Econometrica, Vol. 73, No. 4 (July, 2005), 1351–1365

PROJECTION-BASED STATISTICAL INFERENCE IN LINEAR STRUCTURAL MODELS WITH POSSIBLY WEAK INSTRUMENTS

BY JEAN-MARIE DUFOUR AND MOHAMED TAAMOUTI[1]

It is well known that standard asymptotic theory is not applicable or is very unreliable in models with identification problems or weak instruments. One possible way out consists of using a variant of the Anderson–Rubin ((1949), AR) procedure. The latter 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, but not for individual coefficients. This problem may in principle be overcome by using projection methods (Dufour (1997), Dufour and Jasiak (2001)). At first sight, however, this technique requires the application of costly numerical algorithms. In this paper, we give a general necessary and sufficient condition that allows one to check whether an AR-type confidence set is bounded. Furthermore, we provide an analytic solution to the problem of building projection-based confidence sets from AR-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 needed to build the confidence intervals.

KEYWORDS: Simultaneous equations, structural model, instrumental variable, weak instrument, confidence interval, testing, projection, quadric, exact inference, asymptotic theory.

1. INTRODUCTION

ONE OF THE CLASSIC PROBLEMS of econometrics consists of making inference on the coefficients of structural models. Recently, the statistical problems raised by such models 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 corresponds to situations where the structural parameters are close to being not identifiable (given the instruments used). The literature on this topic is now considerable; see the reviews by Stock, Wright, and Yogo (2002) and Dufour (2003).

In view of the unreliability of asymptotic arguments in such setups, we focus here on procedures for which finite-sample pivotallity obtains under standard assumptions. The oldest one appears to be the statistic proposed by Anderson and Rubin ((1949), henceforth AR). The latter is a limited-information

[1]The authors thank Laurence Broze, John Cragg, Jean-Pierre Florens, Christian Gouriéroux, Joanna Jasiak, Frédéric Jouneau, Lynda Khalaf, Nour Meddahi, Benoît Perron, Tim Vogelsang, Eric Zivot, three anonymous referees, and a co-editor for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence (program on Mathematics of Information Technology and Complex Systems), the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de Recherche sur la Société et la Culture (Québec), and the Fonds de Recherche sur la Nature et les Technologies (Québec). Taamouti was also supported by a fellowship from the Canadian International Development Agency. Dufour holds the Canada Research Chair (Econometrics).

method that allows one to test a hypothesis by setting 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. It is completely robust to the presence of weak instruments. Other potential pivots aimed at being robust to weak instruments have recently been suggested by Wang and Zivot (1998), Kleibergen (2002), and Moreira (2003). However, only asymptotic distributional theories have been supplied for these statistics, so that the level of the procedures may not be controlled in finite samples, even under restrictive Gaussian distributional assumptions. It is important to note here that the quality of asymptotic approximations typically depends on the exogenous regressors (or the instruments) involved: no error bound is available and, of course, simulation evidence never can be viewed as a substitute for an analytical theory. Even under the parametric distributional assumptions that underlie the AR Fisher distribution, this appears to be the case. By contrast, the null distribution of the AR statistic is invariant to the numerical values of the instruments. In our view, this is a fundamental difference, especially in dealing with weak instrument problems where large-sample arguments can be especially misleading (for further discussion of these issues, see Dufour (1997, 2003)).[2]

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 one to test linear restrictions on the vector β. A general solution to this problem is the *projection* technique described by Dufour (1990, 1997), Wang and Zivot (1998), and Dufour and Jasiak (2001). However, a drawback of the projection approach comes from the fact that it can be numerically costly: for example, in Dufour and Jasiak (2001), such confidence intervals were derived for an empirical example, but nonlinear optimization methods (based on Fortran IMSL routines) had to be used.

In this paper, we study some general geometric features of AR-type confidence sets and we provide a closed-form solution to the problem of building projection-based confidence sets from AR sets. *First*, we observe that AR-type confidence sets can be described as *quadrics*, a class of geometric figures that

[2]As a *limited-information* method, the AR procedure may involve an efficiency loss with respect to *full-information* methods, but does allow for a less complete specification of the model and more robustness (for further discussion of this point, see Dufour and Taamouti (2004)). Note also that proposed exact or asymptotic pivots in this context typically take for granted a number of structural restrictions that characterize the specification of the structural equation. If the assumptions suggested by the structural model are relaxed, e.g., by considering the corresponding unrestricted reduced form, the AR statistic as well as most other pivots may cease to be pivotal (see Forchini and Hillier (2003)). Here, as in most of the literature on weak instruments, we focus on the situation where the structural restrictions are maintained.

