Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 2649–2661
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Instrument endogeneity and identification-robust tests: Some analytical results

Firmin Doko[a], Jean-Marie Dufour[b,*,1]

[a]Département de Sciences Économiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, QC, Canada H3C 3J7
[b]Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec, Canada H3A 2T7
Available online 12 March 2008

Abstract

When some explanatory variables in a regression are correlated with the disturbance term, instrumental variable methods are typically employed to make reliable inferences. Furthermore, to avoid difficulties associated with weak instruments, identification-robust methods are often proposed. However, it is hard to assess whether an instrumental variable is valid in practice because instrument validity is based on the questionable assumption that some of them are exogenous. In this paper, we focus on structural models and analyze the effects of instrument endogeneity on two identification-robust procedures, the Anderson–Rubin (1949, AR) and the Kleibergen (2002, K) tests, with or without weak instruments. Two main setups are considered: (1) the level of "instrument" endogeneity is fixed (does not depend on the sample size) and (2) the instruments are *locally exogenous*, i.e. the parameter which controls instrument endogeneity approaches zero as the sample size increases. In the first setup, we show that both test procedures are in general consistent against the presence of invalid instruments (hence asymptotically invalid for the hypothesis of interest), whether the instruments are "strong" or "weak". We also describe cases where test consistency may not hold, but the asymptotic distribution is modified in a way that would lead to size distortions in large samples. These include, in particular, cases where the 2SLS estimator remains consistent, but the AR and K tests are asymptotically invalid. In the second setup, we find (non-degenerate) asymptotic non-central chi-square distributions in all cases, and describe cases where the non-centrality parameter is zero and the asymptotic distribution remains the same as in the case of valid instruments (despite the presence of invalid instruments). Overall, our results underscore the importance of checking for the presence of possibly invalid instruments when applying "identification-robust" tests.

Keywords: Simultaneous equations; Instrumental variables; Locally weak instruments; Invalid instruments; Locally exogenous instruments

1. Introduction

The last decade shows growing interest for so-called *weak-instrument* problems in the econometric literature, i.e. situations where "instruments" are poorly correlated with endogenous explanatory variables; see the reviews of Dufour (2003) and Stock et al. (2002). More generally, these can be viewed as situations where model parameters are not identified or close not to being *identifiable*, as meant in the econometric literature (see Dufour and Hsiao, 2008).

* Corresponding author. Tel.: +1 514 398 8879; fax: +1 514 398 4938.
E-mail addresses: f.doko.tchatoka@umontreal.ca (F. Doko), jean-marie.dufour@mcgill.ca (J.-M. Dufour)
URL: http://www.jeanmariedufour.com (J.-M. Dufour).
[1] William Dow Professor of Economics, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ), McGill University.

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.03.006

When instruments are weak, the limiting distributions of standard test statistics—like Student, Wald, likelihood ratio and Lagrange multiplier criteria in structural models—often depend heavily on nuisance parameters; see e.g. Phillips (1989), Bekker (1994), Dufour (1997), Staiger and Stock (1997) and Wang and Zivot (1998). In particular, standard Wald-type procedures based on the use of asymptotic standard errors are very unreliable in the presence of weak identification. As a result, several authors have worked on proposing more reliable statistical procedures that would be applicable in such contexts.

Interestingly, in the early days of simultaneous-equation econometrics, Anderson and Rubin (1949, AR) proposed a procedure which is completely robust to weak instruments as well as to other difficulties such as missing instruments (see Dufour, 2003; Dufour and Taamouti, 2005, 2006). But the AR procedure may suffer from power losses when too many instruments are used. So alternative methods largely try to palliate this difficulty, for example: pseudo-pivotal LM-type and LR-type statistics (Wang and Zivot, 1998; Kleibergen, 2002; Moreira, 2003), sample-splitting methods (Dufour and Jasiak, 2001), approximately optimal instruments (Dufour and Taamouti, 2003), systematic search methods for identifying relevant instruments and excluding unimportant instruments (Hall et al., 1996; Hall and Peixe, 2003; Dufour and Taamouti, 2003; Donald and Newey, 2001).

However, all these procedures—including the AR method—rely on the availability of valid (exogenous) instruments. This raises the question: what happens to these procedures when some of the instruments are endogenous? In particular, what happens if an invalid instrument is added to a set of valid instruments? How robust are these inference procedures to instrument endogeneity? Do alternative inference procedures behave differently? If yes, what is their relative performance in the presence of instrument endogeneity?

We view the problem of instrument endogeneity as important because it is hard in practice to assess whether an instrumental variable is valid, i.e. whether it is uncorrelated with the disturbance term. Instrument validity or orthogonality tests are built on the availability of a number of undisputed valid instruments, at least as great as the number of coefficients to be estimated, whereas the validity of those initial instruments is not testable.

In the econometric literature, little is known about test procedures when some instruments are both invalid and weak. Hahn and Hausman (2003) deal with both instrument endogeneity and weakness, but they focus on estimation. Ashley (2006) proposed a sensitivity analysis of IV estimators when instruments are imperfect; his results, however, are only applicable if the covariance between the structural error term and some instruments is known, which is not necessarily the case as it is shown in this paper. Analyzing the effect of instrument invalidity on the limiting and empirical distribution of IV estimators, Kiviet and Niemczyk (2006) conclude that for the accuracy of asymptotic approximations, instrument weakness is much more detrimental than instrument invalidity and that the realizations of IV estimators based on strong but possibly invalid instruments seem usually much closer to the true parameter values than those obtained from valid but weak instruments. However, this finding of Kiviet and Niemczyk leaves open crucial questions: is it really possible to make reliable inference with endogenous instruments? Is instrument endogeneity really more detrimental than its weakness for inference procedures like a general family of Anderson–Rubin-type procedures? Swanson and Chao (2005) proposed a weak-instrument unified framework, but they do not take into account possible invalidity of some instruments. Finally, Small (2007) recently studied the properties of tests for identifying restrictions (Sargan, 1958; Kadane and Anderson, 1977), which can be sensitive to the use of "endogenous instruments", and he proposed a sensitivity analysis to assess the importance of the issue. These results, however, do not allow for weak identification.

