + f_\alpha), \tag{3.7}$$

which is a **corrected** difference between the sum of squared residuals from the regression of Y_i on X_1 and the sum of squared residuals from the regression of Y_i on $X = [X_1, X_2]$. This difference may be viewed as a measure of the importance of X_2 in explaining Y_i, *i.e.* the relevance of X_2 as an instrument for Y_i. A necessary condition for matrix A to be positive definite is that the instruments X_2 should provide sufficient additional explanatory power for Y (with respect to X_1). Similarly, $c = y'Hy$ is a corrected difference between the sum of squared residuals from the regression of y on X_1 and the sum of squared residuals from the regression of y on $X = [X_1, X_2]$. For the vector b, a typical element is given by

$$b_i = -2\{[M(X_1)Y_i]'[M(X_1)y] - [M(X)Y_i]'[M(X)y](1 + f_\alpha)\}. \tag{3.8}$$

The first term [multiplied by $-1/(2T)$] is the sample covariance between the residuals of the regression of Y_i on X_1 and the residuals of the regression of y on X_1, while the second term gives the same covariance with X_1 replaced by $X = [X_1, X_2]$.

4.   Geometry of quadric confidence sets

The locus of points that satisfy an equation of the form

$$\beta'A\beta + b'\beta + c = 0, \tag{4.1}$$

where A is a symmetric $G \times G$ matrix, b is a $G \times 1$ vector and c is a scalar, is known in the mathematical literature as a *quadric* surface [Shilov (1961, Chapter 11), Pettofrezzo and Marcoantonio (1970)]. Consequently, we shall call a confidence set of the form

$$C_\beta = \{\beta : \beta'A\beta + b'\beta + c \leq 0\} \tag{4.2}$$

a *quadric confidence set*. A quadric is characterized by the sum a quadratic form $(\beta'A\beta)$ and an affine transformation $(b'\beta + c)$. Depending on the values of A, b and c, it may take several forms. In this section, we examine some general properties of quadric confidence sets, especially the conditions under which such sets are bounded or unbounded. In particular, we will see that the eigenvalues of the A matrix play a central role in these properties and that larger eigenvalues are associated with more "concentrated" (or "smaller") quadric confidence sets. For these reasons, we call A the *concentration matrix* of the quadric.

In the sequel of this section, it will be convenient to distinguish between two basic cases: the

one where A is nonsingular, and the one where it is singular. We adopt the convention that an empty set is bounded.

4.1. Nonsingular concentration matrix

If A is nonsingular, we can write:

$$
\begin{aligned}
\beta' A\beta + b'\beta + c &= \left(\beta + \frac{1}{2}A^{-1}b\right)' A\left(\beta + \frac{1}{2}A^{-1}b\right) - \left(\frac{1}{4}b'A^{-1}b - c\right) \\
&= (\beta - \tilde{\beta})' A(\beta - \tilde{\beta}) - d
\end{aligned}
\tag{4.3}
$$

where $\tilde{\beta} = -\frac{1}{2}A^{-1}b$ and $d = \frac{1}{4}b'A^{-1}b - c$. Since A is a real symmetric matrix, we have:

$$
A = P'DP
\tag{4.4}
$$

where P is an orthogonal matrix and D is a diagonal matrix whose elements are the eigenvalues of A. Inequality (3.1) may then be reexpressed as

$$
\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \le d
\tag{4.5}
$$

where the λ_i's are the eigenvalues of A and $z = P(\beta - \tilde{\beta})$. The transformation $z = P(\beta - \tilde{\beta})$ represents a translation followed by a rotation of β, so it is clear that

$$
C_\beta \text{ is bounded } \Leftrightarrow C_z \text{ is bounded}
\tag{4.6}
$$

where

$$
\begin{aligned}
C_\beta &\equiv \{\beta : \beta' A\beta + b'\beta + c \le 0\} = \{\beta : (\beta - \tilde{\beta})' A(\beta - \tilde{\beta}) \le d\} \\
&= \{\beta : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \le d \text{ and } z = P(\beta - \tilde{\beta})\}, \tag{4.7} \\
C_z &\equiv \{z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \le d\}. \tag{4.8}
\end{aligned}
$$

Again it will be convenient to distinguish between three cases according to the signs of the eigenvalues of A, namely: (1) all the eigenvalues of A are positive ($\lambda_i > 0$, $i = 1, \ldots, G$), *i.e.* A is positive definite; (2) all the eigenvalues of A are negative ($\lambda_i < 0$, $i = 1, \ldots, G$), *i.e.* A is negative definite; (3) A has both positive and negative values, *i.e.* A is neither positive nor negative definite.

4.1.1. Positive definite concentration matrix

If $\lambda_i > 0$, $i = 1, \ldots, G$, the inequality (4.5) can be reexpressed as

$$
\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \le d
\tag{4.9}
$$

10

where $\gamma_i = \sqrt{1/\lambda_i}$, $i = 1, \ldots, G$. If $d = 0$, we have $C_z = \{0\}$ and C_β reduces to $\{\tilde{\beta}\}$. If $d < 0$, C_z and C_β are empty. If $d > 0$, C_z is the area inside an ellipsoid, hence it is a compact set. Consequently, C_z and C_β are bounded.

4.1.2. Negative definite concentration matrix

If $\lambda_i < 0$, $i = 1, \ldots, G$, the set C_z is the set of all values of z that satisfy

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \geq -d \tag{4.10}$$

where $\gamma_i = \sqrt{-1/\lambda_i}$, or equivalently, the set *not inside* the open ellipsoid defined by

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 < -d. \tag{4.11}$$

C_z and C_β are thus unbounded sets. In particular, if $d \geq 0$, we have $C_\beta = C_z = \mathbb{R}^G$.

4.1.3. Concentration matrix not positive or negative definite

If A has both positive and negative eigenvalues, we can assume, without loss of generality, that $\lambda_i > 0$ for $i = 1, \ldots, p$, and $\lambda_i < 0$, for $i = p+1, \ldots, G$, where $1 \leq p < G$. Inequality (4.5) may then be rewritten:

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_p}{\gamma_p}\right)^2 - \left(\frac{z_{p+1}}{\gamma_{p+1}}\right)^2 - \cdots - \left(\frac{z_G}{\gamma_G}\right)^2 - d \leq 0 \tag{4.12}$$

where p is the number of positive eigenvalues of A, $\gamma_i = \sqrt{1/\lambda_i}$ for $i = 1, \ldots, p$, and $\gamma_i = \sqrt{-1/\lambda_i}$ for $i = p+1, \ldots, G$.

In this case C_z and C_β are unbounded. This is easy to see: for arbitrary given values of $z_1, \ldots, z_p$ and d, it is clear that inequality (4.12) will hold if any of the values $z_i, p+1 \leq i \leq G$, is small enough (as $|z_i| \to \infty$). Consequently, each component of z is unbounded in C_z and similarly for each component of β in C_β.

4.2. Singular concentration matrix

We now consider the case where A is singular with rank r ($r < G$). Without loss of generality, we can assume that the first r diagonal elements of D in the decomposition $A = P'DP$ (the first r eigenvalues of A) used in (4.5) are different from zero, while the $G - r$ other ones are equal to zero. Then we can write (see the details in the Appendix):

$$\beta'A\beta + b'\beta + c = \sum_{i=1}^{r} \lambda_i z_i^2 + \sum_{i=r+1}^{G} \delta_i z_i - d \tag{4.13}$$

11

where the λ_i are the eigenvalues of A ($\lambda_i \neq 0$, $i = 1, \ldots, r$), $\delta = Pb$, $z = P\beta + \mu$ and

$$d = -c + \sum_{i=1}^{r} \delta_i^2/(4\lambda_i), \quad \mu_i = \begin{cases} \delta_i/(2\lambda_i), & \text{if } \lambda_i \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.14}$$

In the new space given by the transformation $z = P\beta + \mu$, C_z may take many forms following the number of non-zero eigenvalues and their signs. However, this set will always be unbounded. From (4.13) it is clear that we can make any z_i, $i = r + 1, \ldots, G$, arbitrarily large with an opposite sign of its coefficient δ_i, so that the inequality (4.5) will hold.

4.3. Necessary and sufficient condition for bounded quadric confidence set

We can now deduce the conditions under which C_β is bounded. According to results in Dufour (1997), a valid confidence set C_β for β (with level $1 - \alpha$) in model (2.1)-(2.6) must be unbounded with positive probability for any parameter configuration, a probability that should be large (close to $1 - \alpha$) when the matrix Π_2 does not have full rank (or is close to have full column rank). Given the complicated expressions of the random matrix A, the random vector b and the random scalar c, it seems difficult to evaluate this probability. In the following proposition, we give an easy-to-verify necessary and sufficient condition for a confidence set of the form C_β to be bounded.