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 identifiability). *Second*, we derive simple explicit expressions for projection-based confidence intervals in the case of coefficient linear transformations, so that search by nonlinear methods is no longer required.

In Section 2, we present the background model and the basic statistical method considered. Section 3 presents the quadric 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 conclude in Section 6.

2. FRAMEWORK

We consider here a standard simultaneous equations model (SEM)

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

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

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. Furthermore,

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

where $k = k_1 + k_2$. Finally, to get a finite-sample distributional theory for the test statistics, we use the standard assumptions

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

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

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 (i.e., the matrix Π_2 does not have 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 on 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 of inverting Wald-type statistics to obtain confidence sets (for potentially unidentified parameters), is not valid in these situations since the resulting confidence sets are bounded with probability 1. This is related to the fact that such statistics are not pivotal and follow distributions that depend heavily on nuisance parameters. More precisely, for any sample size, the *confidence level* of Wald-type confidence sets (i.e., the minimum value of the coverage probability over all possible values of the nuisance parameters) is equal to zero (for the definition of the level of a confidence set, see Lehmann (1986, Section 3.5)). In particular, as shown in Dufour (1997), this holds even if the "identifying restriction" $\operatorname{rank}(\Pi_2) = G$ is imposed.

A first solution to this problem (see Dufour (1997) and Staiger and Stock (1997)) consists of using the Anderson–Rubin statistic (Anderson and Rubin (1949)). To test $H_0 : \beta = \beta_0$ in equation (2.1), the test statistic is given by

$$\text{(2.6)} \qquad \mathrm{AR}(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)},$$

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. Under the assumptions (2.3)–(2.5), we have under $H_0 : \mathrm{AR}(\beta_0) \sim F(k_2, T - k)$. This test also remains asymptotically valid under weaker distributional assumptions, in the sense that the asymptotic null distribution of $\mathrm{AR}(\beta_0)$ is $\chi^2(k_2)/k_2$; see Dufour and Jasiak (2001) and Staiger and Stock (1997). The distributional result in (2.6) holds irrespective of the rank of the matrix Π_2, which means that tests based on $\mathrm{AR}(\beta_0)$ are *robust to weak instruments*. A confidence set for β with level $1 - \alpha$ can also be obtained by inverting the above test,

$$\text{(2.7)} \qquad C_\beta(\alpha) = \{\beta_0 : \mathrm{AR}(\beta_0) \le F_\alpha(k_2, T - k)\},$$

where $F_\alpha(k_2, T - k)$ is the $1 - \alpha$ quantile of the F distribution with $(k_2, T - k)$ degrees of freedom.

Below, we shall also consider two alternative statistics proposed by Wang and Zivot (1998). The first one is a likelihood ratio (LR)-type statistic and the second is a Lagrange multiplier (LM)-type statistic. Under the assumptions (2.1)–(2.5) 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 overidentified ($k_2 > G$). To test $H_0 : \beta = \beta_0$, these statistics are

$$\text{(2.8)} \qquad \mathrm{LR}_{\mathrm{LIML}}(\beta_0) = T\big\{\ln[\kappa(\beta_0)] - \ln[\kappa(\hat{\beta}_{\mathrm{LIML}})]\big\},$$

$$\text{(2.9)} \qquad \mathrm{LM}_{\mathrm{2SLS}}(\beta_0) = \frac{T(y - Y\beta_0)'P[P[M(X_1)X_2]Y](y - Y\beta_0)}{(y - Y\beta_0)'M(X_1)(y - Y\beta_0)},$$

where $\kappa(\beta_0) = (y - Y\beta_0)'M(X_1)(y - Y\beta_0)/(y - Y\beta_0)'M(X)(y - Y\beta_0)$. Asymptotic and conservative confidence sets for β can be obtained by inverting the latter tests.