In this paper, we focus on structural models and analyze the effects of instrument endogeneity on the Anderson and Rubin (1949) and Kleibergen (2002) tests, in the presence of possibly weak instruments. After formulating a general asymptotic framework which allows one to study these issues in a convenient way, we consider two main setups: (1) the one where the level of "instrument" endogeneity is fixed (i.e. it does not depend on the sample size) and (2) the one where the instruments are *locally exogenous*, i.e. the parameter which controls instrument endogeneity approaches zero (at rate $T^{-1/2}$) as the sample size increases. In the first setup, we show that both test procedures studied are in general consistent against the presence of invalid instruments (hence asymptotically invalid for the hypothesis of interest), whether the instruments are " strong" or "weak". We also observe there are cases where consistency may not hold, but the asymptotic distribution is modified in a way that would lead to size distortions in large samples. In the second setup, asymptotic non-central chi-square distributions are derived, and we give conditions under which the non-centrality parameter is zero and the asymptotic distribution remains the same as in the case of valid instruments (despite the presence of invalid instruments). Overall, our results underscore the importance of checking for the presence of possibly invalid instruments when applying " identification-robust" tests.

The paper is organized as follows. Section 2 formulates the model considered. Section 3 describes briefly the statistics. Section 4 studies the asymptotic distribution of the statistics (under the null hypothesis) when some instruments are invalid. We conclude in Section 5. Proofs are presented in the Appendix.

2. Framework

We consider the following standard simultaneous-equation framework, which has been the basis of much work on inference in model with possibly weak instruments (see the reviews of Dufour, 2003; Stock et al., 2002):

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

$$Y = X\Pi + Z\Gamma + V, \tag{2.2}$$

where y is a $T \times 1$ vector of observations on the dependent variable, $Y = [Y_1, \ldots, Y_T]'$ is a $T \times G$ matrix of observations on explanatory (possibly) endogenous variables ($G \geqslant 1$), Z is a $T \times r$ matrix of observations on the included exogenous variables, $X = [X_1, \ldots, X_T]'$ is a $T \times k (k \geqslant G)$ full-column-rank matrix of observations on (supposedly) "exogenous variables" (instruments) excluded from the structural equation (2.1), $u = [u_1, \ldots, u_T]'$ and $V = [V_1, \ldots, V_T]' = [v_1, \ldots, v_G]$ are, respectively, $T \times 1$ vector and $T \times G$ disturbance matrices, β and γ are $G \times 1$ and $r \times 1$ vectors of unknown coefficients, Π and Γ are $k \times G$ and $r \times G$ matrices of unknown coefficients. The usual necessary and sufficient condition for identification of this model is $\operatorname{rank}(\Pi) = G$.

Since we focus on the parameter β in our analysis, we can simplify the presentation of the results without notable loss of generality by setting $\gamma = 0$ and $\Gamma = 0$, so that Z drops from the model. With this simplification, model (2.1)–(2.2) reduces to

$$y = Y\beta + u, \tag{2.3}$$

$$Y = X\Pi + V. \tag{2.4}$$

We also assume that

$$u_t = V_t' a + \varepsilon_t, \quad t = 1, \ldots, T, \tag{2.5}$$

$$X_t = X_{0t} + W_t, \quad t = 1, \ldots, T, \tag{2.6}$$

$$u_t = W_t' b + e_t, \quad t = 1, \ldots, T, \tag{2.7}$$

where $X_0 = [X_{01}, \ldots, X_{0T}]'$ is a $T \times k$ matrix of exogenous variables, ε_t is uncorrelated with V_t, and e_t are uncorrelated with W_t. V_t and W_t have mean zero and covariance matrices Σ_V and Σ_W, ε_t and e_t have mean zero and variances σ_ε^2 and σ_e^2, respectively, while a and b are $G \times 1$ and $k \times 1$ vectors of unknown coefficients. Eqs. (2.5)–(2.7) can be rewritten in matrix form as

$$u = Va + \varepsilon, \tag{2.8}$$

$$X = X_0 + W, \tag{2.9}$$

$$u = Wb + e, \tag{2.10}$$

where X_0 is uncorrelated with W, V, ε and e, while $W = [W_1, \ldots, W_T]'$ is uncorrelated with e but may be correlated with u (when $b \neq 0$). So a controls the endogeneity of the variable Y, whereas b represents the possible endogeneity of the instruments X. If $b = 0$, the instruments X are valid; otherwise, they are invalid (endogenous). More precisely, if $b \neq 0$, i.e. there exists at least one i such that $b_i \neq 0$, $i = 1, \ldots, k$, and the corresponding variable X_i does not constitute a valid instrument.

We also make the following generic assumptions on the asymptotic behavior of model variables [where $A > 0$ for a matrix A means that A is positive definite (p.d.), and $\to$ refers to limits as $T \to \infty$]:

$$\frac{1}{T}[V \;\; \varepsilon]'[V \;\; \varepsilon] \xrightarrow{\text{p}} \begin{bmatrix} \Sigma_V & 0' \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} > 0, \tag{2.11}$$

$$\frac{1}{T}[X_0 \;\; W]'[X_0 \;\; W] \xrightarrow{\text{p}} \begin{bmatrix} \Sigma_0 & 0' \\ 0 & \Sigma_W \end{bmatrix}, \quad \Sigma_0 > 0, \tag{2.12}$$

$$\frac{1}{T}X_0'[V \;\; \varepsilon \;\; e] \xrightarrow{\text{p}} 0, \tag{2.13}$$

$$\frac{1}{T}X'X \xrightarrow{\text{p}} \Sigma_X, \tag{2.14}$$

$$\frac{1}{T}[W \;\; e]'[W \;\; e] \xrightarrow{\text{p}} \begin{bmatrix} \Sigma_W & 0' \\ 0 & \sigma_e^2 \end{bmatrix}, \tag{2.15}$$

$$\frac{1}{T}W'V \xrightarrow{\text{p}} \Sigma_{WV}, \tag{2.16}$$

$$\frac{1}{\sqrt{T}}\begin{bmatrix} X'e \\ (X'W - \Sigma_W)b \end{bmatrix} \xrightarrow{\text{L}} \begin{bmatrix} S_e \\ S_b \end{bmatrix} \sim \text{N}[0, \Sigma_S], \tag{2.17}$$

$$S_e \sim \text{N}[0, \sigma_e^2\Sigma_X], \quad S_b \sim \text{N}[0, \sigma_e^2\Sigma_b], \tag{2.18}$$

where Σ_V is $G \times G$ fixed matrix, Σ_0 and Σ_W are $k \times k$ fixed matrices and S_e and S_b are $k \times 1$ random vectors. Note that Σ_W may be singular, and S_b may not be independent of S_e.