Theorem 4.1 NECESSARY AND SUFFICIENT CONDITION FOR BOUNDED CONFIDENCE SET. *The set C_β in* (4.2) *is bounded if and only if the matrix A is positive definite.*

Proofs are provided in the Appendix. It is of interest to note here that the case where A is singular is unlikely to be met with AR-type confidence sets such as those described in Section 3, because in this case we have $A = Y'HY$, where Y and H, are a $T \times G$ and a $T \times T$ matrices respectively. If Y follows an absolutely continuous distribution (as assumed in Section 2), A will be nonsingular with probability one as soon as the rank of H is greater than or equal to G.

4.4. Joint confidence sets for β and γ

Finally, we note that the above results also apply to the problem of building joint confidence sets for β and a subvector γ_1 of γ. This can be done by using an appropriate extension of the AR procedure [see Dufour and Jasiak (2001)]. Let $X_1 = [X_{11}, X_{12}]$, $\gamma = (\gamma_1', \gamma_2')'$ and $\Pi_1 = [\Pi_{11}, \Pi_{12}]$ where X_{11}, γ_1 and Π_{11} have dimensions $T \times k_{11}$, $k_{11} \times 1$ and $k_{11} \times G$ respectively ($k_{11} \leq k_1$). From (2.1) - (2.2), we can write:

$$y - Y\beta_0 - X_{11}\gamma_{10} = X_{11}\theta_{11} + X_{12}\theta_{12} + X_2\theta_2 + \xi \tag{4.15}$$

where $\theta_{11} = \Pi_{11}(\beta - \beta_0) + \gamma_1 - \gamma_{10}$, $\theta_{11} = \Pi_{11}(\beta - \beta_0) + \gamma_2$, $\theta_2 = \Pi_2(\beta - \beta_0)$ and $\xi = V(\beta - \beta_0) + u$. We can test

$$H_0 : (\beta, \gamma_1) = (\beta_0, \gamma_{10}) \tag{4.16}$$

by testing $H_0' : \theta_{11} = 0$ and $\theta_2 = 0$, and obtain a joint confidence set for β and γ_1 by inverting the corresponding F-test. After a similar simple calculation, we obtain the same form as before:

$$C_{(\beta, \gamma_1)}(\alpha) = \{(\beta_0', \gamma_{10}')' : (\beta_0', \gamma_{10}')A_1(\beta_0', \gamma_{10}')' + b_1'(\beta_0', \gamma_{10}')' + c_1 \leq 0\} \tag{4.17}$$

where $A_1 = [Y, X_{11}]'H_1[Y, X_{11}]$, $b_1 = -2[Y, X_{11}]'H_1 y$, $c_1 = y'H_1 y$ and

$$H_1 = M(X_{12}) - \left[1 + \frac{k_{11} + k_2}{T - k} F_\alpha(k_{11} + k_2, T - k)\right] M(X). \tag{4.18}$$

5. Confidence sets for transformations of β

5.1. The projection approach

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 vector of parameters θ, this method enables one to deduce confidence sets for general transformations g in $\mathbb{R}^m$ of this vector. Since $x \in E \Rightarrow g(x) \in g(E)$ for any set E, we have

$$\mathsf{P}[\theta \in C_\theta(\alpha)] \geq 1 - \alpha \Rightarrow \mathsf{P}\big[g(\theta) \in g\left[C_\theta(\alpha)\right]\big] \geq 1 - \alpha \tag{5.1}$$

where $g\left[C_\theta(\alpha)\right] = \{x \in \mathbb{R}^m : \exists \theta \in C_\theta(\alpha), \ g(\theta) = x\}$. Hence $g\left[C_\theta(\alpha)\right]$ 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

$$\mathsf{P}[g^L(\alpha) \leq g(\theta) \leq g^U(\alpha)] \geqslant 1 - \alpha \tag{5.2}$$

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 $I_U(\alpha) = \left[g^L(\alpha), g^U(\alpha)\right]\backslash\{-\infty, +\infty\}$ is a confidence interval (CI) 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 β: $g_1(\beta), g_2(\beta), \ldots, g_n(\beta)$. The set $C_{g_1(\beta)}(\alpha) \times C_{g_2(\beta)}(\alpha) \times \cdots \times C_{g_n(\beta)}(\alpha)$ where $C_{g_i(\beta)}(\alpha)$ is the projection-based confidence set for $g_i(\beta)$, $i = 1, \ldots, n$, is a simultaneous confidence set for the vector $\big(g_1(\beta), g_2(\beta), \ldots, g_n(\beta)\big)'$ with level greater than or equal to $1 - \alpha$. More generally, if $\{g_a(\beta) : a \in A\}$ is a set of functions of β, where A is is some index set, then

$$\mathsf{P}\big[g_a(\beta) \in g_a\left[C_\beta(\alpha)\right] \text{ for all } a \in A\big] \geq 1 - \alpha. \tag{5.3}$$

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).

In this section, we build confidence sets for $g(\beta)$ by "projecting" the set $C_\beta(\alpha)$.[4] We study two particular transformations: $g(\beta) = w'\beta$ (a linear combination of the components of β) and $g(\beta) = \beta_i$ (the projection on the axis β_i). We also show that the confidence interval $I_U(\alpha)$ may involve a sizeable loss of information when the two optimization problems have unbounded solutions [*i.e.*, if $g^L(\alpha) = -\infty$ and $g^U(\alpha) = +\infty$] while the appropriate projection is a proper subset of $\mathbb{R}$, hence the importance of studying the set C_β before choosing the way to project it.

If the aim is to test $H_0 : g(\beta) = 0$, we can easily deduce from $C_\beta(\alpha)$ a conservative test. The latter consists in rejecting H_0 when all vectors β_0 that satisfy H_0 are rejected by the AR test, or equivalently when the minimum of $AR(\beta_0)$ subject to the constraint (s.c.) $g(\beta) = 0$ is larger than $F_\alpha(k_2, T - k)$, *i.e.* when $\min\{AR(\beta) : g(\beta) = 0\} \geq F_\alpha(k_2, T - k)$.

5.2. Projection-based confidence sets for scalar linear transformations

We consider now a general confidence set of the form

$$C_\beta = \{\beta_0 : \beta_0' A \beta_0 + b'\beta_0 + c \leq 0\} \tag{5.4}$$

where c is a real scalar, A is a symmetric $G \times G$ matrix, and b is a $G \times 1$ vector. By definition, the associated projection-based confidence interval for the scalar function $g(\beta) = w'\beta$, where w is a nonzero $G \times 1$ vector, is:

$$C_{w'\beta} \equiv g[C_\beta] = \{\delta_0 : \delta_0 = w'\beta_0 \text{ where } \beta_0' A \beta_0 + b'\beta_0 + c \leq 0\} . \tag{5.5}$$

To study the characteristics of $C_{w'\beta}$, we shall distinguish again between the case where A is nonsingular and the case where it is singular.

5.2.1. Nonsingular concentration matrix

When the concentration matrix is nonsingular, all the eigenvalues of A are different from 0. Using the transformation $z = P(\beta - \tilde{\beta})$, $C_{w'\beta}$ may then be written:

$$C_{w'\beta} = \{w'\beta_0 : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d \text{ and } z = P(\beta_0 - \tilde{\beta})\}.$$

Further,

$$w'\beta = w'P'P\beta = w'P'P(\beta - \tilde{\beta}) + w'P'P\tilde{\beta} = a'z + w'\tilde{\beta} \tag{5.6}$$

where $a = Pw$. Setting

$$C_{a'z} = \{a'z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d\} , \tag{5.7}$$

it is then easy to see that, for $x \in \mathbb{R}$,

$$x \in C_{w'\beta} \quad \Leftrightarrow \quad x - w'\tilde{\beta} \in C_{a'z} , \tag{5.8}$$

---

[4]Since $C_{(\beta, \gamma_1)}(\alpha)$ [in (4.17)] and $C_\beta(\alpha)$, have the same form, projections from $C_{(\beta, \gamma_1)}(\alpha)$ can be computed in the same way.

hence: $C_{w'\beta} = \mathbb{R} \iff C_{a'z} = \mathbb{R}$. We will now distinguish three cases depending on the number of negative eigenvalues: (1) all the eigenvalues of A are positive (*i.e.*, A is positive-definite); (2) A has exactly one negative eigenvalue; (3) A has at least two negative eigenvalues.

When A is positive definite, C_β is a bounded set and, correspondingly, its image $g[C_\beta]$ by the continuous function $g(\beta) = w'\beta$ is also bounded [see Abdelkhalek and Dufour (1998, Proposition 2)]. The following proposition provides an explicit form for the projection-based confidence set $C_{w'\beta}$.