A common shortcoming of all these tests is that they require one to specify the entire vector β. 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 $g(\beta) = \beta_i - \beta_{i0}$, 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 derive closed-form confidence sets for the components of β or for linear combinations of these components.[3] We will show that 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*; see Shilov (1961, Chapter 11) and Pettofrezzo and Marcoantonio (1970, Chapters 9 and 10). We will then classify possible cases as functions of A, b, and c, and we will derive analytic expressions for projection-based confidence sets (or intervals) on linear transformations of model parameters.

3. ANDERSON–RUBIN-TYPE CONFIDENCE SETS

Let us first consider the AR statistic. A simple algebraic calculation shows that the inequality $\mathrm{AR}(\beta_0) \le F_\alpha(k_2, T-k)$ may be written in the simple form

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

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

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

We can thus write

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

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

$$H_{\mathrm{LR}} = M(X_1) - M(X)\kappa(\hat{\beta}_{\mathrm{LIML}}) \exp[\chi^2_\alpha(k_2)/T], \tag{3.4}$$

while for $\mathrm{LM}_{\mathrm{2SLS}}(\beta_0)$ it is

$$H_{\mathrm{LM}} = P\big[P[M(X_1)X_2]Y\big] - M(X_1)[\chi^2_\alpha(k_2)/T]. \tag{3.5}$$

[3]This problem was also considered by Stock and Wright (2000), Kleibergen (2001), and Startz, Zivot, and Nelson (2003), but the solutions provided rely on large-sample approximations and require additional identification assumptions.

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 $f_\alpha = \exp[\chi^2_\alpha(k_2)/T] \times \kappa(\hat{\beta}_{\mathrm{LIML}}) - 1$ for the LR statistic. Clearly A is symmetric with diagonal elements of the form $A_{ii} = Y_i'M(X_1)Y_i - Y_i'M(X)Y_i(1 + f_\alpha)$, where A_{ii} 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 can 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. In general, $Y'HY$ is not positive definite and may have both positive and negative eigenvalues. In the next section, we will show that $C_\beta(\alpha)$ is bounded if and only if all the eigenvalues of $Y'HY$ are positive; in particular, negative eigenvalues occur with high probability when identification conditions are not satisfied. 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 $b_i = -2\{[M(X_1)Y_i]'[M(X_1)y] - [M(X)Y_i]'[M(X)y](1 + f_\alpha)\}$. 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 $\beta' A\beta + b'\beta + c = 0$, where A is a symmetric $G \times G$ matrix, b is a $G \times 1$ vector, and c is a scalar, constitutes a *quadric* surface. These include as special cases various figures such as ellipsoids, paraboloids, hyperboloids, and cones. Consequently, we shall call a confidence set of the form

$$C_\beta = \{\beta_0 : \beta_0' A\beta_0 + b'\beta_0 + c \leq 0\} \tag{4.1}$$

a *quadric confidence set*. A quadric is characterized by the sum of a quadratic form ($\beta_0' A\beta_0$) and an affine transformation ($b'\beta_0 + c$). 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") confidence sets. For these reasons, we call A the *concentration matrix at level* α (or the α*-concentration matrix*) associated with β. It will be convenient here 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

$$\beta_0' A\beta_0 + b'\beta_0 + c = (\beta_0 - \tilde{\beta})' A(\beta_0 - \tilde{\beta}) - d, \tag{4.2}$$

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

$$A = P'DP, \tag{4.3}$$

where P is an orthogonal matrix and D is a diagonal matrix whose elements are the eigenvalues of A. The inequality $\beta_0' A\beta_0 + b'\beta_0 + c \leq 0$ may then be reexpressed as

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d, \tag{4.4}$$

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_β is bounded if and only if (iff) C_z is bounded, where $C_\beta = \{\beta : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d \text{ and } z = P(\beta - \tilde{\beta})\}$ and $C_z \equiv \{z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d\}$. Again it will be convenient to distinguish between three cases according to the signs of the eigenvalues of A, namely (a) all the eigenvalues of A are positive ($\lambda_i > 0, i = 1, \ldots, G$), i.e., A is positive definite; (b) all the eigenvalues of A are negative ($\lambda_i < 0, i = 1, \ldots, G$), i.e., A is negative definite; (c) A has both positive and negative eigenvalues, i.e., A is neither positive nor negative definite.