From the above assumptions, it is easy to see that

$$\frac{1}{T}X_0'u \xrightarrow{\text{p}} 0, \quad \frac{1}{T}X_0'e \xrightarrow{\text{p}} 0, \tag{2.19}$$

$$\frac{1}{T}X'u \xrightarrow{\text{p}} \varphi = \Sigma_W b, \quad \frac{X'V}{T} \xrightarrow{\text{p}} \Sigma_{WV}, \tag{2.20}$$

$$\frac{1}{T}[u \;\; V]'[u \;\; V] \xrightarrow{\text{p}} \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0, \tag{2.21}$$

$$\frac{1}{T}[u \;\; W]'[u \;\; W] \xrightarrow{\text{p}} \begin{bmatrix} \sigma_u^2 & \varphi' \\ \varphi & \Sigma_W \end{bmatrix}, \tag{2.22}$$

$$\frac{1}{T}[W \;\; V]'[\varepsilon \;\; e] \xrightarrow{\text{p}} \begin{bmatrix} \delta_{W\varepsilon} & 0 \\ 0 & \delta_{Ve} \end{bmatrix}, \tag{2.23}$$

where

$$\delta = \Sigma_V a, \quad \sigma_u^2 = a'\Sigma_V a + \sigma_\varepsilon^2 = \sigma_e^2 + b'\Sigma_W b, \tag{2.24}$$

$$\Sigma_V a = \Sigma_{WV}' b + \delta_{Ve}, \quad \Sigma_W b = \Sigma_{WV} a + \delta_{W\varepsilon}, \tag{2.25}$$

$$\Sigma_X = \Sigma_0 + \Sigma_W > 0, \quad \Sigma_{XY} = \Sigma_X \Pi + \Sigma_{WV}, \tag{2.26}$$

$$\Sigma_Y = \Pi'\Sigma_X\Pi + \Sigma_V + \Sigma_{WV}'\Pi + \Pi'\Sigma_{WV}. \tag{2.27}$$

Finally, we denote by $\mathcal{N}(\Sigma_W)$ the null set of the linear map on $\mathbb{R}^k$ characterized by the matrix Σ_W:

$$\mathcal{N}(\Sigma_W) = \{x \in \mathbb{R}^k : \Sigma_W x = 0\}. \tag{2.28}$$

If Σ_W is a full-column-rank matrix, then $\mathcal{N}(\Sigma_W) = \{0\}$; otherwise, there is at least one $x_0 \neq 0$ such that $\Sigma_W x_0 = 0$.

The setup described above is quite wide and does allow one to study several questions associated with the possible presence of "invalid" instruments. In particular, an important practical problem consists in studying the effect on inference of adding an "invalid" instrument to a list of valid (possibly identifying) instruments. Note that this problem is distinct from studying the effect of imposing "incorrect" overidentifying restrictions (as done by Small, 2007). To better see the issues studied here, it will be useful to consider a simple example.

Example 2.1. Consider a model with one endogenous explanatory variable ($G = 1$) and two candidate instruments ($k = 2$). Then Y and V are $T \times 1$ vectors, $X = [X_1, X_2]$ and $W = [W_1, W_2]$ are $T \times 2$ matrices, $\Pi = [\pi_1, \pi_2]'$ and $b = [b_1, b_2]'$ are vectors of dimension 2, and

$$Y = X\Pi + V = X_1\pi_1 + X_2\pi_2 + V, \tag{2.29}$$

$$u = Wb + e = W_1 b_1 + W_2 b_2 + e. \tag{2.30}$$

Let us further assume that X_1 is a valid instrument (with $W_1 = 0$), $\mathsf{E}[u|X_1] = 0$, $X_2 = W_2$, $\pi_2 = 0$ and $b_1 = 0$, where e is independent of X_1 and X_2 (with finite mean zero), so that

$$Y = X\Pi + V = X_1\pi_1 + V, \tag{2.31}$$

$$u = Wb + e = W_2 b_2 + e. \tag{2.32}$$

Here W_2 is not a "valid" instrument when $b_2 \neq 0$. But the structural equation (2.3) may in principle be estimated using only X_1 as an instrument, because $\mathsf{E}[u|X_1] = 0$; if X_1 is not a weak instrument ($\pi_1 \neq 0$) and satisfies usual regularity conditions, a consistent estimate of β can be obtained. Among other things, we study below the effect (on some identification-robust tests) of taking X_2 as an instrument when $b_2 \neq 0$, i.e. when X_2 is correlated with u. Note that the condition $E[u|X_1] = 0$ *does not entail* $\mathsf{E}[e|X_1, X_2] = 0$, which is a maintained hypothesis used by Small (2007). So the problem considered here is distinct from the problem of testing overidentifying restrictions (studied, for example, by Sargan, 1958; Kadane and Anderson, 1977; Small, 2007).

3. Test statistics

We consider in this paper the problem of testing

$$\mathrm{H}_0 : \beta = \beta_0, \tag{3.1}$$

where some of the "instruments" used are in fact endogenous ($b \neq 0$). We analyze the behavior of the Anderson–Rubin and Kleibergen statistics. The Anderson and Rubin (1949) test for H_0 in Eq. (2.3) involves considering the transformed equation

$$y - Y\beta_0 = X\Delta + \varepsilon, \tag{3.2}$$

where $\Delta = \Pi(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. H_0 can then be assessed by testing $\mathrm{H}_0' : \Delta = 0$. The AR statistic for H_0' is given by

$$\mathrm{AR}(\beta_0) = \frac{1}{k}\frac{(y - Y\beta_0)' P_X (y - Y\beta_0)}{(y - Y\beta_0)' M_X (y - Y\beta_0)/(T - k)}, \tag{3.3}$$

where $M_B = I - P_B$ and $P_B = B(B'B)^{-1}B'$ is the projection matrix on the space spanned by the columns of B. If $b = 0$, the asymptotic distribution of $\mathrm{AR}(\beta_0)$ is a $\chi^2(k)/k$ under H_0. If furthermore $u \sim \mathrm{N}[0, \sigma^2 I_T]$ and X is independent of u, then $\mathrm{AR}(\beta_0) \sim F(k, T - k)$ under H_0 irrespective of whether the instruments are strong or weak. However, when some instruments are invalid, the distribution of the AR statistic may be affected.