Theorem 5.1 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANSFORMATIONS WHEN THE CONCENTRATION MATRIX IS POSITIVE DEFINITE. *Let C_β be the set defined in (5.4), $d \equiv \frac{1}{4}b'A^{-1}b - c$ and let w be a nonzero vector in $\mathbb{R}^G$. If the matrix A is positive definite and $d \geq 0$, then*

$$C_{w'\beta} = \left[w'\tilde\beta - \sqrt{d\left(w'A^{-1}w\right)}, \; w'\tilde\beta + \sqrt{d\left(w'A^{-1}w\right)} \right] \tag{5.9}$$

where $\tilde\beta = -\frac{1}{2}A^{-1}b$. If $d < 0$, $C_{w'\beta}$ is empty.

In the special case where $w = e_i = (\delta_{1i}, \delta_{2i}, \dots, \delta_{Gi})'$, with $\delta_{ji} = 1$ if $j = i$ and $\delta_{ji} = 0$ otherwise, the set $C_{w'\beta}$ is a confidence interval for the component β_i. This set is given by the following corollary, which is a direct consequence of Proposition **5.1**.

Corollary 5.2 PROJECTION-BASED CONFIDENCE INTERVAL FOR AN INDIVIDUAL COEFFI-CIENT. *Let C_β be the set defined in (5.4) and $d \equiv \frac{1}{4}b'A^{-1}b - c$. Suppose the matrix A in (5.4) is positive definite. If the matrix A is positive definite and $d \geq 0$, then*

$$C_{\beta_i} = \left[\tilde\beta_i - \sqrt{d\left(A^{-1}\right)_{ii}}, \; \tilde\beta_i + \sqrt{d\left(A^{-1}\right)_{ii}} \right]$$

where $\tilde\beta_i = -(A^{-1})_{i.}b/2$ is the i-th element of $\tilde\beta = -\frac{1}{2}'A^{-1}b$, $(A^{-1})_{i.}$ is the i-th row of A^{-1}, $(A^{-1})_{ii}$ is the i-th element of the diagonal of A^{-1}, and $(A^{-1})_{ii} > 0$. Further, if $d < 0$, then C_{β_i} is empty.

It is interesting to note the relationship with Scheffé-type confidence sets. The confidence set for β is based on the F-test of $H_0 : \theta_2 = \Pi_2(\beta - \beta_0) = 0$ in the regression equation:

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon.$$

Following Scheffé (1959) [see also Savin (1984)], this F-test is equivalent to the test which does not reject H_0 when all hypotheses of the form $H_0(a) : a'\theta_2 = 0$ 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 $S(\alpha) = \sqrt{k_2\, F_{k_2, T-k}(\alpha)}$. Since $a'\theta_2 = w(\beta - \beta_0)$ where $w = \Pi_2 a$, this entails that no hypothesis of the form $H_0'(w) : w'\beta = w'\beta_0$, where $w = \Pi_2 a$, is rejected. The projection-based confidence set for $w'\beta$ can be viewed as a Scheffé-type simultaneous confidence interval for $w'\beta$.

Let us now consider the case where A has exactly one negative eigenvalue. The basic result in this case is given by the following proposition.

Theorem 5.3 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANSFORMATIONS WHEN THE CONCENTRATION MATRIX HAS ONE NEGATIVE EIGENVALUE. *Let C_β be the set defined in (5.4), $d \equiv \frac{1}{4}b'A^{-1}b - c$, $w \in \mathbb{R}^G \backslash \{0\}$, and suppose the matrix A is nonsingular with exactly one negative eigenvalue. If $w'A^{-1}w < 0$ and $d < 0$, then*

$$C_{w'\beta} = \left] -\infty \, , \, w'\tilde{\beta} - \sqrt{d\left(w'A^{-1}w\right)} \right] \cup \left[w'\tilde{\beta} + \sqrt{d\left(w'A^{-1}w\right)} \, , \, +\infty \right[. \tag{5.10}$$

If $w'A^{-1}w > 0$ or if $w'A^{-1}w \leq 0$ and $d \geq 0$, then $C_{w'\beta} = \mathbb{R}$. If $w'A^{-1}w = 0$ and $d < 0$, then $C_{w'\beta} = \mathbb{R} \backslash \{w'\tilde{\beta}\}$.

It is interesting to 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 $\mathbb{R}$.

Finally, we consider the case where A has at least two negative eigenvalues. We cover here the case where the matrix A is negative definite, or not negative definite with at least 2 negative eigenvalues. In this case the projection-based confidence set for any linear combination of the components of β is equal to the real line, thus uninformative. This is stated in the following proposition.

Theorem 5.4 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANSFORMATIONS WHEN THE CONCENTRATION MATRIX HAS MORE THAN ONE NEGATIVE EIGENVALUE. *Let C_β be the set defined in (5.4) and $w \in \mathbb{R}^G \backslash \{0\}$. If the matrix A in (5.4) is nonsingular and admits at least two negative eigenvalues, then $C_{w'\beta} = \mathbb{R}$.*

5.2.2. Singular concentration matrix

We will now study the case where the concentration matrix A can be singular. This can occur, for example, if the system studied involves identities. 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 β :

$$\delta = \left[\begin{array}{c} \delta_1 \\ \delta_2 \end{array} \right] = \left[\begin{array}{c} w'\beta \\ R_2\beta \end{array} \right] = R\beta, \quad R = \left[\begin{array}{c} w' \\ R'_2 \end{array} \right] = \left(\begin{array}{cc} w_1 & w'_2 \\ 0 & I_{G-1} \end{array} \right) \tag{5.11}$$

where $w' = [w_1, \, w'_2]$ and $R_2 = [0, \, I_{G-1}]$ is a $(G-1) \times G$ matrix. If $\beta = (\beta_1, \, \beta_2, \, \ldots, \, \beta_G)'$, it is clear form this notation that $\delta_2 = (\beta_2, \, \ldots, \, \beta_G)'$. We study the problem of building a confidence set for δ_1.

The quadric form which defines C_β in (4.2) may be written:

$$\beta'A\beta + b'\beta + c = \delta'\bar{A}\delta + \bar{b}'\delta + c \equiv \bar{Q}(\delta) \tag{5.12}$$

16

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

$$C_{w'\beta} = C_{\delta_1} = \{\delta_1 : \delta = (\delta_1, \, \delta_2')' \text{ satisfies } \bar{Q}(\delta) \le 0\}. \tag{5.13}$$

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

$$\bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{A}'_{21} \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \end{pmatrix} \tag{5.14}$$

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

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

We can then write:

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \delta_2'\bar{A}_{22}\delta_2 + [2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_2 \tag{5.15}$$

where, by convention, the two last terms of (5.15) simply disappear when $G = 1$. For $G \ge 1$, let $r_2 = \text{rank}(\bar{A}_{22})$, where $0 \le r_2 \le G - 1$, and consider the spectral decomposition:

$$\bar{A}_{22} = P_2 D_2 P_2', \quad D_2 = \text{diag}(d_1, \, \dots, \, d_{G-1}) \tag{5.16}$$

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

$$\begin{aligned} d_i \quad &\ne 0, \text{ if } 1 \le i \le r_2, \\ &= 0, \text{ if } i > r_2. \end{aligned} \tag{5.17}$$

Let us also define (whenever the objects considered exist)

$$\tilde{\delta}_2 = P_2'\delta_2, \quad \tilde{A}_{21} = P_2'\bar{A}_{21}, \quad \tilde{b}_2 = P_2'\bar{b}_2, \; D_{2*} = \text{diag}(d_1, \, \dots, \, d_{r_2}), \tag{5.18}$$

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:

$$\tilde{\delta}_{2*} = P_{21}'\delta_2, \quad \tilde{A}_{21*} = P_{21}'\bar{A}_{21}, \quad \tilde{b}_{2*} = P_{21}'\bar{b}_2, \quad P_2 = [P_{21}, \, P_{22}] \tag{5.19}$$

where P_{21} and P_{22} have dimensions $(G-1) \times r_2$ and $(G-1) \times (G-1-r_2)$ respectively. When A may be singular, the form of the set C_{δ_1} is given by the following theorem.