(a) *Positive definite concentration matrix.* If $\lambda_i > 0, i = 1, \ldots, G$, the inequality (4.4) can be reexpressed as

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \leq d, \tag{4.5}$$

where $\gamma_i = \sqrt{1/\lambda_i}, i = 1, \ldots, G$. If $d = 0$, we have $C_z = \{0\}$ and $C_\beta = \{\tilde{\beta}\}$. If $d < 0$, C_z and C_β are empty. If $d > 0$, C_z is the area inside or on an ellipsoid. Thus, C_z and C_β are bounded.

(b) *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

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \geq -d, \tag{4.6}$$

where $\gamma_i = \sqrt{-1/\lambda_i}$. Since (4.6) holds as soon as any $|z_i|$ is large enough, C_z and C_β are unbounded sets. In particular, if $d \geq 0$, we have $C_\beta = C_z = \mathbb{R}^G$.

(c) *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 (4.4) may then be rewritten

$$(4.7) \qquad \left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_p}{\gamma_p}\right)^2 - \left(\frac{z_{p+1}}{\gamma_{p+1}}\right)^2 - \cdots - \left(\frac{z_G}{\gamma_G}\right)^2 \leq d,$$

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 $i = p+1, \ldots, G$. Then, for arbitrary given values of $z_1, \ldots, z_p$ and d, it is clear that inequality (4.7) 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. This entails that C_z and C_β are unbounded.

4.2. *Singular Concentration Matrix*

We now consider the case where A is singular with rank r ($r < G$). First, if $A = 0$ (i.e., $r = 0$), it is easy to see that the only situation where C_β can be bounded is the one where $b = 0$ and $c > 0$ (in which case C_β is empty). So we can focus on the case where $A \neq 0$, hence $r \geq 1$ and $G - r \geq 1$. 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 (4.3) are different from zero, while the $G - r$ other ones are equal to zero. Then we can write

$$(4.8) \qquad Q(\beta) \equiv \beta' A \beta + b'\beta + c = \sum_{i=1}^{r} \lambda_i z_i^2 + \sum_{i=r+1}^{G} \delta_i z_i - d,$$

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

$$(4.9) \qquad d = -c + \sum_{i=1}^{r} \frac{\delta_i^2}{4\lambda_i}, \qquad \mu_i = \begin{cases} \delta_i/(2\lambda_i), & \text{if } \lambda_i \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

If $b = 0$, we have $Q(\beta) = \sum_{i=1}^{r} \lambda_i z_i^2 + c$ and the values of $z_{r+1}, \ldots, z_G$ can be as large as we wish without affecting the value of $Q(\beta)$. Then C_β is either empty (when $c > 0$ and $\lambda_i > 0$, $i = 1, \ldots, r$) or unbounded (in all the other cases). If $b \neq 0$, there is at least one $k \in \{r+1, \ldots, G\}$ such that $\delta_k \neq 0$. Then we can set $z_j = 0$ for $j \neq k$, and choose z_k such that $|z_k|$ is arbitrarily large and the inequality (4.4) is satisfied. This entails that C_β is unbounded.

4.3. *Necessary and Sufficient Condition for a Bounded Quadric Confidence Set*

Following Gleser and Hwang (1987) and Dufour (1997), a valid confidence set C_β for β (with level $1-\alpha$) in model (2.1)–(2.5) 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 not having 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. On putting together the different cases discussed above, we get the following easy-to-verify necessary and sufficient condition for a quadric confidence set to be bounded.

THEOREM 4.1: *If the matrix A is nonsingular, the set C_β in* (4.1) *is bounded if and only if the matrix A is positive definite. If A is singular, the set C_β is bounded only when it is empty, and C_β is empty if and only if A is positive semidefinite, $b=0$, and $c>0$.*

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 $T\times G$ and $T\times T$ matrices, respectively. If Y follows a nondegenerate absolutely continuous distribution (as assumed in Section 2), A will be nonsingular with probability 1 as soon as the rank of H is greater than or equal to G. In the rest of this paper, we will thus focus on the case of a nonsingular concentration matrix.[4]