Kleibergen (2002) proposed a modification of the AR statistic to take into account the fact that this statistic may have low power when there are too many instruments in the model. The modified statistic for testing H_0 can be written as

$$K(\beta_0) = \frac{(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0)}{(y - Y\beta_0)' M_X (y - Y\beta_0)/(T - k)}, \tag{3.4}$$

where

$$\tilde{Y}(\beta_0) = X\tilde{\Pi}(\beta_0), \quad \tilde{\Pi}(\beta_0) = (X'X)^{-1}X'\left[Y - (y - Y\beta_0)\frac{S_{uV}(\beta_0)}{S_{uu}(\beta_0)}\right], \tag{3.5}$$

$$S_{uu}(\beta_0) = \frac{1}{T - k}(y - Y\beta_0)' M_X (y - Y\beta_0), \quad S_{uV}(\beta_0) = \frac{1}{T - k}(y - Y\beta_0)' M_X Y. \tag{3.6}$$

Unlike the AR statistic which projects $y - Y\beta_0$ on the k columns of X, the K statistic projects $y - Y\beta_0$ on the G columns of $X\tilde{\Pi}(\beta_0)$. If the instruments X are exogenous, $\tilde{\Pi}(\beta_0)$ is both a consistent estimator of Π and asymptotically independent of $X'(y - Y\beta_0)$ under H_0, and $K(\beta_0)$ converges to a $\chi^2(G)$. However, if some instruments are invalid ($b \neq 0$), $\tilde{\Pi}(\beta_0)$ may not be asymptotically independent of $X'(y-Y\beta_0)$ and the asymptotic distribution of the K statistic may not be $\chi^2(G)$.[2]

If the model contains only one instrument and one endogenous variable ($G = k = 1$), the AR and K statistics are equivalent and pivotal even in finite samples whenever $b = 0$. When $k > 1$, even if $b = 0$, the K statistic is not pivotal in finite samples but is asymptotically pivotal, whereas the AR statistic is pivotal even in finite samples (when X is independent of u). Following Staiger and Stock (1997), we refer to the *locally weak-instrument* asymptotic setup by considering a limiting sequence of Π where Π is local-to-zero. We also consider a limiting sequence of b where b is local-to-zero. We refer to this later limiting sequence as *locally exogenous* instruments asymptotic.

4. Asymptotic theory with invalid and weak instruments

In this section, we study the large-sample properties of the statistics described above when some of the instruments used are invalid. Two setups are considered. The first is the possibly invalid instrument setup, i.e. the endogeneity parameter b is a fixed vector. The second is the locally exogenous instrument setup, i.e. b is local-to-zero.

4.1. Possibly invalid instruments

We consider first the case where the endogeneity parameter b is a constant vector and we analyze the asymptotic distributions of the statistics. Our results cover both strong and weak-instrument asymptotic. Theorem 4.1 below summarizes the asymptotic behavior of the AR statistic when some instruments may be endogenous. For a random variable S whose distribution depends on the sample size T, the notation $S \xrightarrow{\mathrm{L}} +\infty$ means that $P[S > x] \to 1$ as $T \to \infty$, for any x.

Theorem 4.1 (*Asymptotic distribution of the AR statistic*). *Suppose that assumptions* (2.3)–(2.18) *hold*, *with* $b = b_0$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors. If* $b_0 \notin \mathcal{N}(\Sigma_W)$, *then*

$$\mathrm{AR}(\beta_0) \xrightarrow{\mathrm{L}} +\infty. \tag{4.1}$$

If $b_0 \in \mathcal{N}(\Sigma_W)$, *then*

$$\mathrm{AR}(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{k\sigma_u^2}(S_e + S_b)'\Sigma_X^{-1}(S_e + S_b), \tag{4.2}$$

where S_e *and* S_b *are defined in* (2.17)–(2.18). *If* $b_0 = 0$, *then*

$$\mathrm{AR}(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{k}\chi^2(k). \tag{4.3}$$

In the above theorem, no restriction is imposed on the rank of Π. In particular, the result holds even if Π is not a full-column-rank matrix. When $b_0 \notin \mathcal{N}(\Sigma_W)$, the AR statistic diverges under the null hypothesis H_0. When $b_0 \in \mathcal{N}(\Sigma_W)$, the limiting distribution of the AR statistic does not diverge, but the AR test is not valid unless $S_b = 0$. Of course, when $b_0 = 0$—which is the classical exogenous instrument setup—$S_b = 0$ and the AR test is asymptotically valid.

Theorem 4.2 below summarizes the asymptotic behavior of the K statistic when some instruments are possibly invalid.

Theorem 4.2 (*Asymptotic distribution of the K statistic*). *Suppose that assumptions* (2.3)–(2.18) *hold*, *with* $b = b_0$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors.*

(A) *If* $b_0 \notin \mathcal{N}(\Sigma_W)$ *then*

$$K(\beta_0) \xrightarrow{\mathrm{L}} +\infty \tag{4.4}$$

[2] We do not study in this paper conditional tests such as those proposed by Moreira (2003), because the distributional theory for such tests is considerably more complex and would go beyond the scope of a short paper like the present one.

when at least one of the following two conditions holds: (i) $\Pi = \Pi_0 \neq 0$ *with* $\operatorname{rank}(\tilde{\Sigma}_{XY}) = G$ *or* (ii) $\Pi = \Pi_0/\sqrt{T}$ *with* $\operatorname{rank}(\Sigma^*_{XY}) = G$, *where*

$$\tilde{\Sigma}_{XY} = \Sigma_{XY} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2), \quad \Sigma^*_{XY} = \Sigma_{WV} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2),$$

$$q_{uV} = \delta' - b_0' \Sigma_W \Sigma_X^{-1} \Sigma_{WV}, \quad \bar{\sigma}_u^2 = \sigma_u^2 - b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0.$$

(B) *If* $b_0 \in \mathcal{N}(\Sigma_W)$, *then*

$$K(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{\sigma_u^2} (S_e + S_b)' \Sigma_X^{-1} \Sigma_{XY} (\Sigma'_{XY} \Sigma_X^{-1} \Sigma_{XY})^{-1} \Sigma'_{XY} \Sigma_X^{-1} (S_e + S_b) \tag{4.5}$$

when $\Pi = \Pi_0 \neq 0$ *and* $\operatorname{rank}(\Sigma_{XY}) = G$, *and*

$$K(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{\sigma_u^2} (S_e + S_b)' \Sigma_X^{-1} \Sigma_{WV} (\Sigma'_{WV} \Sigma_X^{-1} \Sigma_{WV})^{-1} \Sigma'_{WV} \Sigma_X^{-1} (S_e + S_b) \tag{4.6}$$

when $\Pi = \Pi_0/\sqrt{T}$ *and* $\operatorname{rank}(\Sigma_{WV}) = G$.