Theorem 5.5 PROJECTION-BASED CONFIDENCE SETS WITH A POSSSIBLY SINGULAR CONCEN-TRATION MATRIX. *Under the assumptions and notations* $(5.12) - (5.19)$, *the set* C_{δ_1} *takes one of the three following forms:*

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

$$C_{\delta_1} = \{\delta_1 : \tilde{a}_1 \delta_1^2 + \tilde{b}_1 \delta_1 + \tilde{c}_1 \leq 0\} \cup S_1 \tag{5.20}$$

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

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

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

$$C_{\delta_1} = \{\delta_1 : \bar{a}_{11} \delta_1^2 + \bar{b}_1 \delta_1 + c \leq 0\} \cup S_2 \tag{5.21}$$

where

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

(c) *if $\bar{A}_{22}$ is not positive semidefinite and $\bar{A}_{22} \neq 0$, then $C_{\delta_1} = \mathbb{R}$.*

In all the cases covered by the latter theorem the joint confidence set C_β is unbounded if A is singular [by Theorem **4.1**]. However, we can see from Theorem **5.5** that confidence intervals for some parameters (or linear transformations of β) can be bounded. This depends on the values of the coefficients of the second-order polynomials in (5.20) and (5.21). Specifically, it is easy to see that the quadratic set $\tilde{C}_{\delta_1} = \{\delta_1 : \tilde{a}_1 \delta_1^2 + \tilde{b}_1 \delta_1 + \tilde{c}_1 \leq 0\}$ in (5.20) can have the following forms: setting $\tilde{\Delta}_1 \equiv \tilde{b}_1^2 - 4\tilde{a}_1 \tilde{c}_1$,

$$\tilde{C}_{\delta_1} = \begin{cases} \left[\frac{-\tilde{b}_1 - \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}, \frac{-\tilde{b}_1 + \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1} \right], & \text{if } \tilde{a}_1 > 0 \text{ and } \tilde{\Delta}_1 \geq 0, \\ ]-\infty, \frac{-\tilde{b}_1 + \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}] \cup [\frac{-\tilde{b}_1 - \sqrt{\tilde{\Delta}_1}}{2\tilde{a}_1}, \infty[, & \text{if } \tilde{a}_1 < 0 \text{ and } \tilde{\Delta}_1 \geq 0, \\ ]-\infty, -\tilde{c}_1/\tilde{b}_1], & \text{if } \tilde{a}_1 = 0 \text{ and } \tilde{b}_1 > 0, \\ [-\tilde{c}_1/\tilde{b}_1, \infty[, & \text{if } \tilde{a}_1 = 0 \text{ and } \tilde{b}_1 < 0, \\ \mathbb{R}, & \text{if } (\tilde{a}_1 < 0 \text{ and } \tilde{\Delta}_1 < 0) \\ & \quad \text{or } (\tilde{a}_1 = \tilde{b}_1 = 0 \text{ and } \tilde{c}_1 \leq 0), \\ \emptyset, & \text{if } (\tilde{a}_1 > 0 \text{ and } \tilde{\Delta}_1 < 0) \\ & \quad \text{or } (\tilde{a}_1 = \tilde{b}_1 = 0 \text{ and } \tilde{c}_1 > 0). \end{cases} \tag{5.22}$$

Of course, a similar result holds for the quadratic set in (5.21).

The results presented in sections 5.2.1-5.2.2 are important for two main reasons. First, they allow one to obtain confidence sets in situations where no other solution has been proposed to date in the literature. Second, the explicit expressions found avoid one the use of costly numerical methods as used in the papers cited previously. This is much more important given the nature of the problems to be solved numerically. We tried many of the standard software as GAUSS and GAMS,

18

and they seem to have difficulties to find the solutions, unless the starting point is chosen near the solution (which is naturally unknown). However, Fortran-based IMSL routines appear to perform quite well.

5.3. A Wald-type interpretation of the projection-based confidence sets

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 Proposition **5.1**) is similar to the standard form: $[\hat{\theta} - \hat{\sigma}z(\alpha), \hat{\theta} + \hat{\sigma}z(\alpha)]$. Since $\theta = w'\beta$, 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 $1 - \alpha$, $\sqrt{d}/\hat{\sigma}$ should correspond to a quantile of an order greater than or equal $1 - \alpha$ 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, multiplying (2.1) by $(HY)'$, we get

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

Taking the matrix HY as a matrix of instrumental variables for Y, we get:

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

HY is asymptotically uncorrelated with the disturbances u under Assumption (5.23) bellow. Moreover, when C_β is obtained from inverting the AR statistic, then under the usual assumptions,

$$\left(\frac{X'X}{T}, \frac{X'u}{T}, \frac{X'V}{T} \right) \xrightarrow[T\to\infty]{p} (Q_{\bar{X}\bar{X}}, 0, 0), \quad \frac{X'u}{\sqrt{T}} \xrightarrow[T\to\infty]{L} N(0, \sigma_u^2 Q_{\bar{X}\bar{X}}), \tag{5.23}$$

it is easy to show that if Π_2 is of full rank. Then

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow[T\to\infty]{L} N\left[0, \sigma_u^2 \plim_{T\to\infty}\left(\frac{1}{T}A\right)^{-1}\right] \tag{5.24}$$

where $\plim_{T\to\infty} \frac{1}{T}A = \Pi_2'\left[Q_{X_2X_2} - Q_{X_2X_1}Q_{X_1X_1}^{-1}Q_{X_2X_1}'\right]\Pi_2$ and $Q_{X_iX_j} = \plim_{T\to\infty} \frac{1}{T}X_i'X_j$.

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

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

This is the expression of the well-known Theil's k-class estimator [see Davidson and MacKinnon (1993, page 649)] 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. The later may be written:

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

19

Hence when Π_2 is of full rank and the eigenvalues of A are positive, the projection-based confidence set for $w'\beta$ may be interpreted as a Wald-type confidence interval based on the statistic (which is asymptotically pivotal):

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

6. Monte Carlo evaluation

In this section, we study projection-based statistical inference through Monte Carlo simulations. We especially focus on the evaluation of the degree of conservatism of the projection-based confidence sets (CS) and we compare the confidence sets obtained on the basis of different statistics. These statistics are the Anderson-Rubin statistic (AR) given by (2.9), the asymptotic AR statistic (ARS) given by (2.9) but without assumption (2.6) (it follows asymptotically a $\chi^2(k_2)/k_2$ distribution) and the LR and LM statistics proposed by Wang and Zivot (1998) and given by (2.13) - (2.14). We also study the behavior of the Wald statistic based on 2SLS.

The data generating process is:

$$y = Y_1\beta_1 + Y_2\beta_2 + X_1\gamma + u\,, \tag{6.1}$$

$$(Y_1, Y_2) = X_2\Pi_2 + X_1\Pi_1 + (V_1, V_2)\,, \tag{6.2}$$

$$(u_t, V_{1t}, V_{2t})' \overset{i.i.d.}{\sim} N(0, \Sigma)\,, \quad \Sigma = \begin{pmatrix} 1 & .8 & .8 \\ .8 & 1 & .3 \\ .8 & .3 & 1 \end{pmatrix}\,, \tag{6.3}$$

with $k_1 = 1$, $G = 2$, $\beta_1 = \frac{1}{2}$, $\beta_2 = 1$, $\gamma = 2$, and $\Pi_1 = (0.1, 0.2)$. 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 Π_2 is such that $\Pi_2 = C/\sqrt{T}$. 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 the results for $C = 0$ (complete unidentification), Table 3 presents the results for a matrix C with components c_{ij} such that $1 < c_{ij} < 5$ (weak identification), and Table 4 with c_{ij} such that $10 \leq c_{ij} \leq 20$. The nominal confidence level for all tables is 95%.

We begin with the behavior of the classical Wald statistic (Table 1), 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 tends to the exact confidence level $1 - \alpha$ of the confidence set for β. [5] **Thus the projection-based confidence sets become less conservative as the number of relevant instruments increases.** This

---

[5]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 1. Empirical coverage rate of 2SLS-based Wald confidence sets

T	k	$C_{ij} = 0$	$1 \leq C_{ij} \leq 5$	$10 \leq C_{ij} \leq 20$
	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
50	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
	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
100	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
	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
200	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

suggests use of a number of relevant instruments as large as possible. But on the other hand, as noted by Dufour and Taamouti (2001b) and Kleibergen (2002), a large number of instruments will induce loss of power for the Anderson-Rubin test for β. Obviously, this should not be true in the extreme case of $C = 0$, X_2 vanishes from equation (6.2).