5. CONFIDENCE SETS FOR TRANSFORMATIONS OF β

We consider now a general confidence set of the form

$$C_\beta = \{\beta_0 : \beta_0' A\beta_0 + b'\beta_0 + c \le 0\}, \tag{5.1}$$

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 $g(\beta) = w'\beta$ is

$$C_{w'\beta} \equiv g[C_\beta] = \{\delta_0 : \delta_0 = w'\beta_0 \text{ where } \beta_0' A\beta_0 + b'\beta_0 + c \le 0\}, \tag{5.2}$$

where w is a nonzero $G\times 1$ vector. 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:

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

[4]The case where the concentration matrix is singular is discussed in a companion working paper (Dufour and Taamouti (2004)).

Furthermore,

$$w'\beta = w'P'P\beta = w'P'P(\beta - \tilde{\beta}) + w'P'P\tilde{\beta} = a'z + w'\tilde{\beta}, \tag{5.3}$$

where $a = Pw$. Setting

$$C_{a'z} = \{a'z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d\}, \tag{5.4}$$

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

$$x \in C_{w'\beta} \quad \Leftrightarrow \quad x - w'\tilde{\beta} \in C_{a'z}, \tag{5.5}$$

hence, $C_{w'\beta} = \mathbb{R} \Leftrightarrow C_{a'z} = \mathbb{R}$. We will now distinguish three cases that depend on the number of negative eigenvalues: (i) all the eigenvalues of A are positive (i.e., A is positive definite); (ii) A has exactly one negative eigenvalue; (iii) 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. The following proposition then provides an explicit form for the projection-based confidence set $C_{w'\beta}$.

THEOREM 5.1: *Let C_β be the set defined in* (5.1), $d \equiv \frac{1}{4}b'A^{-1}b - c$, *let w be a nonzero vector in $\mathbb{R}^G$, and suppose the matrix A is positive definite. If $d \geq 0$, then*

$$C_{w'\beta} = \left[w'\tilde{\beta} - \sqrt{d(w'A^{-1}w)},\, w'\tilde{\beta} + \sqrt{d(w'A^{-1}w)}\right], \tag{5.6}$$

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

Proofs are provided in the Appendix. Note the case where A is positive definite is one where the instruments X_2 provide additional explanatory power for Y (with respect to X_1): the number of strong instruments is sufficient to pin down all parameters (which suggests a traditional identification condition holds). Let us now consider the case where A has exactly one negative eigenvalue.

THEOREM 5.2: *Let C_β be the set defined in* (5.1), $d \equiv \frac{1}{4}b'A^{-1}b - c$, $w \in \mathbb{R}^G\backslash\{0\}$, *and suppose the matrix A is nonsingular with exactly one negative eigenvalue. If $w'A^{-1}w < 0$ and $d < 0$, then*

$$C_{w'\beta} = \left]-\infty, w'\tilde{\beta} - \sqrt{d(w'A^{-1}w)}\right] \cup \left[w'\tilde{\beta} + \sqrt{d(w'A^{-1}w)}, +\infty\right[. \tag{5.7}$$

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 that $C_{w'\beta}$ can remain informative even if it is unbounded. In particular, if we wish to test $H_0 : w'\beta = r$ and consider a decision rule that rejects H_0 when $r \notin C_{w'\beta}$, H_0 will be rejected for all values of r outside the interval $C_{w'\beta}$ in (5.7). This situation suggests that the rank condition for identification fails, but some parameters may still be identifiable, i.e., some components (or linear transformations) of β are identifiable while others are not.

Finally, we consider the case where A has at least two negative eigenvalues.

THEOREM 5.3: *Let C_β be the set defined in* (5.1) *and let $w \in \mathbb{R}^G\backslash\{0\}$. If the matrix A in* (5.1) *is nonsingular and has at least two negative eigenvalues*, *then* $C_{w'\beta} = \mathbb{R}$.

In the latter case, the projection-based confidence set for $w'\beta$ is equal to the real line, thus uninformative. No linear combination of the elements of β appears to be identifiable.