(C) *If* $b_0 = 0$, *then*

$$K(\beta_0) \xrightarrow{\mathrm{L}} \chi^2(G) \tag{4.7}$$

when at least one of the following two conditions holds: (i) $\Pi = \Pi_0 \neq 0$ *with* $\operatorname{rank}(\Sigma_{XY}) = G$ *or* (ii) $\Pi = \Pi_0/\sqrt{T}$ *with* $\operatorname{rank}(\Sigma_{WV}) = G$.

Unlike Theorem 4.1 for the AR statistic, Theorem 4.2 requires an additional rank assumption. When $b_0 \notin \mathcal{N}(\Sigma_W)$, the null limiting distribution of the K statistic diverges. This means that the K test often rejects H_0 asymptotically when $b_0 \notin \mathcal{N}(\Sigma_W)$. Furthermore, when $b_0 \in \mathcal{N}(\Sigma_W)$, the K test is not asymptotically valid unless $S_b = 0$. As expected, if $b_0 = 0$ (i.e. $S_b = 0$), the K statistic converges to a $\chi^2(G)$. It is worthwhile to note that the case where the rank assumption fails (e.g. the partial identification of β) is not covered in this paper.

Finally, it is interesting to observe that the limiting value of the two-stage least-squares (2SLS) estimator of β,

$$\tilde{\beta} = (\hat{Y}'\hat{Y})^{-1} \hat{Y}' y = [Y'X(X'X)^{-1}X'Y]^{-1} Y'X(X'X)^{-1}X'y, \tag{4.8}$$

is given by

$$\operatorname*{plim}_{T \to \infty} \tilde{\beta} = \beta + [\Sigma'_{XY} \Sigma_X^{-1} \Sigma_{XY}]^{-1} \Sigma'_{XY} \Sigma_X^{-1} \Sigma_W b \tag{4.9}$$

provided $\operatorname{rank}(\Sigma_{XY}) = G$, so that $\tilde{\beta}$ is consistent when $b_0 \in \mathcal{N}(\Sigma_W)$ and Σ_{XY} has full column rank (even if some instruments are invalid). If $b_0 \in \mathcal{N}(\Sigma_W)$ but $b_0 \neq 0$, the asymptotic level of the Anderson–Rubin and Kleibergen tests can be affected.

4.2. *Locally exogenous instruments*

We consider now the case where the endogeneity parameter b is local-to-zero. As in the previous subsection, we analyze the limiting distributions of the statistics. The results also cover two setups: locally exogenous instruments $[\Pi = \Pi_0 \neq 0, b = b_0/\sqrt{T}]$ and weak locally exogenous instruments $[\Pi = \Pi_0/\sqrt{T}, b = b_0/\sqrt{T}]$. Theorems 4.3 and 4.4 below derive the distributions of the statistics for both setups.

Theorem 4.3 (*Asymptotic distributions with locally exogenous instruments*). *Suppose that assumptions* (2.3)–(2.18) *hold, with* $b = b_0/\sqrt{T}$, $\Pi = \Pi_0 \neq 0$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors, and* Π_0 *is a given matrix. If* $b_0 \notin \mathcal{N}(\Sigma_W)$, *then*

$$\mathrm{AR}(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{k} \chi^2(k, \mu_1), \tag{4.10}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G, m'm) \quad \text{if } \operatorname{rank}(\Sigma_{XY}) = G, \tag{4.11}$$

where

$$\mu_1 = \frac{1}{\sigma_e^2} b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0, \quad m = \frac{1}{\sigma_e} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1/2} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_W b_0, \tag{4.12}$$

and Σ_X, Σ_{XY} *and* Σ_W *are given in* (2.11)–(2.27). *If* $b_0 \in \mathcal{N}(\Sigma_W)$, *then*

$$\operatorname{AR}(\beta_0) \xrightarrow{L} \frac{1}{k} \chi^2(k), \tag{4.13}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G) \quad \text{if } \operatorname{rank}(\Sigma_{XY}) = G. \tag{4.14}$$

Theorem 4.4 (*Asymptotic distributions with weak locally exogenous instruments*). *Suppose that assumptions* (2.3)–(2.18) *hold, with* $b = b_0/\sqrt{T}$, $\Pi = \Pi_0/\sqrt{T}$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors, and* Π_0 *is a given matrix* ($\Pi_0 = 0$ *is allowed*). *If* $b_0 \notin \mathcal{N}(\Sigma_W)$, *then*

$$\operatorname{AR}(\beta_0) \xrightarrow{L} \frac{1}{k} \chi^2(k, \mu_1), \tag{4.15}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G, \tilde{m}'\tilde{m}) \quad \text{if } \operatorname{rank}(\Sigma_{WV}) = G, \tag{4.16}$$

where

$$\tilde{m} = \frac{1}{\sigma_e} (\Sigma_{WV}' \Sigma_X^{-1} \Sigma_{WV})^{-1/2} \Sigma_{WV}' \Sigma_X^{-1} \Sigma_W b_0, \tag{4.17}$$

and Σ_X, Σ_{WV}, Σ_W *and* μ_1 *are defined in Theorem* 4.3. *If* $b_0 \in \mathcal{N}(\Sigma_W)$, *then*

$$\operatorname{AR}(\beta_0) \xrightarrow{L} \frac{1}{k} \chi^2(k), \tag{4.18}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G) \quad \text{if } \operatorname{rank}(\Sigma_{WV}) = G. \tag{4.19}$$

We make the following remarks concerning Theorems 4.3 and 4.4. First, the endogeneity parameter b is local-to-zero, and for $b_0 \in \mathcal{N}(\Sigma_W)$ the AR and K tests are asymptotically valid. However, unlike the AR test, the validity of the K test is established under an additional rank assumption (the case where this additional rank assumption fails is not covered in this paper). So, when $b_0 \in \mathcal{N}(\Sigma_W)$, inference with locally exogenous instruments using the AR and K tests is feasible (at least in large samples). Second, if $b_0 \notin \mathcal{N}(\Sigma_W)$, the results in both theorems are different from those of Theorems 4.1 and 4.2 because the limiting distributions of both statistics do not diverge. Third, even though the AR and K statistics have non-central chi-square limiting distributions when $b_0 \notin \mathcal{N}(\Sigma_W)$, they are not pivotal since the non-centrality parameters depend on nuisance parameters. In addition, the limiting distributions of both statistics cannot be bounded by any pivotal distribution.

It will be useful to see how the above theorems apply in a simple example.