The proportions of unbounded confidence sets and confidence sets equal to the real line are nearly zero when identification holds (Table 4). When we approach nonidentification (tables 3 and 2), 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 statistic ARS behaves in the same way as the statistic 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 statistics, the main observation is that they produce confidence sets much more conservative than those based on AR or ARS, and unlike the AR statistic, the degree of conservatism of the resulting confidence sets increases with the number of instruments k_2. The coverage

Table 2. Characteristics of AR and ARS projection-based confidence sets – $C_{ij} = 0$

T	k	AR					ARS				
		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	95.06	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
	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
100	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
	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
200	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
	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

22

Table 2 (continued). Characteristics of LR and LM projection-based confidence sets – $C_{ij} = 0$

T	k	LM					LR				
		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 3. Characteristics of AR and ARS projection-based confidence sets _ $1 \leq C_{ij} \leq 5$

T	k	AR Coverage rate for β	AR Coverage rate for β_1	AR Unbounded CS	AR Empty CS	AR CS = $\mathbb{R}$	ARS Coverage rate for β	ARS Coverage rate for β_1	ARS Unbounded CS	ARS Empty CS	ARS 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 3 (continued). Characteristics of LR and LM projection-based confidence sets _ $1 \leq C_{ij} \leq 5$

T	k	LM					LR				
		Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = ℝ	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = 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 4. Characteristics of AR and ARS projection-based confidence sets $-10 \le C_{ij} \le 20$

T	k	AR					ARS				
		Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = ℝ	Coverage rate for β	Coverage rate for β_1	Unbounded CS	Empty CS	CS = 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 4 (continued). Characteristics of LR and LM projection-based confidence sets _ $10 \leq C_{ij} \leq 20$

T	k	LM Coverage rate for β	LM Coverage rate for β_1	LM Unbounded CS	LM Empty CS	LM CS $= \mathbb{R}$	LR Coverage rate for β	LR Coverage rate for β_1	LR Unbounded CS	LR Empty CS	LR 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 5. Comparison between AR and LR projection-based confidence sets
when they are bounded

		$1 \le C_{ij} \le 5$			$10 \le C_{ij} \le 20$		
T	k_2	AR shorter	CI mean length		AR shorter	CI mean length	
		than LR (%)	AR	LR	than LR (%)	AR	LR
	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
50	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
	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
100	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
	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
200	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

rate of the confidence sets based on 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 exact identification case, the statistic LR have a small size downward distortion.

In Table 5, we present a comparison between AR and LR projection-based confidence sets when they are bounded. The first column gives the percentages of AR confidence sets shorter than LR ones, columns 2 and 3 give the mean length of these intervals. We do not consider the case $C = 0$, since the intervals are nearly always unbounded in this case. As can be seen clearly, except in the case of exact identification, the AR-based confidence sets are shorter than the LR-based ones.

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 6 and Figure 1 where we present estimates of $P[\text{rejecting } H_0 : \beta_1 = 0.5 \,|\, \beta_1 = \beta_1^i]$ with a decision rule consisting of rejecting H_0 if 0.5 is excluded from the

Table 6. Power of tests induced by projection-based confidence sets _ $H_0 : \beta_1 = 0; T = 100$

| β_1	$k_2 = 2$				$k_2 = 4$				$k_2 = 10$				$k_2 = 20$			
	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	0.12	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
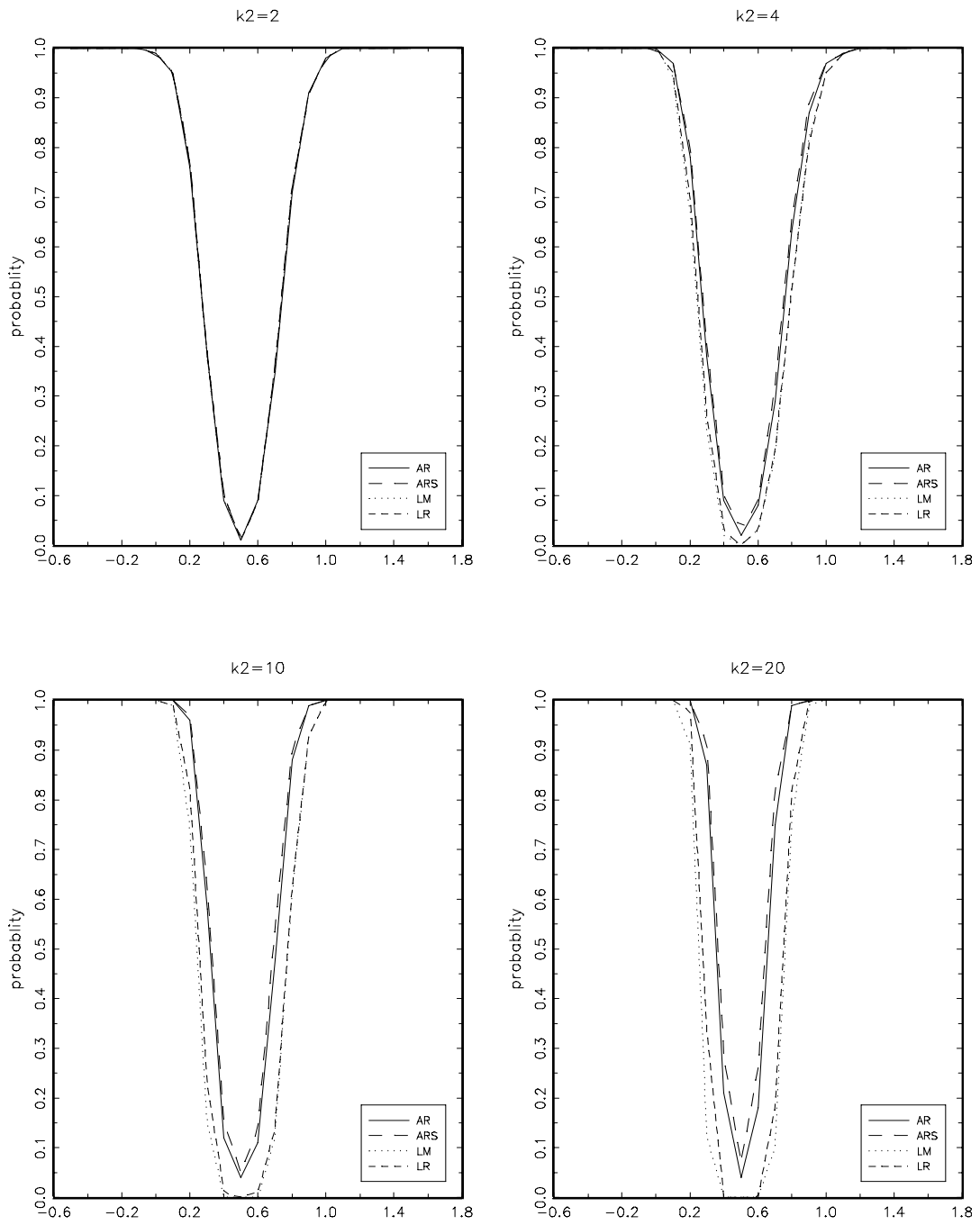
Figure 1. Power of tests induced by projection-based confidence sets
$H_0 : \beta_1 = 0.5$

30

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.

For $k_2 = 2$, the three tests have the same power but as k_2 increases, it appears clearly that the LM and LR based tests have less power and tends to reject few often. On the other hand, when k_2 increases, the test ARS appears to be powerful but in fact its size becomes greater than α.

7. 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.

7.1. Trade and growth

A large number of cross-country studies in the macroeconomics literature have looked at the relationship between standards of living and openness. 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, 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

$$y_i = a + bT_i + c_1 N_i + c_2 A_i + u_i \tag{7.1}$$

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 lists for each country, the trade share in 1985, the area and population (1985), and its income per person in 1985.[6] The authors focus on two samples. The first is the full 150 countries

---

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

Table 7. Confidence sets for the coefficients of the Frankel-Romer income-trade equation

A. Bivariate joint confidence sets (size $= 95\%$)

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 (size $\geq 95\%$)

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

covered by the Penn World Table, and the second sample is the 98-country sample considered by Mankiw, Romer, and Weil (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

$$T_i = \alpha + \beta Z_i + \gamma_1 N_i + \gamma_2 A_i + \varepsilon_i \tag{7.2}$$

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

To draw inference on the coefficients of the structural equation (7.1), 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 want to build confidence sets for the other parameters of (7.1), 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.[7] 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:

$$C_b(\alpha) = \{b : 0.963b^2 - 4.754b + 1.274 \leq 0\} = [0.284\,, \ 4.652]\,. \tag{7.3}$$

---

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

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

$$(\hat{b}_{IV} - 2\hat{\sigma}_{\hat{b}_{IV}}, \; \hat{b}_{IV} + 2\hat{\sigma}_{\hat{b}_{IV}}) = [-0.01, \; 3.95], \tag{7.4}$$

which is not very different from the AR-based confidence set. Interestingly, in contrast with (7.3), (7.4) does not exclude zero and may suggest that b is not significantly different from zero.