In summary, our recommended method for constructing confidence intervals for a single coefficient can be summarized as follows:

1. Compute A, b, and c as defined in (3.1), and find the eigenvalues of A.
2. If all the eigenvalues of A are positive (which entails that C_β is bounded), use (5.6) to compute the confidence intervals of interest $C_{w'\beta}$.
3. If A has exactly one negative eigenvalue, use Theorem 5.2 to compute $C_{w'\beta}$.
4. If A has more than one negative eigenvalue, set $C_{w'\beta} = \mathbb{R}$.

6. CONCLUSION

Recent research in econometrics has shown that weak instruments are quite widespread and should be carefully addressed. Techniques that 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 subvectors). In this paper, we studied this problem from a finite-sample limited-information viewpoint and focused on AR-type tests and confidence sets.

We observed that AR-type confidence sets belong to a class of sets defined by 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 closed-form 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 in the case with nonsingular quadratic form. The confidence sets so obtained turn out to be as easy to compute as standard Wald-type 2SLS-based confidence intervals. Simulation evidence on the performance of projection-based confidence sets as well as empirical illustrations are available in Dufour and Taamouti (2004).

Departement de Sciences Économiques, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7; and Centre Interuniversitaire de Recherche en Analyse des Organisations and Centre Interuniversitaire de Recherche en Économie Quantitative, Université de Montréal; jean.marie.dufour@umontreal.ca

and

Centre Interuniversitaire de Recherche en Économic Quantitative, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7; and INSEA, Rabat, Morocco; taamouti@insea.ac.ma.

Manuscript received June, 2003; final revision received November, 2004.

APPENDIX: PROOFS

PROOF OF THEOREM 5.1: Consider again the decomposition $A = P'DP$ in (4.3). By (5.5), 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 there is a vector $z \in \mathbb{R}^G$ such that

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

Furthermore, there is a z that verifies (A.1) iff the solution of the problem

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

verifies the constraint (A.1). If $d < 0$, it is clear there is no solution that verifies (A.1)—for D is positive definite—and consequently $C_{a'z} = C_{w'\beta} = \emptyset$. Let $d \geq 0$. 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 $2Dz = \mu a$ and $a'z = x$; hence, $\mu = 2x/(a'D^{-1}a)$, $z = x/(a'D^{-1}a)$, and $z'Dz = \mu x/2 = x^2/(a'D^{-1}a)$. Thus

$$\begin{aligned} x \in C_{a'z} \quad &\Leftrightarrow \quad \frac{x^2}{a'D^{-1}a} \leq d \quad \Leftrightarrow \quad |x| \leq \sqrt{d(a'D^{-1}a)} \\ &\Leftrightarrow \quad |x_0 - w'\tilde{\beta}| \leq \sqrt{d(a'D^{-1}a)}. \end{aligned}$$

On noting that $a'D^{-1}a = w'A^{-1}w$, this entails that the confidence set for $w'\beta$ is given by (5.6). *Q.E.D.*

PROOF OF THEOREM 5.2: As in the proof of Proposition 5.1, let us consider again the decomposition (4.3), the equivalence $x_0 \in C_{w'\beta} \Leftrightarrow x_0 - w'\tilde{\beta} \in C_{a'z}$, and 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

$$a'z = a_1z_1 + \cdots + a_{G-1}z_{G-1} + a_Gz_G = x, \tag{A.3}$$

$$z'Dz = \lambda_1 z_1^2 + \cdots + \lambda_{G-1}z_{G-1}^2 - |\lambda_G|z_G^2 \leq d, \tag{A.4}$$

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 $D_{(G)} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{G-1})'$.

If $a_G = 0$, then $a_{(G)} \neq 0$ (because $w \neq 0$ entails $a \neq 0$) 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 $a_1z_1 + \cdots + a_{G-1}z_{G-1} = x$ 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 $a_G \neq 0$. Then, the conditions (A.3) and (A.4) are equivalent to

$$z_G = \frac{x - a'_{(G)}z_{(G)}}{a_G}, \tag{A.5}$$

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

where the latter inequality can also be written as

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

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 $z_{(G)}$ that satisfies inequality (A.7). Furthermore, 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

$$\min_{z_{(G)}} z'_{(G)}D_{(G)}z_{(G)} \quad \text{s.c.} \quad a'_{(G)}z_{(G)} = s, \tag{A.8}$$

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 $2D_{(G)}z_{(G)} = \mu a_{(G)}$ and $a'_{(G)}z_{(G)} = s$; hence,