Example 4.1. Consider again model (2.29)–(2.30), which involves one endogenous explanatory variable and two instruments. If the matrix Σ_W is invertible, then $\mathcal{N}(\Sigma_W) = \{0\}$, and Theorem 4.1 entails that $\operatorname{AR}(\beta_0) \xrightarrow{L} +\infty$ under the null hypothesis $\beta = \beta_0$. Similarly, if $\tilde{\Sigma}_{XY} \neq 0$, then $\operatorname{rank}(\tilde{\Sigma}_{XY}) = G = 1$ and Theorem 4.2 entails that $K(\beta_0) \xrightarrow{L} +\infty$ when $\beta = \beta_0$. If X_1 is a valid instrument (with $W_1 = 0$) and $X_2 = W_2$ with $W_2' W_2 / T \xrightarrow{p} \sigma_{W_2}^2 > 0$, we have

$$\Sigma_W = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{W_2}^2 \end{bmatrix}, \tag{4.20}$$

which is a matrix of rank one, and $\mathcal{N}(\Sigma_W) = \{(x_1, x_2)' : x_2 = 0\}$. If $b_1 = 0$, then $b_0 \in \mathcal{N}(\Sigma_W)$ and Theorem 4.1 entails that the asymptotic distribution given by (4.2) holds for $\operatorname{AR}(\beta_0)$, while for $K(\beta_0)$ part B of Theorem 4.2 is applicable. Of course, when $b_0 = 0$, $\operatorname{AR}(\beta_0)$ follows the usual $\chi^2(2)/2$ asymptotic distribution, while $K(\beta_0)$ follows a $\chi^2(1)$ distribution. For locally exogenous instruments, Theorems 4.3 and 4.4 can be applied in a similar way.

5. Conclusion

In this paper, we have established conditions under which the AR and K tests are asymptotically valid even if some instruments used are endogenous. We have also shown that when these conditions fail, the limiting distributions of both statistics may diverge. Furthermore, when these conditions fail, under locally exogenous instruments setup, the limiting distributions of the statistics depend on nuisance parameters and cannot be bounded by any pivotal distribution. In consequence, the weak-instrument procedure proposed by Wang and Zivot (1998), the unified weak-instrument framework of Swanson and Chao (2005) and the inference with imperfect instruments suggested by Ashley (2006) are not applicable. Overall, our results underscore the importance of checking for the presence of possibly invalid instruments when applying "identification-robust" tests. They also suggest that sensitivity analyses where different sets of instruments are considered (Ashley, 2006; Small, 2007) can be quite useful for the interpretation of empirical results based on instrumental variables.

Acknowledgements

The authors thank Tze Leung Lai and Ingram Olkin for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], 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), and a Killam Fellowship (Canada Council for the Arts).

Appendix A. Proofs

Proof of Theorem 4.1. Note first that

$$\frac{(y - Y\beta_0)' M_X (y - Y\beta_0)}{T-k} = \frac{u'u}{T-k} - \frac{T}{T-k}\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right), \tag{A.1}$$

where, by assumptions (2.3)–(2.18),

$$\frac{u'u}{T-k} \xrightarrow{\text{p}} \sigma_u^2 > 0, \qquad \frac{X'X}{T} \xrightarrow{\text{p}} \Sigma_X > 0, \qquad \frac{X'u}{T} = \frac{X_0'u}{T} + \frac{W'W}{T} b_0 + \frac{W'e}{T} \xrightarrow{\text{p}} \Sigma_W b_0, \tag{A.2}$$

$$\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \xrightarrow{\text{p}} b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0, \tag{A.3}$$

$$\frac{(y - Y\beta_0)' M_X (y - Y\beta_0)}{T-k} \xrightarrow{\text{p}} \bar{\sigma}_u^2 = \sigma_u^2 - b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0 \geqslant 0.. \tag{A.4}$$

(A) Suppose now that $b_0 \notin \mathcal{N}(\Sigma_W)$. Then $b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0 > 0$ and the numerator of the AR statistic diverges:

$$(y - Y\beta_0)' P_X (y - Y\beta_0) = T\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \xrightarrow{\text{L}} +\infty, \tag{A.5}$$

hence

$$\text{AR}(\beta_0) \xrightarrow{\text{L}} +\infty. \tag{A.6}$$

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$ and $\bar{\sigma}_u^2 = \sigma_u^2$. Further,

$$X'u = X'(e + Wb_0) = X'e + X'Wb_0, \tag{A.7}$$

$$\frac{1}{\sqrt{T}}X'u = \frac{1}{\sqrt{T}}[X'u - \Sigma_W b_0] = \frac{1}{\sqrt{T}}X'e + \frac{1}{\sqrt{T}}(X'W - \Sigma_W)b_0 \xrightarrow{\mathrm{L}} S = S_e + S_b. \tag{A.8}$$

Then,

$$(y - Y\beta_0)'P_X(y - Y\beta_0) = \left(\frac{u'X}{\sqrt{T}}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{\sqrt{T}}\right) \xrightarrow{\mathrm{L}} S'\Sigma_X^{-1}S, \tag{A.9}$$

$$\frac{(y - Y\beta_0)'M_X(y - Y\beta_0)}{T - k} \xrightarrow{\mathrm{p}} \sigma_u^2, \tag{A.10}$$

hence

$$\mathrm{AR}(\beta_0) \xrightarrow{L} \frac{1}{k\sigma_u^2}S'\Sigma_X^{-1}S. \tag{A.11}$$

(C) Finally, if $b_0 = 0$, we have $b_0 \in \mathcal{N}(\Sigma_W)$, with the extra restrictions $u = e$, $\sigma_u^2 = \sigma_e^2$,

$$S = \frac{1}{\sqrt{T}}X'u = \frac{1}{\sqrt{T}}X'e \xrightarrow{\mathrm{L}} \mathrm{N}[0, \sigma_e^2\Sigma_X],$$

hence

$$\mathrm{AR}(\beta_0) \xrightarrow{\mathrm{L}} \frac{1}{k\sigma_e^2}S_e'\Sigma_X^{-1}S_e \sim \frac{1}{k}\chi^2(k). \qquad \square \tag{A.12}$$

Proof of Theorem 4.2. We note first, as in (A.1)–(A.4), that

$$S_{uu}(\beta_0) = \frac{(y - Y\beta_0)'M_X(y - Y\beta_0)}{T - k} \xrightarrow{\mathrm{p}} \bar{\sigma}_u^2, \quad \frac{X'X}{T} \xrightarrow{\mathrm{p}} \Sigma_X > 0, \quad \frac{X'u}{T} \xrightarrow{\mathrm{p}} \Sigma_W b_0. \tag{A.13}$$