The joint confidence sets obtained by applying the methods of section 4.4 to each pair obtained by putting the trade share coefficient and each one of the other coefficients of (7.1) are given in Table 7A. All the confidence sets are bounded, a natural outcome since we do not have a serious problem of identification in this model. From this confidence sets we can obtain projection-based confidence intervals for each one of the parameters. Using Proposition **5.1**, we obtain the results presented in Table 7B. Even if 0 is included in the confidence sets for the coefficient of openness it is likely that the true value of the coefficient is positive. 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 6), but if we compare them to those obtained from t-statistics, they are not really larger.

7.2. Education and earnings

The second application considers the well known problem of 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, Jaeger, and Baker (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, for individuals born during 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 explain very little of the variation in education; see Bound, Jaeger, and Baker (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:

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

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

Table 8. 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	$\mathbb{R}$	$\mathbb{R}$	$[-0.111, -0.075]$

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 her the AR procedure. If we were only interested by the coefficient of education, we could compute the quadratic confidence set for β_1. But if we are also interested in other coefficients, for example the age coefficient (say, γ_1), the only way to get a confidence interval is to compute the AR joint confidence set for (β_1, γ_1) and then deduce by projection a confidence set for age. Obviously since the instruments are weak, we should expect large, if not completely uninformative, intervals. Table 8 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 equation of education. As expected many of the valid confidence sets are unbounded while Wald-type confidence sets are always bounded but likely invalid.

For the coefficient β_1 measuring returns to education, the AR-based quadratic confidence inter-

val of confidence level 95% is given by

$$AR\_IC_\alpha(\beta_1) = [-0.86,\ 0.77]\,.\tag{7.5}$$

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 $W\_IC_\alpha(\beta_1) = [0.0031,\ 0.1167]$.

7.3. 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:

$$Y_{it} = F(K_{it}, L_{it}, E_{it}, M_{it})\tag{7.6}$$

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)]:

$$\Delta y_{it} = \rho \Delta x_{it} + \Delta a_{it}\tag{7.7}$$

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.[8] 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:

$$\Delta y_{it} = \rho \Delta x_{it} + \eta \Delta x_t + u_{it}\tag{7.8}$$

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

---

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

35

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.[9]

Given this identification problem, we expect unbounded confidence sets. Using the same data set as Burnside (1996), we obtained the results presented in Table 9. 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 are obtained from joint confidence sets for (ρ, η) of level 90%.[10]

The average estimation over all industries of the coefficients ρ and η are of the same order as those obtained by Burnside (1996).[11] 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.

8. Conclusion

Recent research in econometrics has shown that weak instruments are quite widespread and should be carefully addressed. Techniques which are robust to weak instruments typically require one to consider first joint inference problem on all or, at least, some subvector of model parameters. This leads to the problem of drawing inference on individual coefficients (or lower dimensional

---

[9]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.

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

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

Table 9. Confidence sets for the returns to scale and externality coefficients in different U.S. industries (size $\geq 90\%$)

Industry	Returns to scale		Externalities	
	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	

subvectors). In this paper, we considered this problem from a finite-sample limited-information viewpoint and focused on AR-type tests and confidence sets. Important reasons for this choice stem from the fact that AR statistics are free of nuisance parameters in finite samples under standard assumptions, as well as robust to excluded instruments and more generally to the specification of a model for the endogenous explanatory variables.

We observed that AR-type confidence sets belong to a class of sets defined through quadric curves (which include ellipsoids as a special case). A simple condition for deciding whether such confidence sets are bounded was derived. On observing that a projection technique does provide finite-sample confidence sets for individual coefficients in such contexts (indeed, the only procedure for which a finite-sample theory is currently available), we derived a complete analytic solution to the problem of building projection-based confidence sets for individual structural coefficients (or linear combinations of the latter) when the joint confidence set has a quadric structure. The confidence sets so obtained turn out to be as easy to compute as standard Wald-type 2SLS-based confidence intervals. When they take the form of a closed interval, they can be interpreted as Wald-type confidence intervals based on k-class estimators using a pseudo "standard error" which depend on the AR critical value. The confidence sets so obtained have the additional feature of being simultaneous in the sense of Scheffé, so an unlimited number of such confidence sets can be built without losing control of their overall level. We also provided simulation results showing that projection-based confidence sets perform reasonably well in terms of accuracy. Since no alternative finite-sample procedure that enjoys the same robustness features appears to be available, it is currently the best solution to the problem at hand.

We think that 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 [see Dufour and Taamouti (2001b)] and dynamic models. Such extensions would go beyond the scope of the present paper. Another issue not discussed here consists in choosing the instruments to be used for the purpose of performing AR-type tests. It is easy to see that the power of AR-type tests may decline as the number of instruments increases, especially if they have little relevance. We study in detail the problem of selecting optimal instruments and reducing the number of instruments in two companion papers [Dufour and Taamouti (2001b, 2001a)].

A. Appendix: Proofs

PROOF OF EQUATION (4.13)   Setting $x = P\beta$ and $\delta = Pb$, we can write:

$$
\begin{aligned}
\beta'A\beta + b'\beta + c \;=\;& \beta'P'DP\beta + b'P'P\beta + c = (P\beta)'D(P\beta) + (Pb)'(P\beta) + c \\
=\;& \sum_{i=1}^{r}\lambda_i x_i^2 + \sum_{i=1}^{G}\delta_i x_i + c = \sum_{i=1}^{r}\lambda_i(x_i^2 + \frac{\delta_i x_i}{\lambda_i}) + \sum_{i=r+1}^{G}\delta_i x_i + c \\
=\;& \sum_{i=1}^{r}\lambda_i(x_i + \frac{\delta_i}{2\lambda_i})^2 + \sum_{i=r+1}^{G}\delta_i x_i + c - \sum_{i=1}^{r}\frac{\delta_i^2}{4\lambda_i} = \sum_{i=1}^{r}\lambda_i z_i^2 + \sum_{i=r+1}^{G}\delta_i z_i - d
\end{aligned}
$$

where $d = -\big(c - \sum_{i=1}^{r}\delta_i^2/(4\lambda_i)\big)$, $z = P\beta + u$, with $u_i = \frac{\delta_i}{2\lambda_i}$, if $\lambda_i \neq 0$, and $u_i = 0$, otherwise. ∎

PROOF OF THEOREM **4.1**   The proof of this proposition is a direct consequence of the study of the characteristics of C_β in Section 4. If the eigenvalues of A are all positive (see Section 4.1.1), C_β is bounded. If the eigenvalues of A are negative, C_β is unbounded (see Section 4.1.2). If they are different from 0 but of different signs, C_β is always unbounded whatever the sign of d (see Section 4.1.3). If A is singular, C_β is unbounded by (4.13); see Section 4.2. ∎

PROOF OF THEOREM **5.1**   Consider again the decomposition $A = P'DP$ as in (4.4). By (5.8), we have, for any $x_0 \in \mathbb{R}$, $x_0 \in C_{w'\beta} \Leftrightarrow x_0 - w'\tilde{\beta} \in C_{a'z}$, where $a = Pw$. Let $x = x_0 - w'\tilde{\beta}$. By definition, $x \in C_{a'z}$ iff (if and only if) there is a vector $z \in \mathbb{R}^G$ such that

$$z'Dz \le d \text{ and } a'z = x. \tag{A.1}$$

Further, there is a z verifying (A.1) iff the solution of the problem

$$\min_z z'Dz \quad \text{s.c. } a'z = x \tag{A.2}$$

verifies the constraint (A.1). If $d < 0$, it is clear there is no solution verifying (A.1) _ for D is positive definite _ and consequently $C_{a'z} = C_{w'\beta} = \varnothing$. Let $d \ge 0$. The Lagrangian of the problem (A.2) is $\mathcal{L} = z'Dz + \mu(x - a'z)$. Since D is positive definite, the first order conditions are necessary and sufficient. These are:

$$2Dz = \mu a, \quad a'z = x,$$

hence

$$\mu = \frac{2x}{a'D^{-1}a}, \quad z = \frac{x}{a'D^{-1}a}D^{-1}a, \quad z'Dz = \frac{1}{2}\mu x = \frac{x^2}{a'D^{-1}a}.$$

Thus

$$x \in C_{a'z} \;\Leftrightarrow\; \frac{x^2}{a'D^{-1}a} \le d \;\Leftrightarrow\; |x| \le \sqrt{d\,(a'D^{-1}a)} \;\Leftrightarrow\; |x_0 - w'\tilde{\beta}| \le \sqrt{d\,(a'D^{-1}a)}.$$

39

On noting that $a'D^{-1}a = w'A^{-1}w$, this entails that the confidence set for $w'\beta$ is given by (5.9). ∎

PROOF OF THEOREM **5.3**   As in the proof of Proposition **5.1**, we consider again the decomposition (4.4), the equivalence $x_0 \in C_{w'\beta} \Leftrightarrow x_0 - w'\tilde{\beta} \in C_{a'z}$, and we set $x = x_0 - w'\tilde{\beta}$ and $a = Pw$. Now, $x \in C_{a'z}$ iff there is a value of $z \in \mathbb{R}^G$ such that

$$
\begin{align}
a'z &= a_1 z_1 + \cdots + a_{G-1} z_{G-1} + a_G z_G = x, \tag{A.3}\\
z'Dz &= \lambda_1 z_1^2 + \cdots + \lambda_{G-1} z_{G-1}^2 - |\lambda_G| z_G^2 \le d, \tag{A.4}
\end{align}
$$

where (without loss of generality) we assume that λ_G is the negative eigenvalue. Let $a_{(G)} = (a_1, a_2, \ldots, a_{G-1})'$, $z_{(G)} = (z_1, z_2, \ldots, z_{G-1})'$, and $D_{(G)} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{G-1})'$.