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

$$z_{(G)} = \frac{s}{a'_{(G)} D^{-1}_{(G)} a_{(G)}} D^{-1}_{(G)} a_{(G)},$$

$$z'_{(G)} D_{(G)} z_{(G)} = \frac{s^2}{a'_{(G)} D^{-1}_{(G)} a_{(G)}},$$

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

$$\text{(A.9)} \qquad qs^2 - (2|\lambda_G|x)s + (|\lambda_G|x^2 + da_G^2) \geq 0,$$

where $q = |\lambda_G| - [a_G^2/a'_{(G)} D^{-1}_{(G)} a_{(G)}] = \delta_G(w'A^{-1}w)$ and $\delta_G \equiv |\lambda_G|/a'_{(G)} \times D^{-1}_{(G)} a_{(G)} > 0$. Thus, $x \in C_{a'z}$ iff there is a value of s such that (A.9) holds. The discriminant of this second degree equation is $\Delta = 4\lambda_G^2 x^2 - 4q(|\lambda_G|x^2 + da_G^2) = 4\delta_G a_G^2[x^2 - d(w'A^{-1}w)]$.

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

1. 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.9) 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).

2. If $w'A^{-1}w < 0$ and $d < 0$, then $q < 0$ and (A.9) has a (real) solution iff $\Delta \geq 0$ or, equivalently, $x^2 \geq d(w'A^{-1}w) > 0$. Consequently,

$$\text{(A.10)} \qquad C_{a'z} = \left]-\infty, -\sqrt{d(w'A^{-1}w)}\right] \cup \left[\sqrt{d(w'A^{-1}w)}, +\infty\right[,$$

$$\text{(A.11)} \qquad C_{w'\beta} = \left]-\infty, w'\tilde{\beta} - \sqrt{d(w'A^{-1}w)}\right] \cup \left[w'\tilde{\beta} + \sqrt{d(w'A^{-1}w)}, +\infty\right[.$$

3. If $w'A^{-1}w = 0$ and $d < 0$, (A.9) can be satisfied for any $x \neq 0$; hence, $C_{a'z} = \mathbb{R}\backslash\{0\}$ and $C_{w'\beta} = \mathbb{R}\backslash\{w'\tilde{\beta}\}$.

4. Finally, if $d \geq 0$, (A.9) 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. *Q.E.D.*

PROOF OF THEOREM 5.3: 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 the components of z such that (i) $z_1 = (x - a_{i_2} z_{i_2})/a_1$; (ii) $z_i = 0$ for $i > 1$, $i \neq i_2$; (iii) $\lambda_1 z_1^2 + \lambda_{i_2} z_{i_2}^2 \leq d$. Since λ_{i_1} and λ_{i_2} are negative, z_{i_2} does exist. The vector z verifies (4.4) 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 (i) $z_1 = x/a_1$; (ii) $z_i = 0$ for $i \neq i_1, i \neq i_2$, and $i > 1$; (iii) $\lambda_{i_1} z_{i_1}^2 + \lambda_{i_2} z_{i_2}^2 \leq d - \lambda_1(x/a_1)^2$ 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, $x \in C_{a'z}$. *Q.E.D.*

REFERENCES

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

DUFOUR, J.-M. (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.

——— (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767–808.

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 M. TAAMOUTI (2004): "Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments," Discussion Paper, Département de Sciences Économiques, Université de Montréal.

FORCHINI, G., AND G. HILLIER (2003): "Conditional Inference for Possibly Unidentified Structural Equations," *Econometric Theory*, 19, 707–743.

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

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, 1781–1803.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses* (Second Ed.). New York: Wiley.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

PETTOFREZZO, A. J., AND M. L. MARCOANTONIO (1970): *Analytic Geometry with Vectors*. Glenview, IL: Scott, Foresman.

SHILOV, G. E. (1961): *An Introduction to the Theory of Linear Spaces*. Englewood Cliffs, NJ: Prentice-Hall.

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

STARTZ, R., E. ZIVOT, AND C. R. NELSON (2003): "Improved Inference in Weakly Identified Instrumental Variables Regression," Discussion Paper, Department of Economics, University of Washington, Seattle.

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 & Economic Statistics*, 20, 518–529.

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