(A) Suppose that $b_0 \notin \mathcal{N}(\Sigma_W)$.
(i) Let $\Pi = \Pi_0 \neq 0$. Then, we have

$$S_{uV}(\beta_0) = \frac{1}{T - k}(y - Y\beta_0)'M_X Y \xrightarrow{\mathrm{p}} q_{uV} = \delta' - b_0'\Sigma_W\Sigma_X^{-1}\Sigma_{WV}, \tag{A.14}$$

$$\tilde{\Pi}(\beta_0) = \left(\frac{X'X}{T}\right)^{-1}\frac{X'Y}{T} - \left(\frac{X'X}{T}\right)^{-1}\frac{X'u}{T}\frac{S_{uV}(\beta_0)}{S_{uu}(\beta_0)} \xrightarrow{\mathrm{p}} \Sigma_X^{-1}\tilde{\Sigma}_{XY}, \tag{A.15}$$

where $\tilde{\Sigma}_{XY} = \Sigma_{XY} - \Sigma_W b_0(q_{uV}/\bar{\sigma}_u^2)$, and

$$\frac{\tilde{Y}(\beta_0)'u}{T} = \tilde{\Pi}(\beta_0)'\frac{X'u}{T} \xrightarrow{\mathrm{p}} \tilde{\Sigma}_{XY}'\Sigma_X^{-1}\Sigma_W b_0, \tag{A.16}$$

$$\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{\mathrm{p}} \tilde{\Sigma}_{XY}'\Sigma_X^{-1}\tilde{\Sigma}_{XY}. \tag{A.17}$$

If $\mathrm{rank}(\tilde{\Sigma}_{XY}) = G$, then $\tilde{\Sigma}_{XY}'\Sigma_X^{-1}\tilde{\Sigma}_{XY} > 0$ and $\Sigma_X^{-1}\tilde{\Sigma}_{XY}\Sigma_W b_0 \neq 0$ for $b_0 \notin \mathcal{N}(\Sigma_W)$, hence

$$\frac{u'\tilde{Y}(\beta_0)}{T}\left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T}\right]^{-1}\frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{\mathrm{p}} b_0'\Sigma_W\Sigma_X^{-1}\tilde{\Sigma}_{XY}(\tilde{\Sigma}_{XY}'\Sigma_X^{-1}\tilde{\Sigma}_{XY})^{-1}\tilde{\Sigma}_{XY}'\Sigma X^{-1}\Sigma_W b_0 > 0.$$

Consequently, the numerator of the K statistic diverges:

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = T \frac{u'\tilde{Y}(\beta_0)}{T} \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{\text{p}} +\infty \tag{A.18}$$

and

$$K(\beta_0) \xrightarrow{\text{L}} +\infty. \tag{A.19}$$

(ii) Let $\Pi = \Pi_0/\sqrt{T}$. Then

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = T \frac{u'\tilde{Y}(\beta_0)}{T} \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \frac{\tilde{Y}(\beta_0)'u}{T}, \tag{A.20}$$

where

$$\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{\text{L}} \Sigma_{XY}^{*\prime} \Sigma_X^{-1} \Sigma_{XY}^*, \qquad \frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{\text{p}} \Sigma_X^{-1} \Sigma_{XY}^* \Sigma_W b_0,$$

with $\Sigma_{XY}^* = \Sigma_{WV} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2)$. If $\text{rank}(\Sigma_{XY}^*) = G$, then the numerator of the K statistic diverges, and $K(\beta_0) \xrightarrow{\text{L}} +\infty$.

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$, $\bar{\sigma}_u^2 = \sigma_u^2$ and $(1/\sqrt{T})X'u \xrightarrow{\text{L}} S = S_e + S_b$ as in (A.7)–(A.8).

(i) If $\Pi = \Pi_0 \neq 0$, the denominator of the K statistic satisfies

$$\frac{1}{T}(y - Y\beta_0)' M_X (y - Y\beta_0) \xrightarrow{\text{p}} \sigma_u^2, \tag{A.21}$$

while the denominator can be written as

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = \frac{u'X}{\sqrt{T}} \tilde{\Pi}(\beta_0) \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \tilde{\Pi}(\beta_0)' \frac{X'u}{\sqrt{T}}, \tag{A.22}$$

where

$$\tilde{\Pi}(\beta_0) \xrightarrow{\text{p}} \Sigma_X^{-1} \Sigma_{XY}, \qquad \frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{\text{p}} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY}, \qquad \frac{\tilde{Y}(\beta_0)'u}{\sqrt{T}} \xrightarrow{\text{p}} \Sigma_X^{-1} \Sigma_{XY} S. \tag{A.23}$$

If $\text{rank}(\Sigma_{XY}) = G$, we have $\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY} > 0$, hence

$$K(\beta_0) \xrightarrow{\text{L}} \frac{1}{\sigma_u^2} S' \Sigma_X^{-1} \Sigma_{XY} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1} \Sigma_{XY}' \Sigma_X^{-1} S. \tag{A.24}$$

(ii) If $\Pi = \Pi_0/\sqrt{T}$, the numerator of the K statistic is

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = \frac{u'X}{\sqrt{T}} \tilde{\Pi}(\beta_0) \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \tilde{\Pi}(\beta_0)' \frac{X'u}{\sqrt{T}}, \tag{A.25}$$

hence

$$\tilde{\Pi}(\beta_0) \xrightarrow{\text{p}} \Sigma_X^{-1} \Sigma_{WV}, \qquad \frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{\text{p}} \Sigma_{WV}' \Sigma_X^{-1} \Sigma_{WV}, \qquad \frac{\tilde{Y}(\beta_0)'u}{\sqrt{T}} \xrightarrow{\text{p}} \Sigma_X^{-1} \Sigma_{WV} S. \tag{A.26}$$

If $\text{rank}(\Sigma_{WV}) = G$, then

$$K(\beta_0) \xrightarrow{\text{L}} \frac{1}{\sigma_u^2} S' \Sigma_X^{-1} \Sigma_{WV} (\Sigma_{WV}' \Sigma_X^{-1} \Sigma_{WV})^{-1} \Sigma_{WV}' \Sigma_X^{-1} S. \tag{A.27}$$