If $a_G = 0$, then $a_{(G)} \ne 0$ (because $w \ne 0$ entails $a \ne 0$), and $w'A^{-1}w = a'D^{-1}a > 0$. In this case, for any $x \in \mathbb{R}$, we can choose z such that $a_1 z_1 + \cdots + a_{G-1} z_{G-1} = x$ and z_G is sufficiently large to ensure that (A.4) holds. Hence $C_{a'z} = \mathbb{R}$ and $C_{w'\beta} = \mathbb{R}$.

We will now suppose that $a_G \ne 0$. Then, the conditions (A.3) - (A.4) are equivalent to:

$$
z_G = (x - a_{(G)}' z_{(G)})/a_G, \tag{A.5}
$$

$$
|\lambda_G| \left(\frac{x - a_{(G)}' z_{(G)}}{a_G} \right)^2 \ge -d + z_{(G)}' D_{(G)} z_{(G)}, \tag{A.6}
$$

where the latter inequality can also be written as

$$
\left[|\lambda_G| s_{(G)}^2 - a_G^2 (z_{(G)}' D_{(G)} z_{(G)}) \right] - 2|\lambda_G| s_{(G)} x + \left[|\lambda_G| x^2 + d a_G^2 \right] \ge 0 \tag{A.7}
$$

where $s_{(G)} = a_{(G)}' z_{(G)}$. Since (A.5) always allows one to obtain (A.3) once the vector $z_{(G)}$ is given, a necessary and sufficient condition for $x \in C_{a'z}$ is the existence of a vector $z_{(G)}$ which satisfies inequality (A.7). Further, such a vector $z_{(G)}$ does exist iff we can find a value s such that the supremum (with respect to $z_{(G)}$) of the left-hand side of (A.7) subject to the restriction $a_{(G)}' z_{(G)} = s$ is larger than zero. Consequently, we consider the problem:

$$
\min_z \ z_{(G)}' D_{(G)} z_{(G)} \quad \text{s.c. } a_{(G)}' z_{(G)} = s \tag{A.8}
$$

where s is some real number. Since $D_{(G)}$ is positive definite, the first order conditions are necessary and sufficient to characterize a solution of (A.8). The Lagrangian for this problem is given by $\mathcal{L} = z_{(G)}' D_{(G)} z_{(G)} - \mu(a_{(G)}' z_{(G)} - s)$ and the corresponding first order conditions are:

$$
2D_{(G)} z_{(G)} = \mu a_{(G)}, \quad a_{(G)}' z_{(G)} = s, \tag{A.9}
$$

hence

$$
\mu = \frac{2s}{a_{(G)}' D_{(G)}^{-1} a_{(G)}}, \quad z_{(G)} = \frac{s}{a_{(G)}' D_{(G)}^{-1} a_{(G)}} D_{(G)}^{-1} a_{(G)}, \quad z_{(G)}' D_{(G)} z_{(G)} = \frac{s^2}{a_{(G)}' D_{(G)}^{-1} a_{(G)}}
$$

where $a'_{(G)} D^{-1}_{(G)} a_{(G)} > 0$. Substituting the solution of (A.8) into (A.7), we get:

$$q\, s^2 - (2|\lambda_G|x)s + (|\lambda_G|x^2 + da_G^2) \geq 0 \tag{A.10}$$

where $q = |\lambda_G| - [a_G^2/a'_{(G)} D^{-1}_{(G)} a_{(G)}] = \delta_G(w'A^{-1}w)$ and $\delta_G \equiv |\lambda_G|/a'_{(G)} D^{-1}_{(G)} a_{(G)} > 0$. Thus, $x \in C_{a'z}$ iff there is a value of s such that (A.10) holds. The discriminant of this second degree equation is

$$\Delta = 4\lambda_G^2 x^2 - 4Q(|\lambda_G|x^2 + da_G^2) = 4\delta_G a_G^2 \left[x^2 - d(w'A^{-1}w)\right].$$

We will now consider in turn each possible case for the signs of $w'A^{-1}w$ and d.

If $w'A^{-1}w > 0$, then $q > 0$ and, for any x, we can find a (sufficiently large) value of s such that (A.10) will hold. Consequently, $C_{a'z} = C_{w'\beta} = \mathbb{R}$. Thus, $w'A^{-1}w > 0$ entails $C_{a'z} = C_{w'\beta} = \mathbb{R}$, irrespective of the value of a_G (the case $a_G = 0$ was considered at the beginning of the proof).

If $w'A^{-1}w < 0$ and $d < 0$, then $q < 0$ and (A.10) has a (real) solution iff $\Delta \geq 0$ or, equivalently, $x^2 \geq d\,(w'A^{-1}w) > 0$. Consequently,

$$C_{a'z} \;=\;]-\infty,\, -\sqrt{d\,(w'A^{-1}w)}\,] \cup [\sqrt{d\,(w'A^{-1}w)},\, +\infty\,[, \tag{A.11}$$

$$C_{w'\beta} \;=\;]-\infty,\, w'\tilde{\beta} - \sqrt{d\,(w'A^{-1}w)}\,] \cup [w'\tilde{\beta} + \sqrt{d\,(w'A^{-1}w)},\, +\infty\,[. \tag{A.12}$$

If $w'A^{-1}w = 0$ and $d < 0$, (A.10) can be satisfied for any $x \neq 0$, hence $C_{a'z} = \mathbb{R}\backslash\{0\}$ and $C_{w'\beta} = \mathbb{R}\backslash\{w'\tilde{\beta}\}$.

Finally, if $d \geq 0$, (A.10) is satisfied for any x (on taking $s = 0$), and we have $C_{a'z} = C_{w'\beta} = \mathbb{R}$. All possible cases have been covered. ∎

PROOF OF THEOREM **5.4**   We need to show that $C_{a'z} = \mathbb{R}$. To see this, let λ_{i_1} and λ_{i_2} be the two negative eigenvalues of the matrix A, and (without loss of generality) suppose $a_1 \neq 0$. For any real x, we will show that $x \in C_{a'z}$, which entails that $C_{w'\beta} = C_{a'z} = \mathbb{R}$.

If λ_{i_1} or λ_{i_2} is associated with z_1 (say it is λ_{i_1}), we can set set the components of z such that: (1) $z_1 = (x - a_{i_2} z_{i_2})/a_1$; (2) $z_i = 0$, for $i > 1$, $i \neq i_2$; (3) $\lambda_1 z_1^2 + \lambda_{i_2} z_{i_2}^2 \leq d$. Since λ_{i_1} and λ_{i_2} are negative, z_{i_2} does exist. The vector z verifies (4.5) and $a'z = x$, hence $x \in C_{a'z}$.

If none of λ_{i_1} and λ_{i_2} is associated with z_1, we can set z so that: (1) $z_1 = x/a_1$; (2) $z_i = 0$, for $i \neq i_1$, $i \neq i_2$ and $i > 1$; (3) $\lambda_{i_1} z_{i_1}^2 + \lambda_{i_2} z_{i_2}^2 \leq d - \lambda_1 (x/a_1)^2$ and $a_{i_1} z_{i_1} + a_{i_2} z_{i_2} = 0$. Since λ_{i_1} and λ_{i_2} are negative, appropriate values of z_{i_1} and z_{i_2} always exist, hence $x \in C_{a'z}$. ∎

PROOF OF THEOREM **5.5**   **(a)** Consider first the case where $\bar{A}_{22}$ is positive semidefinite with $\bar{A}_{22} \neq 0$. To cover this situation, it will be convenient to distinguish between 2 subcases: (a.1) $r_2 = G - 1$;   (a.2) $1 \leq r_2 < G - 1$.