(C) Finally, if $b_0 = 0$, we have $b_0 \in \mathcal{N}(\Sigma_W)$, with the extra restrictions $u = e$, $\sigma_u^2 = \sigma_e^2$,

$$S = \frac{1}{\sqrt{T}} X'u \xrightarrow{\text{L}} \text{N}[0, \sigma_e^2 \Sigma_X],$$

hence, if $\Pi = \Pi_0 \neq 0$,

$$K(\beta_0) \xrightarrow{\text{L}} \frac{1}{\sigma_e^2} S_e' \Sigma_X^{-1} \Sigma_{XY} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1} \Sigma_{XY}' \Sigma_X^{-1} S_e \sim \chi^2(G), \tag{A.28}$$

and if $\Pi = \Pi_0/\sqrt{T}$ (where $\Pi_0 = 0$ is allowed),

$$K(\beta_0) \xrightarrow{\text{L}} \frac{1}{\sigma_e^2} S_e' \Sigma_X^{-1} \Sigma_{WV} (\Sigma_{WV}' \Sigma_X^{-1} \Sigma_{WV})^{-1} \Sigma_{WV}' \Sigma_X^{-1} S_e \sim \chi^2(G). \qquad \square \tag{A.29}$$

Proof of Theorem 4.3. Since b is now local-to-zero, we have

$$\frac{X'u}{\sqrt{T}} \xrightarrow{\text{L}} S_e + \Sigma_W b_0, \quad \frac{X'X}{T} \xrightarrow{\text{p}} \Sigma_X, \quad \frac{X'u}{T} \xrightarrow{\text{p}} 0, \quad \frac{(y - Y\beta_0)' M_X (y - Y\beta_0)}{T-k} \xrightarrow{\text{p}} \sigma_u^2 > 0. \tag{A.30}$$

Further, we have

$$\begin{aligned} \frac{u'u}{T-k} &= \frac{\left(e + W\frac{b_0}{\sqrt{T}}\right)' \left(e + W\frac{b_0}{\sqrt{T}}\right)}{T-k} \\ &= \frac{e'e}{T-k} + \frac{b_0' W' e}{\sqrt{T}(T-k)} + \frac{e' W b_0}{\sqrt{T}(T-k)} + \frac{b_0' e' W b_0}{T(T-k)} \xrightarrow{\text{p}} \sigma_e^2 = \sigma_u^2. \end{aligned} \tag{A.31}$$

(A) Let $b_0 \notin \mathcal{N}(\Sigma_W)$. Then,

$$\text{AR}(\beta_0) \xrightarrow{\text{L}} \frac{1}{k\sigma_e^2} (S_e + \Sigma_W b_0)' \Sigma_X^{-1} (S_e + \Sigma_W b_0) \sim \frac{1}{k} \chi^2(k, \mu_1), \tag{A.32}$$

where

$$\mu_1 = \frac{1}{\sigma_e^2} b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0 \neq 0.$$

Similarly, we have

$$\frac{\tilde{Y}(\beta_0)' \tilde{Y}(\beta_0)}{T} \xrightarrow{\text{p}} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY}$$

and

$$\frac{\tilde{Y}(\beta_0)' u}{\sqrt{T}} \xrightarrow{\text{L}} \Sigma_X^{-1} \Sigma_{XY} (S_e + \Sigma_W b_0).$$

So, if $\text{rank}(\Sigma_{XY}) = G$, we have

$$K(\beta_0) \xrightarrow{\text{L}} \frac{1}{\sigma_e^2} (S_e + \Sigma_W b_0)' \Sigma_X^{-1} \Sigma_{XY} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1} \Sigma_{XY}' \Sigma_X^{-1} (S_e + \Sigma_W b_0) \sim \chi^2(G, m'm), \tag{A.33}$$

where

$$m = \frac{1}{\sigma_e} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1/2} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_W b_0 \neq 0.$$

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$. Then $\mu_1 = 0$ and $m = 0$, hence

$$\text{AR}(\beta_0) \xrightarrow{\text{L}} \frac{1}{k} \chi^2(k)$$

and $K(\beta_0) \xrightarrow{\text{L}} \chi^2(G)$. $\square$

Proof of Theorem 4.4. The proof of Theorem 4.3 for the AR statistic covers Theorem 4.4. The proof for the K statistic is similar to the one in Theorem 4.3. □

References

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Ann. of Math. Statist. 20 (1), 46–63.

Ashley, R., 2006. Assessing the credibilility of instrumental variables inference with imperfect instruments via sensitivity analysis. Technical Report, Economics Department, Virginia Polytechnic Institute Blacksburg, Virginia.

Bekker, P., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian J. Econom. 36 (4), 767–808.

Dufour, J.-M., Hsiao, C., 2008. Identification. In: Blume, L.E., Durlauf, S.N. (Eds.), The New Palgrave Dictionary of Economics, second ed. Palgrave Macmillan, Basingstoke, Hampshire, England, to appear.

Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. Internat. Econom. Rev. 42, 815–843.

Dufour, J.-M., Taamouti, M., 2003. Point-optimal instruments and generalized Anderson–Rubins procedures for nonlinear models. Technical Report, C.R.D.E., Université de Montréal.

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

Dufour, J.-M., Taamouti, M., 2006. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. J. Econometrics 139 (1), 133–153.

Hahn, J., Hausman, J.A., 2003. IV estimation with valid and invalid instruments. Technical Report, Department of Economics, Masschusetts Institute of Technology, Cambridge, MA.

Hall, A.R., Peixe, F.P.M., 2003. A consistent method for the selection of relevant instruments. Econometric Rev. 2 (3), 269–287.

Hall, A.R., Rudebusch, G.D., Wilcox, D.W., 1996. Judging instrument relevance in instrumental variables estimation. Internat. Econom. Rev. 37, 283–298.

Kadane, J.B., Anderson, T.W., 1977. A comment on the test of overidentifying restrictions. Econometrica 45 (4), 1027–1031.

Kiviet, J.F., Niemczyk, J., 2006. On the limiting and empirical distribution of IV estimators when some of the instruments are invalid. Technical Report, Department of Quantitative Economics, University of Amsterdam Amsterdam, The Netherlands.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70 (5), 1781–1803.

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

Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26 (3), 393–415.

Small, D.S., 2007. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. J. Amer. Statist. Assoc. 102 (479), 1049–1058.

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

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. J. Business and Econom. Statist. 20 (4), 518–529.

Swanson, N.R., Chao, J.C., 2005. Notes and comments: consistent estimation with a large number of weak instruments. Econometrica 73, 1673–1692.

Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66 (6), 1389–1404.