(a.1)   If $r_2 = G - 1$, $\bar{A}_{22}$ is positive definite. From (5.15), we can write $\bar{Q}(\delta) = \bar{Q}(\delta_1, \delta_2)$. $\delta_1 \in C_{\delta_1}$ 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 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:

$$\frac{\partial \bar{Q}}{\partial \delta_2} = 2\bar{A}_{22}\delta_2 + 2\bar{A}_{21}\delta_1 + \bar{b}_2 = 0, \qquad \frac{\partial^2 \bar{Q}}{\partial \delta_2 \partial \delta_2'} = 2\bar{A}_{22}. \tag{A.13}$$

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

$$\tilde{\delta}_2 = -\frac{1}{2}\bar{A}_{22}^{-1}\left[2\bar{A}_{21}\delta_1 + \bar{b}_2\right] = -\bar{A}_{22}^{-1}\bar{A}_{21}\delta_1 - \frac{1}{2}\bar{A}_{22}^{-1}\bar{b}_2. \tag{A.14}$$

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

$$\bar{Q}(\delta_1, \tilde{\delta}_2) = \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \tag{A.15}$$

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 (5.20) holds with $S_1 = \emptyset$.

(a.2) If $1 \leq r_2 < G - 1$, we get, using (5.15), (5.17) - (5.19):

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

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:

$$\bar{Q}(\delta) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \tilde{\delta}_{2*}'D_{2*}\tilde{\delta}_{2*} + \left[2\tilde{A}_{21*}\delta_1 + \tilde{b}_{2*}\right]'\tilde{\delta}_{2*}. \tag{A.16}$$

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

$$\delta_1 \in C_{\delta_1} \text{ iff } \tilde{a}_1\delta_1^2 + \tilde{b}_1\delta_1 + \tilde{c}_1 \leq 0 \tag{A.17}$$

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)]:

$$\bar{A}_{22}^{+} = P_2 \begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0 \end{bmatrix} P_2' = [P_{21}, P_{22}] \begin{bmatrix} D_{2*}^{-1} & 0 \\ 0 & 0 \end{bmatrix} [P_{21}, P_{22}]' = P_{21}D_{2*}^{-1}P_{21}', \tag{A.18}$$

hence

$$\begin{aligned}
\bar{A}_{21*}'D_{2*}^{-1}\bar{A}_{21*} &= \bar{A}_{21}'P_{21}D_{2*}^{-1}P_{21}'\bar{A}_{21} = \bar{A}_{21}'\bar{A}_{22}^{+}\bar{A}_{21}, &(A.19) \\
\bar{A}_{21*}'D_{2*}^{-1}\bar{b}_{2*} &= \bar{A}_{21}'P_{21}D_{2*}^{-1}P_{21}'\bar{b}_2 = \bar{A}_{21}'\bar{A}_{22}^{+}\bar{b}_2, &(A.20) \\
\bar{b}_{2*}'D_{2*}^{-1}\bar{b}_{2*} &= \bar{b}_2'P_{21}D_{2*}^{-1}P_{21}'\bar{b}_2 = \bar{b}_2'\bar{A}_{22}^{+}\bar{b}_2. &(A.21)
\end{aligned}$$

(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 $\bar{Q}(\delta_1, \delta_2) < 0$,

42

which entails that $\delta_1 \in C_{\delta_1}$. Putting together the conclusions drawn in (i) and (ii) above, we see that

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

and (5.20) 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 $G = 1$ or $\bar{A}_{22} = 0$, we can write:

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

where we set $\bar{A}_{21} = \bar{b}_2 = 0$ when $G = 1$. 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 $G = 1$. 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 $\bar{A}_{22}$ is not positive semidefinite and $\bar{A}_{22} \neq 0$, this means we can find a vector δ_{20} such that $\delta'_{20}\bar{A}_{22}\delta_{20} \equiv q_0 < 0$. Now, for any scalar Δ_0, we have:

$$
\bar{Q}(\delta_1, \Delta_0\delta_{20}) = \bar{a}_{11}\delta_1^2 + \bar{b}_1\delta_1 + c + \Delta_0^2 q_0 + \Delta_0[2\bar{A}_{21}\delta_1 + \bar{b}_2]'\delta_{20}. \quad\quad\quad\quad \text{(A.24)}
$$

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

References

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

ANGRIST, J. D., G. W. IMBENS, AND A. B. KRUEGER (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earning?," *Quarterly Journal of Economics*, CVI, 979–1014.

ANGRIST, J. D., AND A. B. KRUEGER (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225–235.

BASU, S., AND J. G. FERNALD (1995): "Are Apparent Productive Spillovers a Figment of Specification Error?," *Journal of Monetary Economics*, 36, 165–188.

——— (1997): "Returns to Scale in U.S. Production : Estimates and Implications," *Journal of Political Economy*, 105(2), 249–283.

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

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

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 R. K. LYONS (1989): "The Role of External Economies in US Manufacturing," Discussion Paper 3033, National Bureau of Economic Research, Cambridge, Massachusetts.

——— (1992): "External Effects in U.S. Procyclical Productivity," *Journal of Monetary Economics*, 29, 209–225.

CHAO, J., AND N. R. SWANSON (2000): "Bias and MSE of the IV Estimators under Weak Identification," Discussion paper, Department of Economics, University of Maryland.

CHOI, I., AND P. C. B. PHILLIPS (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 S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press, New York.

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

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

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

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

DUFOUR, J.-M., AND J. JASIAK (1993): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," Discussion paper, C.R.D.E., Université de Montréal, 38 pages.

DUFOUR, J.-M., AND J. JASIAK (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 J. F. KIVIET (1998): "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrica*, 66, 79–104.

DUFOUR, J.-M., AND M. TAAMOUTI (2001a): "On Methods for Selecting Instruments," Discussion paper, C.R.D.E., Université de Montréal.

——— (2001b): "Point-Optimal Instruments and Generalized Anderson-Rubin Procedures for Nonlinear Models," Discussion paper, C.R.D.E., Université de Montréal.

FRANKEL, J. A., AND D. ROMER (1996): "Trade and Growth: An Empirical Investigation," Discussion Paper 5476, National Bureau of Economic Research, Cambridge, Massachusetts.

——— (1999): "Does Trade Cause Growth?," *The American Economic Review*, 89(3).

HAHN, J., AND J. HAUSMAN (2002a): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

——— (2002b): "Notes on Bias in Estimators for Simultaneous Equation Models," *Economics Letters*, 75, 237–241.

HALL, A. R., AND F. P. M. PEIXE (2000): "A Consistent Method for the Selection of Relevant Instruments," Discussion paper, Department of Economics, North Carolina State University, Raleigh, North Carolina.

HALL, A. R., G. D. RUDEBUSCH, AND D. W. WILCOX (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 *Growth, Productivity, Employment*, ed. by P. Diamond, pp. 71–112. The MIT Press, Cambridge, Massachusetts.

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.

IRWIN, A. D., AND M. TERVIO (2002): "Does Trade Raise Income? Evidence from the Twentieth Century," *Journal of International Economics*, 58, 1–18.

JORGENSON, D. W., F. M. GOLLOP, AND B. M. FRAUMENI (1987): *Productivity and U.S. Economic Growth*. Harvard University Press, Cambridge, Massachusetts.

KLEIBERGEN, F. (2001): "Testing Subsets of Structural Coefficients in the IV Regression Model," Discussion paper, Department of Quantitative Economics, University of Amsterdam.

——— (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

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

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

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

MANKIW, G. N., D. ROMER, AND D. N. WEIL (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107(2), 407–37.

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

MOREIRA, M. J. (2001): "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

——— (2002): "A Conditional Likelihood Ratio Test for Structural Models," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

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

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

PERRON, B. (1999): "Semi-Parametric Weak Instrument Regressions with an Application to the Risk Return Trade-Off," Discussion Paper 0199, C.R.D.E., Université de Montréal.

PETTOFREZZO, A. J., AND M. L. MARCOANTONIO (1970): *Analytic Geometry with Vectors*. Scott, Fosman and Company, Glenview, Illinois.

RODRIK, D. (1995): "Trade and Industrial Policy Reform," in *Handbook of Development Economics*, ed. by J. Behrman, and T. N. Srinivasan, vol. 3A. Elsevier Science, Amsterdam.

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, chap. 14, pp. 827–879. North-Holland, Amsterdam.

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

STARTZ, R., C. R. NELSON, AND E. ZIVOT (1999): "Improved Inference for the Instrumental Variable Estimator," Discussion paper, Department of Economics, University of Washington.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (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 M. YOGO (2002): "Testing for Weak Instruments in Linear IV Regression," Discussion Paper 284, N.B.E.R., Harvard University, Cambridge, Massachusetts.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.